Marion Schindler/Bernhard Baumgartner/Harald Hruschka*

Nonlinear Effects in Brand Choice Models: Comparing Heterogeneous Latent Class To Homogeneous Nonlinear Models**

Abstract

We investigate whether the latent class multinomial logit choice model with segment-specific linear utility functions implies effects that are similar to those of parametric homogeneous nonlinear models given that this latent class model performs at least as well. The two nonlinear models have higher-order polynomial (i.e. quadratic and cubic) and piecewise linear utility functions, respectively. Piecewise linear functions are represented by linear splines and can reproduce threshold, saturation and asymmetric effects. We evaluate models and their variants using a tenfold cross-validation. As criterion we use the geometric mean of choice probabilities across all purchases for the brand actually chosen. We measure the similarity of effects between two models by the absolute differences of choice probabilities implied by these models for varying values of a predictor. Logits of choice probabilities provide a more detailed insight into the effects implied by models. For the data set we analyze, the latent class model with linear utility is clearly superior to the two homogeneous nonlinear models. Overall, the effects implied by the latent class models are similar to those of the two parametric nonlinear models.

JEL-Classification: C35, M31.

Keywords: Brand Choice; Latent Class Models; Nonlinear Effects.

1 Introduction

In marketing, the majority of models used to explain the effects of predictors on the behavior of individual customers or households have linear or loglinear functional forms.

* Marion Schindler, Database Manager, Versandhaus Robert Klingel GmbH & Co KG, Sachsenstrasse 23, D-75177 Pforzheim, Bernhard Baumgartner, Associate Professor, Faculty of Economics, University of Regensburg, Universitätsstraße 31, D-93053 Regensburg, and Harald Hruschka (corresponding author), Chaired Professor of Marketing, Faculty of Economics, University of Regensburg, Universitätsstraße 31, D-93053 Regensburg, e-mail: harald.hruschka@wiwi.uni-regensburg.de.

** The authors thank two anonymous reviewers who suggested to take heterogeneity into account and to compare the piecewise linear model to models with quadratic and cubic utility functions. We also acknowledge that one reviewer suggested to focus on the similarity of effects implied by the latent class choice model to those of parametric homogeneous nonlinear models.

But it is well known that in practice these effects are often nonlinear. Several earlier empirical studies, which we discuss in Section 2, demonstrate that predictors such as price, reference price, and brand loyalty affect the brand choice of individual households in a nonlinear way. Nevertheless, most brand choice models found in the literature are based on linear deterministic utility functions, i.e., they specify utility as linear combination of predictors (Gensch and Recker (1979); Guadagni and Little (1983); Winer (1986); Gupta (1988); Lattin and Bucklin (1989); Kalwani et al. (1990); Chintagunta (1992); Allenby and Lenk (1994); McCulloch and Rossi (1994); Allenby and Ginter (1995)). However, these models cannot deal with nonlinear effects and result in biased estimates.

There are several alternative ways to deal with nonlinear effects. First, researchers can work with homogeneous parametric nonlinear functions that do not vary across households. The literature on brand choice models discusses different ways of specifying homogeneous parametric nonlinear utility functios. We find parametric nonlinear utility functions with higher-order polynomial terms (e.g., quadratic or cubic in Pedryck and Zufryden (1991)), fixed transformations (e.g., semilog in Krishnamurthi and Raj (1992)) or piecewise linear functions of predictors (e.g., Kalyanaram and Little (1994); Wedel and Leeflang (1998)).

Latent class models with class-specific linear utility functions constitute another alternative method for dealing with nonlinear effects. Latent class models are well known in the marketing literature, which interprets latent classes as market segments. So far, latent class models have mainly been used to adress the latent heterogeneity of households. To the best of our knowledge, an investigation of the capability of these models to deal with nonlinearity is still lacking in the marketing literature. The ability of latent class models to reproduce nonlinear effects is due to the fact that they determine choice probability as a convex combination of class-specific choice probabilities (see Section 4).

These circumstances lead to our research question: which of these alternative approaches performs better, and whether the one that performs better can capture the nonlinearities. Or, to put the question in another way, whether the latent class choice model with segment-specific linear utility functions implies effects that are similar to those of parametric homogeneous nonlinear models given that this latent class model performs at least as well.

If the answer to this question turns out to be positive, we can recommend that researchers use this latent class model instead of a parametric homogeneous nonlinear model. Since standard software for this latent class model is readily available, a positive result would benefit researchers, because the estimation of nonlinear models requires some programming effort.

In Section 2 we give an overview on the evidence of nonlinear effects contained in empirical studies. In Section 3 we motivate the functional forms considered in the empirical study and introduce the reader to linear splines and bivariate tensor products that constitute the mathematical concepts we use to represent piecewise linear utility functions. In Section 4 we provide an overview of the basic form of the choice models we study, their latent class extensions, and an explanation of predictors. In Section 5 we discuss estimation, specification, and evaluation of models. In Section 6 we provide descriptive statistics, estimation results, and a comparison of the effects implied by the models studied.

2 Evidence of Nonlinear Effects

Several empirical and theoretical studies give evidence that predictors such as price, reference price, and brand loyalty have nonlinear effects on brand choice, or, more precisely, on deterministic utility of a brand (see Section 4).

In a cognitive study, Monroe (1973) obtains threshold effects for prices, i.e. utility is affected only if price changes by at least a certain amount. An experimental study by Gupta and Cooper (1992) finds both threshold and saturation effects. Saturation means that price changes have very little or no effect on the behavior of consumers after price passes a certain vaue.

Winer (1988) defines reference prices as internal prices to which households compare observed prices. Reference prices reflect the expected price level of a brand. Therefore, high reference prices are associated with lower choice probabilities. Observed prices below the reference price, which households perceive as gains, stimulate purchases, i.e. increase choice probability. Observed prices above the reference price, which households perceive as losses, may deter buyers from purchasing and therefore decrease choice probability. Prospect theory predicts asymmetric effects with consumers responding more to losses than to gains (Kahnemann and Tversky (1979); Winer (1986); Winer (1988)). Assimilation-contrast theory postulates if prices still lie in the range of acceptable (as a rule medium) prices covering the reference price, then changes have no effect (Winer (1988)).

Kalyanaram and Little (1994) postulate such a medium price-insensitive range and effects that obey prospect theory. Their estimation results corroborate assimilation-contrast theory, but not prospect theory, because the asymmetric effects of gains and losses are not significant.

By using a generalized additive version of the multinomial logit model, Abe (1998; 1999) finds three different price regions. These regions support both assimilation-contrast theory and prospect theory. Price changes in the medium price range have the smallest effect, and prices above the reference price have more effect than do prices below the reference price. Utility increases progressively with brand loyalty, which, following Guadagni and Little (1983), is measured by exponentially smoothing past purchases Y_{ijt_i} (one if brand j is purchased, else zero) of brand j for each household i at purchase occasion $t_i = 1, 2, \ldots$ using smoothing constant α which lies in the interval [0, 1]:

$$LOY_{ijt_i} = \alpha LOY_{ijt_i} - 1 + (1 - \alpha)\ Y_{ijt_i - 1} \tag{1}$$

On the contrary, the model estimated in Abe et al. (2000) determines a weakly nonlinear effect of loyalty and a linear effect of price.

An application of the semiparametric model of Briesch (1996) results in almost linear effects of both price and loyalty. On the other hand, utility is shown to be concave w.r.t. price reduction. Marginal effects decrease with the amount of price reduction.

Hruschka et al. (2002) develop a neural-net enhanced multinomial logit model and find several nonlinear effects in their empirical study. Utility decreases with higher reference prices, and in most cases takes an inverse S-shape. As a rule, utility increases degressively with brand loyalty. Utility levels are higher if loyalty is high or sales promotion is on, with the effect of loyalty being more pronounced.

On the basis of these empirical studies we can expect the following kinds of nonlinear effects:

- an inverse S-shape for reference prices;
- threshold effects for very low prices;
- very small effects for medium prices in the medium price range;
- a concave shape for brand loyalty;
- stronger effects for prices above reference price as opposed to prices below reference price.

3 Functional Forms of Deterministic Utilities

As discussed in Section 2, the empirical evidence suggests that we extend choice models so that they are able to reproduce nonlinear effects. We can achieve this extension by specifying deterministic utility as parametric function. Appropriate parametric functions either consist of higher-order polynomial terms (i.e., quadratic, cubic), fixed transformations (e.g., semilog), or piecewise linear functions of predictors.

Of these alternatives we consider functions with higher-order terms and piecewise linear functions, because both are able to reproduce the nonlinear effects that earlier empirical studies have demonstrated, i.e., (inverse) S-shapes, threshold effects, saturation effects, and very small effects in the medium range of a predictor. We do not study concave transformations, such as semilog, because they prohibit such effects.

Linear piecewise functions constitute another alternative way to reproduce such nonlinear effects. So far, piecewise linear utility functions contained in brand choice models are univariate, i.e. they relate to one predictor only (Kalyanaram and Little (1994); Wedel and Leeflang (1998)). The model we introduce here allows interactions of predictors and can be seen as multivariate generalization of the univariate approach. Therefore, it is more powerful than the generalized additive model of Abe (1999) and similar to the semiparametric model of Briesch et al. (1996) and the neural net approach of Bentz and Merunka (2000) or Hruschka et al. (2002). Of course, this model is restricted to piecewise linear functions, but this limitation may be outweighed by the fact that both estimation and interpretation turn out to be straightforward.

These piecewise functions should obey continuity restrictions for different sections of the total value range of a predictor (i.e. the linear pieces must be connected). Mathematically, linear splines and bivariate tensor products can represente linear piecewise functions with these properties. This representation turns out to be advantageous for estimation. Conventional (log) linear estimation procedures can be used after the value range of a predictor is divided into at least two sections by fixing one or more knot positions.

Linear splines are defined only for metric predictors. The linear spline $S(x)$ of a metric predictor x consists of piecewise linear functions with knots $x_0^{(\kappa)}$, ..., $x_K^{(\kappa)}$ (Schumaker (1981); Nürnberger (1989)):

$$S(x) = \begin{cases} s_1(x) = a_0^{(1)} + a_1^{(1)}x & \text{if } x \in [x_0^{(\kappa)}; x_1^{(\kappa)}] \\ \vdots & \vdots \\ s_K(x) = a_0^{(K)} + a_1^{(K)}x & \text{if } x \in [x_{K-1}^{(\kappa)}; x_K^{(\kappa)}] \end{cases} \tag{2}$$

The linear spline may also be written as:

$$S(x) = \beta_0 + \beta_1 x + \sum_{k=1}^{K-1} \beta_{k+1}(x - x_K^{(\kappa)})_+ \tag{3}$$

with the truncated identity function for $k = 1, ..., K - 1$:

$$(x - x_K^{(\kappa)})_+ = \begin{cases} (x - x_K^{(\kappa)}) & \text{if } x > x_K^{(\kappa)} \\ 0 & \text{if } x \leq x_K^{(\kappa)} \end{cases} \tag{4}$$

A linear spline may also be expressed as linear combination of $(K + 1)$ basis functions $B^{(k)}$:

$$S(x) = \sum_{k=0}^{K} \beta_k B^{(k)} \text{ with } x \in [x_0^{(\kappa)}; x_K^{(\kappa)}] \tag{5}$$

The basis functions are:

$$\begin{aligned} B^{(0)} &= 1 \\ B^{(1)} &= x \\ B^{(2)} &= (x - x_1^{(\kappa)})_+ \\ &\vdots \\ B^{(K)} &= (x - x_{K-1}^{(\kappa)})_+ \end{aligned}$$

We note that these basis functions also include the constant term $B^{(0)}$ and the linear term of the predictor $B^{(1)}$.

The bivariate tensor product of splines for two predictors $x_1,\ x_2$ with one knot each (i.e. $x_{11}^{(\kappa)}$and $x_{21}^{(\kappa)}$, respectively) consists entirely of basis functions (Prenter (1975); Greub (1978); Schumaker (1981)):

$$\begin{aligned} S(x_1,\ x_2) = {} & \beta_{00} + \beta_{10}x_1 + \beta_{01}x_2 + \beta_{11}x_1x_2 + \\ & \beta_{20}(x_1 - x_{11}^{(\kappa)})_+ + \beta_{02}(x_2 - x_{21}^{(\kappa)})_+ + \beta_{21}(x_1 - x_{11}^{(\kappa)})_+ x_2 + \\ & \beta_{12}x_1(x_2 - x_{21}^{(\kappa)})_+ + \beta_{22}(x_1 - x_{11}^{(\kappa)})_+(x_2 - x_{21}^{(\kappa)})_+ \end{aligned} \tag{6}$$

As we can see from this expression linear terms and one spline for each predictor as well as the product of the linear terms of the predictors (i.e. the conventional bivariate interaction effect known from linear models) form part of the bivariate tensor product. Extensions to splines with more knots are obvious (for statistical applications of splines to regression problems see Friedman and Silverman (1989); Friedman (1991); Stone et al. (1997)).

A model with linear splines can reproduce both threshold and saturation effects, both of which imply that over a certain value range, the effect of the respective predictor is constant. In the case of a saturation effect, the coefficient for this value range must compensate the sum of coefficients of preceding value ranges (given by the linear part of the model and preceding knots). In the case of a threshold effect, the model consists of only a constant term (the coefficient for this value range equals zero), and coefficients for values of the predictor greater than the threshold are different from zero. Moreover, because it allows different marginal effects of any predictor for different value ranges, the model with linear splines can also reproduce asymmetric effects.

4 Choice Models

We investigate brand choice models because the empirical studies discussed in Section 2 give evidence on nonlinear effects on brand choice. The models we use are all variants of the multinomial logit (MNL) model which is the most widespread choice model in marketing. The basic assumptions of the MNL model are: (1)buyers select out of a choice set the brand that they perceive to have maximum utility, (2) utility is conceived to be additively made up by a deterministic component and a random term, and (3) the random term is iid type I extreme value distributed (McFadden (1973)).

If we suppress indices of purchase occasions to keep notation simple, then we can write the conditional choice (purchase) probability p_{ij} of brand j for household i according to a MNL model as (McFadden (1973)):

$$p_{ij} = \frac{\exp(V_{ij})}{\Sigma_{j'} \exp{(V_{ij'})}} \tag{7}$$

V_{ij} denotes deterministic utility of brand j for household i. The denominator in expression 7 sums over all brands available at the respective purchase occasion of household i.

In the piecewise linear model deterministic utility is a linear combination of basis functions. It can be decomposed into main effects and pairwise interaction effects:

$$\begin{aligned} V_{ij} &= \underbrace{\beta_{j0} + \sum_{n=1}^{N} f_n(x_n)}_{\text{main effects}} + \underbrace{\sum_{n \neq 0} f_{nl}(x_n,x_l)}_{\text{pairwise interaction effects}} \\ \text{with} \quad f_n(x_n) &= \sum_{p_n=1}^{P_n} \beta_{p_n} B^{(p_n)}(x_n) \\ f_{nl}(x_n,\ x_l) &= \sum_{p_{nl}=1}^{P_n} \beta_{p_{nl}} B^{(p_{nl})}(x_n,x_l) \end{aligned} \tag{8}$$

P_n number of univariate basis functions of predictor n

P_{nl} number of bivariate basis function of predictors n and l

The choice models we study here have the following predictors (abbreviated terms appear in parentheses):

- brand specific constants;
- reference price (RPRICE);
- for models with nonlinear utility, we define the price deviation (PDEV) of observed price (PRICE) from the reference price as RPRICE – PRICE, so that a positive (negative) price deviation indicates a gain (loss);
- for models with linear utility, we replace price deviation with both loss defined as max(PRICE – RPRICE,0) and gain defined as max(RPRICE – PRICE,0);
- brand loyalty (LOY) following Guadagni-Little (1983);
- display (D), i.e. POS advertising (binary);
- feature (F), i.e. local newspaper ads or flyers (binary).

With the exception of brand constants, feature, and display these predictors are metric.

By introducing separate predictors, we obtain asymmetric effects of gains and losses even in a linear utility model given that their coefficients differ. On the other hand, separate predictors are not necessary for models with nonlinear utility. These models can identify asymmetric effects of gains and losses, since they allow different marginal effects w.r.t. the price deviation variable, whereas linear utility models are characterized by constant marginal effects.

The reference price mechanism we study is of the extrapolative expectations type (Winer (1988)). Reference price depends on prices observed for the three previous purchases and a trend term in the following manner:

$$RPRICE_{ijt_i} = \gamma_0 + \sum_{l=1}^{3} \gamma_l PRICE_{ijt_i - l} + \gamma_4 TREND_{it_i} \tag{9}$$

As in most studies, this reference price mechanism is homogeneous, i.e. coefficients of equation 9 do not vary across households. In other words, as in almost all relevant studies (Winer (1986); Lattin (1989); Kalwani et al. (1990); Kalyanaram and Little (1994); Mayhew and Winer (1992); Mazumdar and Papatla (1995); Chang et al. (1999)) we allow for latent heterogeneity only in choice models, but parameters of the reference price model are equal across households. An advantageous consequence of this approach is the fact that the predictors reference price, price deviation, gain, and loss assume the same values for each household and purchase occasion, no matter which model is estimated. This property also applies to latent class models which differ w.r.t. the number of segments.

We do not consider observed price as predictor of brand choice. Observed price is equal to the difference of reference price and price deviation, and price deviation is equal to the difference of gain and loss. Therefore, observed price is collinear if either price deviation or gain and loss are predictors. On the other hand, reference price is uncorrelated with price deviation, if determined on the basis of a model estimated by OLS (Chow (1985)).

Thus, we replace observed price with reference price. This collinearity problem shows up in the empirical data used in this study. Regressing observed price and reference price against the remaining predictors reproduces 53.18% and 10.18% of variance, respectively.

All the choice models we consider have brand constants and coefficients for feature and display. We compare the piecewise linear utility model with interactions to the following less complex models. These models do not include interactions and therefore have a smaller number of parameters:

- a conventional linear utility function that consists of linear terms for each remaining predictor;
- a quadratic utility function that consists of linear and quadratic terms for each remaining predictor;
- a cubic utility function that consists of linear, quadratic, and cubic terms for each remaining predictor;
- the piecewise linear utility function with main effects only, i.e., without pairwise interaction effects.

To answer the main research question of our paper we need latent class extensions of the basic choice models described above (the latent class extension of the linear utility model was introduced by Kamakura and Russel (1989)). Latent class models conceive the population of households as finite mixture of classes, which can be interpreted as market segments (Wedel and Kamakura (1998)).

In latent class extensions of MNL models, we follow Kamakura and Russel (1989), who specify class-specific choice probabilities p_{cij} of household i for brand j as:

$$p_{cij} = \frac{\exp(V_{cij})}{\Sigma_{j'} \exp(V_{cij'})} \tag{10}$$

Conditional choice probabilities for latent class MNL models follow from multiplying these class-specific choice probabilities with posterior probabilities π_c and summing across C classes:

$$p_{ij} = \sum_{c=1}^{C} \pi_c \, p_{cij} \tag{11}$$

The capability of the latent class model with linear utility to reproduce nonlinear effects is due to the fact that marginal efects on utility differ across classes, although they are constant within each class.

Contrary to homogeneous MNL models, coefficients of latent class MNL models vary across classes. The number of parameters, P, of latent class models is equal to the number of classes, C, times the number of coefficients of any class-specific MNL model, P_c, plus the number of free a posteriori class probabilities:

$$P = CP_c + (C - 1) \tag{12}$$

5 Estimation, Specification, and Evaluation

We estimate parameters of choice models by maximizing the log likelihood across purchases. To this end, we use the BFGS algorithm (Seber and Wild (1989)) for homogeneous choice models. For latent class choice models we use an EM algorithm (McLachlan and Krishnan (1997); Wedel and Kamakura (1998)).

For C classes the likelihood function can be written as (Wedel and Kamakura (1998)):

$$L = \prod_{i=1}^{I} \sum_{c=1}^{C} \pi_c \prod_{T=1}^{T_i}{}_c \prod_{j=1}^{J} {}^{Y_{ijt}}_{p_{cijt}} \tag{13}$$

I denotes the number of households, C the number of classes, π_c the a posteriori probability of class c. T_i is the number of purchases of household i and J is the number of brands. Y_{ijt} is the purchase indicator of household i of brand j at occasion t (equal to one if household i purchases brand j at occasion t, else zero).

For homogeneous choice models, we can simplify the likelihood function by setting $C = 1$, $\pi_1 = 1$ and $p_{1ijt} = p_{ijt}$ in equation 13:

$$L = \prod_{i=1}^{I} \prod_{t=1}^{T_i}{}_c \prod_{j=1}^{J} {}^{Y_{ijt}}_{p_{ijt}} \tag{14}$$

The piecewise linear utility function, which always contains brand constants, is automatically specified by using three types of basis functions: each predictor, truncated identity functions of each metric predictor with up to seven knot positions (determined by appropriate quantiles), and bivariate tensor products of predictors or truncated identity functions.

We generate different models by making a specification search consisting of stepwise additions of selectable basis functions followed by stepwise eliminations as described in Kooperberg et al. (1995) and Kooperberg et al. (1997). From the set of all the models generated, we select the model attaining the best BIC value (see equation 15).

As the piecewise linear utility function is linear in parameters, the log likelihood function has only one global optimum given knot positions. We estimate the loyalty smoothing constant using the approach introduced by Fader et al. (1992) and we estimate parameters of the reference price by OLS.

We compare the following model types:

- a homogeneous MNL model with linear utility;
- a homogeneous MNL model with piecewise linear utility;
- a homogeneous MNL model with quadratic utility;
- a homogeneous MNL model with cubic utility;
- a latent class MNL model with linear utility;
- a latent class MNL model with piecewise linear utility;
- a latent class MNL model with quadratic utility;
- a latent class MNL model with cubic utility.

We base our comparisons of model types on both Bayesian Information Criterion (BIC) and cross-validated mean choice probability (CVMCP). Given log likelihood $ln(L)$, number of parameters P, and total number of purchases T, we follow Schwarz (1979) and compute BIC as:

$$BIC = -2\ ln(L) + P\ ln(T) \tag{15}$$

We evaluate the out-of-sample performance of models by tenfold cross-validation (Stone (1974), Wahba and Wold (1975)). We randomly divide the whole data set into ten disjunctive subsets, each with the same number of households. Each of these ten subsets serves once as validation data set, for which we compute choice probabilities based on the model that we estimate by using the remaining nine subsets.

We define CVMCP as the geometric mean of cross-validated choice probabilities across all purchases for the brand actually chosen. Obviously, a model with higher CVMCP is preferable to a model with lower CVMCP, since the model with higher CVMCP predicts better. In our opinion, mean choice probability is easier to interpret than is the log likelihood to which it is related. We compute the mean choice probability of a model by:

$$\left(\prod_{i=1}^{I} \prod_{t=1}^{T_i}{}_{c} \prod_{j=1}^{J} p_{ijt}^{Y_{ijt}} \right)^{1/\Sigma_{i=1}^{I} T_i} \tag{16}$$

We use CVMCP as our main evaluation criterion, because we consider out-of-sample performance to be more important than in-sample performance, especially for more complex models, which may suffer from overfitting problems.

6 Empirical Study

6.1 Data

Table 1: Descriptive Statistics

Brand	Number of Purchases	Mean	Price Standard Dev.	Min	Max	Percentage with Display	Feature
1	4422	$1.65	$0.25	$0.99	$2.90	9.93	25.76
2	2388	$1.71	$0.19	$1.07	$4.73	7.16	24.58
3	2666	$1.76	$0.17	$1.14	$2.29	2.55	7.39
4	2615	$1.66	$0.24	$0.99	$5.21	9.79	24.36
5	1902	$1.70	$0.21	$0.99	$5.90	8.83	28.02
6	1335	$1.76	$0.19	$1.14	$5.61	4.12	9.74
Aggregate	15328	$1.71	$0.22	$0.99	$5.90	7.55	21.03

Table 1 contains descriptive statistics on purchases, prices (US dollars) and shares of purchases with display and feature for a database that consists of purchases acquired from a household scanner panel. The data refer to the six largest brands in terms of market share and the most frequently purchased package size across all brands in the product category of peanut butter. We select 960 households with at least ten purchases across 123 weeks. We use 30 weeks for the initialization of brand loyalty and 93 weeks for estimation.

We choose this database because of high variation of prices across purchases which can be seen from the columns for minimum and maximum prices of *table 1*. A high variation of prices makes nonlinear effects detectable. On the other hand, a low variation of prices suppresses nonlinear effects and leads to better performance of linear models.

6.2 Estimation Results

Table 2 presents the coefficients of the predictors of the extrapolative reference price model that we estimate by OLS. The positive effect of lagged prices decreases in time, with price at the most recent purchase being most important. The positive coefficient of trend indicates that reference prices increase with time. In addition to the predictors already mentioned in Section 4, we add a binary dummy variable that we set to one after the 97th week. We do so to reproduce a general increase of price level of the product category analyzed.

Table 2: Reference Price Model

	Coefficient
Constant	8.262
PRICE$_{t-1}$	0.269
PRICE$_{t-2}$	0.056
PRICE$_{t-3}$	0.067
TREND	0.490
WEEK > 97	0.893

Because for each utility function, models with both gains and losses among the predictors perform worse w.r.t. both BIC and CVMCP, here we present only results for choice models with price deviation. Grid searches over the value range of the smoothing constant in the loyalty equation of Guadagni and Little (1983) demonstrate that it has a unique global maximum w.r.t. log likelihood values. Metric predictors and their quadratics or cubics are z-transformed to achieve better-conditioned estimation matrices.

Table 3: Evaluation of Homogeneous Choice Models

Q	(Number of) Selected Knots: Reference Price	Price Deviation	Loyalty	P	BIC	CVMCP
Linear						
				10	26523.62	0.3243
Piecewise Linear without Interactions						
1	0	0	1	11	26225.66	0.3305
2	0	2	2	14	26140.31	0.3320
3	0	2 (2/3)	2 (1/3)	14	26098.67	0.3326
4	0	2 (2/4)	3 (1/3/4)	15	26072.07	0.3331
5	0	2 (3/5)	3 (1/4/5)	15	26056.05	0.3333
6	0	2 (3/6)	3 (1/5/6)	15	26054.31	0.3333
7	0	2 (3/7)	2 (1/6)	14	26061.23	0.3331
Piecewise Linear with Interactions						
1	0	1	1	15	26204.28	0.3331
2	0	2	2	15	26115.30	0.3324
3	0	2 (2/3)	2 (1/3)	18	26060.45	0.3334
4	0	2 (2/4)	3 (1/3/4)	19	26033.41	0.3340
5	0	2 (3/5)	3 (1/4/5)	17	26012.75	0.3341
6	0	2 (3/6)	3 (1/5/6)	17	26011.37	0.3341
7	0	2 (3/7)	4 (1/5/6/7)	18	26017.85	0.3341
Quadratic						
				13	26183.51	0.3313
Cubic						
				16	26002.33	0.3341

3 (1/5/6) means that 3 knots, i.e. the first, fifth and sixth knot, are selected

Tables 3 and *4* contain estimation results for homogeneous and latent class models, respectively. As mentioned in Section 5, CVMCP serves as our main criterion for evaluating models. In terms of CVMCP, the two best nonlinear homogeneous choice models are the piecewise linear utility model with interactions which starts from seen quantiles for each metric predictor (note that differences to variants with fewer nodes are very small), and the cubic utility model. In the following, we consider only these two nonlinear homogeneous choice models. We also estimate latent class extensions of the linear, piecewise linear with interactions, and the cubic utility models. For each of these three models, we set the number of classes to the value at which CVMCP attains its maximum.

Table 4: Evaluation of Latent Class Choice Models

Utility Function	Number of Classes	P	BIC	CVMCP
Linear	5	54	25546.06	0.3386
Piecewise Linear	3	53	25746.35	0.3341
Quadratic	3	41	25651.82	0.3414
Cubic	3	50	25725.54	0.3415

We take the latent class model with linear utility as standard of comparison. The homogeneous MNL model with linear utility is clearly worse. Its CVMCP is lower by 4.21% than the CVMCP of the latent class model with linear utility. The CVMCPs of the homogeneous models with linear piecewise and cubic utility are lower by 1.31% and 1.33%, respectively. Therefore the latent class model with linear utility is clearly superior to the two homogeneous nonlinear models, although differences are not large. This result also demonstrates that the two homogeneous nonlinear models attain similar performance levels, and that neither of these two models can be preferred over the other.

The latent class model with linear piecewise utility performs worse than the latent class model with linear utility. Compared to the latter model, the latent class models with quadratic and cubic utility increase CVMCP by only 0.83% and 0.86%, respectively. In view of these results we ignore these more complex latent class models in the following and restrict our attention to the latent class model with linear utility.

6.3 Comparison of Effects Implied by Models

Because of the superior performance of the latent class MNL model with segment-specific linear utility functions shown in subsection 6.2, we can now adress the main research question of whether effects implied by the latent class choice model are similar to those implied by parametric homogeneous nonlinear models.

To measure the similarity of effects between two models, we compute absolute differences of choice probabilities implied by these models. We call effects of a predictor implied by models A and B to be more similar than effects implied by models A and C if absolute differences of choice probabilities between models A and B are smaller than absolute differences of choice probabilities between models A and C.

We compare absolute differences of choice probabilities between each of the two nonlinear models and the linear utility model, and between each of the two nonlinear models and the latent class model.

If the absolute differences of choice probabilities between each nonlinear model and the latent class model are smaller than are absolute differences between each nonlinear model and the linear utility model, then we say that the latent class choice model implies effects that are similar to those of parametric homogeneous nonlinear models.

We consider the effects of each metric predictor for which the piecewise linear utility model indicates nonlinear effects. As explained in section 3, nonlinearity is achieved by one or several knots that divide the value range of a predictor into different sections. This property applies to the predictors price deviation and loyalty (see *table 3*).

We compute the choice probabilities of one brand for each of the models considered by setting values of predictors as follows:

- We vary the price deviation (loyalty) of the respective brand over a grid of 201 equidistant values. We set the other predictor with nonlinear effects to its brand-specific median value.
- The reference price, price deviation and loyalty of the remaining brands assume their brand-specific median values.
- We set feature and display to one for all brands.

For homogeneous choice models, we obtain the conditional choice probability p_{ij} by computing deterministic utilities for each brand on the basis of predictor values and inserting deterministic utilities into equation 7.

In the case of the latent class model with linear utility, we obtain class-specific deterministic utilities for each brand on the basis of predictor values. They lead to class-specific choice probabilities p_{cijt} as given by equation 10. We determine the (overall) conditional choice probability of brand j by inserting class-specific choice probabilities and posterior class probabilities into equation 11.

The arithmetic means of absolute differences of choice probabilities for both nonlinear models and the latent class model are clearly smaller than are the analogous values for both nonlinear models and the homogeneous linear utility model, for both varying price deviation and for varying loyalty (see *table 5*). These arithmetic means also indicate that there is more agreement between the latent class and the cubic utility model. Paired sample t-tests are highly significant. These results demonstrate that the latent class model implies effects that resemble those of the two nonlinear homogeneous choice models.

To provide more details on similarities between the latent class model and homogeneous nonlinear models we plot logits as functon of price deviation and loyalty, respectively. A logit (also called log odd) can be interpreted as log of the chance of a choice of brand 1

Table 5: Absolute Differences of Choice Probabilities between Model Pairs Piecewise Linear Piecewise Linear and Linear and Latent Class

	Piecewise Linear and Linear		Piecewise Linear and Latent Class
for varying price deviation			
Mean	0.0975		0.0454
Variance	0.0021		0.0018
Paired samples *t*-test		35.75	
for varying loyalty			
Mean	0.1307		0.0616
Variance	0.0149		0.0024
Paired samples *t*-test		12.34	
	Cubic and Cubic and Linear		**Cubic and Latent Class**
for varying price deviation			
Mean	0.0899		0.0375
Variance	0.0001		0.0001
Paired samples *t*-test		35.48	
for varying loyalty			
Mean	0.1100		0.0445
Variance	0.0098		0.0011
Paired samples *t*-test		11.63	

by household i relative to choices of the other brands. It is defined as transformation of the conditional choice probability p_{i1} (Greene (2003)):

$$logit_{i1} = ln\ p_{i1} - ln\,(1 - p_{i1}) \tag{17}$$

We compute conditional choice probabilities of brand 1 w.r.t. both price devation and loyalty, which are necessary to determine logits, in the same way as explained at the beginning of this subsection.

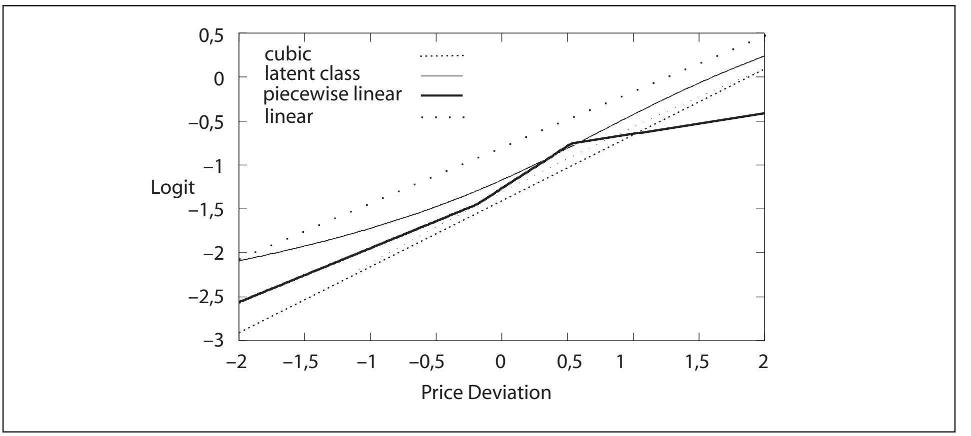
Starting from equation 7 we show that logits of the choice of brand 1 for homogeneous models can be computed as follows :

$$logit_{i1} = V_{i1} - ln\Big(\sum_{j \neq 1} \exp V_{ij} \Big) \quad (18)$$

Therefore, logits for the homogeneous model with linear utility depend linearly on each predictor of the respective brand. But for the other models studied here, the dependence of logits on predictors is nonlinear. Therefore, logits provide insight into nonlinear effects implied by the latent class model, which a comparison of linear class-specific utilities and utilities implied by homogeneous nonlinear models does not give.

We show one plot for price deviation and loyalty, setting the other predictors to median values. Plots, which are based on the other quartiles of the remaining predictors, have similar shapes. *Figure 1* shows logits as a function of price deviation we obtain for four different choice models. For all models, logits increase as price deviation becomes greater. Logits for the latent class model with linear utility develop in a nonlinear way. For price devations in the interval [–0.80, 1.16], logits of the latent class model differ more from logits of the linear utility model than from logits of models with piecewise linear or cubic utility. For price deviations in the interval [–1.26, 1.16], logits of the latent class model are more like those of the piecewise linear utility model. For price deviations greater than 1.10, logits of the latent class model are more like those of the cubic utility model. Only for price deviations lower than –1.26 do logits of the latent class model differ less from those of the linear utility model compared to the two nonlinear models.

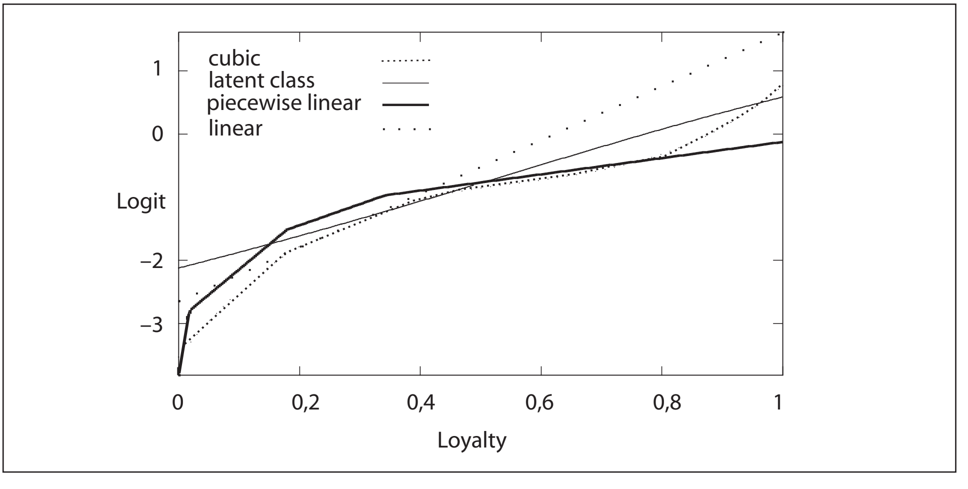
Figure 1: Logits vs. price deviation for different MNL models



For the cubic utility model, the marginal effects (i.e., first derivatives) of price deviations stay pretty much the same across their observed value range, but change for both the piecewise linear model and the latent class model. Therefore, the latter two models indicate asymmetric effects, but these asymmetric effects differ across these two models. The piecewise linear model shows lower marginal effects at higher positive price deviations (i.e., at higher gains), which explains why the piecewise linear model is more in accordance with prospect theory. On the other hand, marginal effects of price deviation increase for the latent class model and therefore imply higher marginal effects of positive price deviations (i.e. gains) compared to those of negative price deviations (i.e. losses). Neither model provides evidence for the medium price insensitive range as postulated by assimilation-contrast theory.

Figure 2 shows logits as function of loyalty according to the four different choice models studied. For all models, logits increase as loyalty becomes greater. For loyalties greater than 0.40, logits of the latent class model differ more from logits of the linear utility model than from logits of models with piecewise linear and cubic utility, respectively. If loyalties lie in the interval [0.11, 0.21], then logits of the latent class model are similar to those of the cubic utility model and differ from those of the linear utility model. The marginal effects are almost constant for the latent class model. For models with cubic utility the marginal effects decrease for loyalties in the intervals [0.15, 0.44]. For models with piecewise linear utility marginal effects decrease for loyalties greater equal 0.03.

Figure 2: Logits vs. loyalty for different MNL models



Figures 1 and *2* indicate that the cubic utility, piecewise linear utility and latent class models do not completely agree on the effects of the predictors (especially w.r.t. marginal effects). However, these figures show that the effects implied by the two homogeneous

nonlinear models are similar to the effects implied by the latent class model. This result is in accordance with the comparisons of absolute differences of choice probabilities between the model pairs discussed ealier. The conclusion we draw is that the latent class choice model implies effects that are similar to those of homogeneous nonlinear models. This result leads us to recommend the use of the latent class model instead of a parametric homogeneous nonlinear model.

7 Conclusions

Earlier empirical studies show that brand choices of individual households are subject to nonlinear effects for predictors such as price, reference price, and brand loyalty. Nevertheless, most brand choice models found in the literature are based on linear deterministic utility functions and therefore may result in biased estimates.

Alternative approaches to adress nonlinearity are either parametric homogeneous nonlinear functions that do not vary across hoseholds, or latent class models with class-specific linear utility functions. We deal with the question of whether the latent class choice model with segment-specific linear utility functions implies effets that are similar to those of parametric homogeneous nonlinear models, given that the latent class model performs at least as well.

For the data set we analyze, the latent class model with linear utility is clearly superior to the two homogeneous nonlinear models (with piecewise linear and cubic utility, respectively), although differences are not large. Therefore, we are justified in investigating whether the effects of predictors implied by this latent class model function resemble the effects implied by the nonlinear models.

On the basis of varying values of each of two predictors, price deviation and loyalty, the latent class model turns out to be similar to each of the two nonlinear models in terms of choice probabilities. This conclusion can be also be drawn by looking at logits of choice probabilities as functions of each of these two predictors. Overall, our results confirm that the the latent class model implies effects that are similar to those of the two nonlinear models considered.

Our results suggest that researchers should use the latent class model instead of a parametric homogeneous nonlinear model. Unlike nonlinear models, standard software for this latent class model is readily available and thus constitutes an additional benefit for researchers.

References

Abe, Makoto (1998), Measuring Consumer, Nonlinear Brand Choice Response to Price, *Journal of Retailing* 74, 541-568.

Abe, Makoto (1999), A Generalized Additive Model for Discrete Choice Data, *Journal of Business & Economic Statistics* 17, 271-284.

Abe, Makoto, Yasemin Boztuğ, and Lutz Hildebrandt (2000), *Investigation of the Stochastic Utility Maximization Process of Consumer Brand Choice by Semiparametric Modeling*, Discussion Paper, Humboldt-Universität zu Berlin.

Allenby, Greg M. and James L. Ginter (1995), The Effects of In-Store Display and Feature Advertising on Consideration Sets, *International Journal of Research in Marketing* 12, 67-80.

Allenby, Greg M. and Peter J. Lenk (1994), Modeling Household Purchase Behavior with Logistic Normal Regression, *Journal of the American Statistal Association* 89, 1218-1231.

Amemiya, Takeshi (1985), *Advanced Econometrics*, Oxford, UK: Basil Blackwell.

Bentz, Yves and Dwight Merunka (2000), Neural Networks and the Multinomial Logit for Brand Choice Modelling: A Hybrid Approach, *Journal of Forecasting* 19, 177-200.

Briesch, Richard A., Pradeep Chintagunta, and Rosa L. Matzkin (1996), *A Nonparametric Investigation into the Differential Consumer Response to Price and Deal Discounts*, Working Paper, Stern School of Business, New York University, New York.

Chang, Kwangpil, S. Siddarth, and Charles B. Weinberg, (1999), The Impact of Heterogeneity in Purchase Timing and Price Responsiveness on Estimates of Sticker Shock Effects, *Marketing Science* 18, 178-192.

Chintagunta, Pradeep K. (1992), Estimating a Multinomial Probit Model of Brand Choice Using the Method of Simulated Moments, *Marketing Science* 11, 386-407.

Chow, Gregory C. (1985), *Econometrics*, Auckland: McGraw Hill.

Fader, Peter, James M. Lattin, and John D.C. Little (1992), Estimating Nonlinear Parameters in the Multinomial Logit Model, *Marketing Science* 11, 372-385.

Friedman, Jerome H. and Bernard W. Silverman (1989), Flexible Parsimonious Smoothing and Additive Modeling, *Techonometrics* 31, 3-39.

Friedman, Jerome H. (1991), Multivariate Adaptive Regression Splines, *The Annals of Statistics* 19, 1-141.

Gensch, Dennis and W.W. Recker (1979), The Multinomial Multiattribute Logit Choice Model, *Journal of Marketing Research* 16, 124-132.

Greene, William H. (2003), *Econometric Analysis*, 5th Edition, Upper Saddle River, NJ: Prentice Hall.

Greub, Werner (1978) *Multilinear Algebra,* 2nd Edition, New York: Springer.

Guadagni, Peter M. and John D.C. Little (1983), A Logit Model of Brand Choice Calibrated on Scanner Data, *Marketing Science* 2, 203-238.

Gupta, Sunil (1988), The Impact of Sales Promotions on When, What, and How Much to Buy, *Journal of Marketing Research* 25, 342-356.

Gupta, Sunil and Lee G. Cooper (1992), The Discounting of Discounts and Promotion Thresholds, *Journal of Consumer Research* 19, 401-411.

Hruschka, Harald, Werner Fettes, Markus Probst, and Christian Mies, (2002), A Flexible Brand Choice Model Based On Neural Net Methodology. A Comparison to the Linear Utility Multinomial Logit Model and Its Latent Class Extension, *OR Spectrum* 24, 127-143.

Kahneman, Daniel and Amos Tversky (1979), An Analysis of Decision Under Risk, *Econometrica* 47, 363-391.

Kalwani, Manohar O., Chi K. Yim, Heikki J. Rinne, and Yoshi Sugita (1990), A Price Expectations Model of Customer Brand Choice, *Journal of Marketing Research* 27, 251-262.

Kalyanaram, Gurumurthy and John D.C. Little (1994), An Empirical Analysis of Latitude of Price Acceptance in Consumer Package Goods, *Journal of Consumer Research* 21, 408-418.

Kamakura, Wagner A. and Gary J. Russel (1989), A Probabilistic Choice Model for Market Segmentation and Elasticity Structure, *Journal of Marketing Research* 26, 379-390.

Kooperberg, Charles, Smarajit Bose, and Charles J. Stone (1997), Polychotomous Regression, *Journal of the American Statistical Association* 92, 117-127.

Kooperberg, Charles, Charles J. Stone, and Young K. Truong (1995), Hazard Regression, *Journal of the American Statistical Association* 90, 78-94.

Krishnamurthi, Lakshman and S.P. Raj (1992), A Model of Brand Choice and Purchase Quantity-Price Sensitivities, *Marketing Science* 7, 1-20.

Lattin, James M. and Randolph E. Bucklin (1989), Reference Effects of Price and Promotion on Brand Choice Behavior, *Journal of Marketing Research* 26, 299-310.

Mayhew, Glenn E. and Russel S. Winer (1992), An Empirical Analysis of Internal and External Reference Prices Using Scanner Data, *Journal of Consumer Research* 19, 62-70.

Mazumdar, Tridib and Purushottam Papatla (1995), Loyalty Differences in the Use of Internal and External Reference Prices*, Marketing Letters* 6, 111-122.

McCulloch, Robert and Peter E. Rossi (1994), An Exact Likelihood Analysis of the Multinomial Probit Model, *Journal of Econometrics* 64, 207-240.

McFadden, Daniel (1973), Conditional Logit Analysis of Qualitative Choice Behavior, in Zarembka, Paul (ed.), *Frontiers in Econometrics*, New York: Academic Press, 105-142.

McLachlan, Geoffrey J. and Thriyambakam Krishnan (1997), *The EM Algorithm and Extensions*, New York: Wiley.

Monroe, Kent B. (1973), Buyers' Subjective Perceptions of Price, *Journal of Marketing Research* 10, 70-80.

Nürnberger, Günther (1989), *Approximation by Spline Functions*, Berlin: Springer.

Pedryck, James H. and Fred S. Zufryden (1991), Evaluating the Impact of Advertising Media Plans: A Model of Consumer Purchase Dynamics Using SingleSource Data, Marketing Science 10, 111-130.

Prenter, P. M. (1975), *Splines and Variational Methods*, New York: Wiley.

Schumaker, Larry L. (1981), *Spline Functions*, New York: Wiley.

Schwarz, Gideon (1979), Estimating the Dimension of a Model, *Annals of Statistics* 6, 461-464.

Seber, George A.F. and C.J Wild (1989), *Nonlinear Regression*, New York: Wiley

Stone, M. (1974), Cross-Validatory Choice and Assessment of Statistical Predictions, *Journal of the Royal Statistical Society* B 36, 111-147.

Stone, Charles J., Mark H. Hansen, Charles Kooperberg, and Young K. Truong (1997), Polynomial Splines and Their Tensor Products in Extended Linear Modeling, *The Annals of Statistics* 25, 1371-1470.

Wahba G. and S. Wold (1975), A Completely Automatic French Curve: Fitting Spline Functions by Cross-Validation, *Communications in Statistics* A 4, 1-17.

Wedel, Michel and Wagner A. Kamakura (1998). *Market Segmentation, Conceptual and Methodological Foundations*, Dordrecht: Kluwer.

Wedel, Michel and Peter S.H. Leeflang (1998), A Model for the Effects of Psychological Pricing in Gabor-Granger Price Studies, *Journal of Economic Psychology* 19, 237-260.

Winer, Russel S. (1986), A Reference Price Model of Brand Choice for Frequently Purchased Products, *Journal of Consumer Research* 13, 250-256.

Winer, Russel S. (1988), Behavioral Perspective on Pricing, in Devinney, Timothey M. (ed.), *Issues in Pricing*, Lexington, MA: Lexington Books, 35-57.

Marion Schindler/Bernhard Baumgartner/Harald Hruschka*

Nonlinear Effects in Brand Choice Models: Comparing Heterogeneous Latent Class To Homogeneous Nonlinear Models**

Abstract

We investigate whether the latent class multinomial logit choice model with segment-specific linear utility functions implies effects that are similar to those of parametric homogeneous nonlinear models given that this latent class model performs at least as well. The two nonlinear models have higher-order polynomial (i.e. quadratic and cubic) and piecewise linear utility functions, respectively. Piecewise linear functions are represented by linear splines and can reproduce threshold, saturation and asymmetric effects. We evaluate models and their variants using a tenfold cross-validation. As criterion we use the geometric mean of choice probabilities across all purchases for the brand actually chosen. We measure the similarity of effects between two models by the absolute differences of choice probabilities implied by these models for varying values of a predictor. Logits of choice probabilities provide a more detailed insight into the effects implied by models. For the data set we analyze, the latent class model with linear utility is clearly superior to the two homogeneous nonlinear models. Overall, the effects implied by the latent class models are similar to those of the two parametric nonlinear models.

JEL-Classification: C35, M31.

Keywords: Brand Choice; Latent Class Models; Nonlinear Effects.

1 Introduction

In marketing, the majority of models used to explain the effects of predictors on the behavior of individual customers or households have linear or loglinear functional forms.

* Marion Schindler, Database Manager, Versandhaus Robert Klingel GmbH & Co KG, Sachsenstrasse 23, D-75177 Pforzheim, Bernhard Baumgartner, Associate Professor, Faculty of Economics, University of Regensburg, Universitätsstraße 31, D-93053 Regensburg, and Harald Hruschka (corresponding author), Chaired Professor of Marketing, Faculty of Economics, University of Regensburg, Universitätsstraße 31, D-93053 Regensburg, e-mail: harald.hruschka@wiwi.uni-regensburg.de.

** The authors thank two anonymous reviewers who suggested to take heterogeneity into account and to compare the piecewise linear model to models with quadratic and cubic utility functions. We also acknowledge that one reviewer suggested to focus on the similarity of effects implied by the latent class choice model to those of parametric homogeneous nonlinear models.

But it is well known that in practice these effects are often nonlinear. Several earlier empirical studies, which we discuss in Section 2, demonstrate that predictors such as price, reference price, and brand loyalty affect the brand choice of individual households in a nonlinear way. Nevertheless, most brand choice models found in the literature are based on linear deterministic utility functions, i.e., they specify utility as linear combination of predictors (Gensch and Recker (1979); Guadagni and Little (1983); Winer (1986); Gupta (1988); Lattin and Bucklin (1989); Kalwani et al. (1990); Chintagunta (1992); Allenby and Lenk (1994); McCulloch and Rossi (1994); Allenby and Ginter (1995)). However, these models cannot deal with nonlinear effects and result in biased estimates.

There are several alternative ways to deal with nonlinear effects. First, researchers can work with homogeneous parametric nonlinear functions that do not vary across households. The literature on brand choice models discusses different ways of specifying homogeneous parametric nonlinear utility functios. We find parametric nonlinear utility functions with higher-order polynomial terms (e.g., quadratic or cubic in Pedryck and Zufryden (1991)), fixed transformations (e.g., semilog in Krishnamurthi and Raj (1992)) or piecewise linear functions of predictors (e.g., Kalyanaram and Little (1994); Wedel and Leeflang (1998)).

Latent class models with class-specific linear utility functions constitute another alternative method for dealing with nonlinear effects. Latent class models are well known in the marketing literature, which interprets latent classes as market segments. So far, latent class models have mainly been used to adress the latent heterogeneity of households. To the best of our knowledge, an investigation of the capability of these models to deal with nonlinearity is still lacking in the marketing literature. The ability of latent class models to reproduce nonlinear effects is due to the fact that they determine choice probability as a convex combination of class-specific choice probabilities (see Section 4).

These circumstances lead to our research question: which of these alternative approaches performs better, and whether the one that performs better can capture the nonlinearities. Or, to put the question in another way, whether the latent class choice model with segment-specific linear utility functions implies effects that are similar to those of parametric homogeneous nonlinear models given that this latent class model performs at least as well.

If the answer to this question turns out to be positive, we can recommend that researchers use this latent class model instead of a parametric homogeneous nonlinear model. Since standard software for this latent class model is readily available, a positive result would benefit researchers, because the estimation of nonlinear models requires some programming effort.

In Section 2 we give an overview on the evidence of nonlinear effects contained in empirical studies. In Section 3 we motivate the functional forms considered in the empirical study and introduce the reader to linear splines and bivariate tensor products that constitute the mathematical concepts we use to represent piecewise linear utility functions. In Section 4 we provide an overview of the basic form of the choice models we study, their latent class extensions, and an explanation of predictors. In Section 5 we discuss estimation, specification, and evaluation of models. In Section 6 we provide descriptive statistics, estimation results, and a comparison of the effects implied by the models studied.

2 Evidence of Nonlinear Effects

Several empirical and theoretical studies give evidence that predictors such as price, reference price, and brand loyalty have nonlinear effects on brand choice, or, more precisely, on deterministic utility of a brand (see Section 4).

In a cognitive study, Monroe (1973) obtains threshold effects for prices, i.e. utility is affected only if price changes by at least a certain amount. An experimental study by Gupta and Cooper (1992) finds both threshold and saturation effects. Saturation means that price changes have very little or no effect on the behavior of consumers after price passes a certain vaue.

Winer (1988) defines reference prices as internal prices to which households compare observed prices. Reference prices reflect the expected price level of a brand. Therefore, high reference prices are associated with lower choice probabilities. Observed prices below the reference price, which households perceive as gains, stimulate purchases, i.e. increase choice probability. Observed prices above the reference price, which households perceive as losses, may deter buyers from purchasing and therefore decrease choice probability. Prospect theory predicts asymmetric effects with consumers responding more to losses than to gains (Kahnemann and Tversky (1979); Winer (1986); Winer (1988)). Assimilation-contrast theory postulates if prices still lie in the range of acceptable (as a rule medium) prices covering the reference price, then changes have no effect (Winer (1988)).

Kalyanaram and Little (1994) postulate such a medium price-insensitive range and effects that obey prospect theory. Their estimation results corroborate assimilation-contrast theory, but not prospect theory, because the asymmetric effects of gains and losses are not significant.

By using a generalized additive version of the multinomial logit model, Abe (1998; 1999) finds three different price regions. These regions support both assimilation-contrast theory and prospect theory. Price changes in the medium price range have the smallest effect, and prices above the reference price have more effect than do prices below the reference price. Utility increases progressively with brand loyalty, which, following Guadagni and Little (1983), is measured by exponentially smoothing past purchases Y_{ijt_i} (one if brand j is purchased, else zero) of brand j for each household i at purchase occasion $t_i = 1, 2, \ldots$ using smoothing constant α which lies in the interval [0, 1]:

$$LOY_{ijt_i} = \alpha LOY_{ijt_i} - 1 + (1 - \alpha)\, Y_{ijt_i - 1} \tag{1}$$

On the contrary, the model estimated in Abe et al. (2000) determines a weakly nonlinear effect of loyalty and a linear effect of price.

An application of the semiparametric model of Briesch (1996) results in almost linear effects of both price and loyalty. On the other hand, utility is shown to be concave w.r.t. price reduction. Marginal effects decrease with the amount of price reduction.

Hruschka et al. (2002) develop a neural-net enhanced multinomial logit model and find several nonlinear effects in their empirical study. Utility decreases with higher reference prices, and in most cases takes an inverse S-shape. As a rule, utility increases degressively with brand loyalty. Utility levels are higher if loyalty is high or sales promotion is on, with the effect of loyalty being more pronounced.

On the basis of these empirical studies we can expect the following kinds of nonlinear effects:

- an inverse S-shape for reference prices;
- threshold effects for very low prices;
- very small effects for medium prices in the medium price range;
- a concave shape for brand loyalty;
- stronger effects for prices above reference price as opposed to prices below reference price.

3 Functional Forms of Deterministic Utilities

As discussed in Section 2, the empirical evidence suggests that we extend choice models so that they are able to reproduce nonlinear effects. We can achieve this extension by specifying deterministic utility as parametric function. Appropriate parametric functions either consist of higher-order polynomial terms (i.e., quadratic, cubic), fixed transformations (e.g., semilog), or piecewise linear functions of predictors.

Of these alternatives we consider functions with higher-order terms and piecewise linear functions, because both are able to reproduce the nonlinear effects that earlier empirical studies have demonstrated, i.e., (inverse) S-shapes, threshold effects, saturation effects, and very small effects in the medium range of a predictor. We do not study concave transformations, such as semilog, because they prohibit such effects.

Linear piecewise functions constitute another alternative way to reproduce such nonlinear effects. So far, piecewise linear utility functions contained in brand choice models are univariate, i.e. they relate to one predictor only (Kalyanaram and Little (1994); Wedel and Leeflang (1998)). The model we introduce here allows interactions of predictors and can be seen as multivariate generalization of the univariate approach. Therefore, it is more powerful than the generalized additive model of Abe (1999) and similar to the semiparametric model of Briesch et al. (1996) and the neural net approach of Bentz and Merunka (2000) or Hruschka et al. (2002). Of course, this model is restricted to piecewise linear functions, but this limitation may be outweighed by the fact that both estimation and interpretation turn out to be straightforward.

These piecewise functions should obey continuity restrictions for different sections of the total value range of a predictor (i.e. the linear pieces must be connected). Mathematically, linear splines and bivariate tensor products can represente linear piecewise functions with these properties. This representation turns out to be advantageous for estimation. Conventional (log) linear estimation procedures can be used after the value range of a predictor is divided into at least two sections by fixing one or more knot positions.

Linear splines are defined only for metric predictors. The linear spline $S(x)$ of a metric predictor x consists of piecewise linear functions with knots $x_0^{(\kappa)}$, ..., $x_K^{(\kappa)}$ (Schumaker (1981); Nürnberger (1989)):

$$S(x) = \begin{cases} s_1(x) = a_0^{(1)} + a_1^{(1)}x & \text{if } x \in [x_0^{(\kappa)}; x_1^{(\kappa)}] \\ \vdots & \vdots \\ s_K(x) = a_0^{(K)} + a_1^{(K)}x & \text{if } x \in [x_{K-1}^{(\kappa)}; x_K^{(\kappa)}] \end{cases} \tag{2}$$

The linear spline may also be written as:

$$S(x) = \beta_0 + \beta_1 x + \sum_{k=1}^{K-1} \beta_{k+1}(x - x_K^{(\kappa)})_+ \tag{3}$$

with the truncated identity function for $k = 1, ..., K - 1$:

$$(x - x_K^{(\kappa)})_+ = \begin{cases} (x - x_K^{(\kappa)}) & \text{if } x > x_K^{(\kappa)} \\ 0 & \text{if } x \leq x_K^{(\kappa)} \end{cases} \tag{4}$$

A linear spline may also be expressed as linear combination of $(K + 1)$ basis functions $B^{(k)}$:

$$S(x) = \sum_{k=0}^{K} \beta_k B^{(k)} \text{ with } x \in [x_0^{(\kappa)}; x_K^{(\kappa)}] \tag{5}$$

The basis functions are:

$$\begin{aligned} B^{(0)} &= 1 \\ B^{(1)} &= x \\ B^{(2)} &= (x - x_1^{(\kappa)})_+ \\ &\vdots \\ B^{(K)} &= (x - x_{K-1}^{(\kappa)})_+ \end{aligned}$$

We note that these basis functions also include the constant term $B^{(0)}$ and the linear term of the predictor $B^{(1)}$.

The bivariate tensor product of splines for two predictors $x_1,\ x_2$ with one knot each (i.e. $x_{11}^{(\kappa)}$and $x_{21}^{(\kappa)}$, respectively) consists entirely of basis functions (Prenter (1975); Greub (1978); Schumaker (1981)):

$$\begin{aligned} S(x_1,\, x_2) = {} & \beta_{00} + \beta_{10}x_1 + \beta_{01}x_2 + \beta_{11}x_1x_2 + \\ & \beta_{20}(x_1 - x_{11}^{(\kappa)})_+ + \beta_{02}(x_2 - x_{21}^{(\kappa)})_+ + \beta_{21}(x_1 - x_{11}^{(\kappa)})_+ x_2 + \\ & \beta_{12}x_1(x_2 - x_{21}^{(\kappa)})_+ + \beta_{22}(x_1 - x_{11}^{(\kappa)})_+(x_2 - x_{21}^{(\kappa)})_+ \end{aligned} \tag{6}$$

As we can see from this expression linear terms and one spline for each predictor as well as the product of the linear terms of the predictors (i.e. the conventional bivariate interaction effect known from linear models) form part of the bivariate tensor product. Extensions to splines with more knots are obvious (for statistical applications of splines to regression problems see Friedman and Silverman (1989); Friedman (1991); Stone et al. (1997)).

A model with linear splines can reproduce both threshold and saturation effects, both of which imply that over a certain value range, the effect of the respective predictor is constant. In the case of a saturation effect, the coefficient for this value range must compensate the sum of coefficients of preceding value ranges (given by the linear part of the model and preceding knots). In the case of a threshold effect, the model consists of only a constant term (the coefficient for this value range equals zero), and coefficients for values of the predictor greater than the threshold are different from zero. Moreover, because it allows different marginal effects of any predictor for different value ranges, the model with linear splines can also reproduce asymmetric effects.

4 Choice Models

We investigate brand choice models because the empirical studies discussed in Section 2 give evidence on nonlinear effects on brand choice. The models we use are all variants of the multinomial logit (MNL) model which is the most widespread choice model in marketing. The basic assumptions of the MNL model are: (1)buyers select out of a choice set the brand that they perceive to have maximum utility, (2) utility is conceived to be additively made up by a deterministic component and a random term, and (3) the random term is iid type I extreme value distributed (McFadden (1973)).

If we suppress indices of purchase occasions to keep notation simple, then we can write the conditional choice (purchase) probability p_{ij} of brand j for household i according to a MNL model as (McFadden (1973)):

$$p_{ij} = \frac{\exp(V_{ij})}{\Sigma_{j'} \exp(V_{ij'})} \tag{7}$$

V_{ij} denotes deterministic utility of brand j for household i. The denominator in expression 7 sums over all brands available at the respective purchase occasion of household i.

In the piecewise linear model deterministic utility is a linear combination of basis functions. It can be decomposed into main effects and pairwise interaction effects:

$$
\begin{aligned}
V_{ij} &= \underbrace{\beta_{j0} + \sum_{n=1}^{N} f_n(x_n)}_{\text{main effects}} + \underbrace{\sum_{n \neq 0} f_{nl}(x_n, x_l)}_{\text{pairwise interaction effects}} \\
\text{with} \quad f_n(x_n) &= \sum_{p_n=1}^{P_n} \beta_{p_n} B^{(p_n)}(x_n) \qquad (8) \\
f_{nl}(x_n,\ x_l) &= \sum_{p_{nl}=1}^{P_n} \beta_{p_{nl}} B^{(p_{nl})}(x_n, x_l)
\end{aligned}
$$

P_n number of univariate basis functions of predictor n

P_{nl} number of bivariate basis function of predictors n and l

The choice models we study here have the following predictors (abbreviated terms appear in parentheses):

- brand specific constants;
- reference price (RPRICE);
- for models with nonlinear utility, we define the price deviation (PDEV) of observed price (PRICE) from the reference price as RPRICE – PRICE, so that a positive (negative) price deviation indicates a gain (loss);
- for models with linear utility, we replace price deviation with both loss defined as max(PRICE – RPRICE,0) and gain defined as max(RPRICE – PRICE,0);
- brand loyalty (LOY) following Guadagni-Little (1983);
- display (D), i.e. POS advertising (binary);
- feature (F), i.e. local newspaper ads or flyers (binary).

With the exception of brand constants, feature, and display these predictors are metric.

By introducing separate predictors, we obtain asymmetric effects of gains and losses even in a linear utility model given that their coefficients differ. On the other hand, separate predictors are not necessary for models with nonlinear utility. These models can identify asymmetric effects of gains and losses, since they allow different marginal effects w.r.t. the price deviation variable, whereas linear utility models are characterized by constant marginal effects.

The reference price mechanism we study is of the extrapolative expectations type (Winer (1988)). Reference price depends on prices observed for the three previous purchases and a trend term in the following manner:

$$RPRICE_{ijt_i} = \gamma_0 + \sum_{l=1}^{3} \gamma_l PRICE_{ijt_i - l} + \gamma_4 TREND_{it_i} \quad (9)$$

As in most studies, this reference price mechanism is homogeneous, i.e. coefficients of equation 9 do not vary across households. In other words, as in almost all relevant studies (Winer (1986); Lattin (1989); Kalwani et al. (1990); Kalyanaram and Little (1994); Mayhew and Winer (1992); Mazumdar and Papatla (1995); Chang et al. (1999)) we allow for latent heterogeneity only in choice models, but parameters of the reference price model are equal across households. An advantageous consequence of this approach is the fact that the predictors reference price, price deviation, gain, and loss assume the same values for each household and purchase occasion, no matter which model is estimated. This property also applies to latent class models which differ w.r.t. the number of segments.

We do not consider observed price as predictor of brand choice. Observed price is equal to the difference of reference price and price deviation, and price deviation is equal to the difference of gain and loss. Therefore, observed price is collinear if either price deviation or gain and loss are predictors. On the other hand, reference price is uncorrelated with price deviation, if determined on the basis of a model estimated by OLS (Chow (1985)).

Thus, we replace observed price with reference price. This collinearity problem shows up in the empirical data used in this study. Regressing observed price and reference price against the remaining predictors reproduces 53.18% and 10.18% of variance, respectively.

All the choice models we consider have brand constants and coefficients for feature and display. We compare the piecewise linear utility model with interactions to the following less complex models. These models do not include interactions and therefore have a smaller number of parameters:

- a conventional linear utility function that consists of linear terms for each remaining predictor;
- a quadratic utility function that consists of linear and quadratic terms for each remaining predictor;
- a cubic utility function that consists of linear, quadratic, and cubic terms for each remaining predictor;
- the piecewise linear utility function with main effects only, i.e., without pairwise interaction effects.

To answer the main research question of our paper we need latent class extensions of the basic choice models described above (the latent class extension of the linear utility model was introduced by Kamakura and Russel (1989)). Latent class models conceive the population of households as finite mixture of classes, which can be interpreted as market segments (Wedel and Kamakura (1998)).

In latent class extensions of MNL models, we follow Kamakura and Russel (1989), who specify class-specific choice probabilities p_{cij} of household i for brand j as:

$$p_{cij} = \frac{\exp(V_{cij})}{\Sigma_{j'} \exp(V_{cij'})} \tag{10}$$

Conditional choice probabilities for latent class MNL models follow from multiplying these class-specific choice probabilities with posterior probabilities π_c and summing across C classes:

$$p_{ij} = \sum_{c=1}^{C} \pi_c \, p_{cij} \tag{11}$$

The capability of the latent class model with linear utility to reproduce nonlinear effects is due to the fact that marginal efects on utility differ across classes, although they are constant within each class.

Contrary to homogeneous MNL models, coefficients of latent class MNL models vary across classes. The number of parameters, P, of latent class models is equal to the number of classes, C, times the number of coefficients of any class-specific MNL model, P_c, plus the number of free a posteriori class probabilities:

$$P = CP_c + (C - 1) \tag{12}$$

5 Estimation, Specification, and Evaluation

We estimate parameters of choice models by maximizing the log likelihood across purchases. To this end, we use the BFGS algorithm (Seber and Wild (1989)) for homogeneous choice models. For latent class choice models we use an EM algorithm (McLachlan and Krishnan (1997); Wedel and Kamakura (1998)).

For C classes the likelihood function can be written as (Wedel and Kamakura (1998)):

$$L = \prod_{i=1}^{I} \sum_{c=1}^{C} \pi_c \prod_{T=1}^{T_i}{}_c \prod_{j=1}^{J} {}^{Y_{ijt}}_{p_{cijt}} \tag{13}$$

I denotes the number of households, C the number of classes, π_c the a posteriori probability of class c. T_i is the number of purchases of household i and J is the number of brands. Y_{ijt} is the purchase indicator of household i of brand j at occasion t (equal to one if household i purchases brand j at occasion t, else zero).

For homogeneous choice models, we can simplify the likelihood function by setting $C = 1$, $\pi_1 = 1$ and $p_{1ijt} = p_{ijt}$ in equation 13:

$$L = \prod_{i=1}^{I} \prod_{t=1}^{T_i}{}_c \prod_{j=1}^{J} {}^{Y_{ijt}}_{p_{ijt}} \tag{14}$$

The piecewise linear utility function, which always contains brand constants, is automatically specified by using three types of basis functions: each predictor, truncated identity functions of each metric predictor with up to seven knot positions (determined by appropriate quantiles), and bivariate tensor products of predictors or truncated identity functions.

We generate different models by making a specification search consisting of stepwise additions of selectable basis functions followed by stepwise eliminations as described in Kooperberg et al. (1995) and Kooperberg et al. (1997). From the set of all the models generated, we select the model attaining the best BIC value (see equation 15).

As the piecewise linear utility function is linear in parameters, the log likelihood function has only one global optimum given knot positions. We estimate the loyalty smoothing constant using the approach introduced by Fader et al. (1992) and we estimate parameters of the reference price by OLS.

We compare the following model types:

- a homogeneous MNL model with linear utility;
- a homogeneous MNL model with piecewise linear utility;
- a homogeneous MNL model with quadratic utility;
- a homogeneous MNL model with cubic utility;
- a latent class MNL model with linear utility;
- a latent class MNL model with piecewise linear utility;
- a latent class MNL model with quadratic utility;
- a latent class MNL model with cubic utility.

We base our comparisons of model types on both Bayesian Information Criterion (BIC) and cross-validated mean choice probability (CVMCP). Given log likelihood $ln(L)$, number of parameters P, and total number of purchases T, we follow Schwarz (1979) and compute BIC as:

$$BIC = -2\ ln(L) + P\ ln(T) \tag{15}$$

We evaluate the out-of-sample performance of models by tenfold cross-validation (Stone (1974), Wahba and Wold (1975)). We randomly divide the whole data set into ten disjunctive subsets, each with the same number of households. Each of these ten subsets serves once as validation data set, for which we compute choice probabilities based on the model that we estimate by using the remaining nine subsets.

We define CVMCP as the geometric mean of cross-validated choice probabilities across all purchases for the brand actually chosen. Obviously, a model with higher CVMCP is preferable to a model with lower CVMCP, since the model with higher CVMCP predicts better. In our opinion, mean choice probability is easier to interpret than is the log likelihood to which it is related. We compute the mean choice probability of a model by:

$$\left(\prod_{i=1}^{I} \prod_{t=1}^{T_i}{}_c \prod_{j=1}^{J} {}^{Y_{ijt}}_{p_{ijt}} \right)^{1/\Sigma_{i=1}^{I}\ T_i} \tag{16}$$

We use CVMCP as our main evaluation criterion, because we consider out-of-sample performance to be more important than in-sample performance, especially for more complex models, which may suffer from overfitting problems.

6 Empirical Study

6.1 Data

Table 1: Descriptive Statistics

Brand	Number of Purchases	Price Mean	Standard Dev.	Min	Max	Percentage with Display	Feature
1	4422	$1.65	$0.25	$0.99	$2.90	9.93	25.76
2	2388	$1.71	$0.19	$1.07	$4.73	7.16	24.58
3	2666	$1.76	$0.17	$1.14	$2.29	2.55	7.39
4	2615	$1.66	$0.24	$0.99	$5.21	9.79	24.36
5	1902	$1.70	$0.21	$0.99	$5.90	8.83	28.02
6	1335	$1.76	$0.19	$1.14	$5.61	4.12	9.74
Aggregate	15328	$1.71	$0.22	$0.99	$5.90	7.55	21.03

Table 1 contains descriptive statistics on purchases, prices (US dollars) and shares of purchases with display and feature for a database that consists of purchases acquired from a household scanner panel. The data refer to the six largest brands in terms of market share and the most frequently purchased package size across all brands in the product category of peanut butter. We select 960 households with at least ten purchases across 123 weeks. We use 30 weeks for the initialization of brand loyalty and 93 weeks for estimation.

We choose this database because of high variation of prices across purchases which can be seen from the columns for minimum and maximum prices of *table 1*. A high variation of prices makes nonlinear effects detectable. On the other hand, a low variation of prices suppresses nonlinear effects and leads to better performance of linear models.

6.2 Estimation Results

Table 2 presents the coefficients of the predictors of the extrapolative reference price model that we estimate by OLS. The positive effect of lagged prices decreases in time, with price at the most recent purchase being most important. The positive coefficient of trend indicates that reference prices increase with time. In addition to the predictors already mentioned in Section 4, we add a binary dummy variable that we set to one after the 97th week. We do so to reproduce a general increase of price level of the product category analyzed.

Table 2: Reference Price Model

	Coefficient
Constant	8.262
PRICE_{t-1}	0.269
PRICE_{t-2}	0.056
PRICE_{t-3}	0.067
TREND	0.490
WEEK > 97	0.893

Because for each utility function, models with both gains and losses among the predictors perform worse w.r.t. both BIC and CVMCP, here we present only results for choice models with price deviation. Grid searches over the value range of the smoothing constant in the loyalty equation of Guadagni and Little (1983) demonstrate that it has a unique global maximum w.r.t. log likelihood values. Metric predictors and their quadratics or cubics are z-transformed to achieve better-conditioned estimation matrices.

Table 3: Evaluation of Homogeneous Choice Models

Q	(Number of) Selected Knots Reference Price	Price Deviation	Loyalty	P	BIC	CVMCP
Linear						
				10	26523.62	0.3243
Piecewise Linear without Interactions						
1	0	0	1	11	26225.66	0.3305
2	0	2	2	14	26140.31	0.3320
3	0	2 (2/3)	2 (1/3)	14	26098.67	0.3326
4	0	2 (2/4)	3 (1/3/4)	15	26072.07	0.3331
5	0	2 (3/5)	3 (1/4/5)	15	26056.05	0.3333
6	0	2 (3/6)	3 (1/5/6)	15	26054.31	0.3333
7	0	2 (3/7)	2 (1/6)	14	26061.23	0.3331
Piecewise Linear with Interactions						
1	0	1	1	15	26204.28	0.3331
2	0	2	2	15	26115.30	0.3324
3	0	2 (2/3)	2 (1/3)	18	26060.45	0.3334
4	0	2 (2/4)	3 (1/3/4)	19	26033.41	0.3340
5	0	2 (3/5)	3 (1/4/5)	17	26012.75	0.3341
6	0	2 (3/6)	3 (1/5/6)	17	26011.37	0.3341
7	0	2 (3/7)	4 (1/5/6/7)	18	26017.85	0.3341
Quadratic						
				13	26183.51	0.3313
Cubic						
				16	26002.33	0.3341

3 (1/5/6) means that 3 knots, i.e. the first, fifth and sixth knot, are selected

Tables 3 and *4* contain estimation results for homogeneous and latent class models, respectively. As mentioned in Section 5, CVMCP serves as our main criterion for evaluating models. In terms of CVMCP, the two best nonlinear homogeneous choice models are the piecewise linear utility model with interactions which starts from seen quantiles for each metric predictor (note that differences to variants with fewer nodes are very small), and the cubic utility model. In the following, we consider only these two nonlinear homogeneous choice models. We also estimate latent class extensions of the linear, piecewise linear with interactions, and the cubic utility models. For each of these three models, we set the number of classes to the value at which CVMCP attains its maximum.

Table 4: Evaluation of Latent Class Choice Models

Utility Function	Number of Classes	P	BIC	CVMCP
Linear	5	54	25546.06	0.3386
Piecewise Linear	3	53	25746.35	0.3341
Quadratic	3	41	25651.82	0.3414
Cubic	3	50	25725.54	0.3415

We take the latent class model with linear utility as standard of comparison. The homogeneous MNL model with linear utility is clearly worse. Its CVMCP is lower by 4.21% than the CVMCP of the latent class model with linear utility. The CVMCPs of the homogeneous models with linear piecewise and cubic utility are lower by 1.31% and 1.33%, respectively. Therefore the latent class model with linear utility is clearly superior to the two homogeneous nonlinear models, although differences are not large. This result also demonstrates that the two homogeneous nonlinear models attain similar performance levels, and that neither of these two models can be preferred over the other.

The latent class model with linear piecewise utility performs worse than the latent class model with linear utility. Compared to the latter model, the latent class models with quadratic and cubic utility increase CVMCP by only 0.83% and 0.86%, respectively. In view of these results we ignore these more complex latent class models in the following and restrict our attention to the latent class model with linear utility.

6.3 Comparison of Effects Implied by Models

Because of the superior performance of the latent class MNL model with segment-specific linear utility functions shown in subsection 6.2, we can now adress the main research question of whether effects implied by the latent class choice model are similar to those implied by parametric homogeneous nonlinear models.

To measure the similarity of effects between two models, we compute absolute differences of choice probabilities implied by these models. We call effects of a predictor implied by models A and B to be more similar than effects implied by models A and C if absolute differences of choice probabilities between models A and B are smaller than absolute differences of choice probabilities between models A and C.

We compare absolute differences of choice probabilities between each of the two nonlinear models and the linear utility model, and between each of the two nonlinear models and the latent class model.

If the absolute differences of choice probabilities between each nonlinear model and the latent class model are smaller than are absolute differences between each nonlinear model and the linear utility model, then we say that the latent class choice model implies effects that are similar to those of parametric homogeneous nonlinear models.

We consider the effects of each metric predictor for which the piecewise linear utility model indicates nonlinear effects. As explained in section 3, nonlinearity is achieved by one or several knots that divide the value range of a predictor into different sections. This property applies to the predictors price deviation and loyalty (see *table 3*).

We compute the choice probabilities of one brand for each of the models considered by setting values of predictors as follows:

- We vary the price deviation (loyalty) of the respective brand over a grid of 201 equidistant values. We set the other predictor with nonlinear effects to its brand-specific median value.
- The reference price, price deviation and loyalty of the remaining brands assume their brand-specific median values.
- We set feature and display to one for all brands.

For homogeneous choice models, we obtain the conditional choice probability p_{ij} by computing deterministic utilities for each brand on the basis of predictor values and inserting deterministic utilities into equation 7.

In the case of the latent class model with linear utility, we obtain class-specific deterministic utilities for each brand on the basis of predictor values. They lead to class-specific choice probabilities p_{cijt} as given by equation 10. We determine the (overall) conditional choice probability of brand j by inserting class-specific choice probabilities and posterior class probabilities into equation 11.

The arithmetic means of absolute differences of choice probabilities for both nonlinear models and the latent class model are clearly smaller than are the analogous values for both nonlinear models and the homogeneous linear utility model, for both varying price deviation and for varying loyalty (see *table 5*). These arithmetic means also indicate that there is more agreement between the latent class and the cubic utility model. Paired sample t-tests are highly significant. These results demonstrate that the latent class model implies effects that resemble those of the two nonlinear homogeneous choice models.

To provide more details on similarities between the latent class model and homogeneous nonlinear models we plot logits as functon of price deviation and loyalty, respectively. A logit (also called log odd) can be interpreted as log of the chance of a choice of brand 1

Table 5: Absolute Differences of Choice Probabilities between Model Pairs Piecewise Linear Piecewise Linear and Linear and Latent Class

	Piecewise Linear and Linear		Piecewise Linear and Latent Class
for varying price deviation			
Mean	0.0975		0.0454
Variance	0.0021		0.0018
Paired samples *t*-test		35.75	
for varying loyalty			
Mean	0.1307		0.0616
Variance	0.0149		0.0024
Paired samples *t*-test		12.34	
	Cubic and Cubic and Linear		Cubic and Latent Class
for varying price deviation			
Mean	0.0899		0.0375
Variance	0.0001		0.0001
Paired samples *t*-test		35.48	
for varying loyalty			
Mean	0.1100		0.0445
Variance	0.0098		0.0011
Paired samples *t*-test		11.63	

by household i relative to choices of the other brands. It is defined as transformation of the conditional choice probability p_{i1} (Greene (2003)):

$$logit_{i1} = ln\ p_{i1} - ln\,(1 - p_{i1}) \qquad (17)$$

We compute conditional choice probabilities of brand 1 w.r.t. both price devation and loyalty, which are necessary to determine logits, in the same way as explained at the beginning of this subsection.

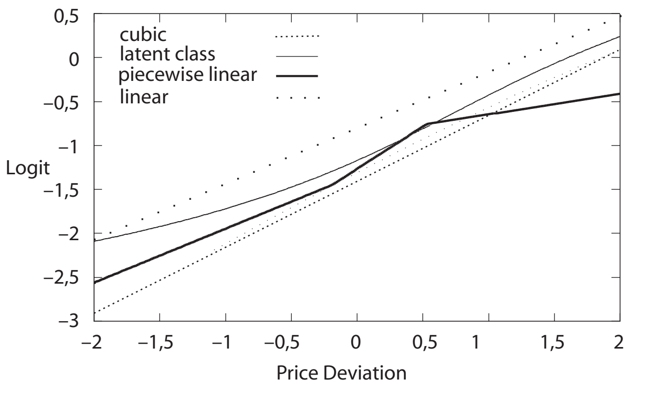
Starting from equation 7 we show that logits of the choice of brand 1 for homogeneous models can be computed as follows :

$$logit_{i1} = V_{i1} - ln\left(\sum_{j \neq 1} \exp V_{ij}\right) \tag{18}$$

Therefore, logits for the homogeneous model with linear utility depend linearly on each predictor of the respective brand. But for the other models studied here, the dependence of logits on predictors is nonlinear. Therefore, logits provide insight into nonlinear effects implied by the latent class model, which a comparison of linear class-specific utilities and utilities implied by homogeneous nonlinear models does not give.

We show one plot for price deviation and loyalty, setting the other predictors to median values. Plots, which are based on the other quartiles of the remaining predictors, have similar shapes. *Figure 1* shows logits as a function of price deviation we obtain for four different choice models. For all models, logits increase as price deviation becomes greater. Logits for the latent class model with linear utility develop in a nonlinear way. For price devations in the interval [–0.80, 1.16], logits of the latent class model differ more from logits of the linear utility model than from logits of models with piecewise linear or cubic utility. For price deviations in the interval [–1.26, 1.16], logits of the latent class model are more like those of the piecewise linear utility model. For price deviations greater than 1.10, logits of the latent class model are more like those of the cubic utility model. Only for price deviations lower than –1.26 do logits of the latent class model differ less from those of the linear utility model compared to the two nonlinear models.

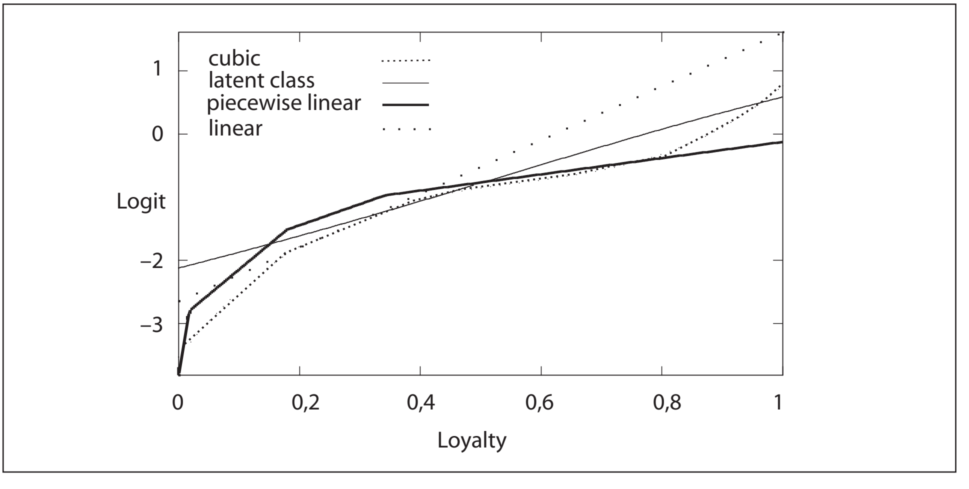
Figure 1: Logits vs. price deviation for different MNL models



For the cubic utility model, the marginal effects (i.e., first derivatives) of price deviations stay pretty much the same across their observed value range, but change for both the piecewise linear model and the latent class model. Therefore, the latter two models indicate asymmetric effects, but these asymmetric effects differ across these two models. The piecewise linear model shows lower marginal effects at higher positive price deviations (i.e., at higher gains), which explains why the piecewise linear model is more in accordance with prospect theory. On the other hand, marginal effects of price deviation increase for the latent class model and therefore imply higher marginal effects of positive price deviations (i.e. gains) compared to those of negative price deviations (i.e. losses). Neither model provides evidence for the medium price insensitive range as postulated by assimilation-contrast theory.

Figure 2 shows logits as function of loyalty according to the four different choice models studied. For all models, logits increase as loyalty becomes greater. For loyalties greater than 0.40, logits of the latent class model differ more from logits of the linear utility model than from logits of models with piecewise linear and cubic utility, respectively. If loyalties lie in the interval [0.11, 0.21], then logits of the latent class model are similar to those of the cubic utility model and differ from those of the linear utility model. The marginal effects are almost constant for the latent class model. For models with cubic utility the marginal effects decrease for loyalties in the intervals [0.15, 0.44]. For models with piecewise linear utility marginal effects decrease for loyalties greater equal 0.03.

Figure 2: Logits vs. loyalty for different MNL models



Figures 1 and *2* indicate that the cubic utility, piecewise linear utility and latent class models do not completely agree on the effects of the predictors (especially w.r.t. marginal effects). However, these figures show that the effects implied by the two homogeneous

nonlinear models are similar to the effects implied by the latent class model. This result is in accordance with the comparisons of absolute differences of choice probabilities between the model pairs discussed ealier. The conclusion we draw is that the latent class choice model implies effects that are similar to those of homogeneous nonlinear models. This result leads us to recommend the use of the latent class model instead of a parametric homogeneous nonlinear model.

7 Conclusions

Earlier empirical studies show that brand choices of individual households are subject to nonlinear effects for predictors such as price, reference price, and brand loyalty. Nevertheless, most brand choice models found in the literature are based on linear deterministic utility functions and therefore may result in biased estimates.

Alternative approaches to adress nonlinearity are either parametric homogeneous nonlinear functions that do not vary across hoseholds, or latent class models with class-specific linear utility functions. We deal with the question of whether the latent class choice model with segment-specific linear utility functions implies effets that are similar to those of parametric homogeneous nonlinear models, given that the latent class model performs at least as well.

For the data set we analyze, the latent class model with linear utility is clearly superior to the two homogeneous nonlinear models (with piecewise linear and cubic utility, respectively), although differences are not large. Therefore, we are justified in investigating whether the effects of predictors implied by this latent class model function resemble the effects implied by the nonlinear models.

On the basis of varying values of each of two predictors, price deviation and loyalty, the latent class model turns out to be similar to each of the two nonlinear models in terms of choice probabilities. This conclusion can be also be drawn by looking at logits of choice probabilities as functions of each of these two predictors. Overall, our results confirm that the the latent class model implies effects that are similar to those of the two nonlinear models considered.

Our results suggest that researchers should use the latent class model instead of a parametric homogeneous nonlinear model. Unlike nonlinear models, standard software for this latent class model is readily available and thus constitutes an additional benefit for researchers.

References

Abe, Makoto (1998), Measuring Consumer, Nonlinear Brand Choice Response to Price, *Journal of Retailing* 74, 541-568.

Abe, Makoto (1999), A Generalized Additive Model for Discrete Choice Data, *Journal of Business & Economic Statistics* 17, 271-284.

Abe, Makoto, Yasemin Boztuğ, and Lutz Hildebrandt (2000), *Investigation of the Stochastic Utility Maximization Process of Consumer Brand Choice by Semiparametric Modeling*, Discussion Paper, Humboldt-Universität zu Berlin.

Allenby, Greg M. and James L. Ginter (1995), The Effects of In-Store Display and Feature Advertising on Consideration Sets, *International Journal of Research in Marketing* 12, 67-80.

Allenby, Greg M. and Peter J. Lenk (1994), Modeling Household Purchase Behavior with Logistic Normal Regression, *Journal of the American Statistal Association* 89, 1218-1231.

Amemiya, Takeshi (1985), *Advanced Econometrics*, Oxford, UK: Basil Blackwell.

Bentz, Yves and Dwight Merunka (2000), Neural Networks and the Multinomial Logit for Brand Choice Modelling: A Hybrid Approach, *Journal of Forecasting* 19, 177-200.

Briesch, Richard A., Pradeep Chintagunta, and Rosa L. Matzkin (1996), *A Nonparametric Investigation into the Differential Consumer Response to Price and Deal Discounts*, Working Paper, Stern School of Business, New York University, New York.

Chang, Kwangpil, S. Siddarth, and Charles B. Weinberg, (1999), The Impact of Heterogeneity in Purchase Timing and Price Responsiveness on Estimates of Sticker Shock Effects, *Marketing Science* 18, 178-192.

Chintagunta, Pradeep K. (1992), Estimating a Multinomial Probit Model of Brand Choice Using the Method of Simulated Moments, *Marketing Science* 11, 386-407.

Chow, Gregory C. (1985), *Econometrics*, Auckland: McGraw Hill.

Fader, Peter, James M. Lattin, and John D.C. Little (1992), Estimating Nonlinear Parameters in the Multinomial Logit Model, *Marketing Science* 11, 372-385.

Friedman, Jerome H. and Bernard W. Silverman (1989), Flexible Parsimonious Smoothing and Additive Modeling, *Techonometrics* 31, 3-39.

Friedman, Jerome H. (1991), Multivariate Adaptive Regression Splines, *The Annals of Statistics* 19, 1-141.

Gensch, Dennis and W.W. Recker (1979), The Multinomial Multiattribute Logit Choice Model, *Journal of Marketing Research* 16, 124-132.

Greene, William H. (2003), *Econometric Analysis*, 5th Edition, Upper Saddle River, NJ: Prentice Hall.

Greub, Werner (1978) *Multilinear Algebra,* 2nd Edition, New York: Springer.

Guadagni, Peter M. and John D.C. Little (1983), A Logit Model of Brand Choice Calibrated on Scanner Data, *Marketing Science* 2, 203-238.

Gupta, Sunil (1988), The Impact of Sales Promotions on When, What, and How Much to Buy, *Journal of Marketing Research* 25, 342-356.

Gupta, Sunil and Lee G. Cooper (1992), The Discounting of Discounts and Promotion Thresholds, *Journal of Consumer Research* 19, 401-411.

Hruschka, Harald, Werner Fettes, Markus Probst, and Christian Mies, (2002), A Flexible Brand Choice Model Based On Neural Net Methodology. A Comparison to the Linear Utility Multinomial Logit Model and Its Latent Class Extension, *OR Spectrum* 24, 127-143.

Kahneman, Daniel and Amos Tversky (1979), An Analysis of Decision Under Risk, *Econometrica* 47, 363-391.

Kalwani, Manohar O., Chi K. Yim, Heikki J. Rinne, and Yoshi Sugita (1990), A Price Expectations Model of Customer Brand Choice, *Journal of Marketing Research* 27, 251-262.

Kalyanaram, Gurumurthy and John D.C. Little (1994), An Empirical Analysis of Latitude of Price Acceptance in Consumer Package Goods, *Journal of Consumer Research* 21, 408-418.

Kamakura, Wagner A. and Gary J. Russel (1989), A Probabilistic Choice Model for Market Segmentation and Elasticity Structure, *Journal of Marketing Research* 26, 379-390.

Kooperberg, Charles, Smarajit Bose, and Charles J. Stone (1997), Polychotomous Regression, *Journal of the American Statistical Association* 92, 117-127.

Kooperberg, Charles, Charles J. Stone, and Young K. Truong (1995), Hazard Regression, *Journal of the American Statistical Association* 90, 78-94.

Krishnamurthi, Lakshman and S.P. Raj (1992), A Model of Brand Choice and Purchase Quantity-Price Sensitivities, *Marketing Science* 7, 1-20.

Lattin, James M. and Randolph E. Bucklin (1989), Reference Effects of Price and Promotion on Brand Choice Behavior, *Journal of Marketing Research* 26, 299-310.

Mayhew, Glenn E. and Russel S. Winer (1992), An Empirical Analysis of Internal and External Reference Prices Using Scanner Data, *Journal of Consumer Research* 19, 62-70.

Mazumdar, Tridib and Purushottam Papatla (1995), Loyalty Differences in the Use of Internal and External Reference Prices*, Marketing Letters* 6, 111-122.

McCulloch, Robert and Peter E. Rossi (1994), An Exact Likelihood Analysis of the Multinomial Probit Model, *Journal of Econometrics* 64, 207-240.

McFadden, Daniel (1973), Conditional Logit Analysis of Qualitative Choice Behavior, in Zarembka, Paul (ed.), *Frontiers in Econometrics*, New York: Academic Press, 105-142.

McLachlan, Geoffrey J. and Thriyambakam Krishnan (1997), *The EM Algorithm and Extensions*, New York: Wiley.

Monroe, Kent B. (1973), Buyers' Subjective Perceptions of Price, *Journal of Marketing Research* 10, 70-80.

Nürnberger, Günther (1989), *Approximation by Spline Functions*, Berlin: Springer.

Pedryck, James H. and Fred S. Zufryden (1991), Evaluating the Impact of Advertising Media Plans: A Model of Consumer Purchase Dynamics Using SingleSource Data, Marketing Science 10, 111-130.

Prenter, P. M. (1975), *Splines and Variational Methods*, New York: Wiley.

Schumaker, Larry L. (1981), *Spline Functions*, New York: Wiley.

Schwarz, Gideon (1979), Estimating the Dimension of a Model, *Annals of Statistics* 6, 461-464.

Seber, George A.F. and C.J Wild (1989), *Nonlinear Regression*, New York: Wiley

Stone, M. (1974), Cross-Validatory Choice and Assessment of Statistical Predictions, *Journal of the Royal Statistical Society* B 36, 111-147.

Stone, Charles J., Mark H. Hansen, Charles Kooperberg, and Young K. Truong (1997), Polynomial Splines and Their Tensor Products in Extended Linear Modeling, *The Annals of Statistics* 25, 1371-1470.

Wahba G. and S. Wold (1975), A Completely Automatic French Curve: Fitting Spline Functions by Cross-Validation, *Communications in Statistics* A 4, 1-17.

Wedel, Michel and Wagner A. Kamakura (1998). *Market Segmentation, Conceptual and Methodological Foundations*, Dordrecht: Kluwer.

Wedel, Michel and Peter S.H. Leeflang (1998), A Model for the Effects of Psychological Pricing in Gabor-Granger Price Studies, *Journal of Economic Psychology* 19, 237-260.

Winer, Russel S. (1986), A Reference Price Model of Brand Choice for Frequently Purchased Products, *Journal of Consumer Research* 13, 250-256.

Winer, Russel S. (1988), Behavioral Perspective on Pricing, in Devinney, Timothey M. (ed.), *Issues in Pricing*, Lexington, MA: Lexington Books, 35-57.

Order form –
Order now!

Verlagsgruppe Handelsblatt GmbH
Abo-Service Ausland
Postfach 10 27 53
40018 Düsseldorf
Germany

Fon: 0049 211 887 1730
Fax: 0049 211 887 1738
e-mail: abo-service@vhb.de
Internet: www.sbr-online.com

simply the best research.

Use this form to order your free sample copy and to subscribe to sbr!

Free sample copy

☐ Please send me a free sample copy of **sbr**

PB-ZFSBRPH1

Subscription

☐ Open ended subscription*

☐ One-Year subscription

PB-ZFSBRO15

* In case of open-ended subscription an invoice will be issued at the end of each subscription year to cover the next year. Cancellation within a period of at least 21 days before the new subscription year begins.

Subscription rates**

Schmalenbach Business Review (**sbr**),
ISSN: 1439-2917, Quaterly

Institutions:	☐ $ 95.00	☐ £ 60.00	☐ € 91.00
Individuals:	☐ $ 48.00	☐ £ 30.00	☐ € 45.00
Students*:	☐ $ 24.00	☐ £ 50.00	☐ € 21.00

* Student rate only accepted with copy of validated ID.
** Postage rates are – depending on the currency you want to be charged in – $ 14, £ 8, € 12.

Payment

Payment is due within 14 days on receipt of invoice. You will receive the invoice directly from Verlagsgruppe Handelsblatt GmbH in Düsseldorf.

Address

Institute/Company

Position/Department

First and Surname

Street and Number

Zip Code City

State Country

Fon

Fax

e-mail

✘

Signature Date