Nonparametric modeling of buying behavior in fast moving consumer goods markets

Boztug, Yasemin; Hildebrandt, Lutz

NONPARAMETRIC MODELING OF BUYING BEHAVIOR IN FAST MOVING CONSUMER GOODS MARKETS

YASEMIN BOZTUĞ AND LUTZ HILDEBRANDT

1. Introduction

Much empirical research of consumer behavior has been done using discrete choice modeling. In general, model selection in this area is a decision between complexity of the model and the simplicity of estimation (Ben-Akiva et al., 1997). The well-known multinomial logit model (MNL) (Guadagni and Little, 1983), which is often used on the basis of scanner panel data, especially shows this dilemma. It is easy to estimate but has many restrictive assumptions. Relaxing assumptions however leads to estimation problems. Much research has been done to improve the parametric MNL model (e.g. Kannan and Wright, 1991; Chintagunta, 1992; Fader et al., 1992; Chintagunta, 1993; Gupta and Chintagunta, 1994; Erdem and Keane, 1996; Erdem, 1996; Papatla, 1996). The estimation of nonparametric and semiparametric variants of the MNL model may offer useful alternatives to circumvent its intrinsic constraints. These models encompass a large latitude for modeling, and are based on statistical theory that allows for relatively simple estimation.

This article starts with a short explanation of the logit model in Section 2, where the assumptions are also discussed. The theoretical aspects of nonparametric density estimation with respect to discrete choice models are investigated in Section 3. Accordingly the case of kernel density estimation for binary data is shown. Also, we document the semiparametric approach with a typical data structure of mixed binary and continuous explanatory variables. In addition, we will demonstrate the advantages and benefits of pursuing nonparametric and semiparametric methods. Section 4 presents an application of a semiparametric method to a real panel data set, estimated by two different algorithms.

2. Logit models in marketing

2.1 General description of a logit model

The multinomial logit model (MNL) captures the individual choice behavior between several alternatives. Here, the theorem of utility maximization for the consumers is assumed (Ben-Akiva and Lerman, 1985). That means that a consumer n chooses the alternative which maximizes his utility U_n. The choice set C_n is comprised of the alternatives $i = 1,\ldots, I$. The probability $Pr_n(i)$ for individual n to choose the alternative i is described by

$$Pr_n(i) = Pr(U_{in} \geq U_{jn}, \forall j \in C_n, i \neq j), \qquad (1)$$

with U_{in} and U_{jn} the utilities of the alternatives i and j, where U_{jn} specifies all alternative utilities to U_{in}. Usually, the utility function can be separated additively into a systematic part (V_{in}) and a random part (ε_{in}) with $U_{in} = V_{in} + \varepsilon_{in}$. In the MNL, the systematic utility function is assumed to be linear in the parameters ($V_{in} = \beta^T x_{in}$, where β is a parameter vector to be estimated and x_{in} a characteristic of alternatives in the opinion of the individual n). With these assumptions, the choice probability has the form

$$\begin{aligned} Pr_n(i) &= Pr(U_{in} \geq U_{jn}, \forall j \in C_n, j \neq i) \\ &= Pr(V_{in} + \varepsilon_{in} \geq V_{jn} + \varepsilon_{jn}, \forall j \in C_n, j \neq i) \\ &= Pr(\varepsilon_{jn} - \varepsilon_{in} \leq V_{in} - V_{jn}, \forall j \in C_n, j \neq i). \end{aligned} \qquad (2)$$

A further supposition of the MNL model is that the differences of the errors ε_{jn} - ε_{in} from equation (2) are i.i.d. logistic distributed. Following McFadden (1974) and using the information about the distribution of the errors, equation (2) can also be written as

$$\begin{aligned} Pr_n(i) &= Pr(U_{in} \geq U_{jn}, \forall j \in C_n, i \neq j) \\ &= \frac{exp(V_{in})}{\sum_{j \in C_n} exp(V_{jn})} \\ &= \frac{exp(\beta^T x_{in})}{\sum_{j \in C_n} exp(\beta^T x_{jn})}. \end{aligned} \qquad (3)$$

A maximum-likelihood estimator can obtain the parameter values contained in β. Here, the sign and also the absolute values of the elements of β are of interest. The sign informs us of the direction of connection, e.g. if the sign for a specific explanatory variable of a brand is positive, than a decrease of this variable means also a decrease in the probability of choice for that brand and the same holds for negative signs in the other direction. If all variables are scaled the same, the absolute value of β can be interpreted regarding the strength of the connection between the explanatory and the dependent variables.

2.2 Assumptions of the logit model

The logit model follows assumptions that restrict the interpretation of the estimation results, and also the application of the model is limited. The main criticism of the logit model is related to the IIA assumption (independence of irrelevant alternatives), which implies that the relative utility of one alternative to a second one is independent of the existence of a third one.

The second weak point of the logit model is the assumption about the logistic distribution of the differences of its error terms. It is not obvious, why the differences should follow this distribution, because there are several possibilities for modeling the error terms.

The third weakness relates to the assumption of linear formulation of the utility function. Due to this assumption, a lot of possibilities to describe the utility are excluded. Also, there is no economic justification for the linearity.

To address the first two weak points of the logit model often probit models are used. They use an i.i.d. normal distribution assumption for the differences of the error terms. But for specifying a probit model for the multidimensional case it is difficult to estimate because higher order integrals must be solved. An alternative is to test the appropriateness of the logit model (Bartels et al., 2000). Another solution to deal with the strong assumptions of the logit model is a semi- or nonparametric formulation of the model, which will be presented in the next section.

3. Non- and semiparametric discrete choice models

A nonparametric or semiparametric discrete choice modeling approach allows a nonparametric systematic utility component and/or a distribution-free random component. With this type of modeling one has much more flexibility to specify a choice process. To classify the different characteristics of model types (parametric, non- and semiparametric), see Table 1.

Table 1: Different model types for parametric and nonparametric choice models.

Model type	**Systematic utility function**	**Random component**	**Possible resulting method**
Parametric	parametric	parametric	MNL, probit, etc.
semiparametric I	parametric	distribution free	various
semiparametric II	nonparametric	parametric	Generalized Additive Model (GAM)
nonparametric	nonparametric	distribution free	Nonparametric Density Estimation (NDE)

Parametric models could be e.g. the MNL, or a probit model (Manrai, 1995). The type of “semiparametric I” is described by a parametric utility function, but a distribution free random term. These kind of models are discussed by a wide group of researchers, e.g. Horowitz et al. (1994); Horowitz and Härdle (1996); Matzkin (1991); Chintagunta and Honore (1996). Because of some disadvantages for practical use (see below) this approach will not be described here. The “semiparametric II” type includes a nonparametric utility function, but a parametric random term. The Generalized Additive Models (GAM), introduced by Hastie and Tibshirani (1986), could be used to specify these models. The Nonparametric Density Estimation (NDE) is one possibility to model a pure nonparametric approach. Abe (1995) introduced this approach in the marketing context. The NDE approach will be discussed in the next section in more detail.

3.1 Nonparametric density estimation

The availability of scanner panel data sets has made the application of the nonparametric methods in marketing a feasible alternative to the existing parametric models. Nonparametric methods are based only on very few assumptions, so that they have a lot more structural freedom than the models in the parametric model class, e.g. the MNL. On the other hand, they need sufficient large data sets to produce a good fit of the data. One popular nonparametric method is nonparametric density estimation.

3.1.1 General formulation of the density estimation

The choice decision formulated with a nonparametric method is usually described as a conditional expectation, where the condition is the actual marketing-mix situation at the purchase time. So the conditional expectation $E[y|x]$ is needed, with x the marketing-mix condition and y the choice decision. Usually, the choice is coded binary, so that the following identity holds $E\lfloor y \mid x \rfloor \equiv P(y \mid x)$ with the assumption of $0 \leq E\lfloor y \mid x \rfloor \leq 1$ (e.g.

Hosmer and Lemeshow, 1989). Also the restriction that the sum over the choice probabilities is 1 for given covariates x is needed. Using Bayes' theorem it follows that

$$\mathrm{E}\left[y \mid x\right] = \frac{Pr(y) f(x \mid y)}{f(x)} \tag{6}$$

under the assumption of y binary and $f(\cdot)$ a probability density function. For the estimation of $\mathrm{E}\lfloor y \mid x \rfloor$ through $\mathrm{E}\lfloor y \mid x \rfloor$, the estimator can be partitioned into several different components. One is the estimator $\hat{P}r(y)$ of $Pr(y)$, which can simply be described as

$$\hat{P}r(y) = \frac{\text{number of choices made for the inspected brand}}{\text{number of all possible choices for all brands}}.$$

The components remaining to complete the expectation of equation (6) are the densities functions $f(x)$ and $f(x|y)$. A kernel density estimator can estimate them. The density estimation f_h at x for the density function f of the n continuous multivariate i.i.d. random variables $X_1, \ldots, X_n$ with d dimensions can be described as

$$\hat{f}_h(x) = \frac{1}{nh^d} \sum_{i=1}^{n} K\left(\frac{x - X_i}{h}\right) = \frac{1}{nh^d} \sum_{i=1}^{n} K(x, h) \tag{7}$$

with $K(\cdot)$ a kernel function, which gives the smoothing instruction for the inner distance from $(\cdot)$ as described in equation (7). The distance measure is usually chosen Euclidic for continuous variables. The kernel function also has to fulfil the following conditions (e.g., Silverman, 1986)

- $\int_{R^d} K(x)\, dx = 1$
- $K(\cdot)$ is symmetric and positive
- $K(\cdot)$ is l times continuously differentiable

The parameter h is called the bandwidth of the kernel. It determines with the window width h, in other words which observations are included in the calculation. It is also a measure for the balance between the bias and the variance of the estimation. There are different methods to determine the value for h as well as different expressions for the kernel function that will not be discussed further here. For a detailed explanation the reader is referred to Härdle (1991). It is shown in several statistical research projects that the choice of the bandwidth h is important, whereas the choice of the kernel $K(\cdot)$ influences the estimates only slightly (Härdle, 1991; Silverman, 1986; Fan and Marron, 1992).

The conditional expectation can also be seen as a kernel regression on a 0-1 binary response variable. Therefore, the conditional expectation with the estimation of the densities as in equation (7) can be expressed as

$$\hat{E}[y \mid x] = \frac{\sum_i y_i K\left(\frac{x - X_i}{h}\right)}{\sum_i K\left(\frac{x - X_i}{h}\right)} = \frac{\sum_i y_i K(x, X, h)}{\sum_i K(x, X, h)}, \quad (8)$$

with y a multivariate binary variable with y_i a vector of J elements with

$$y_{ij} = \begin{cases} 1 & \text{if alternative } j \text{ is choosen} \\ 0 & \text{otherwise} \end{cases}$$

Up to now a general description of a kernel density estimation has been given. In the marketing context some particularities exist, e.g. the fact that the explanatory variables have a discrete or binary character.

3.1.2 Extension of the kernel density estimation to marketing data

Usually kernel density estimation is applied for continuous, and most of the time multivariate, data. In the application of this nonparametric method in marketing we have the special situation of mixed (binary and continuous) explanatory variables. We therefore need a suitable modification of the estimator for this case. The use of kernel density estimation for binary data has been well known for a long time (e.g., Aitchison and Aitken, 1976; Silverman, 1986). The idea is to build a kernel similar to the spherical normal kernel. In the k-dimensional binary space $B^k = \{0,1\}^k$, the needed kernel for binary variables is defined as

$$K(x, X, b) = b^{k-d(x,X)} (1-b)^{d(x,X)} \quad (9)$$

where *x* is the variable at which the density should be estimated, *X* the explanatory variable and *b* the smoothing parameter with $\frac{1}{2} \leq b \leq 1$. The distance function *d*(*x,X*) is defined as a measure of the disagreements between *x* and *X* with

$$d(x,X) = (x-X)^T (x-X). \qquad (10)$$

The analyst should determine the value of the smoothing parameter *b*. It describes the weight that a binary observation is given that lies not completely at the point where the estimate is wanted. A value close to 1 gives those not totally equal observations little importance and a value close to ½ produces a uniform weighting of right and wrong specified observations. The value close to 1 produces the effect of a small, but existing, weighting of values "close" to the one wanted. In a classic continuous kernel approach, the value of *b* would equal 1, which leads to a very strict separation of "right" and "wrong". This is not wanted here for the binary variables.

Now we have a kernel density estimator for strict continuous (equation (7)) or strict binary (equation (9)) data, and so both expressions must be combined to a mixed nonparametric density estimator. This model is given in the form of

$$\begin{aligned} K(x,X,b,h) &= b^{k_1 - d_1(x,X)} \bullet (1-b)^{d_1(x,X)} \\ &\bullet \frac{1}{h^{d_2}} \bullet \frac{15}{16} \left\{ 1 - \left[\frac{1}{h} d_2(x,X) \right]^2 \right\} \bullet I \left\{ \left| \frac{d_2(x,X)}{h} \right| < 1 \right\} \end{aligned} \qquad (11)$$

where $d_1(x,X)$ is the distance function for the binary data as defined before in equation (10), and $d_2(x,X)$ is the Euclidic distance of the continuous variables. The value of k_1 describes the number of binary components and k_2 the number of the continuous components in the explanatory variables. For the continuous components of the observation the Quartic kernel was chosen (e.g., Härdle, 1991), which is a common choice for a kernel.

Now the conditional expectation can be written with respect to the mixed kernel of equation (11) and due to equation (8) as

$$\hat{E}[y \mid x] = \frac{\sum_i y_i K(x, X, b, h)}{\sum_i K(x, X, b, h)}$$

to model the brand choice decision. The conditional expectation is modeled for each brand separately, and from there the choice probabilities dependent on the explanatory variables can be obtained. No assumption like the "utility maximization" for the MNL is made. Moreover, it is not necessary to make an assumption about the distribution of the random component as in the MNL. These two points represent the main differences between nonparametric density estimation and parametric models (e.g., the MNL). The nonparametric approach affords a wide field of applications but often some parameters are wanted to calculate market shares or to make predictions. With a pure nonparametric model it is not possible to give these parameter values. This leads us to another possible brand choice model formulation, a semiparametric approach.

3.2 The semiparametric approach

Brand choice models can also be specified and estimated in a semiparametric way. This type of model is closer to the MNL model than the purely nonparametric methods. Semiparametric models specify one part of the MNL in the common parametric way (in our approach the error term distribution), but the other component (the utility function) is formulated by a nonparametric method. The resulting semiparametric model is called the "Generalized Additive Model" (GAM) (Hastie and Tibshirani, 1986; 1987; 1990). In modeling brand choice, Abe (1997) introduced the method to the choice modeling research area. He worked with direct additive components in the specification of the utility function. Our model has a more general formulation with nonparametric modeling of the explanatory variables.

3.2.1 The general formulation of a GAM

The general form of a GAM can be written as

$$E[y \mid x] = G\left(\sum_{k=1}^{K} f_k(x_k) \right). \qquad (12)$$

Here, $G(\cdot)$ is a logistic link function with

$$G(x) = \frac{1}{1 + exp(-x)}, \qquad (13)$$

which implies a parametric description of the error term. The terms of f_k describe one-dimensional nonparametric functions, which must be estimated, and x_k are the explanatory variables. In this formulation, the GAM is close to the classic MNL approach with an additive, but nonparametric, utility function f_k and with the same error term distribution as in the MNL, a logistic i.i.d. one for the error term differences. There exist many alternatives to model the influence of the explanatory variables in a nonparametric way. In the statistical context, modeling by the GAM approach is often used, because it gives the benefit of a large model class including a proved theory.

3.2.2 Application of the GAM theory to brand choice models

The general formulation of the GAM has the limitation of being only formulated for continuous explanatory variables. But the usual data sets in the choice-modeling context include discrete, e.g. binary, explanatory variables as well as continuous ones.

For this kind of data structure a new approach is needed that allows binary variables in the model formulation. This extended approach (e.g., Hastie and Tibshirani, 1990) has the form of

$$E[y \mid x] = G\left(\sum_{k=1}^{K} f_k(x_k) + \beta^T x_l \right). \qquad (14)$$

Here, β describes the parameter vector of the linear part of the model, which must be estimated. The terms f_k and x_k have the same meaning as described before. All binary explanatory variables must be included in x_l. Continuous variables can be modeled parametrically by inclusion in x_l, or nonparametrically by inclusion in x_k. The use of the linear formulation of some explanatory variables (in x_l) has the advantage of using much less computational time for estimation, because the nonparametric functions f_k needs most of the estimating time. The extended GAM supplies a good starting point to model choice behavior in a semiparametric approach. Modeling the nonparametric functions f_k and the parameter β supplies two possibilities to act with the resulting estimates. One way of using the results is to calculate the conditional expectation E[$y|x$] as described in equation (14). The other possibility is to get ideas of a functional form of the underlying explanatory variables x_k described by nonparametric functions f_k. Plotting the explanatory variable versus its nonparametric representation should provide some clues. This functional form supplies one possibility of how the explanatory variable could be included in the classic MNL approach. These two possibilities of using the estimates from the extended GAM open up a wide field of research opportunities.

3.2.3 Estimation algorithm for the GAM

Estimation of the GAM can be performed by the backfitting algorithm, introduced by several authors (e.g., Friedman and Stuetzle, 1981; Hastie and Tibshirani, 1986; Buja et al., 1989; Hastie and Tibshirani, 1990). This method works with variance decomposition of the additive part of the model as described in equation (14). It can also be described as a projection from a Hilbert space into a lower dimensional subspace. The backfitting algorithm deals with the assumption of an additive form of the “true” underlying model and splits the whole variance into components. But if the underlying model is misspecified (e.g. by omitting variables or interactions), the estimates for the nonparametric functions f_k can only be interpreted with care, because you do not know exactly which other additional parts are in the estimated f_k. Here you deal with a bias which cannot be attached to the separate explanatory variables. Another problem of this method consists in the lack of statistical measures, e.g. standard errors or other asymptotic properties. Also the exact behavior of the algorithm is unknown, even the convergence to the correct nonparametric functions is not established (Hastie and Tibshirani, 1990, pp. 117-118). The advantage of this estimation method is that it is an established algorithm, which is implemented in many common statistical software packages (e.g., S-Plus, R, XploRe etc.). And if the underlying data structure is specified correctly, the method gives usually quickly correct estimates for f_k.

There exists a second method for estimating a GAM, the marginal integration estimator (Chen et al., 1995; Linton and Nielsen, 1995; Linton and Härdle, 1996; Sperlich et al., 1997; Nielsen and Linton, 1998). This method estimates the marginal influence of the additive components specified in equation (14). It works on integration over a product kernel, usually the Nadaraya-Watson estimator. Because this method belongs to the well-known field of kernel estimation, all common statistical measures, e.g., bias, variance, confidence intervals and other asymptotic properties are available. Also, the estimated functional form of f_k is correct, even if the model is to a certain degree misspecified for the underlying data set. But if strong interaction effects exist in the data, the estimates for the f_k of the marginal effects are not sufficient for interpreting the results. The method is not well known and due to this, it is only implemented in one statistical software package, XploRe (Härdle et al., 2000).

If a GAM is estimated by both methods, and the estimation results for the functional forms of the explanatory variables differ greatly, then this is usually due to interaction terms, which are not taken into consideration by the backfitting algorithm. If different results are observed usually the underlying model structure is misspecified. Here, more precise analysis is needed. First approaches to estimating interaction terms via the marginal integration estimator are made by Sperlich et al. (1998).

4. Application of the model to panel data of consumer brand choice

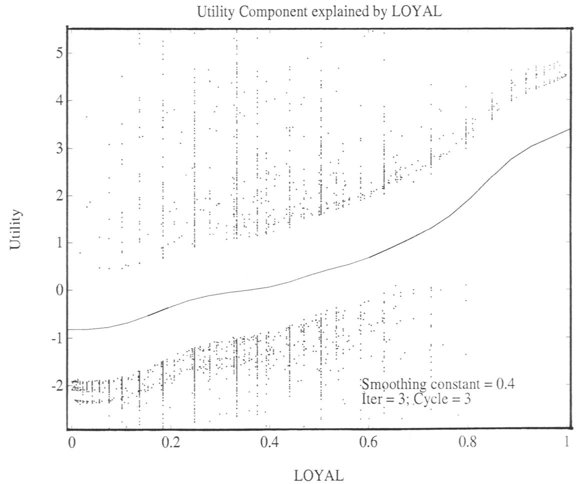
In this part of the paper we apply the two estimation algorithms for a GAM to a real data set in a consumer research setting. GfK, Germany, provided the data. They contain panel purchase records at one store over a period of 104 weeks. Also included are price and binary promotion indicator variables (feature and display) for each brand. We create a subset of the data by extracting purchases of panelists who have bought only one of the three leading brands. This results in a database with 2651 purchases made by 964 households.

The aim of the analysis is first to discover what kind of shape the response function of the marketing instrument price has, controlling for other impact variables (loyalty), instruments (promotion) and the activities of the competing brands using backfitting. The structure of the model is comparable to a model of van Heerde et al. (1998) of the shape of the price response with store level data. Second, we use marginal integration estimation to estimate these effects again, but allowing for interaction between the key variables. If the results are different we are able to conclude that the model based only on direct (order 1) effects may be misspecified.

For the model specification, we use two continuous explanatory variables, *PRICE* and *LOYALTY*, and one binary explanatory variable, *PROMOTION*. *LOYALTY* is defined comparable to Guadagni and Little (1983), a weighted geometric sum over past purchases, to capture household heterogeneity through the purchase history. *PROMOTION* is defined to be 0 if neither feature nor display occurred and 1 otherwise. This is defined this way due to high correlation between the two promotional activities feature and display.

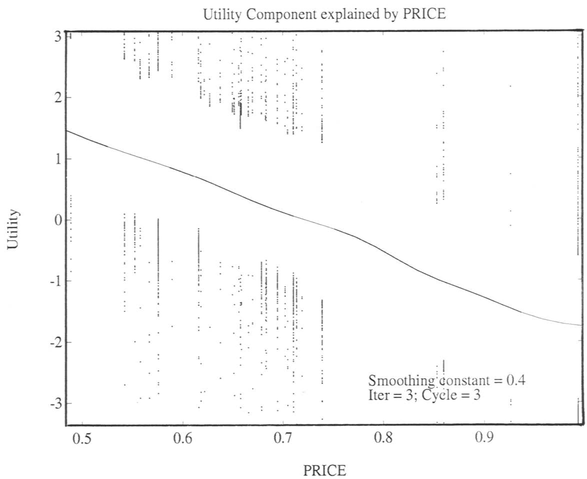
First, we estimated this semiparametric model with backfitting. As shown in Figures 1a and 1b, utility increases with *LOYALTY* in a slightly nonlinear fashion and decreases linearly with *PRICE*.

Figure 1a: Estimation results of the continuous variables made by backfitting



Source: ZUMA data of GfK Consumer Panel 1995, own calculations

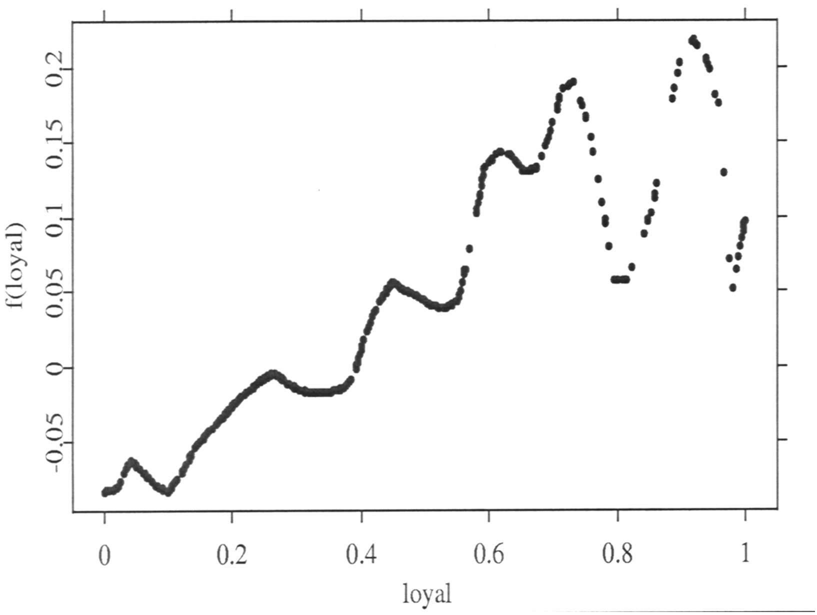
Figure 1b: Estimation results of the continuous variables made by backfitting



Source: ZUMA data of GfK Consumer Panel 1995, own calculations

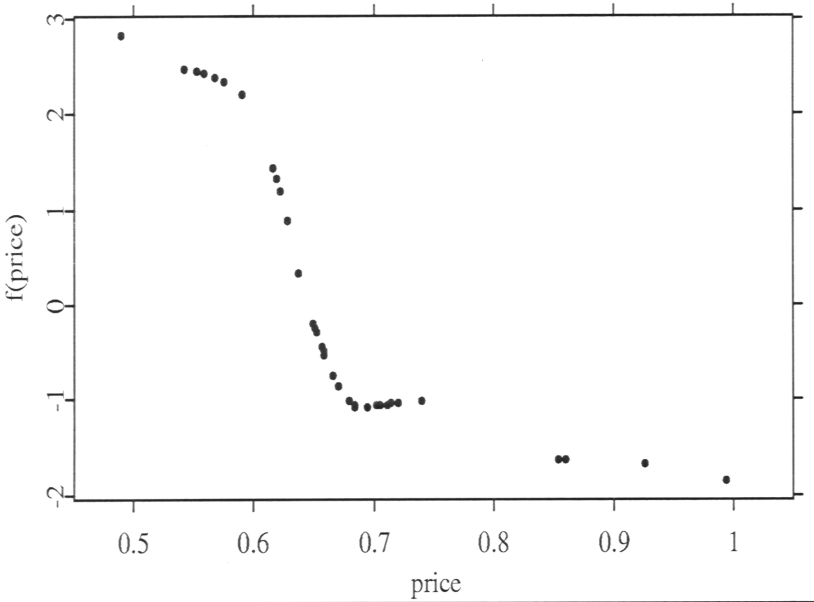
The model was also estimated with marginal integration. The estimation of this small model took a much larger computation time (50x). The results are shown in Figures 2a and 2b. The functional form of *LOYALTY* is again described in a linear increasing way (the slopes at the right side are due to bandwidth effects). But the estimated functional form of *PRICE* indicates a polynomial form of order 3, which is far from a linear form. The form could also be interpreted as a "loss and gain" function, which is often assumed for a price process. This estimation result is different from the one estimated by the backfitting algorithm which leads to the assumption that the underlying data structure may have an interaction effect of *PRICE* with other variables in the model. Therefore the estimation of the model by the backfitting procedure might produce results, which in this case are not valid and may be misleading if the interaction is ignored.

Figure 2a: Estimation results of the continuous variables made by marginal integration



Source: ZUMA data of GfK Consumer Panel 1995, own calculations

Figure 2b: Estimation results of the continuous variables made by marginal integration



Source: ZUMA data of GfK Consumer Panel 1995, own calculations

We have to keep in mind that the marginal integration estimation indicated an additional effect due to the different shapes of the functions of the price. To get a valid model a search for and integration of an additional effect might be necessary.

5. Summary

From the statistical point of view a nonparametric formulation of a brand choice model (NDE) is a powerful alternative to the logit model. But in the marketing context, researchers in general want to have parameter values to make predictions or to estimate market shares. This leads to a semiparametric model (GAM) formulation with two possible ways of using the results. One is to perform estimation of choice probabilities, but there one is confronted with the same problem as in the nonparametric approach, because no parameters are estimated for the nonparametric part of the model. The second possibility of a semiparametric model formulation overcomes this problem. In addition, with the estimation results a modified parametric model formulation can be estimated. This also gives the possibility to work with the parameter values to estimate market shares or make predictions. Especially for this use of modeling, the underlying data structure should be detected correctly. Therefore, two different estimation algorithms for a GAM were presented and the application of the semiparametric model to a real data set was reported. The estimations were made by the two common algorithms, backfitting and marginal integration, and are compared to each other. An interaction effect in the variable price in the data set was discovered, which leads to the need of additional studies of the data set.

Contact

Yasemin Boztuğ and Lutz Hildebrandt
Humboldt University of Berlin
Institute of Marketing; SFB 373
Spandauer Str. 1
D-10178 Berlin
boztug@wiwi.hu-berlin.de

References

Abe, Makoto, 1995, A Nonparametric Density Estimation Method for Brand Choice using Scanner Data. In: *Marketing Science*, 14 (3), 300-325.

Abe, Makoto, 1997, A Generalized Additive Model for Discrete Choice Data. Technical report, The University of Illinois at Chicago.

Aitchison, J., C. G. G. Aitken, 1976, Multivariate binary discrimination by the kernel method. In: *Biometrika*, 63, 413-420.

Bartels, Knut, Yasemin Boztuğ, Marlene Müller, 2000, Testing the multinomial logit model. In: Wolfgang Gaul, Reinhold Decker (Hg.), *Classification and Information Processing at the Turn of the Millennium*, Heidelberg: Springer.

Ben-Akiva, Moshe, Steven R. Lerman, 1985, *Discrete Choice Analysis*. Cambridge: The MIT Press.

Ben-Akiva, Moshe, Daniel McFadden, Makoto Abe, Ulf Böckenholt, Denis Bolduc, Dinesh Gopinath, Takayuki Morikawa, Venkatram Ramawamy, David Revelt, Dan Steinberg, 1997, Modeling Methods for Discrete Choice Analysis. In: *Marketing Letters*, 3 (3), 273-286.

Buja, Andreas, Trevor Hastie, Robert Tibshirani, 1989, Linear Smoothers and Additive Models. In: *The Annals of Statistics*, 17 (2), 453-555 (with discussion).

Chen, R., Wolfgang Härdle, Oliver B. Linton, E. Severance-Lossin, 1995, Nonparametric Estimation of Additive Separable Regression Models. Discussion Paper, Sonderforschungsbereich 373.

Chintagunta, Pradeep K., 1992, Estimating a multinomial probit model of brand choice using the method of simulated moments. In: *Marketing Science*, 3 (3), 386-407.

Chintagunta, Pradeep K., 1993, Investigating Purchase Incidence, Brand Choice and Purchase Quantity Decisions of Households. In: *Marketing Science*, 12 (2), 184-208.

Chintagunta, Pradeep. K., Bo E. Honore, 1996, Investigating the effect of marketing variables and unobserved heterogeneity in a multinomial probit model. In: *International Journal of Research in Marketing*, 13, 1-15.

Erdem, Tülin, 1996, A Dynamic Analysis of Market Structure Based on Panel Data. In: *Marketing Science*, 15 (4), 359-378.

Erdem, Tülin, Michael P. Keane, 1996, Decision-Making Under Uncertainty: Capturing Dynamic Brand Choice Processes in Turbulent Consumer Goods Market. In: *Marketing Science*, 15 (1), 1-20.

Fader, Peter S., James M. Lattin, John D. C. Little, 1992, Estimating Nonlinear Parameters in the Multinomial Logit Model. In: *Marketing Science*, 11(4), 372-385.

Fan, Jianqing, James S. Marron, 1992, Best Possible Constant for Bandwidth Selection. In: *The Annals of Statistics*, 29 (4), 2057-2070.

Friedman, Jerome H., Werner Stuetzle, 1981, Projection pursuit regression. In: *Journal of the American Statistical Association*, 76, 817-823.

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

Gupta, Sachin, Pradeep K. Chintagunta, 1994, On Using Demographic Variables to Determine Segment Membership in Logit Mixture Models. In: *Journal of Marketing Research*, 31, 128-136.

Härdle, Wolfgang, 1991, *Smoothing Techniques With Implementation in S*. New York: Springer.

Härdle, Wolfgang, Sigbert Klinke, Marlene Müller, 2000, *XploRe Learning Guide*. Berlin: Springer.

Hastie, Trevor J., Robert J. Tibshirani, 1986, Generalized Additive Models. In: *Statistical Science*, 1 (3), 297-318.

Hastie, Trevor J., Robert J. Tibshirani, 1987, Generalized Additive Models: Some Applications. In: *Journal of the American Statistical Association*, 82 (398), 371-386.

Hastie, Trevor J., Robert J. Tibshirani, 1990, *Generalized Additive Models*. London: Chapman and Hall.

Heerde, Harald J. van, Peter S. H. Leeflang, Dick R. Wittink, 1998, Semiparametric analysis to estimate the deal effect curve. Working Paper, University of Groningen.

Horowitz, Joel L., Denis Bolduc, Suresh Divakar, John Geweke, Füsun Gönül, Vassilis Hajivassiliou, Frank S. Koppelman, Michael P. Keane, Rosa L. Matzkin, Peter Rossi, Paul Ruud, 1994, Advances in Random Utility Models. In: *Marketing Letters*, 5, 311-322.

Horowitz, Joel L., Wolfgang Härdle, 1996, Direct Semiparametric Estimation of Single-Index Models with Discrete Covariates. In: *Journal of the American Statistical Association*, 91 (436), 1632-1640.

Hosmer, David W., Stanley Lemeshow, 1989, *Applied Logistic Regression*. New York: Wiley.

Kannan, P. K., Gordon P. Wright, 1991, Modeling and Testing Structured Markets: A Nested Logit Approach. In: *Marketing Science*, 10 (1), 58-82.

Linton, Oliver B., Wolfgang Härdle, 1996, Estimation of additive regression models with known links. In: *Biometrika*, 83 (3), 529-540.

Linton, Oliver B., Jens Perch Nielsen, 1995, A kernel method of estimating structured nonparametric regression based on marginal integration. In: *Biometrika*, 82 (1), 93-100.

Manrai, Ajay K., 1995, Mathematical models of brand choice behavior. In: *European Journal of Operational Research*, 82, 1-17.

Matzkin, Rosa L., 1991, Semiparametric Estimation of Monotone and Concave Utility Functions for Polychotomous Choice Models. In: *Econometrica*, 59 (5), 1315-1327.

McFadden, Daniel, 1974, Conditional logit analysis of qualitative choice behavior. In: Zarembka, P. (Hg.): *Frontiers in Econometrics*, New York, 105-142.

Nielsen, Jens Perch, Oliver B. Linton, 1998, An optimization interpretation of integration and back-fitting estimators for separable nonparametric models. In: *Journal of the Royal Statistical Society, Series B*, 60 (1), 217-222.

Papatla, Purushottam, 1996, A Multiplicative Fixed-effects Model of Consumer Choice. In: *Marketing Science*, 15 (3), 243-261.

Silverman, B. W., 1986, *Density Estimation for Statistics and Data Analysis*. London: Chapman and Hall.

Sperlich, Stefan, Oliver B. Linton, Wolfgang Härdle, 1997, A Simulation Comparison between Integration and Backfitting Methods of Estimation Separable Nonparametric Regression Models. Discussion Paper, Sonderforschungsbereich 373.

Sperlich, Stefan, Dag Tjøstheim, Lijian Yang, 1998, Nonparametric Estimation and Testing of Interaction in Additive Models. Discussion Paper, Sonderforschungsbereich 373.