Nicht- und semiparametrische Markenwahlmodelle im Marketing

Yasemin Boztuğ
Lutz Hildebrandt

Humboldt-Universität zu Berlin
Institut für Marketing II
SFB 373

13. November 1998

1 Einführung

Ein zentrales Untersuchungsziel der Konsumentenverhaltensforschung im Marketing ist die Erklärung des Kaufverhaltens bei Produkten und Marken von Gütern des täglichen Bedarfs (Trommsdorff, 1993). In die quantitative Modellierung dieses Verhaltens fließen Inputvariablen wie Käufermerkmale und Marketinginstrumente, intervenierende Variablen wie z. B. Einstellungen und/oder Präferenzen, und die Reaktionen der Käufer wie Kauf, Kaufzeitpunkt u. ä., ein. Hierbei kann die Käuferschaft als Gesamtheit (aggregierte Nachfragemodelle) oder auf individueller Basis (disaggregierte Nachfragemodelle) betrachtet werden. Liegen Scanner Panel Daten einzelner Kaufhandlungen vor, wird im allgemeinen auf individueller Basis vorgegangen. Die existierenden Modelle liefern Auskunft über Zusammenhänge zwischen dem Kaufverhalten und einer Vielzahl von Einflußgrößen (Hruschka, 1996). Der vorliegende Beitrag stellt das Markenwahlverhalten in den Mittelpunkt der Modellbildung.

Bei den Markenwahlmodellen besteht für den Forscher das Problem der Entscheidung zwischen der Komplexität des Modells und der Einfachheit seiner Schätzung (Ben-Akiva et al. 1997). Ein Beispiel für eine relativ einfach durchzuführende parametrische Schätzung bildet das konditionale Logit Modell (CLM) (Guadagni & Little, 1983). Der Preis der Einfachheit ist jedoch, daß sehr viele, zum Teil unrealistische Annahmen über die Struktur des zugrundeliegenden Wahlmodells und der Nutzenfunktion getroffen werden müssen. Diese Annahmen schränken die Aussagefähigkeit des Modells stark ein. Die Forschung der vergangenen Jahre hat sich deshalb fast ausschließlich mit der Überwindung dieser Annahmen in parametrischen Modellen befaßt (z. B. Kannan & Wright, 1991; Chintagunta, 1992; Fader, Lattin & Little, 1992; Chintagunta, 1993; Gupta & Chintagunta, 1994; Erdem & Keane, 1996; Erdem, 1996; Papatla, 1996). Eine Lösung diesem Dilemma zu entgehen können nicht- bzw. semiparametrische Modellansätze liefern. Sie bieten einen großen Spielraum in der Formulierung des Modells, basieren aber trotzdem auf einer statistischen Theorie, die eine relativ einfache Schätzung erlaubt. Der Begriff der nicht- bzw. semiparametrischen Modelle ergibt sich durch die Aufteilung des Modells (in Anlehnung an die CLM-Begrifflichkeit) in eine systematische Nutzenfunktion ($f(x)$) und eine zufällige Komponente (ϵ) mit

$$E[y|x] = f(x) + \epsilon.$$

Je nach Spezifikation der Komponenten des Gesamtmodells liegt entweder ein nicht- oder semiparametrisches Modell vor. Eine schematische Darstellung dieser Unterteilung ist in Tabelle 1 zu sehen.

Modelltyp	systematische Nutzenfunktion	zufällige Komponente	resultierende Methode
parametrisch	parametrisch	parametrisch	CLM
semiparametrisch I	parametrisch	verteilungsfrei	verschiedene
semiparametrisch II	nichtparametrisch	parametrisch	GAM
nichtparametrisch	nichtparametrisch	verteilungsfrei	NDE

Tabelle 1: Modellierungsmöglichkeiten der Markenwahl

Für die Modellierung der Markenwahl durch den Typ 'semiparametrisch I', d. h. mit einer parametrischen Nutzenfunktion und einer zufälligen Komponente, die verteilungsfrei modelliert wird, liegt eine umfassende Literatur aus der statistisch-ökonomischen Methodenforschung vor (z. B. Horowitz et al. 1994; Horowitz & Härdle, 1996; Matzkin, 1991; Chintagunta & Honore, 1996).

'Semiparametrisch II' behandelt den Fall, in dem die Nutzenfunktion nichtparametrisch, aber die zufällige Komponente parametrisch modelliert wird. In der Modellierung für 'nichtparametrisch' werden sowohl die systematische Funktion als auch die zufällige Komponente nichtparametrisch spezifiziert. Erste Ansätze dieser Modelltypen im Marketing finden sich in Abe (1995; 1997) und Matzkin (1993). Aus den Typen der Modelle ergibt sich dann die jeweils zu verwendende Schätzmethode. 'CLM' steht dabei für das 'Conditionale Logit Model', 'GAM' für 'Generalized Additive Models' und 'NDE' für 'Nonparametric Density Estimation'.

Die nichtparametrischen Ansätze werden im folgenden detailiert beschrieben und Weiterentwicklungen vorgestellt. Der Artikel ist wie folgt aufgebaut: Als erstes wird das konditionale Logit Modell (CLM) als Ausgangsmodell kurz erläutert. Dann wird die Modellierung eines Markenwahlmodells mit Hilfe der nichtparametrischen

Dichteschätzung (NDE) vorgestellt. Außerdem wird untersucht, wie diese allgemeine Methode auf die Datenstruktur angepaßt werden kann, die bei Markenwahlmodellen vorliegt (erklärende Variablen sind sowohl stetig als auch diskret). Anschließend folgt eine Erörterung des semiparametrischen Ansatzes der 'Generalized Additive Models' (GAM).

Auf der Grundlage von Scanner Panel Daten der GfK wird gezeigt, wie die Modellbildung der verschiedenen Typen durchgeführt werden kann und was sie leistet. Hierbei werden die Schätzergebnisse der unterschiedlichen Modellierungen verglichen und bewertet. Die Umsetzung der Modelle erfolgte mit dem interaktiven statistischen Software-Paket XploRe (Härdle, Klinke & Turlach, 1995).

2 Das Logit Modell

2.1 Das binäre Logit Modell

Das binäre Logit Modell modelliert das individuelle Wahlverhalten zwischen zwei Alternativen. Bei der Produkt- bzw. Markenwahl wird dem Konsumenten als Entscheidungsregel die Nutzenmaximierung unterstellt (Ben-Akiva & Lerman, 1985). Dies bedeutet, daß ein Konsument n die Alternative wählt, die seinen Nutzen U_n maximiert. Die Alternativen i und j bilden die Wahlmenge C_n mit $C_n = \{i, j\}$. Die Wahrscheinlichkeit $P_n(i)$, daß das Individuum n die Alternative i wählt, ist

$$P_n(i) = P(U_{in} \geq U_{jn}), \tag{1}$$

mit U_{in}, U_{jn} dem Nutzen der Alternativen i und j. Die Nutzenfunktion U_{in} aus Gleichung (1) läßt sich aufteilen in $U_{in} = V_{in} + \epsilon_{in}$ mit V_{in} als systematische Nutzenfunktion und ϵ_{in} als stochastische Komponente. Die Nutzenfunktion V_{in} wird linear spezifiziert mit

$$V_{in} = \beta' x_{in}, \tag{2}$$

wobei β den zu schätzenden Parameter und x_{in} die erklärende Variable (im allgemeinen die Merkmale einer Alternative i im Urteil des Individuums n) darstellt. Gleichung (1) kann nun bei Nutzenmaximierung folgendermaßen umgeschrieben werden

$$\begin{aligned} P_n(i) &= P(U_{in} \geq U_{jn}) \\ &= P(V_{in} + \epsilon_{in} \geq V_{jn} + \epsilon_{jn}) \\ &= P(\epsilon_{jn} - \epsilon_{in} \leq V_{in} - V_{jn}). \end{aligned} \tag{3}$$

Die Differenz der stochastischen Komponenten $\epsilon_{jn} - \epsilon_{in}$ aus Gleichung (3) wird in Logit Modellen als i.i.d. (unabhängig identisch) und logistisch verteilt angenommen. Unter den zuvor getroffenen Verteilungsannahmen über die Differenzen der stochastischen Komponenten läßt sich die Wahlwahrscheinlichkeit aus Gleichung (3) unter Verwendung von Gleichung (2) als Verhältnis vom Nutzen der Alternativen darstellen als

$$\begin{aligned} P_n(i) &= P(U_{in} \geq U_{jn}) \\ &= \frac{1}{1 + \exp(V_{jn} - V_{in})} \\ &= \frac{\exp(V_{in})}{\exp(V_{in}) + \exp(V_{jn})} \\ &= \frac{\exp(\beta' x_{in})}{\exp(\beta' x_{in}) + \exp(\beta' x_{jn})}. \end{aligned} \tag{4}$$

Die Werte von β lassen sich über eine Maximum-Likelihood-Schätzung bestimmen (Ben-Akiva & Lerman, 1985). Dabei sind sowohl die Vorzeichen als auch die absoluten Werte der Schätzung von großem Interesse. Das Vorzeichen gibt Auskunft über den Richtungszusammenhang, d. h. bei positivem Vorzeichen bedeutet eine Erhöhung der Einflußkomponente einer Marke eine größere Kaufwahrscheinlichkeit für diese Marke, bei negativem Vorzeichen kehrt sich dieser Einfluß um. Der Wert des Absolutbetrages des β-Vektors gibt bei der gleichen Skalierung der erklärenden Variablen Auskunft über die Stärke des Zusammenhangs, je größer der Wert ist, desto stärker ist der Zusammenhang zwischen der erklärenden und der zu erklärenden Variablen. Der Ansatz des binären Logit Modells kann auf J Alternativen, den multinomialen Fall, übertragen werden.

2.2 Das multinomiale Logit Modell

Das Prinzip der Nutzenmaximierung wird ebenfalls für den multinomialen Fall angenommen. Die Wahlmenge C_n enthält J Alternativen. Jedes Individuum wählt aus dieser Menge die Alternative aus, die für ihn den maximalen Nutzen besitzt. Damit kann der multinomiale Fall auf den binären Fall zurückgeführt werden, indem die Alternative, die den Nutzen maximiert, gegen die restlichen Alternativen gestellt wird. Damit verschmelzen sie als 'Restmenge' zu einer Alternative. Formal läßt sich das multinomiale Logit Modell beschreiben durch

$$\begin{aligned} P_n(i) &= P\left(U_{in} \geq \max_{\substack{j \in C_n \\ j \neq i}} U_{jn}\right) \qquad (5) \\ &= P(U_{in} \geq U_{jn}, \forall j \in C_n, j \neq i) \\ &= P(V_{in} + \epsilon_{in} \geq V_{jn} + \epsilon_{jn}, \forall j \in C_n, j \neq i) \\ &= P(\epsilon_{jn} - \epsilon_{in} \leq V_{in} - V_{jn}, \forall j \in C_n, j \neq i). \end{aligned}$$

Mit Gleichung (5) und der Annahme, daß die Differenzen der stochastischen Terme $\epsilon_{jn} - \epsilon_{in}$ wieder i.i.d. und logistisch verteilt sind, läßt sich die Wahrscheinlichkeit, daß der n-te Konsument Marke i wählt, gemäß Gleichung (4) darstellen als

$$\begin{aligned} P_n(i) &= \frac{\exp(V_{in})}{\sum_{j \in C_n} \exp(V_{jn})} \qquad (6) \\ &= \frac{\exp(\beta' x_{in})}{\sum_{j \in C_n} \exp(\beta' x_{jn})}. \end{aligned}$$

Die Werte für β werden wiederum über die Maximum-Likelihood-Methode bestimmt. Auch für die gewonnenen Schätzwerte gilt die Interpretion gemäß dem binären Logit Modell. Bis jetzt wurde für die systematische Nutzenfunktion V_{in} nur eine lineare Form angenommen. Im konditionalen Logit Modell wird die Struktur der Einflußgrößen differenzierter spezifiziert.

2.3 Das konditionale Logit Modell

Das konditionale Logit Modell stellt eine Alternative zum multinomialen Ansatz dar (Greene, 1990; Collett, 1991). Die Nutzenfunktion V_{in} kann auf verschiedene

Aspekte, die die Wahlentscheidung eines Individuums beeinflussen, zurückgeführt werden. So kann davon ausgegangen werden, daß Individuen bei einer Markenwahlentscheidung durch haushaltsspezifische Eigenschaften (individuenspezifisch) und die Produkteigenschaften (alternativenspezifisch) beeinflußt werden. Soll dies beachtet werden, läßt sich die erklärende Variable x_{in} aus Gleichung (2) in zwei Einflüsse $x_{in} = [z_{in}, w_n]$ mit z_{in} als alternativenspezifische und w_n als individuenspezifische Variable zerlegen. w_n variiert nicht über die Alternativen, da Merkmale wie die Anzahl der Kinder im Haushalt oder die Haushaltsgröße unabhängig von einer getroffenen Markenwahl sind. Werden nur diese individuenspezifischen Variablen zur Bestimmung einer Markenwahlwahrscheinlichkeit mit

$$P_n(i) = \frac{\exp(\alpha_i' w_n)}{\sum_{k=0}^{K-1} \exp(\alpha_k' w_n)} \tag{7}$$

berücksichtigt, so wird in der einschlägigen Literatur wiederum die Bezeichnung multinomiales Logit Modell (MNL) verwendet (Greene, 1990, S. 697). K ist die Anzahl der möglichen Ausprägungen der individuellen Eigenschaft. Damit die Schätzwerte für α bei der Schätzung identifizierbar sind, muß ein α-Wert a priori festgelegt werden (Greene, 1990), üblicherweise wählt man $\alpha_0 = 0$.

Gehen nur die alternativenspezifischen Variablen z_{in} als erklärende Variablen in die Schätzung ein mit

$$P_n(i) = \frac{\exp(\beta' z_{in})}{\sum_{j \in C_n} \exp(\beta' z_{jn})}, \tag{8}$$

so handelt es sich um das rein konditionale Logit Modell. Auch eine Kombination beider erklärender Variablen (individuenspezifisch und alternativenspezifisch) ist möglich, das Modell wird als multinomial-konditionales Logit Modell bezeichnet mit

$$P_n(i) = \frac{\exp(\beta' z_{in} + \alpha_i' w_n)}{\sum_{\substack{j \in C_n \\ k=0}}^{K-1} \exp(\beta' z_{jn} + \alpha_k' w_n)}. \tag{9}$$

Die Schätzung der Werte für β und α werden mit Hilfe einer konditionalen Maximum-Likelihood-Methode bestimmt (McFadden, 1974). Die Interpretation für die gewonnenen Ergebnisse für die Werte der alternativenspezifischen (β) und für die individuenspezifischen Parameterwerte (α) erfolgt analog des binären bzw. multinominalen Logit Modells.

2.4 Beschränkungen des Logit Modells

Die dargestellten Logit Modelle sind mit einigen Annahmen verbunden, die den Interpretations- und Anwendungsbereich des Modelltyps zum Teil erheblich einschränken.

Im folgenden werden einzelne Beschränkungen des Modells erörtert, die den Kern der Kritik am Logit Ansatz bilden. Die IIA-Annahme (Independence of Irrelevant Alternatives) unterstellt, daß der relative Nutzen einer Alternative gegenüber einer zweiten unabhängig seien sollte von der Existenz einer dritten. Dies ist anzuzweifeln: Angenommen, es gibt zwei Alternativen, die jeweils von der Hälfte der Individuen gewählt werden, also der relative Nutzen gleich $\frac{1}{2}$ beträgt. Wird eine farbliche Variante der ersten Alternative eingeführt, so würde sich nach der IIA-Annahme das relative Verhältnis der Nutzen zwischen den ersten beiden Alternativen gleich verändern und $\frac{1}{3}$ betragen. Wenn nun nicht mehr mit relativen Wahrscheinlichkeiten, sondern mit absoluten gerechnet wird, stellt sich das Problem der IIA nicht mehr.

Eine weitere Einschränkung stellt die Annahme der logistische Verteilung der Differenzen der stochastischen Komponenten dar. Es ist nicht unbedingt ersichtlich, warum die Differenzen der Störterme gerade dieser Verteilung folgen sollten, da es eine Vielzahl von Möglichkeiten für die Modellierung von Differenzen stochastischer Größen gibt.

Eine dritte Schwäche im Modellansatz der Logit Modelle besteht in der Annahme, daß die Nutzenfunktion einen linearen Charakter aufweist. Dadurch wird die Modellspezifikation sehr stark eingeschränkt. Auch hier läßt sich keine ökonomische Rechtfertigung für die Linearität anführen.

Es gibt verschiedene Ansätze, die im Kontext der parametrischen Modelle die Schwächen bzw. die beiden zuletzt genannten Modellannahmen versuchen zu umgehen, wie z. B. die Probit Modelle. Sie beruhen auf der Annahme normalverteilter Differenzen der Störgrößen. Probit Modelle sind jedoch bei multinomialler Spezifizierung rechentechnisch außerordentlich aufwendig und verlangen große Stichproben und sind von geringer Handhabbarkeit. Eine alternative Möglichkeit besteht in der expliziten Überprüfung der Annahmen des Logit Modells über nichtparametrische

Verfahren, wie z. B. in einer nichtparametrischen Modellierung des Fehlerterms. Oder es erfolgt die Prüfung der Form der Nutzenfunktion mit einer nichtparametrischen Schätzung.

Um die beiden wichtigsten Kritikpunkte am CLM ausschließen zu können (die logistische Modellierung der Differenzen der Fehlerterme und die Linearität der Nutzenfunktion), wird im folgenden der Einsatz des nichtparametrischen Modellansatzes vorgestellt.

3 Die nichtparametrische Dichteschätzung (NDE) für ein Wahlmodell

3.1 Der grundlegende Ansatz

Für die Anwendung nichtparametrischer Techniken werden große Beobachtungszahlen benötigt, deshalb ist es erst mit dem Aufbau Elektronischer Märkte und der Scannertechnik möglich geworden, diese Methoden auch im Marketing, speziell in der Markenwahlmodellierung, einzusetzen.

Bei der nichtparametrischen Schätzung wird die Wahlentscheidung als bedingter Erwartungswert formuliert, wobei als Bedingung die in der aktuellen Kaufsituation vorhandene Marketing-Mix-Konstellation angesehen wird. Die Marketing-Mix-Konstellation erfaßt praktisch den Vektor der Merkmale einer Produktwahlalternative. Es wird der bedingte Erwartungswert $E[y|x]$ gesucht, wobei x die zuvor beschriebene Marketing-Mix-Bedingung darstellt und y die Wahlentscheidung beschreibt. Diese Wahlentscheidung wird binär kodiert, d. h. der bedingte Erwartungswert $E[y|x]$ wird für jede Marke einzeln bestimmt. Da y binär ist, gilt die folgende Identität:

$$E[y_j|x] \equiv P(y_j|x) \qquad \text{mit} \quad y_j = \begin{cases} 1 & \text{Marke } j \text{ wurde gekauft} \\ 0 & \text{sonst} \end{cases}$$

Mit der Regel von Bayes läßt sich der bedingte Erwartungswert für die Wahl jeder

Marke darstellen als

$$E[y|x] = \frac{p(y)f(x|y)}{f(x)}, \tag{10}$$

wobei in Gleichung (10) bereits die Binärität von y berücksichtigt wurde, sonst müßte für $p(y)$ ein $f(y)$ stehen.

Die Wahrscheinlichkeit für die Wahl von y mit $p(y)$ läßt sich empirisch leicht bestimmen über

$$p(y) = \frac{\text{Anzahl getätigter Käufe}}{\text{Anzahl möglicher Käufe}}.$$

Die noch fehlenden Größen zur Bestimmung des bedingten Erwartungswertes $E[y|x]$ sind die Dichten $f(x|y)$ und $f(x)$. Eine Dichte $f(x)$ läßt sich mit Hilfe eines nichtparametrischen Dichteschätzers bestimmen. Der allgemeine Kerndichteschätzer $\hat{f}(x)$ für $f(x)$ lautet

$$\hat{f}_h(x) = \frac{1}{nh^d} \sum_{i=1}^{n} K\left(\frac{x - X_i}{h}\right) = \frac{1}{nh^d} \sum_{i=1}^{n} K(x, h) \tag{11}$$

mit n als Anzahl der Beobachtungen, X_i sei die i-te Beobachtung der d-dimensionalen erklärenden Variablen X (d entspricht hier der Anzahl der Komponenten in x), x bezeichne die Stelle, an der der Dichteschätzer bestimmt werden soll, h sei der Glättungsparameter und $K(x, h)$ eine Kernfunktion.

h als Glättungsparameter beschreibt die Balance zwischen dem Bias und der Varianz der Schätzung. Mit h wird festgelegt, welche Spannweite von Beobachtungen jeweils für die Schätzung von $\hat{f}$ eingestzt wird. Der Wert von h kann über verschiedene Methoden vorgegeben werden. Als wichtigste sind zu nennen die 'Rule of thumb' (Silverman, 1986), die 'Cross-Validierung' und 'Plug-in Methoden' (Härdle, 1991; Fan & Marron, 1992).

Die in Gleichung (11) eingeführte Kernfunktion $K(\cdot)$ beschreibt, in welcher funktionalen Form die in $(\cdot)$ bestimmten Distanzen geglättet werden. Sie muß verschiedene Bedingungen erfüllen: Sie sei nichtnegativ, besitze Ableitungen höherer Ordnung und erfülle noch weitere allgemeine Glattheitsbedingungen. Es hat sich gezeigt, daß nicht die Wahl des Kerns, sondern daß die Wahl der Bandweite (Glättungsparameter) die Hauptrolle für die Güte der Anpassung spielt (Härdle, 1991; Silverman, 1986; Fan & Marron, 1992).

Das Modell aus Gleichung (10) kann auch als Kernregression auf eine 0-1 kodierte binäre Response-Variable aufgefaßt werden. Mit der Schätzung $\hat{f}_h(x)$ aus Gleichung (11) läßt sich die bedingte Wahrscheinlichkeit bzw. der bedingte Erwartungswert darstellen als

$$\hat{P}(y|x) \equiv \hat{E}[y|x] = \frac{\sum_i y_i K\left(\frac{x-X_i}{h}\right)}{\sum_i K\left(\frac{x-X_i}{h}\right)} = \frac{\sum_i y_i K(x,h)}{\sum_i K(x,h)}. \tag{12}$$

Gleichung (12) beschreibt die allgemeine Modellform, mit der der gesuchte bedingte Erwartungswert geschätzt werden kann.

3.2 Erweiterung der NDE auf die spezielle Struktur von Marketingdaten mit binären und stetigen erklärenden Variablen

Die in Abschnitt 3.1 dargestellte Methodik der Kerndichteschätzung wird üblicherweise für stetige erklärende Variablen verwendet. Jedoch ist es in der Problematik der Markenwahlmodelle notwendig, daß als erklärende Variablen auch binäre (wie z. B. der Einsatz von Display ja/nein als Variable) zulässig sind. Diese binären Variablen sind nun in den zuvor dargestellten allgemeinen Kerndichteschätzer aus Gleichung (12) zu integrieren.

In Silverman (1986) werden unter anderem Kerne für rein binäre erklärende Variablen bzw. für gemischte (binäre und stetige) vorgestellt. Ein möglicher Kern für k_1 binäre und k_2 stetige Komponenten der erklärenden Variable lautet danach

$$\begin{aligned} K(x,b,h) &= b^{k_1-d_1(x,X)}(1-b)^{d_1(x,X)} \\ &\quad * \frac{1}{h^{k_2}} * \frac{15}{16}\left[1-\left(\frac{1}{h}d_2(x,X)\right)^2\right]^2 \\ &\quad * I\left[\left|\frac{1}{h}d_2(x,X)\right| \leq 1\right]. \end{aligned} \tag{13}$$

b sei der Glättungparameter des Kerns für die diskreten Komponenten mit $0,5 \leq b \leq 1$. b wird meistens nahe bei 1 gewählt, denn auf diese Weise werden den Beobachtungen, die nahe am ‘wahren‘ Wert liegen, die größere Bedeutung zugemessen.

h sei wiederum der Glättungsparameter der Kernfunktion der stetigen erklärenden Variablen. Für die stetigen Komponenten wurde der Quartic Kern (Härdle, 1991) gewählt. x sei die Stelle, an der die Schätzung bestimmt werden soll, und X stellt die erklärende Variable dar.

In Gleichung (13) stehen d_1 und d_2 für zwei Distanzfunktionen der binären respektive stetigen Komponenten der erklärenden Variablen. d_1 wird beschrieben durch die Unterschiede bzw. Übereinstimmungen der beobachteten Werte zu den 'wahren', für d_2 wird zumeist die Euklidische Distanz gewählt.

Der bedingte Erwartungswert mit dem in Gleichung (13) vorgestellten gemischten Kern lautet dann

$$\hat{E}[y|x] = \frac{\sum_i y_i K(x, b, h)}{\sum_i K(x, b, h)}. \tag{14}$$

Mit der Berechnungsvorschrift in Gleichung (14) kann die Markenwahl modelliert werden, wobei für jede Marke i der bedingte Erwartungswert $\hat{E}[y|x]$ getrennt berechnet wird.

Aus den bedingten Erwartungswerten für die einzelnen Marken lassen sich ihre Kaufwahrscheinlichkeiten in Abhängigkeit der erklärenden Variablen bestimmen.

Hier tritt auch einer der problematischen Aspekte der nichtparametrischen Verfahren gegenüber den parametrischen Verfahren in der Anwendungspraxis auf. Es können keine Parameterwerte zur direkten Prognose geliefert werden. Dies ist ein Problem, da in der klassischen Markenwahlmodellierung insbesondere die Einflußstärken von Marketing-Mix-Variablen und von Persönlichkeitsmerkmalen über die dort gewonnenen Parameterwerte versucht werden zu ermitteln, um die Markenwahl zu prognostizieren und Marktanteile zu schätzen. Das Verfahren des NDE ist dagegen eher zur Datenmodellierung geeignet. Durch Änderung einzelner Größen im Modell wird es nur implizit möglich, den Einfluß zugrundeliegender Einflußvariablen auf die Markenwahlwahrscheinlichkeit zu bestimmen.

Ein Kompromiß zwischen dem nichtparametrischen Modellansatz und den parametrischen Modellansätzen bildet die semiparametrische Vorgehensweise, auf die im folgenden eingegangen wird.

4 Die semiparametrische Schätzung für ein Wahlmodell

4.1 Der grundlegende Ansatz

Geht man nun von der in Abschnitt 2 dargestellten Struktur des multinomialen-konditionalen Logit Modell aus, so ergeben sich durch die Modellkomponenten zwei Möglichkeiten, ein rein parametrisches Wahlmodell zu einem semiparametrischen umzuformen. Bei einer Aufteilung eines Markenwahlmodells in eine systematische Nutzenfunktion und eine zufällige Komponente kann jeweils eine dieser beiden Modellteile nichtparametrisch geschätzt werden. Das resultierende Gesamtmodell hat dann die Form des semiparametrischen Markenwahlmodells, bei dem für den nichtparametrischen Teil über die Vorgehensweise zu entscheiden ist. Die Modellierung der Zufallskomponente durch nichtparametrische Methoden wird in einer Vielzahl von Artikeln behandelt (Härdle, 1991; Silverman, 1986; Fan & Marron, 1992).

Eine Formulierung eines Modells mit einer nichtparametrischen Nutzenfunktion wurde zum ersten Mal in Abe (1997) aufgegriffen. Dort wird ein semiparametrisches Modell mit direkten additiven Komponent als Spezifikation der Nutzenfunktion bestimmt und basierend auf dem Ansatz von Hastie & Tibshirani (1990) geschätzt. Die Modellformulierung bleibt nah an der des CLM durch die Formulierung einzelner Nutzenfunktionen pro Marke.

Im folgenden wird ein allgemeinerer semiparametrischer Ansatz zur Schätzung von Kaufwahrscheinlichkeiten vorgestellt. Wie auch in Gleichung (10) wird der bedingte Erwartungswert $E[y|x]$ mit y als Wahlentscheidung und x als Marketing-Mix-Variable gesucht. Für die Modellierung der nichtparametrischen Nutzenfunktion kann auf den Ansatz der 'Generalized Additive Models' (GAM) von Hastie & Tibshirani (1986; 1987; 1990) zurückgegriffen werden. In der allgemeinen Formulierung eines GAMs wird der bedingte Erwartungswert beschrieben durch

$$E[y|x] = G\left(\sum_p f_p(x_p)\right). \tag{15}$$

f_p sind eindimensionale nichtparametrische Funktionen und G sei eine Link-Funktion. Diese verbindet die geschätzten nichtparametrischen additiven Komponenten. Wenn für G ein logistischer Link gewählt wird mit

$$G(x) = \frac{1}{1 + \exp(-x)}, \tag{16}$$

werden gleichzeitig die Differenzen der zufälligen Komponenten des Markenwahlmodells als logistisch verteilt festgelegt. Damit liefert das Modell aus Gleichung (15) einem semiparametrischen Ansatz zur Schätzung des bedingten Erwartungswertes.

Es existieren zwei verschieden Schätzmethoden zur Bestimmung der eindimensionalen nichtparametrischen Funktionen f_p. Der Ansatz von Hastie & Tibshirani (1990) basiert auf der Methode des Backfitting. Hierbei wird eine Varianzzerlegung auf die einzelnen erklärenden Variablen vorgenommen. Diese Methode hat den Vorteil, daß sie in jedem statistischen Standard-Software-Paket integriert ist, und die Schätzung schnell und einfach durchführbar ist. Da jedoch der Backfitting-Algorithmus die Gesamtvarianz auf die im Modell vorgegebene Struktur aufteilt, besteht die Gefahr, daß für im Modell spezifizierte erklärende Variablen x_p Schätzungen ihrer f_p bestimmt werden, die nicht allein diesen Variablen zugehörig sind. Dies kann geschehen, da die zugrundeliegende Modellstruktur nicht richtig spezifiziert wurde, und durch die Aufteilung der Gesamtvarianz den erklärenden Variablen x_p mehr Varianz zugeteilt wurde als sie besitzen. Deshalb können in diesen Fällen die Schätzungen der f_p nicht zur Prognose verwendet werden.

Eine weitere Möglichkeit zur Schätzung der f_p bietet der Marginale Integrationsschätzer (Linton & Härdle, 1996; Sperlich, Linton & Härdle, 1997; Nielsen & Linton, 1998). Bei dieser Technik werden die marginalen Einflüsse der erklärenden Variablen auf die Response-Variable bestimmt. Ein wesentlicher Vorteil dieser Methode besteht darin, daß nur die marginalen Einflüsse der erklärenden Variablen, also ohne die Einflüsse von Interaktionen, bestimmt werden. Allerdings haben diese marginalen Effekte eine geringe Aussagekraft, falls die Interaktionseffekte stark sind, denn dann müssen zusätzlich diese Interaktionen bei der Interpretation der Schätzung berücksichtigt werden. Außerdem ist diese Technik noch sehr neu, so daß ihre Implementierung bisher nur im Software-Paket XploRe realisiert wurde.

Zur Datenanalyse sollten beide Schätztechniken verwendet werden. Liefern sie annähernd gleiche Ergebnisse für die Schätzung der f_p, entspricht das spezifizierte Modell dem der zugrundeliegenden Daten. Sollten jedoch die beiden Algorithmen zu verschiedenen Ergebnissen kommen, so ist davon auszugehen, daß die Modellstruktur nicht der der Daten entspricht. In diesem Fall sollten weitere Modellvarianten und Modellspezifizierungen untersucht werden.

In der allgemeinen Formulierung der GAMs ist es nicht vorgesehen, binäre erklärende Komponenten in x_p zuzulassen, denn dadurch würden die klassischen Algorithmen zur Bestimmung der nichtparametrischen Funktionen f_p nicht mehr zu Lösungen kommen. Damit auch binäre erklärende Variablen in die Formulierung eines GAMs mit einfließen können, wird das in Gleichung (15) vorgestellte Grundmodell um eine additive lineare Komponente ergänzt. Dieser erweiterte GAM-Ansatz wird im folgenden Abschnitt dargestellt.

4.2 Die Anwendung der GAMs auf die spezielle Datenstruktur im Marketing mit binären und stetigen erklärenden Variablen

In den erklärenden Variablen eines Markenwahlmodells im Marketing befinden sich meistens auch binäre Größen. Deshalb muß die Formulierung eines GAM-Ansatzes entsprechend angepaßt werden. Eine Möglichkeit dafür besteht darin, daß die binären Komponenten als additiver linearer Teil in das GAM mit eingehen. Dies entspricht auch dem additiven Charakter des Modells, nur daß die erklärenden Variablen sowohl durch nichtparametrische als auch durch parametrische Funktionen modelliert werden.

Damit ergibt sich der bedingte Erwartungswert über

$$E[y|x] = G\left(\sum_p f_p(x_p) + \beta' x\right). \tag{17}$$

β sei der Parametervektor für den linearen Teil des Modells und x der Vektor der linear eingehenden erklärenden Variablen. f_p stellt wiederum eine eindimensionale nichtparametrische Funktion dar, und x_p seien die erklärenden stetigen Variablen, wie sie bereits in Gleichung (15) verwendet wurden. Falls für bestimmte stetige erklärende Variablen bekannt ist, daß sie linear zu modellieren sind, können sie in den parametrischen Teil des Modells integriert werden.

Nachdem der Parametervektor β und die einzelnen nichtparametrischen Funktionen bestimmt worden sind, gibt es zwei verschiedene Möglichkeiten mit den gefundenen Ergebnissen das entgültige Markenwahlmodell zu bestimmen.

Zum einem kann durch eine Addition der berechneten nichtparametrischen Funktionen und des linearen Parts der bedingte Erwartungswert $E[y|x]$ direkt bestimmt werden. Auf die resultierende Summe in Gleichung (17) wird dazu für G eine logistische Linkfunktion angewendet.

Zum anderen ist es möglich, durch die Plots der gefundenen f_p über die funktionale Struktur Vorstellungen zu bekommen mit der die einzelnen erklärenden Variablen in ein parametrisches Modell eingehen sollten. Diese funktionale Struktur könnte dann in einem klassischen CLM-Ansatz berücksichtigt werden und zu einer Verbesserung der Prognosefähigkeit führen.

Die Interpretation der gefundenen β-Werte erfolgt wie im klassischen CLM. Für die gefundenen nichtparametrischen Funktionen gilt ähnliches wie für die aus dem NDE gewonnenen Ergebnisse, so daß für diejenigen, die die Interpretation von Parameterwerten vorziehen, die zweite Möglichkeit zur Verwendung der GAM-Technik zu präferieren ist.

Im folgenden wird die Datenbasis vorgestellt, anschließend werden zur Dokumentation der Leistungsfähigkeit der Methodik Analysen mit den in den vorhergehenden Abschnitten erläuterten Methoden (CLM, NDE und GAM) gerechnet und verglichen.

5 Datenpräsentation

Die in den letzten Abschnitten vorgestellte Theorie soll anhand eines Datensatzes praktisch umgesetzt werden. Die Daten stammen aus dem GfK BehaviorScan. Sie enthalten die Verkäufe von Körperpflegeprodukten über einen Zeitraum von 104 Wochen als Scanner Panel Datensatz.

Im folgenden werden die drei Marken der Marktführer betrachtet, die von 964 Haushalten 2651 mal gekauft wurden. Die Beobachtungen enthalten Informationen zur Markenwahl, zum Preis und eine Auskunft, ob die Werbeaktivitäten Display und Handzettel zum Kaufzeitpunkt vorhanden waren.

Mit diesen Daten werden die Modelle zu den einzelnen Methoden aufgestellt. Aus den erklärenden Variablen Display und Handzettel wurde die neue Variable 'Promotion' gebildet, da die Größen Display und Handzettel stark korrelieren und somit eine Einbeziehung beider Varaiblen zum bekannten Problem der Multikollinearität geführt hätte. 'Promotion' wurde wiederum als binäre Variable folgendermaßen kodiert:

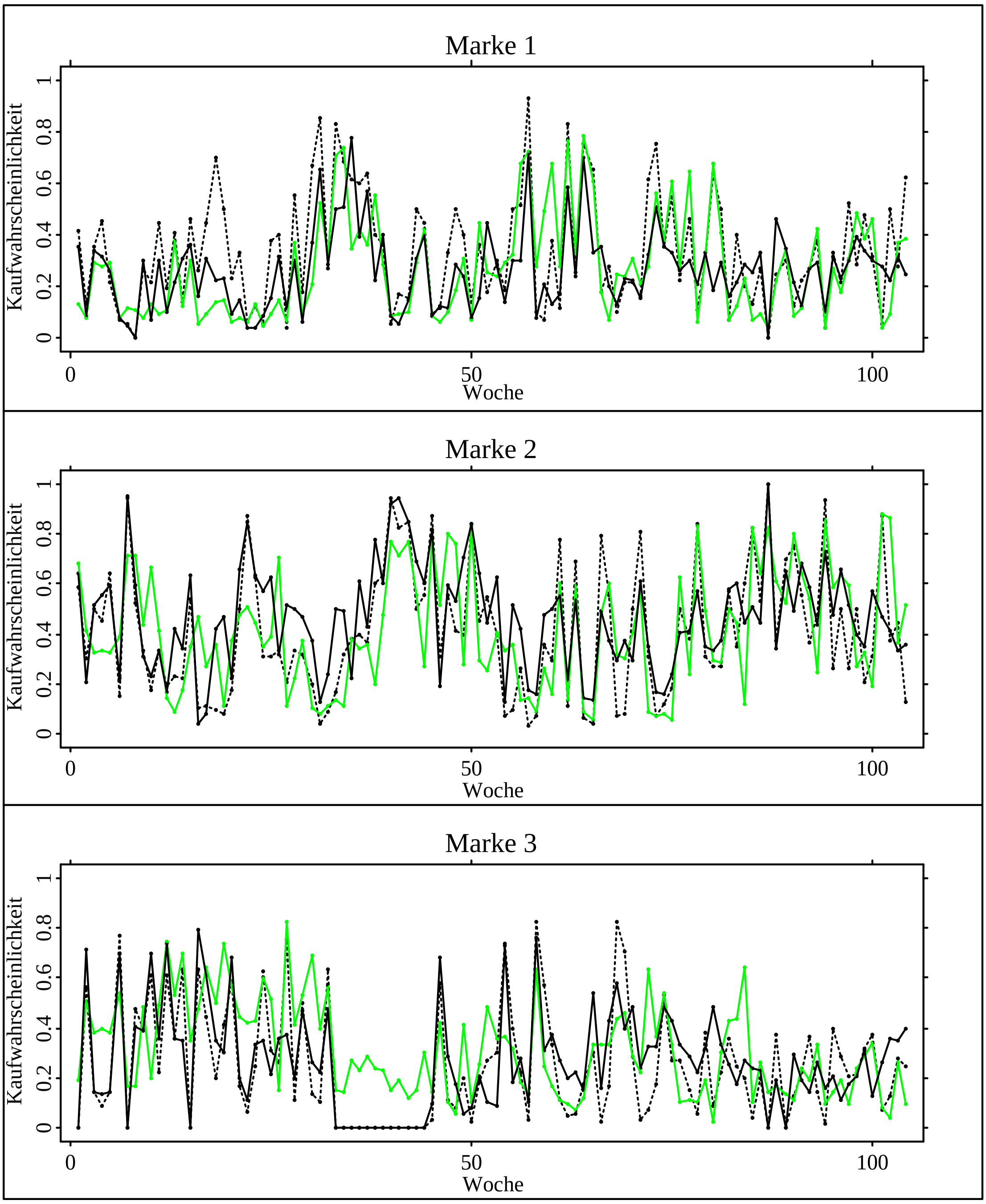
$$\text{Promotion} = \begin{cases} 1 & \text{sonst} \\ 0 & \text{weder Display noch Handzettel vorhanden.} \end{cases}$$

Als weitere neue Variable wurde die 'Loyalität' in Anlehnung an Guadagni & Little (1983) als Folge gewichteter Käufe der Vergangenheit aufgenommen. Sie soll in den vorliegenden Modellen den Feedback-Effekt repräsentieren (Ailawadi, Gedenk & Neslin, 1997). Die Loyalität ist eine stetige Variable, ebenso wie der Preis. Die Modelle der verschiedenen Methoden (nicht- bzw. semiparametrisch sowie CLM) werden mit folgenden erklärenden Variable aufgestellt: Preis, Loyalität und Promotion. Das CLM hat folgende Form

$$\hat{E}[y|x] = \beta_1\text{Loyalität} + \beta_2\text{Preis} + \beta_3\text{Promotion}. \tag{18}$$

Abbildung 1 zeigt für die drei Hauptmarken die Kurvenverläufe für die Kaufwahrscheinlichkeiten über den Beobachtungszeitraum von 104 Wochen bestimmt mit NDE und CLM gegen die realen Werte. Es ist zu sehen, daß das nichtparametrische Modell sich den Daten bedeutend besser anpaßt als das CLM. Vor allem bei Perioden, in denen extreme Situationen vorlagen (so wurde z. B. Marke 3 mehrere Wochen hintereinander nicht gekauft), ist die Anpassung des NDE der des CLM überlegen.

In Abbildung 2 wurde das semiparametrische Modell (GAM) gegen das CLM geplottet. Hier zeigt sich deutlich, daß das GAM sich den Daten unmittelbarer anpaßt als das CLM. Allerdings ist diese Anpassung nicht immer so gut wie beim NDE, wie in Abbildung 3 erkennbar ist. Hier wurde das nichtparametrische (NDE) gegen das semiparametrische (GAM) Modell aufgezeichnet. In der bereits zuvor erwähnten



Legende: - · - real — NDE — CLM

Abbildung 1: Kaufwahrscheinlichkeit der Marken bestimmt mit NDE und CLM gegen die realen Daten über ihren Kaufzeitraum von 104 Wochen

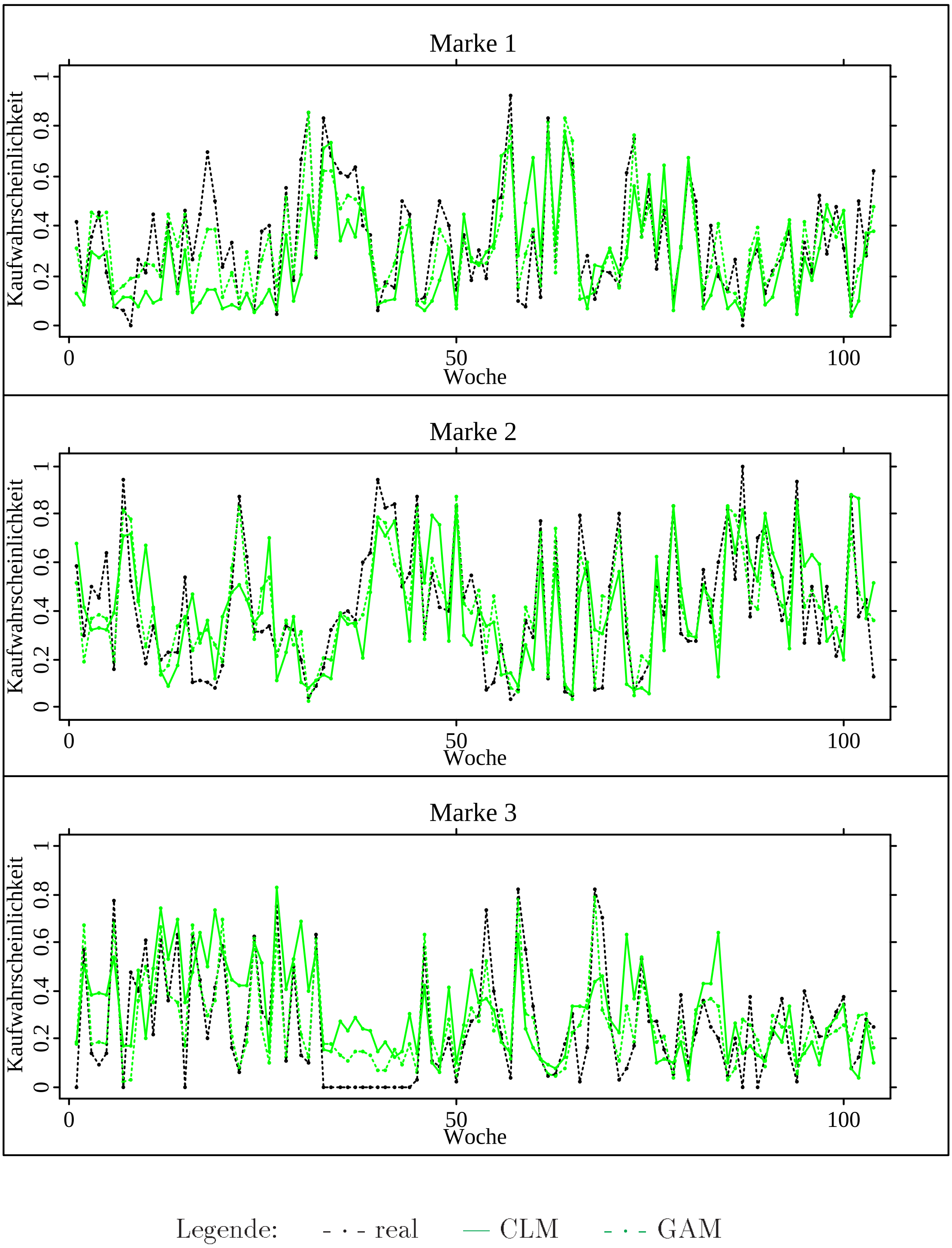


Abbildung 2: Kaufwahrscheinlichkeit der Marken bestimmt mit GAM und CLM gegen die realen Daten über ihren Kaufzeitraum von 104 Wochen

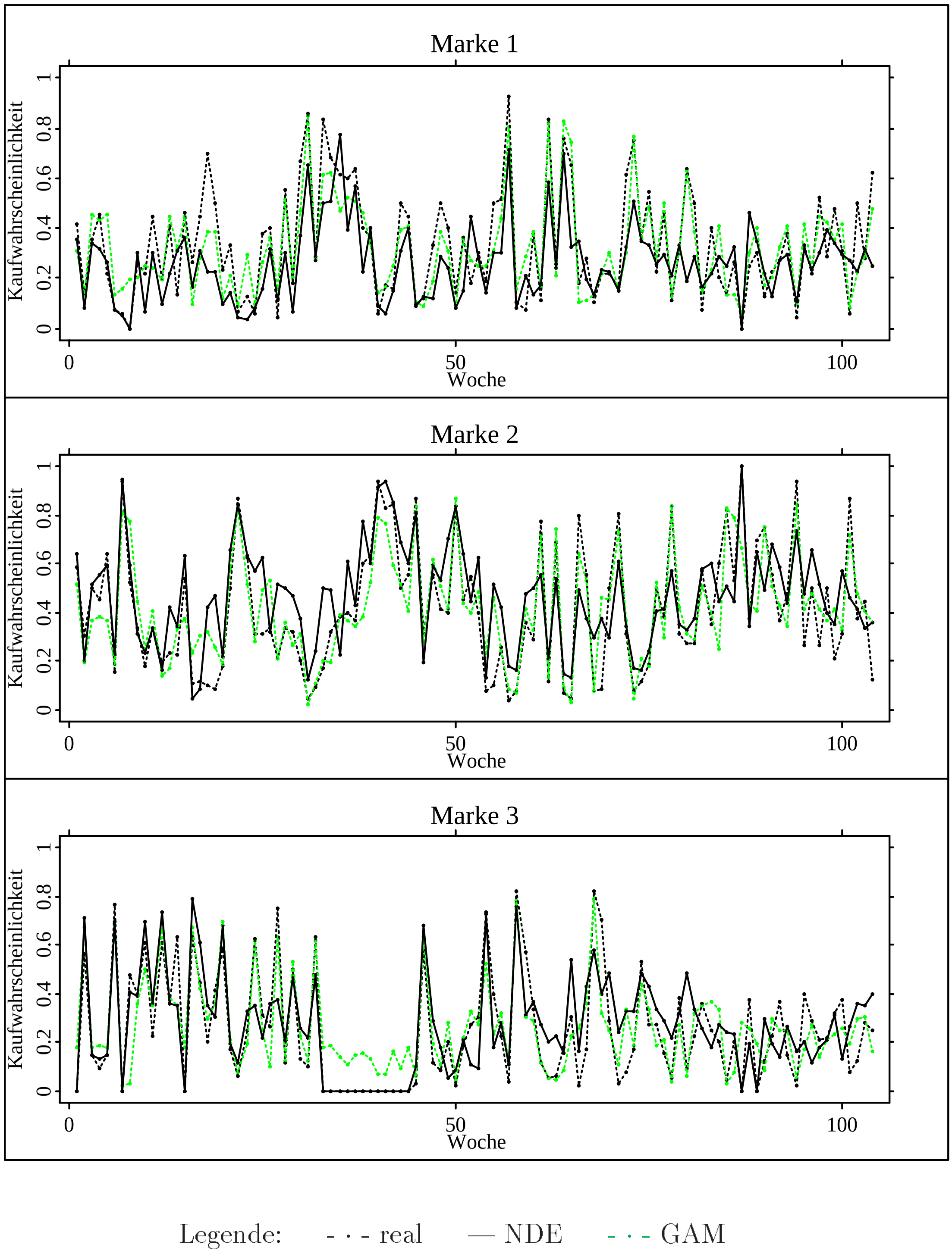


Abbildung 3: Kaufwahrscheinlichkeit der Marken bestimmt mit NDE und GAM gegen die realen Daten über ihren Kaufzeitraum von 104 Wochen

speziellen Situation bzgl. Marke 3 zeigt sich, daß ein GAM die prinzipielle Bewegung der Daten nachvollzieht, diese aber erwartungsgemäß nicht so gut wie ein NDE abbilden kann.

	MSE
CLM	0.189522
NDE	0.157380
GAM	0.121501

Tabelle 2: Vergleich der verschiedenen Modelltypen anhand des MSE

Zur besseren Beurteilung der verschiedenen Modelltypen wurde die mittlere quadratische Abweichung (MSE) der Modelle zu den realen Daten in Tabelle 2 dargestellt. Dabei zeigt sich, daß die GAMs die beste Anpassung liefern. Außerdem bieten sie im Vergleich zum NDE deutlich bessere Möglichkeiten der Interpretation der Ergebnisse. Da auch ihre Schätzung im Bezug zum MSE-Fit sowohl den CLMs als auch denen der NDE überlegen sind, ist nach diesen Ergebnissen die Modellierung eines Markenwahlmodells mit der Methodik der GAMs zu präferieren.

Die nicht- und semiparametrischen Markenwahlmodelle (NDE und GAM) kommen mit sehr wenigen Annahmen aus. Somit sind diese Modelle bedeutend besser als das CLM in der Lage, auch schwierige Marktsituationen adäquat abbilden zu können, was sich in Abbildung 1 bis Abbildung 3 zeigt. Die große Flexibilität wird beim semiparametrischen Ansatz nur durch die Annahme über die Verteilungsstruktur der zufälligen Komponenten eingeschränkt oder aber durch die Annahmen über die Nutzenfunktion. Außerdem sind die Modelle auf Grund ihrer Flexibilität äußerst robust, so daß auch gewisse Mißspezifikationen keine große Rolle bei den Ergebnissen spielen. Auch sind die nicht- und semiparametrischen Modelle leicht an neue Modellsituationen anpaßbar.

Die GAMs wurden sowohl mit dem Backfitting-Algorithmus als auch mit dem Marginalen Integrationsschätzer geschätzt. Da hier die funktionale Form der erklärenden Varilben betrachtet werden soll, sind die Schätzergebnisse des Backfitting-Algorithmuses verwendet worden, denn diese Methode entspricht der Schätzmethode, die zur Bestimmung der Parameterwerte im CLM verwendet wird. Somit können die funktionalen Formen in einen modifizierten CLM-Ansatz übernommen werden. Allerdings liefert der Marginale Integrationsschätzer abweichende Ergebnisse, so daß davon auszugehen ist, daß Interaktionseffekte in den zugrundeliegenden Daten existieren. Die Untersuchung dieser Effekte soll hier nicht weiter behandelt werden, da ein Verbesserung des CLM angestrebt wird.

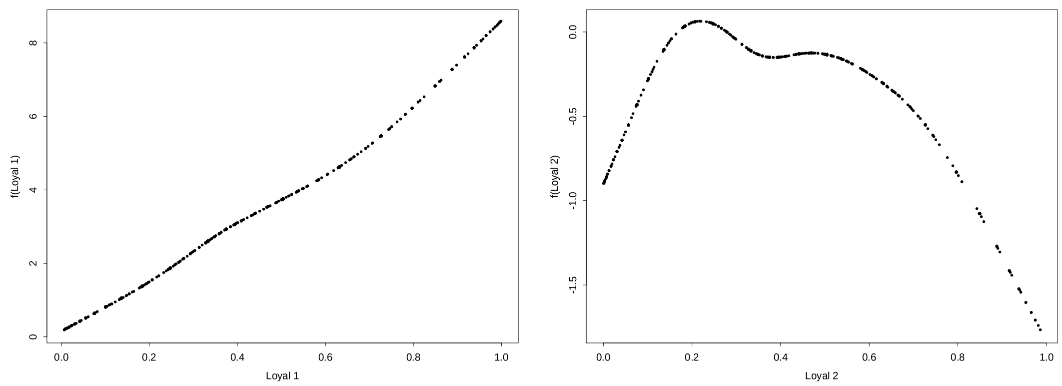


Abbildung 4: Loyalität (x-Achse) für die ersten beiden Marken gegen ihre geschätzte funktionale Form $f_p(x)$ (y-Achse) bei Kauf von Marke 1

Des weiteren werden die einzelnen nichtparametrischen additiven Komponenten eines GAMs in Abbildung 4 und Abbildung 5 betrachtet. Hierbei wurden exemplarisch die Schätzungen für die Loyalität der Marken unter der Bedingung, daß Marke 1 gekauft wurde (Abbildung 4), und die Schätzung für den Preis der Marken unter der Bedingung, daß Marke 1 gekauft wurde (Abbildung 5), ausgewählt. Für die Schätzung der Loyalität wurden nur die Loyalitäten der ersten beiden Marken berücksichtigt, da sich die Loyalitäten über alle Marken zu eins summieren, und somit bei einer Schätzung mit allen drei Loyalitäten das Problem der Multikollinearität aufgetreten wäre. Auf der x-Achse wird jeweils die erklärende Variable (Loyalität bzw. Preis) dargestellt. Die y-Achse ist je ein Anteil an der Summe über die eindimensionalen nichtparametrischen Funktionen aus Gleichung (17) mit $f_p(x)$. Somit sind nicht die absoluten Werte der y-Achse zu verwenden. Hier ist jedoch nur der Kurvenverlauf der eindimensionalen nichtparametrischen Funktionen von Interesse.

In Abbildung 4 ist deutlich eine nichtlineare Form der Kurven erkennbar, die gegen eine lineare Modellierung der Loyalität im CLM spricht. Eine quadratische

Transformation könnte einen linearen Kurvenverlauf produzieren. Bei der Modellierung des Preises (siehe Abbildung 5) erscheint eine kubische Form denkbar zu sein, allerdings widerspricht dies einigen Forschungsansätzen (Krishnamurthi, Raj & Sivakumar, 1995; Krishnamurthi & Raj, 1991; Natter & Hruschka, 1997), die den Preis als nichtlineare S-förmige Sättigungskurve modellieren. Bei den übrigen Graphiken der nichtparametrischen Funktionen des GAM-Ansatzes sind sehr ähnliche Kurvenverläufe erkennbar.

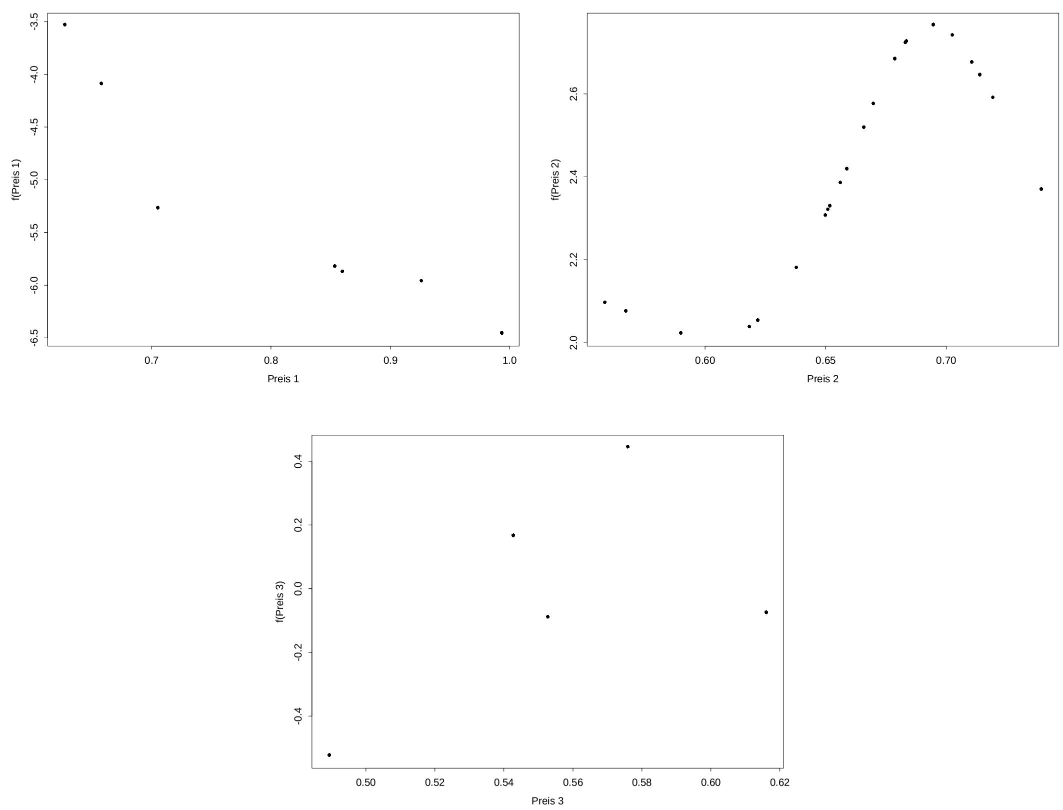
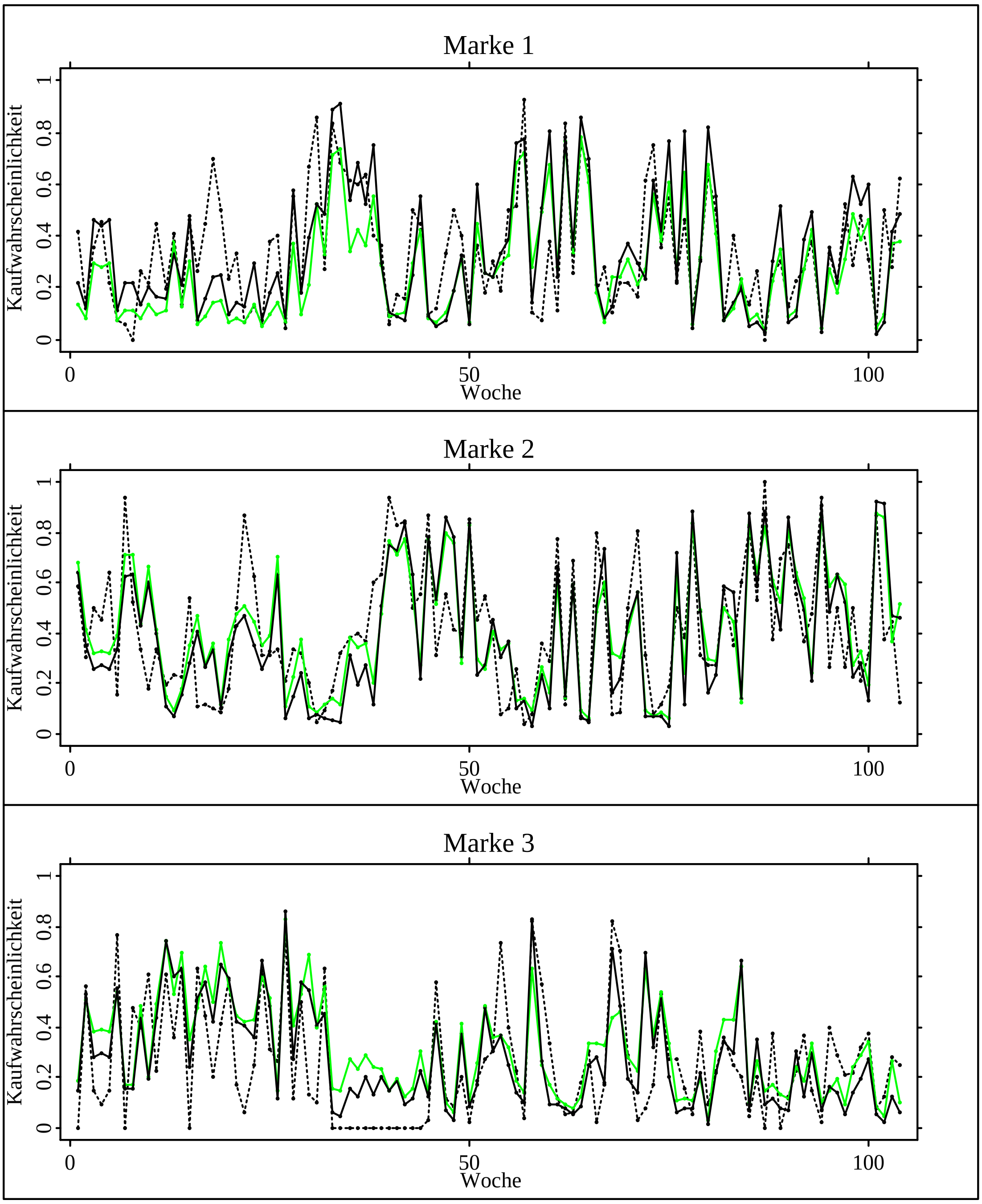


Abbildung 5: Preis (x-Achse) für die 3 Marken gegen ihre geschätzte funktionale Form $f_p(x)$ (y-Achse) bei Kauf von Marke 1

Mit Hilfe der hier gefundenen Ergebnissen des GAM-Ansatzes wird ein 'modifiziertes CLM' gerechnet, in das (siehe auch Abschnitt 4.2) die Loyalität quadratisch und der Preis in einer kubischen Form eingeht. Das Ergebnis dieser neuen Berechnung ist in Abbildung 6 aufgezeichnet.



Legende: - · - real — modifiziertes CLM — CLM

Abbildung 6: Kaufwahrscheinlichkeit der Marken bestimmt mit modifiziertem CLM und CLM gegen die realen Daten über ihren Kaufzeitraum von 104 Wochen

Es wurde das modifizierte CLM gegen den klassischen CLM-Ansatz geplottet. Eine Verbesserung in der Anpassung ist deutlich erkennbar. Dies spiegelt sich auch in den Likelihood-Werten wider ($\text{loglikelihood}_{\text{CLM}} = -2060,37 < \text{loglikelihood}_{\text{mod. CLM}} = -2010,96$), die eine signifikante Verbesserung des modifizierten CLM gegenüber dem klassischen CLM aufweisen.

6 Fazit und Ausblick

Zusammenfassend läßt sich feststellen, daß eine nichtparametrische Modellierung (NDE) eines Markenwahlmodells vom statistischen Standpunkt aus sinnvoll ist, jedoch die im Marketing übliche Interpretation von Parametern nicht möglich macht. Die semiparametrische Modellbeschreibung mittels eines GAMs liefert zwei Anwendungsgebiete. Zum einen kann es ähnlich wie ein NDE zur Bestimmung von Kaufwahrscheinlichkeiten einer Marke benutzt werden. Doch auch hier fehlen die 'Parameter', die nur für die linear modellierten Variablen, insbesondere für die binären, zur Verfügung stehen. Aus diesem Grund ist die zweite Verwendungsmöglichkeit des GAMs, die Betrachtung der eindimensionalen nichtparametrischen Funktionen zur Prüfung der Art der Zusammenhänge zwischen erklärenden und abhängiger Variable, die im Marketingkontext geeignetste Form zur Verwendung semiparametrischer Ansätze. Der daraus resultierende Modellansatz des modifizierten CLM gibt die Möglichkeit mit den vertrauten Parametern Aussagen zu treffen, und doch eine datenorientierte Sichtweise des Modells mit einbeziehen zu können. Zukünftig sollte die Verbesserung des modifizierten CLM im Vordergrund stehen. Außerdem soll untersucht werden, welche Interaktionseffekte einen wichtigen Einfluß auf die Response-Variable haben.

Literatur

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

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

Ailawadi, K. L., Gedenk, K., and Neslin, S. A. (1997). Purchase Event Feedback and Heterogeneity in Choice Models: A Review of Concepts and Methods with Implications for Model Building.

Ben-Akiva, M. and Lerman, S. R. (1985). *Discrete Choice Analysis.* The MIT Press.

Ben-Akiva, M., McFadden, D., Abe, M., Böckenholt, U., Bolduc, D., Gopinath, D., Morikawa, T., Ramawamy, V., Rao, V., Revelt, D., and Steinberg, D. (1997). Modeling Methods for Discrete Choice Analysis. *Marketing Letters 8*(3), 273–286.

Chintagunta, P. K. (1992). Estimating a Multinomial Probit Model of Brand Choice using the Method of Simulated Moments. *Marketing Science 3*(4), 386–407.

Chintagunta, P. K. (1993). Investigating Purchade Incidence, Brand Choice and Purchase Quantity Decisions of Households. *Marketing Science 12*(2), 184–208.

Chintagunta, P. K. and Honore, B. E. (1996). Investigating the effects of marketing variables and unobserved heterogeneity in a multinomial probit model. *International Journal of Research in Marketing 13*, 1–15.

Collett, D. (1991). *Modelling Binary Data.* Chapman & Hall.

Erdem, T. (1996). A Dynamic Analysis of Market Structure Based on Panel Data. *Marketing Science 15*(4), 359–378.

Erdem, T. and Keane, M. P. (1996). Decision-making Under Uncertainty: Capturing Dynamic Brand Choice Processes in Turbulent Consumer Goods Market. *Marketing Science 15*(1), 1–20.

Fader, P. S., Lattin, J. M., and Little, J. D. C. (1992). Estimating Nonlinear Parameters in the Multinomial Logit Model. *Marketing Science 11*(4), 372–385.

Fan, J. and Marron, J. S. (1992). Best Possible Constant for Bandwidth Selection. *The Annals of Statistics 29*(4), 2057–2070.

Greene, W. H. (1990). *Econometric Analysis.* Macmillian Publishing Company.

Guadagni, P. M. and Little, J. D. C. (1983). A Logit Model of Brand Choice Calibrated on Scanner Data. *Marketing Science 2*(3), 203–238.

Gupta, S. and Chintagunta, P. K. (1994). On Using Demographic Variables to Determine Segment Membership in Logit Mixture Models. *Journal of Marketing Research 31*, 128–136.

Hastie, T. and Tibshirani, R. (1986). Generalized Additive Models. *Statistical Science 1*(3), 297–318.

Hastie, T. and Tibshirani, R. (1987). Generalized Additive Models : Some Applications. *Journal of the American Statistical Association 82*(398), 371–386.

Hastie, T. J. and Tibshirani, R. J. (1990). *Generalized Additive Models.* Chapman & Hall.

Horowitz, J. L., Bolduc, D., Divakar, S., Geweke, J., Gönül, F., Hajivassiliou, V., Koppelman, F. S., Keane, M., Matzkin, R., Rossi, P., and Ruud, P. (1994). Advances in Random Utility Models. *Marketing Letters 5*, 311–322.

Horowitz, J. L. and Härdle, W. (1996). Direct Semiparametric Estimation of Single-Index Models with Discrete Covariates. *Journal of the American Statistical Association 91*(436), 1632–1640.

Härdle, W. (1991). *Smoothing Techniques with Implementation in S.* Springer-Verlag.

Härdle, W., Klinke, S., and Turlach, B. A. (1995). *XploRe: An Interactive Statistical Computing Enviroment.* Springer-Verlag.

Hruschka, H. (1996). *Marketing-Entscheidungen.* Verlag Vahlen.

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

Krishnamurthi, L. and Raj, S. P. (1991). An Empirical Analysis of the Relationship between Brand Loyalty and Consumer Price Elasticity. *Marketing Science 10*(2), 172–183.

Krishnamurthi, L., Raj, S. P., and Sivakumar, K. (1995). Unique Inter-Brand Effects of Price on Brand Choice. *Journal of Business Research 34*, 47–56.

Linton, O. B. and Härdle, W. (1996). Estimation of additive regression models with known links. *Biometrika 83*(3), 529–540.

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

Matzkin, R. L. (1993). Nonparametric identification and estimation of polychotomous choice models. *Journal of Econometrics 58*, 137–168.

McFadden, D. (1974). Conditional logit analysis of qualitative choice behavior. In P. Zarembka (Ed.), *Frontiers in Econometrics*, S. 105–142. Academic Press.

Natter, M. and Hruschka, H. (1997). Ankerpreise als Erwartungen oder dynamische latente Variablen in Marktreaktionsmodellen. *Zeitschrift für betriebswirtschaftliche Forschung 49*(9), 747–764.

Nielsen, J. P. and Linton, O. B. (1998). An optimization interpretation of integration and back-fitting estimators for separable nonparametric models. *Journal of the Royal Statistical Society, Series B 60*(1), 217–222.

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

Silverman, B. W. (1986). *Density Estimation for Statistics and Data Analysis.* Chapman & Hall.

Sperlich, S., Linton, O. B., and Härdle, W. (1997). A Simulation Comparison between Integration and Backfitting Methods of Estimating Separable Nonparametric Regression Models. Discussion Paper, SFB 373.

Trommsdorff, V. (1993). *Konsumentenverhalten.* Kohlhammer.