Bartels, Knut

Working Paper

Testen der Spezifikation von multinomialen Logit-Modellen

Statistische Diskussionsbeiträge, No. 17

Provided in Cooperation with:
Wirtschafts- und Sozialwissenschaftliche Fakultät, Universität Potsdam

Suggested Citation: Bartels, Knut (2000) : Testen der Spezifikation von multinomialen Logit-Modellen, Statistische Diskussionsbeiträge, No. 17, Universität Potsdam, Wirtschafts- und Sozialwissenschaftliche Fakultät, Potsdam, https://nbn-resolving.de/urn:nbn:de:kobv:517-opus-8451

This Version is available at:
https://hdl.handle.net/10419/337325

UNIVERSITÄT POTSDAM
Wirtschafts- und Sozialwissenschaftliche Fakultät

STATISTISCHE DISKUSSIONSBEITRÄGE

Nr. 17

Knut Bartels

Testen der Spezifikation von multinomialen Logit-Modellen

Potsdam 2000
ISSN 0949-068X

STATISTISCHE DISKUSSIONSBEITRÄGE

Nr. 17

Knut Bartels

Testen der Spezifikation von multinomialen Logit-Modellen

Herausgeber: Prof. Dr. Hans Gerhard Strohe, Lehrstuhl für Statistik und Ökonometrie
der Wirtschafts- und Sozialwissenschaftliche Fakultät
der Universität Potsdam
Postfach 90 03 27
D-14439 Potsdam
Tel. (+49 331) 977-3225
Fax (+49 331) 977-3210
2000, ISSN 0949-068X

Inhaltsverzeichnis

Zusammenfassung

Eine Klasse von statistischen Tests zur Spezifikation von Logit-Modellen wird vorgestellt. Diese Tests sind Adaptionen von allgemeineren Spezifikationstests nichtlinearer Modelle. Die theoretische Anwendbarkeit der Tests wird dargelegt und die praktische Anwendbarkeit in Simulationsstudien erörtert. Eine konkrete Anwendung auf einen Datensatz und ein multinomiales Logit-Modell zur Produktwahl wird präsentiert.

1 Spezifikationstests parametrischer Modelle

Parametrische Ein- und Mehrgleichungsmodelle sind ein unverzichtbares Werkzeug zur Beschreibung und Erklärung ökonomischer Phänomene. Grundsätzlich werden die Parameter so gewählt, dass ein Modell sich bestmöglich an gegebene Daten anpasst. Anschließend sind Vorzeichen und absoluter Wert der Parameter Gegenstand von Interpretationen oder Grundlage von Entscheidungen.

Der Gehalt dieser Interpretationen und die Qualität dieser Entscheidungen hängen aber unmittelbar von der Güte des anfangs unterstellten Modells ab. Um Trugschlüssen vorzubeugen oder Fehlentscheidungen zu vermeiden ist es daher wünschenswert, wenn nicht unverzichtbar, außer einer auf die Theorie gestützten Rechtfertigung der Form des Modells auch seine Anpassung an die Daten zu prüfen.

Notwendigerweise kann das Prüfen der Anpassungsgüte nur durch den Vergleich des betrachteten Modells mit alternativen Modellen erfolgen. Da ein Modell reale ökonomische Daten nie exakt und vollständig beschreiben oder erklären kann und soll, ist es angebracht, die sich ergebenden Abweichungen als zufällige Residuen zu modellieren. Das geeigneteste Instrument zur Prüfung ist somit ein statistischer Test der Nullhypothese, dass das betrachtete und theoretisch fundierte Modell geeignet ist, also die Modellannahmen empirisch widerspiegelt. Ferner sollten von allen Methoden die Anpassungsgüte zu prüfen solche bevorzugt werden, die nicht selbst wieder von Modellannahmen abhängen und dadurch die Objektivität beeinträchtigen. Daher ist es erstrebenswert, dass die Alternative der anzuwendenden Tests möglichst umfassend, also nichtparametrisch ist.

In diesem Artikel werden solche Testverfahren für Logit-Modelle vorgestellt und auf einen Datensatz zur Produktwahl angewendet. Diese Tests sind die Adaptionen für Logit-Modelle der allgemeineren Verfahren die in Bartels (1999) präsentiert wurden.

Die bisher etablierten Methoden zur Überprüfung von multinomialen Logit-Modellen sind entweder keine statistischen Tests, also informaler, beziehungsweise deskriptiver Natur (McCullagh und Nelder, 1989, S.391f), oder testen nur gegen Alternativen, die aus parametrischen Erweiterungen der Modellklasse bestehen (Fahrmeir und Tutz, 1994, S.119f). Die hier vorgestellten Tests auf diese Modellklasse stellen ein neues und in vieler Hinsicht besseres Werkzeug zur Überprüfung der Spezifikation von Logit-Modellen dar.

In diesem Diskussionsbeitrag werden multinomiale Logit-Modelle in Abschnitt 2 kurz beschrieben. In Abschnitt 3 werden die allgemeinen Tests vorgestellt. Ihre Anpassung an Logit-Modelle wird in Abschnitt 4 ausgeführt. In einer Simulationsstudie wird dann die praktische Anwendbarkeit dargelegt (Abschnitt 5). Schließlich werden die Tests in Abschnitt 6 auf einen realen Datensatz zur Produktwahl angewendet. Dabei wird aufgezeigt, dass multinomiale Logit-Modelle diesen Daten kaum entsprechen. Im Anhang finden sich noch Details zur Theorie der Tests.

2 Multinomiale Logit-Modelle

Logit-Modelle beschreiben die individuelle Wahl zwischen zwei oder mehreren Alternativen. Die Menge der Alternativen sei $\mathcal{A} = \{0, \ldots, m\}$, $m \in \mathbb{N}$. Es wird unterstellt, dass Individuum i die Alternative $j \in \mathcal{A}$ wählt, wenn diese einen Nutzen U_{ij} maximiert. Dieser Nutzen wird in einen systematischen Teil V_{ij} und einen stochastischen Teil ε_{ij} zerlegt:

$$U_{ij} = V_{ij} + \varepsilon_{ij} \quad .$$

Die Wahrscheinlichkeit, dass Individuum i die Alternative j wählt ist nun

$$\begin{aligned}
\mathsf{P}_i[\,j\,] &= \mathsf{P}\big[U_{ij} = \max_{k\in\mathcal{A}} U_{ik}\big] \\
&= \mathsf{P}\big[(U_{ij} - U_{i0}) \geq 0, \ldots, (U_{ij} - U_{im}) \geq 0\big] \\
&= \mathsf{P}\big[\varepsilon_{i0} \leq V_{ij} + \varepsilon_{ij} - V_{i0}, \ldots, \varepsilon_{im} \leq V_{ij} + \varepsilon_{ij} - V_{im}\big] \\
&= \int_{-\infty}^{\infty} \prod_{\nu\neq j} \Phi(V_{ij} - V_{im} + \varepsilon)\phi(\varepsilon)\,\mathrm{d}\varepsilon \quad , \qquad (1)
\end{aligned}$$

wobei $\phi = \Phi'$ die Dichte und Φ die Verteilung von ε_{ij} bezeichnen, und $\varepsilon_{i0}, \ldots, \varepsilon_{im}$ als identisch verteilt und unabhängig angenommen werden.

Wenn nun eine konkrete Verteilung Φ unterstellt und die Form der systematischen Nutzenfunktion V_{ij} festgelegt wird, lässt sich diese Wahrscheinlichkeit bestimmen. Mit der Extremwert-Verteilung $\Phi(x) = \exp(-\exp(-x))$ ergeben sich die Logit-Modelle

$$\mathsf{P}_i[j] = \frac{\exp(V_{ij})}{\sum_{k=0}^{m}\exp(V_{ik})} = \frac{\exp(V_{ij} - V_{i0})}{1 + \sum_{k=1}^{m}\exp(V_{ik} - V_{i0})} \quad .$$

Bei anderen Verteilungsannahmen ergeben sich andere funktionale Formen (Link-Funktionen). Bei der Annahme normalverteilter Störgrößen ε_{ij} ergibt sich zum Beispiel das Probit-Modell.

Im einfachen binomialen Logit-Modell wird die lineare Nutzenfunktion $V_{ij} - V_{i0} = x_{ij}^T\vartheta$ unterstellt und die Wahl des Individuums i zwischen zwei Alternativen $\mathcal{A} = \{0, 1\}$ soll mit den Wahrscheinlichkeiten

$$\mathsf{P}_i[j] = \frac{\exp(x_{ij}^T\vartheta)}{1 + \exp(x_{ij}^T\vartheta)} \qquad (2)$$

$$x \in \mathbb{R}^p \ , \quad \vartheta = (\vartheta_1, \ldots, \vartheta_p)^T \in \Theta_0 := \mathbb{R}^p$$

beschrieben werden, $i = 1 \ldots, n$, $j \in \mathcal{A}$. Die erklärenden Variablen x_{ij} beschreiben dabei Merkmale des Individuums oder der Alternativen, die die Entscheidung möglicherweise beeinflussen.

Wenn mehr als zwei Alternativen zur Wahl stehen, werden die Variablen üblicherweise noch unterteilt in Alternativen-spezifische a_{ij} und Individuen-spezifische b_i, wobei Letztere nun nicht mehr von j abhängen. Also sind $x_{ij} = (a_{ij}, b_i)$ und

$\vartheta = (\alpha, \beta)$, so dass $V_{ij} = x_{ij}^T \vartheta = a_{ij}^T \alpha + b_i^T \beta_j$ gilt. Das allgemeine multinomiale Logit-Modell lautet dann

$$\mathsf{P}_i[j] = \frac{\exp\left(a_{ij}^T \alpha + b_i^T \beta_j\right)}{\sum_{k=0}^{m} \exp\left(a_{ij}^T \alpha + b_i^T \beta_j\right)} \quad . \tag{3}$$

Die Parameter (α, β) können bei gegebenen Daten nach der Maximum-Likelihood-Methode geschätzt und numerisch approximativ bestimmt werden (Ben-Akiva und Lerman, 1985; Fahrmeir und Tutz, 1994, S.97f).

Multinomiale Logit-Modelle haben offensichtliche Schwachstellen. Ein Problem ist die Annahme der logistisch verteilten Störgrößen und der dadurch erfolgten Festlegung der Link-Funktion. Ein weiteres strukturelles Problem ist die Annahme der Linearität für den systematischen Teil der Nutzenfunktion. Ferner wird durch das Modell die unrealistische Annahme getroffen, dass der relative Nutzen einer Alternative im Vergleich zu einer anderen unabhängig von der Existenz einer dritten oder weiterer Alternativen ist (siehe (1), *Unabhängigkeit von irrelevanten Alternativen*). Diese Schwachpunkte sind gute Gründe für die Suche nach Verbesserungen des Modells (Ben-Akiva et al. 1997; Horowitz et al. 1994). Die im Folgenden vorgestellten Tests sollen eine weitere Grundlage für die Beurteilung der Anwendbarkeit von Logit-Modellen geben.

3 Theorie der Tests

Allgemein sei eine Stichprobe $Z_1 = (Y_1, X_1), \ldots, Z_n = (Y_n, X_n)$ unabhängiger Zufallsgrößen mit Verteilung D auf $\mathbb{R}^c \times \mathbb{R}^d$ gegeben. Ein statistisches Modell, welches aus dieser Situation heraus spezifiziert wird, lautet in allgemeiner Form

$$\mathsf{E}[Y|X = x] = f(x, \vartheta_0) \;\text{ für }\; (Y, X) \sim \mathsf{D} \;, \tag{4}$$

beziehungsweise

$$Y_i = f(X_i, \vartheta_0) + U_i \;\text{ mit }\; \mathsf{E}[U_i|X_i] = 0 \;\text{ für } i \in \{1, \ldots, n\} \;,$$

mit einer bekannten Funktion f.

Am besten lässt sich eine korrekte Spezifikation über die zugehörige Klasse von Verteilungen beschreiben. Dazu bezeichne $\mathsf{D}\{g\}$ für eine messbare Funktion $g : \mathbb{R}^d \to \mathbb{R}^c$ die Menge aller Verteilungen D auf $\mathbb{R}^c \times \mathbb{R}^d$, für die die Kovarianzmatrix $\mathsf{Var}[Y]$ existiert und $\mathsf{P}\big\{\mathsf{E}[Y|X] = g(X)\big\} = 1$ ist, wobei die Wahrscheinlichkeit P bezüglich des durch die Randverteilung D_X induzierten Randmaßes genommen ist. Mit diesen Bezeichnungen ist das Modell nun korrekt spezifiziert, falls ein Parameter $\vartheta_0 \in \Theta_0$ existiert, für den $\mathsf{D} \in \mathsf{D}\{f(\cdot, \vartheta_0)\}$ gilt. Das Testproblem lautet somit

$$\mathbf{H}_0 : \quad \mathsf{D} \in \mathcal{D}_0 := \bigcup_{\vartheta \in \Theta_0} \mathsf{D}\{f(\cdot, \vartheta)\} \quad . \tag{5}$$

versus

$$\mathbf{H}_1: \quad \mathsf{D} \in \mathcal{D}_1 := \bigcup_{g \in \mathcal{B}(\mathbb{R}^d, \mathbb{R}^c)} \mathsf{D}\{g\} \setminus \mathcal{D}_0 \quad , \tag{6}$$

wobei $\mathcal{B}(\mathbb{R}^d, \mathbb{R}^c)$ die Menge aller messbaren Funktionen $g : \mathbb{R}^d \to \mathbb{R}^c$ bezeichnet.

Zu diesem Testproblem finden sich in der statistischen und ökonometrischen Literatur erst seit den achtziger Jahren einige Arbeiten, zum Beispiel White (1981), Bierens (1982), Newey (1985) oder Cox, Koh, Wahba und Yandell (1988). Diese Vorschläge führten aber kaum zu anwendbaren Verfahren. Um 1990 wurde, wohl auch durch die wachsende Verfügbarkeit leistungsfähigerer Rechner, die Idee populär, *die parametrische Schätzung mit einer nichtparametrischen Schätzung der Regressionsfunktion zu vergleichen.* Mit sehr unterschiedlichen Ansätzen verfolgten etwa Azzalini, Bowman und Härdle (1989), Eubank und Spiegelman (1990), Staniswalis und Severini (1991), Kozek (1991) sowie Firth, Glosup und Hinkley (1991) diese Grundidee, die sich in kleinen Simulationsstudien als durchaus praktikabel erwies. Allerdings beinhalteten die verwendeten nichtparametrischen Schätzungen einen zusätzlich zu wählenden Bandweiten-Parameter, der asymptotisch verschwinden sollte und dessen Einfluss auf das Verhalten der Tests unbestimmt war. Die Tests von Bierens (1990), Bierens und Ploberger (1997) sowie Diebolt (1995) und Stute (1997), die auf Integralen gewisser empirischer Prozesse beruhen, vermieden dagegen die Wahl eines Bandweiten-Parameters. In weiteren Arbeiten wurden leicht veränderte Testprobleme behandelt: Beispielsweise testeten Härdle und Horowitz (1994) oder Fan und Li (1996) auf eine semiparametrische Form, Su und Wei (1991) oder Rodrigues-Campos, Gonzales Manteiga und Cao (1998) betrachteten verallgemeinerte lineare Modelle, zu denen auch Logit-Modelle gehören.

Ein heuristischer Zugang zur Funktionsweise solcher Tests ist der schon erwähnte Vergleich einer parametrischen Schätzung mit einer nichtparametrischen Schätzung der Modellfunktion $f(\cdot, \vartheta)$. Zur Veranschaulichung seien 25 Beobachtungen $(y_i, x_i) \in \mathbb{R}^2$, $i \in \{1, \ldots, 25\}$ gegeben, die mit dem nichtlinearen Wachstumsmodell

$$f(x, \vartheta) = \frac{1}{1 + \vartheta_2 \exp(-\vartheta_1 x)} \tag{7}$$

erklärt werden sollen ($\vartheta = (\vartheta_1, \vartheta_2)^T$); also

$$y_i = f(x_i, \vartheta) + u_i$$

mit gewissen Fehlern u_i. Für eine Stichprobe aus einer Grundgesamtheit mit einer zur Alternative gehörenden Verteilung schließen die parametrische und nichtparametrische Schätzfunktion mit hoher Wahrscheinlichkeit eine gewisse Fläche ein, wie in Abbildung 1 dargestellt. In Abbildung 2 sind die entsprechenden Schätzungen für eine zur Nullhypothese gehörige Verteilung skizziert. Hier verschwindet die umschlossene Fläche zwischen beiden Schätzungen nahezu. Diese Effekte treten je nach Art der Abweichung und Größe der Stichprobe unterschiedlich deutlich auf.

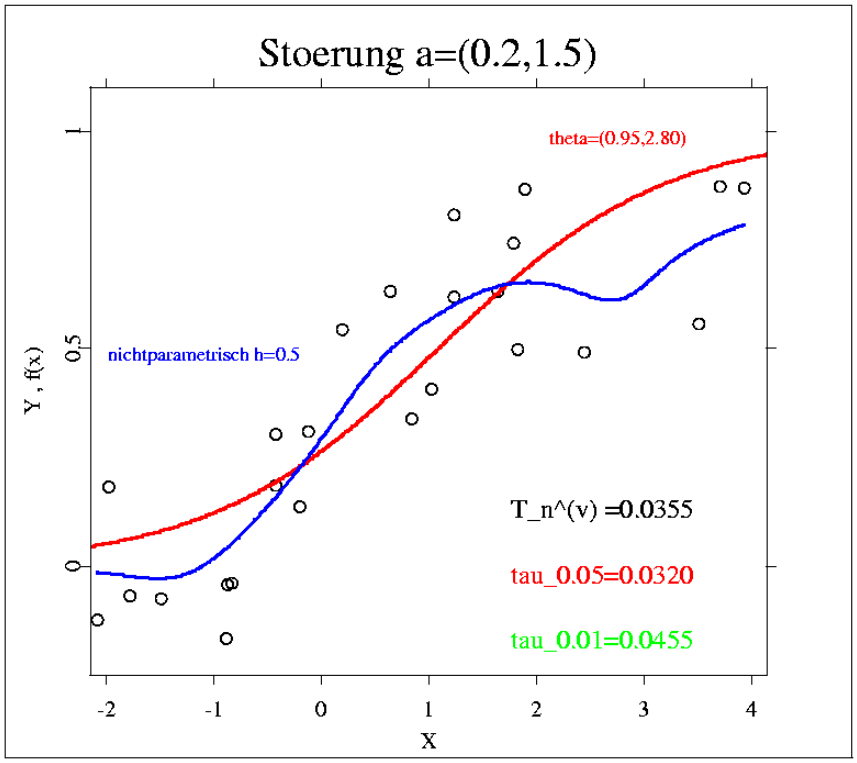

Abbildung 1: Nichtlineares Modell (7), Stichprobe aus $\mathbf{H}_1$, parametrische und nichtparametrische Schätzungen

Die meisten Teststatistiken, die auf einem quadratischen Abstand beruhen, lassen sich in der Form einer U- oder V-Statistik schreiben:

$$\hat{T}_n := \frac{1}{n} \sum_{1 \le i < j \le n} \hat{U}_i \hat{U}_j K_{ijn} \tag{8}$$

und

$$\hat{T}_n^{(v)} := \frac{1}{n} \sum_{1 \le i,j \le n} \hat{U}_i \hat{U}_j K_{ijn} \quad , \tag{9}$$

wobei $\hat{U}_i := u(Y_i, X_i, \hat{\vartheta}_n) = Y_i - f(X_i, \hat{\vartheta}_n)$ für die parametrisch geschätzten Fehler steht und

$$K_{ijn} = k(X_i - X_j, \hat{\vartheta}_n) \tag{10}$$

Gewichte bezeichnen, die von einer festen Kernfunktion k bestimmt werden. Dies heißt, dass die Gewichte nur über die Parameterschätzung $\hat{\vartheta}_n = \hat{\vartheta}(Z_1, \ldots, Z_n)$ von der Stichprobe abhängen dürfen. In den meisten oben zitierten Quellen, insbesondere denen, die einen Vergleich einer parametrischen mit einer nichtparametrischen Kernschätzung behandeln, wurden die Kernfunktionen mit einer vom Stichprobenumfang n abhängigen Bandweite h_n betrachtet, die $h_n \to 0$, $nh_n^d \to \infty$, $n \to \infty$

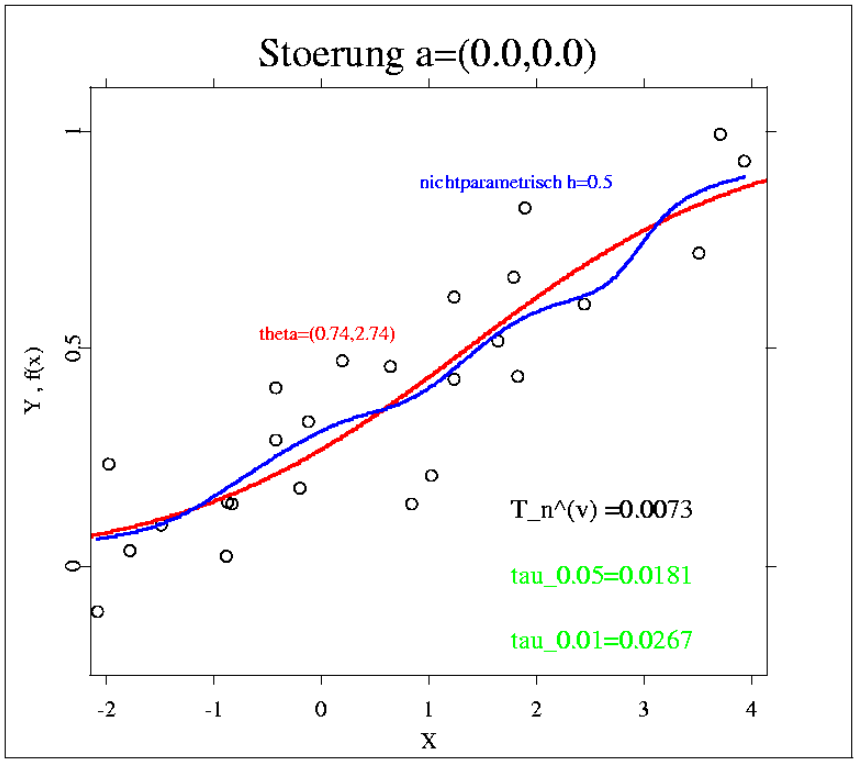

Abbildung 2: Nichtlineares Modell (7), Stichprobe aus $\mathbf{H}_0$, parametrische und nichtparametrische Schätzungen

erfüllen sollte. Unter dieser Annahme sind die Kernfunktionen nicht fest, und die Teststatistiken haben mit entsprechender Normierung eine normale Grenzverteilung, die für Anwendungen aufgrund der langsamen Konvergenz aber unbrauchbar ist.

Unter $\mathbf{H}_0$ und Annahmen **A0**-**A4** verhält sich die mit $\frac{1}{n}$ normierte Teststatistik mit festem Kern (10) asymptotisch wie eine *degenerierte* U-Statistik (Bartels, 1999, S.18f). Sie besitzt daher eine Grenzverteilung der Art

$$\mathcal{L}\Big(\gamma + \sum_{j=1}^{\infty} \lambda_j(\chi_{1j}^2 - 1)\Big) \tag{11}$$

wobei $\gamma \in I\!R$ eine Konstante, $\chi_{11}^2, \chi_{12}^2, \ldots$ unabhängige χ_1^2-verteilte Zufallsvariable und λ_j die Eigenwerte eines linearen Operators sind, der durch die Kernfunktion k festgelegt ist (Gregory, 1977). Unter $\mathbf{H}_1$, der Regularitätsbedingung **A5** und für geeignete (positiv definite) Kernfunktionen k ist $\frac{1}{n}\hat{T}_n$ dagegen bestimmt divergent, da $\hat{T}_n$ nun nicht mehr degeneriert und $\frac{1}{\sqrt{n}}\hat{T}_n$ asymptotisch normalverteilt ist. Somit sind die Tests konsistent und grundsätzlich auf diese Testprobleme anwendbar.

Die Grenzverteilung (11) wird durch die Eigenwerte λ_j bestimmt, die aber unter $\mathbf{H}_0$ mit $\mathsf{D} \in \mathsf{D}\{f(\cdot, \vartheta_0)\}$ von der unbekannten Verteilung der Fehler $u(Z, \vartheta_0)$ und dem unbekannten Parameter ϑ_0 abhängen. Da die kritischen Werte aber Quantile dieser unbekannten Verteilung sein sollten, müssen sie geschätzt werden. Wenn $\hat{\tau}_{\alpha n}^*$

und $\hat{\tau}_{\alpha n}^{(v)*}$ solche Schätzungen für die (1-α)-Quantile der Grenzverteilungen von $\hat{T}_n$ beziehungsweise $\hat{T}_n^{(v)}$ bezeichnen, dann lauten die Tests:

$$\text{„lehne } \mathbf{H}_0 \text{ ab, wenn } \hat{T}_n > \hat{\tau}_{\alpha n}^{*} \text{ ist“} \tag{12}$$

und

$$\text{„lehne } \mathbf{H}_0 \text{ ab, wenn } \hat{T}_n^{(v)} > \hat{\tau}_{\alpha n}^{(v)*} \text{ ist“ .} \tag{13}$$

Die Bestimmung der kritischen Werte $\hat{\tau}_{\alpha n}^{*}$ und $\hat{\tau}_{\alpha n}^{(v)*}$ ist mit Stichprobenwiederholungs-Methoden (Resampling, Bootstrap) in mehreren Varianten möglich. Mit den so bestimmten kritischen Werten sind die Tests im Allgemeinen

- *asymptotisch*, da das Niveau nur für $n \to \infty$ eingehalten wird,
- *adaptiv*, da sich die kritischen Werte über die Schätzung des Parameters ϑ an die Modellklasse und die unbekannte Verteilung der Fehler anpassen, und
- *randomisiert*, da die mit Resampling-Verfahren ermittelten kritischen Werte auf Zufallszahlen beruhen, die von der Stichprobe unabhängig sind.

Das wilde Bootstrap-Verfahren von Wu (1986), das von Härdle und Mammen (1993) schon im Falle einer asymptotisch verschwindenden Bandweite empfohlen wird, ist auch hier, trotz anderer Asymptotik (!), anwendbar, wenn die Annahmen **A6** und **A7** erfüllt sind. Als Alternative wurde in Bartels (1999) eine *Monte-Carlo-Approximation* entwickelt, die asymptotisch gleichwertig ist und den notwendigen Rechenaufwand bei nichtlinearen Modellen erheblich reduziert, da die Berechnung eines Parameterschätzers für jede iterierte Bootstrap-Stichprobe vermieden wird.

Diese Resampling-Verfahren können in der zu erwartenden Weise modifiziert werden, wenn zusätzliche Informationen über die Verteilung der Fehler vorliegen. Für die hier betrachteten Logit-Modelle wird, wie in Abschnitt 2 dargelegt, eine parametrische Form der Verteilung der Fehler angenommen, die unbedingt berücksichtigt werden sollte.

Es lassen sich ferner gewisse obere Schranken für die kritischen Werte angeben, die ohne iterative Verfahren berechenbar sind. Diese können als Vorab-Kriterium dienen und die Durchführung der aufwändigeren Resampling-Verfahren in manchen Anwendungsfällen überflüssig machen (Beispiele finden sich in den Tabellen in Abschnitt 6).

Die Form (8) beziehungsweise (9) der Teststatistiken bietet eine hohe Flexibilität in Bezug auf die Wahl eines Kernes und eines Verfahrens für die Parameterschätzung. Da die Form der U- oder V-Statistik auch im multivariaten Fall $c > 1$ erhalten bleibt, lässt sich auch diese Verallgemeinerung problemlos formulieren.

Die Schätzung des Parameters ϑ kann bei der Berechnung der Teststatistiken und der Bestimmung der kritischen Werte mit verschiedenen Verfahren durchgeführt

werden. Die gebräuchlichsten Schätzverfahren, zum Beispiel Maximum Likelihood oder Kleinste Quadrate, erfüllen die Voraussetzungen hierfür.

Die theoretischen Ergebnisse wurden in Simulationsstudien zu verschiedenen Regressionsmodellen auf ihre praktische Anwendbarkeit untersucht. Dabei bestätigten sich die theoretischen hergeleiteten Ergebnisse und es zeigte sich auch, dass die Tests schon bei relativ kleinen Stichprobenumfängen ordentliche Resultate aufweisen (Bartels, 1999).

4 Anwendbarkeit auf Logit-Modelle

Wir betrachten zunächst das einfache binomiale Logit-Modell (2). Damit die Tests angewendet werden können, muss das Modell aber von der Form (4) sein. Der Nachweis, dass Logit-Modelle sich derart umformulieren lassen, soll hier erbracht werden.

Da die Wahl $j \in \mathcal{A}$ eines Individuums i anhand von Auswahl-Wahrscheinlichkeiten $\mathsf{P}_i[\,j\,]$ beschrieben werden soll, liegt es nahe, eine Zufallsvariable Y zu definieren, die Werte aus der Menge der Alternativen $\mathcal{A}$ annimmt. Wenn das Logit-Modell mit einem Parameter $\vartheta_0 \in \Theta$ zutrifft, dann sind diese Auswahl-Wahrscheinlichkeiten, also die Verteilung D_Y von Y, festgelegt. Das Modell sieht jedoch auch vor, dass exogene Daten x diese Wahrscheinlichkeiten beeinflussen. Es ist angebracht, diese Daten als gegebene Ausprägung eines zufälligen Vektors X zu modellieren. (Modelle mit einem festen Design können zwar nicht so formuliert werden, es ist aber grundsätzlich möglich, sie auf diesen Fall zurückzuführen.) Mit diesen Bezeichnungen gilt für das binomiale Logit-Modell insgesamt

$$\mathsf{P}_i[\,1\,] = \mathsf{P}_i[\,Y = 1\,] = \mathsf{P}[\,Y = 1 \,|\, X = x] = \frac{\exp(x^T\vartheta)}{1 + \exp(x^T\vartheta)} \; .$$

Mit der Bezeichnung $f(x,\vartheta) = \frac{\exp(\vartheta_1 + x\vartheta_2)}{1+\exp(\vartheta_1+x\vartheta_2)}$ ist (2) somit äquivalent zu

$$\mathsf{E}[Y|X = x] = f(x,\vartheta) \; ,$$

also einem Modell der Form (4). Da $\mathsf{P}_i[\,1\,] = 1 - \mathsf{P}_i[\,0\,]$ ist, ist das binomiale Logit-Modell hiermit vollständig festgelegt. Wenn nun noch formal angenommen wird (**A0**), dass die Wahrscheinlichkeiten bezüglich einer gemeinsamen Verteilung D von Y und X erklärt sein sollen, dann entspricht diese Formulierung dem betrachteten Testproblem.

In der bisher betrachteten Nullhypothese $\mathbf{H}_0$ ist über die Verteilung der Fehler U_i Nichts ausgesagt. Wenn ein binomiales Logit-Modell unterstellt wird, dann können die Fehler aber nur die Werte ρ oder $1-\rho$ mit $\rho = \mathsf{P}[\,Y = 1 \,|\, X = x]$ annehmen. Sie sind daher für jedes x binomial verteilt. Daher sollte die Nullhypothese nur aus Verteilungen D bestehen, die diese parametrische Form der Fehler mit sich bringen.

Für eine messbare Funktion g und eine parametrische Klasse von Verteilungen $\mathsf{D}_U\{\Pi, x\}$ definieren wir daher

$$\mathsf{D}''\{g\} := \left\{ \mathsf{D} \in \mathsf{D}\{g\} \;\middle|\; \text{für jedes } x \in I\!R^d \text{ gilt } \; \mathsf{D}_{U|X=x} \in \mathsf{D}_U\{\Pi, x\} \right\}$$

und betrachten die Hypothese

$$\mathbf{H}_0'': \quad \mathsf{D} \in \mathcal{D}_0'' := \bigcup_{\vartheta \in \Theta_0} \mathsf{D}''\{f(\cdot, \vartheta)\} \quad ,$$

versus

$$\mathbf{H}_1: \quad \mathsf{D} \in \mathcal{D}_1 := \bigcup_{g} \mathsf{D}\{g\} \setminus \bar{\mathcal{D}}_0 \ .$$

Die Forderung der parametrischen Verteilung bleibt unter $\mathbf{H}_1$ nicht erhalten und wir testen wieder gegen die ursprüngliche Alternative. Der Test braucht also Verteilungen aus $\bigcup_{\vartheta \in \Theta} \mathsf{D}\{f(\cdot, \vartheta)\} \setminus \mathcal{D}_0''$ nicht als Alternative zu erkennen.

Nun soll dieses Ergebnis auf multinomiale Logit-Modelle (3) erweitert werden. Dies geschieht, indem es auf m binomiale Logit-Modelle zurückgeführt wird, die gemeinsam, also multivariat, zu betrachten sind.

Dazu definieren wir die binäre Zufallsvariable Y_j mit Wert 1 falls die Alternative j gewählt wird und 0 sonst, $j \in \mathcal{A} = \{0, \ldots, m\}$. Mit den Funktionen $f_j(x, \vartheta) := \frac{\exp(x_j^T \vartheta)}{\sum_{j=0}^{m} \exp(x_j^T \vartheta)}$ ist (3) nun äquivalent zu $\mathsf{E}[Y_j | X = x] = f_j(x, \vartheta)$ für alle $j \in \mathcal{A}$. Mit den vektoriellen Größen $\underline{Y} = (Y_0, \ldots, Y_m)^T$ und $\underline{f} = (f_0, \ldots, f_m)^T$ kann (3) daher als

$$\mathsf{E}[\underline{Y} | X = x] = \underline{f}(x, \vartheta) \tag{14}$$

in multivariater Form geschrieben werden. Dann sind auch die Residuen $\hat{U}_i$ Vektoren, aber die Tests können mit entsprechenden (positiv definiten) Matrizen von Gewichten $K_{ijn} \in I\!R^{(m+1)\times(m+1)}$ angewendet werden, wenn in naheliegender Verallgemeinerung

$$\hat{T}_n := \tfrac{1}{n} \sum_{1 \le i < j \le n} \hat{U}_i^T K_{ijn} \hat{U}_j \tag{15}$$

und

$$\hat{T}_n^{(v)} := \tfrac{1}{n} \sum_{1 \le i,j \le n} \hat{U}_i^T K_{ijn} \hat{U}_j \quad , \tag{16}$$

gesetzt werden. Die Tests sind in diesem multivariaten Fall unter denselben Bedingungen und in gleicher Weise anwendbar, wie im univariaten Fall (Bartels, 1999, S.42f).

Die Verteilung der Fehler ist im Logit-Modell (3) multinomial, beziehungsweise in der Schreibweise (14) in jeder Komponente binomial. Deshalb sollte auch hier wieder die entsprechende Hypothese $\mathbf{H}_0''$ gegen $\mathbf{H}_1$ getestet werden.

5 Simulationsstudien

Wir betrachten das binomiale Logit-Modell

$$\begin{aligned} \mathsf{P}[Y=1|X=x] &= \frac{\exp(\vartheta_1 + x\vartheta_2)}{1+\exp(\vartheta_1 + x\vartheta_2)} \\ x \in [0,1]\ , \quad \vartheta &= (\vartheta_1, \vartheta_2)^T \in \Theta_0 := I\!R^2\ . \end{aligned} \tag{17}$$

Mit der Bezeichnung $f(x,\vartheta) = \frac{\exp(\vartheta_1+x\vartheta_2)}{1+\exp(\vartheta_1+x\vartheta_2)}$ ist (17) also äquivalent zu

$$\mathsf{E}[Y|X=x] = f(x,\vartheta)\ .$$

Es wurden 1000 Simulations-Datensätze mit Stichprobenumfang $n = 50$

$$x_{i,1}, \ldots, x_{i,50}\ , \quad y_{i,1}, \ldots, \varepsilon_{i,50}$$

mit $x_{i,j} \sim U[0,1]$ (gleichverteilt) und $y_{i,j} \in \{0,1\}$ mit Verteilung gemäß (17) und den Parameterwerten $\vartheta = (0.5, 3)$ für $i \in \{1, \ldots, 1000\}$ und $j \in \{1, \ldots, 50\}$ erzeugt. Ebenso wurden Alternativen betrachtet, bei denen die Werte von $y_{i,j} \in \{0,1\}$ mit Verteilungen gemäß Tabelle 1 erzeugt wurden (siehe Abbildung 3).

logit	:	$\mathsf{P}[Y=1\|X=x] := \frac{\exp(\vartheta_1+x\vartheta_2)}{1+\exp(\vartheta_1+x\vartheta_2)}$	$, \vartheta = (0.5, 3)^T$
quadlog	:	$\mathsf{P}[Y=1\|X=x] := \frac{\exp(\vartheta_1+x\vartheta_2+x^2\vartheta_3)}{1+\exp(\vartheta_1+x\vartheta_2+x^2\vartheta_3)}$	$, \vartheta = (0.5, -6, 7)^T$
extrem	:	$\mathsf{P}[Y=1\|X=x] := \exp\big(1-\exp(\vartheta_1+x\vartheta_2)\big)$	$, \vartheta = (0.05, 3)^T$
polylog	:	$\mathsf{P}[Y=1\|X=x] := \frac{\exp(\vartheta_1+x\vartheta_2+x^2\vartheta_3+x^3\vartheta_4)}{1+\exp(\vartheta_1+x\vartheta_2+x^2\vartheta_3+x^3\vartheta_4)}$	$, \vartheta = (-0.5, -7, 7, 6)^T$
loglog	:	$\mathsf{P}[Y=1\|X=x] := \frac{\log(1+\vartheta_1+x\vartheta_2)}{1+\log(1+\vartheta_1+x\vartheta_2)}$	$, \vartheta = (0.05, 3)^T$
random	:	$\mathsf{P}[Y=1\|X=x] := p$ mit $p \sim U[\vartheta_1, \vartheta_2]$	$, \vartheta = (0.3, 0.7)^T$

Tabelle 1: Betrachtete Modelle zur Datenerzeugung

Dies ist im Kern dieselbe Studie, die Rodrigues-Campos, Gonzales Manteiga und Cao (1998) für Tests durchgeführt haben, die auf Funktionalen von empirischen Prozessen beruhen. Die Alternativen *polylog*, *loglog* und *random* werden hier zusätzlich betrachtet.

Mit der Nullhypothese des Vorliegens eines Logit-Modells wird impliziert, dass die Fehler bei gegebenem $x \in [0,1]$ binomial verteilt sind. Demnach wird unter der Nullhypothese eine parametrische Verteilung der Fehler unterstellt, so dass ein parametrisches Bootstrap-Verfahren (PBS) bezüglich $\mathbf{H}_0''$ hier angebracht ist. Einige Ergebnisse unter Verwendung des univariaten Gauss-Kerns sind in den Tabellen 2 und 3 aufgeführt.

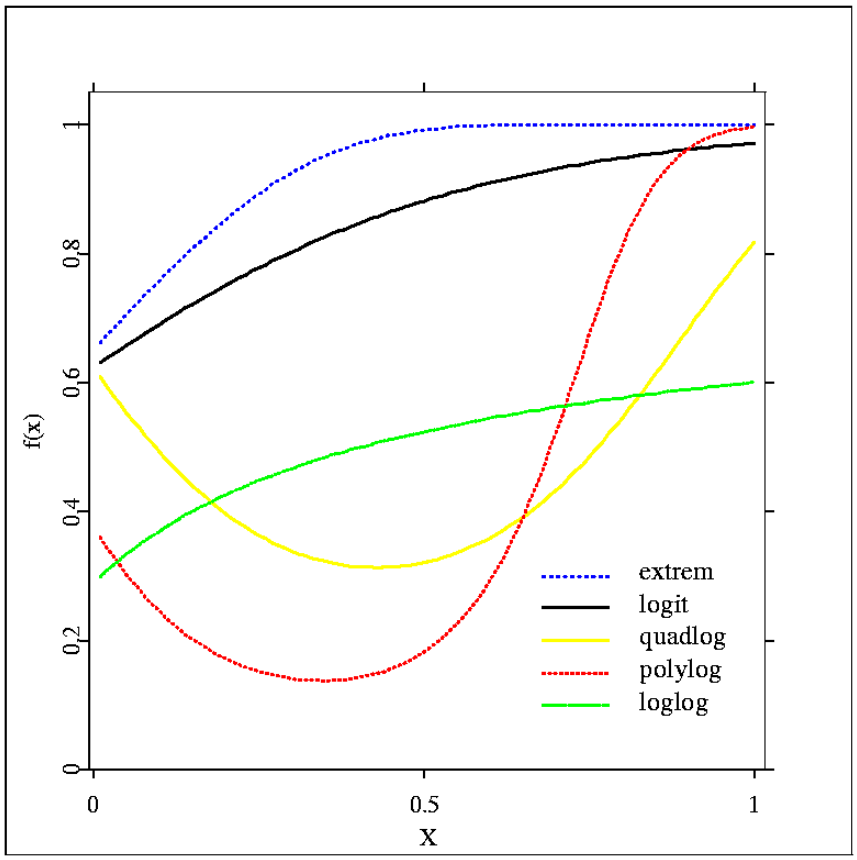

Abbildung 3: Logit-Modell (17), alternative Modelle gemäß Tabelle 1

Zu Tabelle 2:

- *Beobachtung: Das nominale Niveau wird gut eingehalten.*

- *Beobachtung: Die Daten aus* quadlog *und* polylog *werden des Öfteren, diejenigen aus* loglog *und* random *werden kaum als zur Alternative gehörig erkannt. Die Daten aus dem Modell* extrem *werden nur für kleine Bandweiten überhaupt manchmal als Alternative erkannt.*

  Die Modelle *extrem* und *loglog* sind einem Logit-Modell nach (17) mit entsprechenden Parametern sehr ähnlich. Insbesondere sind die Verläufe der Wahrscheinlichkeiten konkav. Hierdurch unterscheiden sich die Modelle *quadlog* und *polylog* deutlich. Die zufällig erzeugten Daten im Modell *random* weisen zwar keine Struktur auf, passen aber in ein Logit-Modell (17) mit Parameter $\vartheta = (0, 0)^T$, so dass die Beobachtungen als zufällige Abweichungen vom Erwartungswert $\frac{1}{2}$ aufgefasst werden. Diese Ähnlichkeiten und Unterschiede spiegeln sich gut in den Resultaten der Simulationen wider. Das besonders schlechte Verhalten gegenüber den Daten aus dem Modell *extrem* haben auch Rodrigues-Campos, Gonzales Manteiga und Cao (1998) beobachtet.

- *Beobachtung: Bei den Daten aus* quadlog *und* polylog *sinkt die empirische*

Empirische Güte zum nominalen Niveau 0.05						
Gauss-Kern, $\hat{T}_n^{(v)}$-PBS, $n = 50$						
	h					
Daten	0.05	0.20	0.40	0.60	0.80	1.00
logit	0.046	0.049	0.049	0.046	0.043	0.041
extrem	0.037	0.006	0.003	0.003	0.003	0.002
quadlog	0.229	0.372	0.431	0.434	0.434	0.435
polylog	0.470	0.711	0.767	0.768	0.766	0.766
loglog	0.057	0.062	0.060	0.061	0.060	0.062
random	0.058	0.060	0.069	0.070	0.071	0.072
Gauss-Kern, $\hat{T}_n$-PBS, $n = 50$						
	h					
Daten	0.05	0.20	0.40	0.60	0.80	1.00
logit	0.060	0.032	0.000	0.000	0.000	0.000
extrem	0.050	0.001	0.000	0.000	0.000	0.000
quadlog	0.220	0.357	0.295	0.085	0.006	0.000
polylog	0.461	0.684	0.489	0.058	0.002	0.000
loglog	0.052	0.051	0.033	0.003	0.000	0.000
random	0.053	0.057	0.033	0.000	0.000	0.000

Tabelle 2: Test auf Logit-Modell (17), Daten gemäß Tabelle 1, Variation der Bandweite h

Güte mit kleiner werdender Bandweite. Die Ergebnisse bei Bandweiten $h \geq 0.40$ unterscheiden sich kaum.

Die Abweichungen der Daten aus *quadlog* und *polylog* oszillieren nicht stark um die jeweils beste Approximation. Daher genügen größere Bandweiten, um mit relativ geringem Fehler zweiter Art gegen diese Alternativen zu testen.

- *Beobachtung: Die Ergebnisse bezüglich $\hat{T}_n$ sind bedeutend schlechter als diejenigen bezüglich $\hat{T}_n^{(v)}$.*

  Die Nullhypothese des Vorliegens eines Logit-Modells bedeutet auch, dass die Varianz der Fehler beschränkt ist. Daher bietet die Diagonale $\hat{T}_n^{(v)} - 2\hat{T}_n$ durchaus ein Kriterium für das Testproblem. Für wachsende Bandweiten h zeigt sich ferner der in anderen Simulationsstudien schon öfter beobachtete Unterschied für Bootstrap-Verfahren bezüglich $\hat{T}_n^{(v)}$ und $\hat{T}_n$. Dieser lässt sich theoretisch gut erklären, indem die Auswirkungen der Bandweite h auf die Eigenwerte des zugehörigen Kernoperators für $h \to 0$ und $h \to \infty$ untersucht werden (Bartels, 1999, S.38f).

Als Ergebnis bleibt festzuhalten, dass für Logit-Modelle stets die Teststatistik $\hat{T}_n^{(v)}$ verwendet werden sollte.

Empirische Güte zum nominalen Niveau 0.05						
Gauss-Kern, $\hat{T}_n^{(v)}$-PBS, $n = 100$						
	h					
Daten	0.02	0.05	0.10	0.20	0.40	0.60
logit	0.028	0.037	0.049	0.051	0.056	0.060
extrem	0.047	0.039	0.026	0.011	0.005	0.003
quadlog	0.283	0.418	0.538	0.629	0.698	0.710
polylog	0.626	0.811	0.900	0.955	0.974	0.976
loglog	0.068	0.074	0.066	0.078	0.089	0.091
random	0.056	0.043	0.049	0.045	0.053	0.056
Gauss-Kern, $\hat{T}_n$-PBS, $n = 100$						
	h					
Daten	0.02	0.05	0.10	0.20	0.40	0.60
logit	0.044	0.045	0.053	0.051	0.009	0.000
extrem	0.070	0.051	0.025	0.004	0.000	0.000
quadlog	0.271	0.411	0.535	0.624	0.653	0.450
polylog	0.617	0.806	0.898	0.953	0.933	0.566
loglog	0.060	0.071	0.066	0.074	0.075	0.023
random	0.052	0.041	0.047	0.045	0.039	0.021

Tabelle 3: Test auf Logit-Modell (17), Daten gemäß Tabelle 1, Variation der Bandweite h

Zu Tabelle 3:

- *Beobachtung: Qualitativ zeigen sich keine Unterschiede zu Tabelle 2. Die zu erwartenden Verbesserungen bei größerem Stichprobenumfang treten ein.*

Insgesamt zeigt sich, dass die Tests schon bei den relativ kleinen Stichprobenumfängen von $n = 50$ und $n = 100$ und bei nur einer Beobachtung für jeden gegebenen Wert von x starke Abweichungen von der Nullhypothese mit akzeptabler Wahrscheinlichkeit erkennen. Daher darf erwartet werden, dass die Tests in Anwendungen mit realen Daten und größerem Stichprobenumfang gut funktionieren. Da multinomiale Modelle sich gemäß (14) als eine Kombination mehrerer binomialer Modelle formulieren lassen, darf diese Erwartung ohne große Bedenken auch auf multinomiale Logit-Modelle übertragen werden.

6 Anwendung auf Daten zur Produktwahl

Wir wenden die Tests nun auf reale Daten an. Die Daten stammen aus dem *GfK BehaviorScan* und beschreiben Käufe eines Kosmetikprodukts verschiedener Marken von 1377 Haushalten während einer Dauer von 104 Wochen. Die Daten enthalten zu jedem der 5532 Käufe, die in diesen Zeitraum fielen, Informationen über die Wahl der Marke ($M \in \mathcal{A} = \{0, \ldots, m\}, m \in \mathbb{N}$), die Preise der Produkte ($P_j, j \in \mathcal{A}$), die Identität des Käufers, das Datum des Kaufs und Angaben zu Marketingaktivitäten zum Kaufzeitpunkt. Aus diesen Informationen wurden für jede der m Marken zwei neue Variablen konstruiert: Werbung (W_j) und Loyalität (L_{ij}). Dabei ist W_j eine binäre Variable mit Wert 1 beim Vorhandensein von Marketingaktivitäten zur Marke $j \in \mathcal{A}$ zum Kaufzeitpunkt und 0 sonst. Die Loyalität L_{ij} eines Haushalts i zur Marke $j \in \mathcal{A}$ ist eine quasi-stetige positive Variable, die gemäß Guadagni und Little (1983) definiert ist und stets $\sum_{j=1}^{m} L_j = 1$ erfüllt.

Um die Dimension des Parameterraumes $p = 3 \cdot (m + 1)$ nicht zu groß werden zu lassen, wurden hieraus zwei konzentriertere Datensätze abgeleitet. In den ersten dieser Datensätze (*10 Marken*, $m = 9$) gehen alle Käufe der 9 meistgekauften Marken unverändert ein und alle anderen Käufe werden zu einer zehnten „Restmarke“ zusammengefasst. Der zweite Datensatz (*3 Marken*, $m = 2$) umfasst nur die von 964 verschiedenen Haushalten getätigten 2651 Käufe der drei in ihrem Preissegment meistgekauften Marken mit den Kennzeichnungen 5, 7 und 8. Diese Datensätze sind in den Tabellen 4 und 5 grob beschrieben.

	Käufe	**Loyalität**		**Preis**		**Werbung**
Marke	(in %)	Mittelwert	(Std.Abw.)	Mittelwert	(Std.Abw.)	(in %)
1	4.79	0.0781	(0.1057)	0.7284	(0.0252)	15.89
2	8.97	0.0944	(0.1408)	0.6629	(0.0328)	14.95
3	6.78	0.0896	(0.1115)	0.5871	(0.0443)	23.83
4	11.59	0.1065	(0.1298)	0.6523	(0.0587)	25.96
5	15.67	0.1304	(0.1849)	0.9033	(0.1153)	34.07
6	3.34	0.0694	(0.0982)	0.6143	(0.0134)	1.14
7	19.11	0.1397	(0.1753)	0.6942	(0.0362)	54.52
8	13.14	0.1169	(0.1457)	0.5781	(0.0281)	39.44
9	14.37	0.1199	(0.1557)	0.6903	(0.0322)	39.15
10	2.24	0.0552	(0.0588)	0.8162	(0.0030)	16.72

Tabelle 4: Deskriptive Statistik für den 10-Marken-Datensatz

Anhand dieser Datensätze soll nun getestet werden, ob sich die Markenwahl anhand eines multinomialen Logit-Modells beschreiben lässt (McFadden, 1974). Im betrachteten Fall lautet dieses Modell mit dem unbekannten Parameter $\vartheta \in \mathbb{R}^3$

$$\mathsf{P}_i[j|X_{ij}] = \frac{\exp(\vartheta^T X_{ij})}{\sum_{j=1}^{m} \exp(\vartheta^T X_{ij})} \,, \tag{18}$$

Marke	Käufe (in %)	Loyalität Mittelwert	Loyalität (Std.Abw.)	Preis Mittelwert	Preis (Std.Abw.)	Werbung (in %)
5	32.71	0.3413	(0.1916)	0.8943	(0.1250)	40.89
7	39.87	0.3451	(0.1737)	0.6864	(0.0401)	56.17
8	27.42	0.3137	(0.1539)	0.5754	(0.0317)	43.30

Tabelle 5: Deskriptive Statistik für den 3-Marken-Datensatz

wobei $X_{ij} := (P_j, L_{ij}, W_j)^T$ ist und $\mathsf{P}_i[j|X_{ij}]$ die Wahrscheinlichkeit dafür bezeichnet, dass Haushalt i unter den Bedingungen X_{ij} die Marke j kauft.

Die Modellgleichung (18) kann wie in Abschnitt 4 ausgeführt so umgeformt werden, dass die Tests anwendbar werden. Dazu definieren wir hier konkret die binäre Zufallsvariable Y_j mit Wert 1 falls die Marke j gekauft wird und 0 sonst. Da der Einfluss des Haushaltes nur über die Loyalität L_{ij} eingeht, kann der Index i im Folgenden weggelassen werden. Mit der Funktion $f_j(x, \vartheta) := \frac{\exp(\vartheta^T x_j)}{\sum_{j=1}^m \exp(\vartheta^T x_j)}$ und der $(3 \times m)$-Matrix

$$x = (x_0, \ldots, x_m) = \begin{pmatrix} p_0 & \ldots & p_m \\ l_0 & \ldots & l_m \\ w_0 & \ldots & w_m \end{pmatrix}$$

ist (18) nun äquivalent zu $\mathsf{E}[Y_j|X = x] = f_j(x, \vartheta)$ für alle $j \in \mathcal{A}$. Mit den vektoriellen Größen $\underline{Y}$ und $\underline{f}$ kann (3) daher wie in Abschnitt 4 als (14) geschrieben werden.

Wir testen das in Abschnitt 4 modifizierte Problem $\mathbf{H}_0''$ gegen $\mathbf{H}_1$. Als Schätzer verwenden wir auch in diesem multinomialen Logit-Modellen den Maximum-Likelihood-Schätzer. Da die Nullhypothese eine parametrische Verteilung der Fehler beinhaltet, deren Varianz beschränkt ist, sollte die Teststatistik $\hat{T}_n^{(v)}$ verwendet werden. Dies hatte sich auch als Folgerung aus den Simulationen ergeben. Die Tabellen 6 und 7 geben die Testergebnisse für Kerne der Form (19) mit verschiedenen Konstellationen der Parameter h und λ an. Die kritischen Werte wurden mit einem parametrischen Bootstrap-Verfahren auf der Basis von 1000 Iterationen ermittelt.

Die Nullhypothese, dass die Daten mit einem multinomialen Logit-Modell erklärt werden können, wird in allen Fällen mit Irrtumswahrscheinlichkeit kleiner $\alpha = 0.01$ abgelehnt. Insbesondere ist der geringe Einfluss der Parameter h und λ auf die Testentscheide zu erkennen. Da der Kern nicht, wie sonst üblich, durch h^{2m} geteilt wurde, sind sogar die Einflüsse auf die absoluten Werte relativ gering. Dies lässt darauf schliessen, dass eine ausgeprägte systematische Abweichung vom Logit-Modell vorliegt.

Die oberen Schranken liegen im 10-Marken-Fall deutlich und im 3-Marken-Fall einige Male über den Werten der Teststatistik. Aber diese Werte berücksichtigen als Maxima über alle Kerne insbesondere die diskrete Struktur der Variablen W in keiner Weise. Sie dienen daher nur als Vergleichsmaßstab oder als Kriterium dafür, ob man auf die Approximation der kritischen Werte mit Resampling-Verfahren hätte

Ergebnisse für 10 Marken					
	Teststatistik	**kritische Werte**		**obere Schranken**	
h , λ	$10^8 \cdot \hat{T}_n^{(v)}$	$\alpha = 0.05$	$\alpha = 0.01$	$\alpha = 0.05$	$\alpha = 0.01$
0.02, 0.80	**1.586**	0.795	0.813	3.213	5.170
0.02, 0.90	**1.580**	0.795	0.812	3.213	5.170
0.02, 0.95	**1.577**	0.795	0.812	3.213	5.170
0.02, 0.99	**1.575**	0.795	0.812	3.216	5.175
0.05, 0.80	**1.594**	0.795	0.815	3.211	5.168
0.05, 0.90	**1.576**	0.795	0.816	3.212	5.168
0.05, 0.95	**1.569**	0.795	0.817	3.212	5.169
0.05, 0.99	**1.565**	0.795	0.817	3.212	5.169
0.10, 0.80	**1.686**	0.799	0.822	3.209	5.165
0.10, 0.90	**1.643**	0.798	0.822	3.210	5.166
0.10, 0.95	**1.631**	0.799	0.823	3.211	5.167
0.10, 0.99	**1.623**	0.798	0.824	3.211	5.167
0.20, 0.80	**2.521**	0.807	0.829	3.203	5.155
0.20, 0.90	**2.378**	0.805	0.831	3.206	5.160
0.20, 0.95	**2.339**	0.805	0.833	3.207	5.161
0.20, 0.99	**2.316**	0.805	0.834	3.208	5.163

Tabelle 6: Tests auf Modell (3) für verschiedene Kerne

Ergebnisse für 3 Marken					
	Teststatistik	**kritische Werte**		**obere Schranken**	
h , λ	$10^8 \cdot \hat{T}_n^{(v)}$	$\alpha = 0.05$	$\alpha = 0.01$	$\alpha = 0.05$	$\alpha = 0.01$
0.02, 0.80	**8.253**	2.182	2.469	7.691	12.378
0.02, 0.90	**7.856**	2.192	2.498	7.697	12.387
0.02, 0.95	**7.692**	2.184	2.509	7.699	12.390
0.02, 0.99	**7.574**	2.185	2.516	7.701	12.393
0.05, 0.80	**9.087**	2.252	2.572	7.625	12.271
0.05, 0.90	**8.534**	2.234	2.582	7.640	12.296
0.05, 0.95	**8.300**	2.242	2.574	7.647	12.306
0.05, 0.99	**8.129**	2.243	2.594	7.651	12.313
0.10, 0.80	**11.280**	2.268	2.664	7.494	12.060
0.10, 0.90	**10.341**	2.308	2.687	7.529	12.116
0.10, 0.95	**9.956**	2.320	2.682	7.543	12.139
0.10, 0.99	**9.681**	2.338	2.667	7.553	12.155
0.20, 0.80	**18.037**	2.365	2.804	7.104	11.433
0.20, 0.90	**16.039**	2.389	2.804	7.195	11.579
0.20, 0.95	**15.259**	2.409	2.846	7.231	11.636
0.20, 0.99	**14.713**	2.430	2.856	7.256	11.677

Tabelle 7: Tests auf Modell (3) für verschiedene Kerne

verzichten können.

Mit dem 3-Marken-Datensatz wurden zur weiteren Analyse folgende fünf alternative Logit-Modelle mit anderen *Index-Funktion* $x_j^T\vartheta$ getestet:

Modell ohne Preis :	$x_j = (l_j, w_j)^T,$	$\vartheta \in I\!R^2$
Modell ohne Loyalität :	$x_j = (p_j, w_j)^T,$	$\vartheta \in I\!R^2$
bivariate Interaktion :	$x_j = (p_j, p_j l_j, l_j, w_j)^T,$	$\vartheta \in I\!R^4$
quadratisch in Preis und Loyalität :	$x_j = (p_j, p_j^2, l_j, l_j^2, w_j)^T,$	$\vartheta \in I\!R^5$
kubisch in Preis und Loyalität :	$x_j = (p_j, p_j^2, p_j^3, l_j, l_j^2, l_j^3, w_j)^T,$	$\vartheta \in I\!R^7$.

Diese Modelle erfüllen die Voraussetzungen für die Tests in gleicher Weise, wie das ursprünglich betrachtete Modell (14). Die kritischen Werte wurden wieder mit dem parametrischen Bootstrap-Verfahren auf der Basis von 1000 Iterationen ermittelt. Einige Testergebnisse sind in den Tabelle 8 bis 10zusammengefasst.

Auch die zu diesen fünf alternativen Modellen gehörenden Nullhypothesen werden stets abgelehnt. Die absoluten Werte der Teststatistiken für das Modell ohne Loyalität und das Modell ohne Preis sind weder untereinander noch mit dem ursprünglichen Modell (3) vergleichbar, da sich durch das Fortlassen einer Variable andere Kerngewichte ergeben. Die anderen drei Modelle sind Erweiterungen des Modells (3) und können den Daten deshalb nicht schlechter angepasst sein als dieses. Die Ergebnisse aus den Tabellen 9 und 10 weisen allerdings kaum eine Verbesserung gegenüber dem ursprünglichen Modell auf. Dies kann darauf deuten, dass vor allem die *Link-Funktionen* f_j selbst schlecht spezifiziert ist.

Die *Link-Funktionen* können isoliert getestet werden, wenn wir zusätzlich davon ausgehen, dass das Modell als Funktion des Indexes $\vartheta^T x$ gebildet werden muss (Su und Wei, 1991; Werwatz, 1997). Wir schreiben $f_j(x,\vartheta) = f_j^{(Ind)}(\xi_1,\ldots,\xi_m)$ mit $\xi_j := x_j^T\vartheta$ für $j \in I\!N_m$ und testen also $\mathbf{H}_0''$ gegen

$$\mathbf{H}_1^{(Ind)}: \quad \mathsf{D} \in \mathcal{D}_1^{(Ind)} := \bigcup_{g\in\mathcal{B}(R^m,R)} \mathsf{D}\{g\} \setminus \bar{\mathcal{D}}_0 \quad ,$$

wobei $\mathcal{B}(I\!R^m, I\!R)$ die Menge der Borel-messbaren Funktionen $g : I\!R^m \to I\!R$ bezeichne. In den Teststatistiken ist dann $\hat{K}_{ij} = k(\hat{\xi}_i, \hat{\xi}_j) = k^\dagger\big((X_i - X_j)^T\hat{\vartheta}_n\big)$ mit $\hat{\xi}_i := X_i^T\hat{\vartheta}_n$. Da der Träger von D_X kompakt ist, bleiben alle Vorausetzungen erfüllt. Testergebnisse von $\mathbf{H}_0''$ gegen $\mathbf{H}_1^{(Ind)}$ sind in der Tabelle 11 aufgeführt. Auch gegen diese Alternative wird $\mathbf{H}_0''$ bei kleinen und mittleren Bandweiten klar abgelehnt. Lediglich bei der größten betrachteten Bandweite $h = 1.00$ kann $\mathbf{H}_0''$ zum Niveau $\alpha = 0.01$ nicht abgelehnt werden. Allerdings ist diese Bandweite schon so groß, dass sich die Güte der Tests auf Abweichungen konzentriert, die lediglich aus einer Addition der Link-Funktion mit einer Konstanten bestehen. Es darf vermutet werden, dass die Link-Funktion in anderer Weise den Daten nicht gut angepasst ist.

Ergebnisse für 3 Marken (1)						
		Teststatistik	**kritische Werte**		**obere Schranken**	
Modell	h , λ	$10^3 \cdot \hat{T}_n^{(v)}$	$\alpha = 0.05$	$\alpha = 0.01$	$\alpha = 0.05$	$\alpha = 0.01$
	0.02, 0.80	**79.982**	41.274	48.997	121.613	195.715
	0.02, 0.90	**76.960**	40.485	48.819	121.868	196.126
	0.02, 0.95	**75.540**	40.212	48.943	122.014	196.360
	0.02, 0.99	**74.455**	40.135	49.150	122.133	196.551
	0.05, 0.80	**81.387**	41.509	48.844	121.701	195.855
	0.05, 0.90	**78.748**	40.688	47.985	121.958	196.270
ohne	0.05, 0.95	**77.498**	39.952	50.292	122.104	196.506
	0.05, 0.99	**76.539**	39.985	50.006	122.223	196.697
	0.10, 0.80	**98.459**	41.495	49.411	120.725	194.285
Preis	0.10, 0.90	**95.687**	40.548	48.036	121.026	194.770
	0.10, 0.95	**94.386**	40.481	48.103	121.190	195.033
	0.10, 0.99	**93.388**	40.226	48.719	121.321	195.244
	0.20, 0.80	**141.419**	42.598	52.199	117.195	188.604
	0.20, 0.90	**140.826**	41.906	52.772	117.661	189.354
	0.20, 0.95	**140.250**	41.998	53.213	117.906	189.749
	0.20, 0.99	**139.697**	41.740	52.936	118.101	190.062
	0.02, 0.80	**542.452**	45.909	52.386	143.873	231.538
	0.02, 0.90	**516.654**	45.341	52.337	144.117	231.931
	0.02, 0.95	**506.429**	45.304	51.853	144.200	232.065
	0.02, 0.99	**499.219**	45.526	51.735	144.253	232.150
	0.05, 0.80	**696.467**	48.274	57.814	138.918	223.565
	0.05, 0.90	**650.578**	47.721	57.934	139.763	224.924
ohne	0.05, 0.95	**631.882**	47.787	57.584	140.068	225.414
	0.05, 0.99	**618.509**	47.806	57.423	140.270	225.740
	0.10, 0.80	**692.936**	49.040	63.417	132.208	212.766
Loyalität	0.10, 0.90	**633.536**	49.475	61.976	134.019	215.680
	0.10, 0.95	**608.846**	49.237	61.149	134.684	216.751
	0.10, 0.99	**591.051**	49.015	61.814	135.133	217.473
	0.20, 0.80	**632.462**	51.497	69.210	125.965	202.718
	0.20, 0.90	**564.440**	51.544	65.993	128.704	207.127
	0.20, 0.95	**537.153**	51.470	65.250	129.735	208.785
	0.20, 0.99	**517.820**	51.528	64.813	130.440	209.920

Tabelle 8: Tests auf alternative Modelle für verschiedene Bandweiten

Ergebnisse für 3 Marken (2)						
		Teststatistik	**kritische Werte**		**obere Schranken**	
Modell	h , λ	$10^3 \cdot \hat{T}_n^{(v)}$	$\alpha = 0.05$	$\alpha = 0.01$	$\alpha = 0.05$	$\alpha = 0.01$
	0.02, 0.80	**8.015**	2.155	2.467	7.602	12.234
	0.02, 0.90	**7.614**	2.150	02.466	7.610	12.247
	0.02, 0.95	**7.450**	2.153	2.476	7.613	12.252
	0.02, 0.99	**7.333**	2.151	2.485	7.616	12.256
	0.05, 0.80	**8.927**	2.241	2.569	7.536	12.127
	0.05, 0.90	**8.356**	2.231	2.574	7.553	12.155
bivariate	0.05, 0.95	**8.117**	2.228	2.539	7.560	12.166
	0.05, 0.99	**7.944**	2.218	2.539	7.565	12.174
	0.10, 0.80	**10.959**	2.288	2.645	7.377	11.872
Interaktion	0.10, 0.90	**10.007**	2.294	2.650	7.416	11.935
	0.10, 0.95	**9.621**	2.279	2.721	7.432	11.960
	0.10, 0.99	**9.346**	2.285	2.691	7.443	11.978
	0.20, 0.80	**17.263**	2.319	2.892	6.923	11.141
	0.20, 0.90	**15.261**	2.329	2.891	7.017	11.292
	0.20, 0.95	**14.488**	2.333	2.953	7.053	11.351
	0.20, 0.99	**13.951**	2.355	2.944	7.078	11.391
	0.02, 0.80	**8.097**	2.133	2.369	7.591	12.216
	0.02, 0.90	**7.698**	2.140	2.410	7.603	12.236
	0.02, 0.95	**7.535**	2.140	2.430	7.608	12.244
	0.02, 0.99	**7.417**	2.138	2.439	7.611	12.249
quadratisch	0.05, 0.80	**9.030**	2.169	2.452	7.479	12.036
	0.05, 0.90	**8.454**	2.171	2.481	7.509	12.085
in	0.05, 0.95	**8.213**	2.182	2.502	7.521	12.104
	0.05, 0.99	**8.038**	2.189	2.500	7.529	12.117
Preis	0.10, 0.80	**11.400**	2.165	2.599	7.260	11.684
	0.10, 0.90	**10.393**	2.177	2.625	7.328	11.793
und	0.10, 0.95	**9.985**	2.182	2.660	7.354	11.835
	0.10, 0.99	**9.695**	2.206	2.641	7.37	2 11.864
Loyalität	0.20, 0.80	**18.521**	2.171	2.615	6.642	10.690
	0.20, 0.90	**16.340**	2.243	2.669	6.815	10.967
	0.20, 0.95	**15.495**	2.279	2.748	6.879	11.071
	0.20, 0.99	**14.907**	2.300	2.803	6.923	11.142

Tabelle 9: Tests auf alternative Modelle für verschiedene Bandweiten

Ergebnisse für 3 Marken (3)						
		Teststatistik	**kritische Werte**		**obere Schranken**	
Modell	h , λ	$10^3 \cdot \hat{T}_n^{(v)}$	$\alpha = 0.05$	$\alpha = 0.01$	$\alpha = 0.05$	$\alpha = 0.01$
	0.02, 0.80	**7.259**	2.083	2.272	7.322	11.783
	0.02, 0.90	**7.048**	2.082	2.308	7.338	11.808
	0.02, 0.95	**6.955**	2.082	2.302	7.344	11.818
	0.02, 0.99	**6.886**	2.087	2.318	7.348	11.825
kubisch	0.05, 0.80	**7.980**	2.100	2.296	7.176	11.549
	0.05, 0.90	**7.632**	2.119	2.344	7.219	11.618
in	0.05, 0.95	**7.477**	2.120	2.370	7.235	11.644
	0.05, 0.99	**7.362**	2.126	2.387	7.247	11.662
Preis	0.10, 0.80	**9.746**	2.103	2.432	6.918	11.133
	0.10, 0.90	**9.084**	2.141	2.474	7.008	11.277
und	0.10, 0.95	**8.813**	2.156	2.501	7.042	11.333
	0.10, 0.99	**8.619**	2.163	2.508	7.066	11.371
Loyalität	0.20, 0.80	**15.066**	2.081	2.520	6.254	10.064
	0.20, 0.90	**13.629**	2.150	2.552	6.450	10.380
	0.20, 0.95	**13.077**	2.174	2.625	6.524	10.500
	0.20, 0.99	**12.695**	2.186	2.631	6.576	10.583

Tabelle 10: Tests auf alternative Modelle für verschiedene Bandweiten

Ergebnisse für 3 Marken					
	Teststatistik	**kritische Werte**		**obere Schranken**	
h	$10 \cdot \hat{T}_n^{(v)}$	$\alpha = 0.05$	$\alpha = 0.01$	$\alpha = 0.05$	$\alpha = 0.01$
0.01	**1.144**	0.000	0.000	0.001	0.001
0.02	**1.257**	0.001	0.001	0.003	0.005
0.05	**1.548**	0.006	0.007	0.019	0.030
0.10	**1.784**	0.024	0.030	0.074	0.119
0.20	**1.846**	0.105	0.126	0.290	0.467
0.30	**2.102**	0.242	0.290	0.638	1.027
0.40	**2.408**	0.440	0.549	1.109	1.785
0.50	**2.683**	0.704	0.881	1.696	2.729
0.60	**2.937**	1.017	1.311	2.393	3.851
0.70	**3.178**	1.387	1.831	3.197	5.145
0.80	**3.398**	1.833	2.437	4.107	6.610
0.90	**3.590**	2.372	3.141	5.123	8.244
1.00	**3.752**	3.028	3.928	6.243	10.048

Tabelle 11: Test der Link-Funktionen für verschiedene Bandweiten

7 Schlussfolgerungen

Die vorgestellte Klasse von Tests stellt ein neues Werkzeug zur Beurteilung der Güte von Logit-Modellen dar. Im Gegensatz zu anderen statistischen Tests ist die Alternative nichtparametrisch. Dies bedeutet, dass keine zusätzlichen Annahmen zur Modellklasse notwendig sind, und dass nicht nur verschiedene Modelle innerhalb einer Klasse verglichen werden. Die Residuen werden nicht transformiert, bevor die Tests angewendet werden. Dies gibt den Tests eine gewisse Objektivität, die andere Verfahren nicht bieten. Darüber hinaus sind die Tests so ausgelegt, dass sie auch bei nur einer Beobachtung pro Zelle angewandt werden können und sich auch für multinomiale Modelle eignen. Eine Version der Tests erlaubt es die Link-Funktion, das heißt die unterstellte Verteilung der Residuen, separat zu testen. Ferner lassen sich aus den Ergebnissen der Tests mit verschiedenen Kernfunktionen k, insbesondere verschiedenen Bandweiten, grundsätzlich Rückschlüsse auf die Art der vorliegenden Abweichung ziehen.

Der Preis für diese Vorteile ist lediglich die Notwendigkeit, die kritischen Werte mit Stichprobenwiederholungsverfahren ermitteln zu müssen. Allerdings ist dies mit heutigen PCs auch bei größeren Datensätzen in vertretbarer Zeit (wenige Minuten) berechenbar.

In der konkreten Anwendung in Abschnitt 6 kann gefolgert werden, dass die Spezifikation eines Logit-Modells für die gegebenen Daten zur Markenwahl, insbesondere die Spezifikation der Link-Funktion, problematisch ist. Dies bedeutet natürlich nicht, dass Logit-Modelle prinzipiell nicht geeignet wären, die Produktwahl zu erklären. Aber dieses Resultat untermauert einige bestehende Bedenken gegen die vorbehaltlose Interpretation von Parametern der Logit-Modelle in ähnlichen Anwendungen. In weiteren Untersuchungen sollten Logit-Modelle für andere Datensätze getestet werden. Auch der direkte Vergleich der gegebenen Daten mit simulierten multinomialen Logit-Modellen auf *demselben* gegebenen Design X kann weitere Einsichten in gegebene Daten und das Verhalten der Tests bringen und sollte Gegenstand zukünftiger Arbeiten sein.

Anhang

A: Annahmen

Annahme A0. $Z_1 = (Y_1, X_1), \ldots, Z_n = (Y_n, X_n)$ ist für jedes $n \in \mathbb{N}$ eine unabhängige und identisch verteilte Stichprobe aus der gemeinsamen Verteilung D auf $\mathbb{R} \times \mathbb{R}^d$ mit $\mathsf{E}[Y_1^2] < \infty$. Die Randverteilung von X_1 wird mit D_X bezeichnet.

Annahme A1. Der Parameterbereich Θ_0 der Nullhypothese ist eine Teilmenge der offenen Menge $\Theta \subseteq \mathbb{R}^p$, $p \geq 1$.

Annahme A2. Die Funktion $f : \mathbb{R}^d \times \Theta \to \mathbb{R}$ ist für jedes feste ϑ Borel-messbar und zweimal stetig partiell differenzierbar bezüglich ϑ. Für $X \sim \mathsf{D}_X$ und jedes feste $\vartheta \in \Theta$ gelten $\mathsf{E}\big[f(X,\vartheta)^2\big] < \infty$ und $\mathsf{E}\big[[f'(X,\vartheta)]_\alpha^2\big] < \infty$ mit $\alpha \in \mathbb{N}_p$. Desweiteren existiert zu jedem $\vartheta \in \Theta$ eine Umgebung $\Psi = \Psi(\vartheta) \subseteq \Theta$ mit $\mathsf{E}\big[\sup_{\tau\in\Psi}[f''(X,\tau)]_{\alpha\beta}^2\big] < \infty$ für alle $\alpha, \beta \in \mathbb{N}_p$.

Annahme A3. Die Kernfunktion $k : \mathbb{R}^d \times \mathbb{R}^d \times \Theta \to \mathbb{R}$ ist bezüglich ϑ stetig partiell differenzierbar. k ist in den ersten beiden Argumenten symmetrisch und k und $[k']_\alpha$ sind für alle $\alpha \in \mathbb{N}_p$ beschränkte Funktionen auf ganz $\mathbb{R}^d \times \mathbb{R}^d \times \Theta$.

Annahme A4. Unter $\mathbf{H}_0$ mit $\mathsf{D} \in \mathsf{D}\{f(\cdot,\vartheta_0)\}$ gilt für den Schätzer $\hat{\vartheta}_n = \hat{\vartheta}(Z_1, \ldots, Z_n)$ die Entwicklung

$$n^{\frac{1}{2}}(\hat{\vartheta}_n - \vartheta_0) = n^{-\frac{1}{2}} \sum_{i=1}^{n} w(Z_i, \vartheta_0) + o_p(1)$$

für eine Borel-messbare und in ϑ stetige Funktion $w : \mathbb{R} \times \mathbb{R}^d \times \Theta \to \mathbb{R}^p$ mit $\mathsf{E}[w(Z_1,\vartheta_0) \,|\, X_1] = 0$. Des Weiteren existiert zu jedem $\vartheta \in \Theta$ eine Umgebung $\Psi = \Psi(\vartheta) \subseteq \Theta$ mit $\mathsf{E}[\sup_{\tau\in\Psi}[w(Z,\tau)]_\alpha^2] < \infty$ für alle $\alpha \in \mathbb{N}_p$.

Annahme A5. Unter $\mathbf{H}_1$ sind alle Häufungspunkte der Folge der Schätzer $\{\hat{\vartheta}_n \,|\, n \in \mathbb{N}\}$ fast sicher Elemente von Θ_0.

Annahme A6. Unter $\mathbf{H_0}$ mit $\mathsf{D} \in \mathsf{D}\{f(\cdot,\vartheta_0)\}$ ist der Schätzer $\hat{\vartheta}_n$ stark konsistent, das heißt es gilt $\hat{\vartheta}_n \xrightarrow{f.s.} \vartheta_0$.

Annahme A7. Die Funktion w in **A4** erfüllt zusätzlich $w(z,\vartheta) = w(y,x,\vartheta) = \tilde{w}(x,\vartheta)\cdot u(z,\vartheta)$ für eine Borel-messbare und in ϑ stetige Funktion $\tilde{w}(\cdot,\vartheta) : \mathbb{R}^d \times \Theta \to \mathbb{R}^p$. Desweiteren existiert zu jedem $\vartheta \in \Theta$ eine Umgebung $\Psi = \Psi(\vartheta) \subseteq \Theta$ mit $\mathsf{E}[\sup_{\tau\in\Psi}[\tilde{w}(X,\tau)]_\alpha^2] < \infty$ für alle $\alpha \in \mathbb{N}_p$.

B: Theoretische Ergänzungen

Nachweis der Annahmen für das binomiale Logit-Modell der Simulationsstudie

Annahme **A0** ist erfüllt, da die Beobachtungen identisch verteilt und unabhängig sind, und da Y als binäre Variable eine Varianz kleiner oder gleich $\frac{1}{2}$ hat. Die Modellfunktion f ist beschränkt und unendlich oft stetig differenzierbar. Mit dem Parameterbereich $\Theta_0 = I\!R^2$ gilt Annahme **A1** trivial. Die Verteilung D_X ist zwar unbekannt, hat aber ihren Träger in der kompakten Menge $[0,1]^{3\times m}$. Damit ist auch Annahme **A2** erfüllt. Als Kern wurde der univariate Gauss-Kern verwendet, der **A3** erfüllt. Als Schätzer wurde der Maximum-Likelihood-Schätzer bezüglich der für gegebenes X modellierten Binomialverteilung eingesetzt. Dieser erfüllt die Annahmen **A4**, **A5** und **A6** unter den vorliegenden Bedingungen (McFadden, 1974; Fahrmeir und Kaufmann, 1985). Somit sind die Tests auf dieses Modell anwendbar.

Nachweis der Annahmen für das multinomiale Logit-Modell für die Daten zur Produktwahl

Die Voraussetzungen im multivariaten Fall verlangen die Gültigkeit der Annahmen nur für jedes univariate Logit-Modell $\mathsf{E}[Y_j|X] = f_j(X,\vartheta)$, $j \in \mathcal{A}$ (Bartels, 1999, S.42). Es genügt daher, die Annahmen nur für f_0 zu prüfen.

Annahme **A0** ist erfüllt, da die Beobachtungen identisch verteilt und unabhängig sind, und da Y_0 als binäre Variable eine Varianz kleiner oder gleich $\frac{1}{2}$ hat. Die Modellfunktion f_0 ist beschränkt und unendlich oft stetig differenzierbar. Als Parameterbereich können wir $\Theta_0 = I\!R^3$ wählen, und Annahme **A1** gilt. Die Verteilung D_X ist zwar unbekannt, hat aber ihren Träger in der kompakten Menge $[0,1]^{3\times(m+1)}$. Damit ist auch Annahme **A2** erfüllt. Als Kern wählen wir die folgende Kombination aus dem Gauss-Kern für die stetigen Variablen und einem diskreten Kern für die binären Variablen w_j:

$$k\big(x^{(1)},x^{(2)}\big) = k_{h,\lambda}\big((p^{(1)},l^{(1)},w^{(1)})^T,(p^{(2)},l^{(2)},w^{(2)})^T\big) := \\ \lambda^{-m}\prod_{j=0}^{m}\Big\{\mathrm{gau}\big(\tfrac{p_j^{(1)}-p_j^{(2)}}{h}\big)\cdot\mathrm{gau}\big(\tfrac{l_j^{(1)}-l_j^{(2)}}{h}\big)\cdot\kappa_\lambda\big(|w_j^{(1)}-w_j^{(2)}|\big)\Big\} \quad (19)$$

mit $\kappa_\lambda(0) = \lambda$ und $\kappa_\lambda(1) = 1-\lambda$ für ein $\lambda \in (0.5,1)$. Da dieser beschränkte Kern nicht von ϑ abhängt, ist Annahme **A3** trivialerweise erfüllt. Ferner sind sowohl der Gauss-Kern als auch κ_λ positiv definite Kerne, so dass auch der multiplikative Kern $k_{h,\lambda}$ für alle Glättungsparameter h,λ positiv definit ist.

Der Maximum-Likelihood-Schätzer erfüllt die Annahmen **A4**, **A5** und **A6** auch unter den hier vorliegenden Bedingungen. Da D_X einen beschränkten Träger hat,

existieren die Schätzer fast sicher, und **A5** gilt. Die starke Konsistenz und die asymptotische Normalverteilung, und damit **A6** und **A4**, hängen von asymptotischen Eigenschaften der Fisher-Informations-Matrix ab, vorwiegend vom Verhältnis des größten zum kleinsten Eigenwert. Für die vorliegende empirische Verteilung D_{nX} sind diese Voraussetzungen für beide Datensätze erfüllt. Außerdem sind die Beobachtungen unabhängig und identisch verteilt, so dass wir diese Annahmen als erfüllt ansehen dürfen (Fahrmeir und Tutz, 1994, S.43). Somit ist die Anwendbarkeit der Tests sichergestellt.

Literatur

Azzalini, A., Bowman, W., und Härdle, W. (1989). On the use of nonparametric regression for model checking. *Biometrika 76*(1), 1–11.

Bartels, K. (1999). *Tests zur Modellspezifikation in der nichtlinearen Regression*. Universität Potsdam: Dissertation.

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

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

Bierens, H. J. (1982). Consistent model specification tests. *Journal of Econometrics 20*, 105–134.

Bierens, H. J. (1990). A consistent conditional moment test of functional form. *Econometrica 58*(6), 1443–1458.

Bierens, H. J. und Ploberger, W. (1997). Asymptotic theory of integrated conditional moment tests. *Econometrica 65*(5), 1129–1152.

Cox, D., Koh, E., Wahba, G., und Yandell, B. S. (1988). Testing the (parametric) null model hypothesis in (semiparametric) partial and generalized spline models. *The Annals of Statistics 16*(1), 113–119.

Diebolt, J. (1995). A nonparametric test for the regression function: Asymptotic theory. *Journal of Statistical Planning and Inference 44*, 1–17.

Eubank, R. L. und Spiegelman, S. (1990). Testing the goodness-of-fit of a linear model via nonparametric regression techniques. *Journal of the American Statistical Association 85*, 387–392.

Fahrmeir, L. und Kaufmann, H. (1985). Consistency and asymptotic normality of the maximum likelihood estimator in generalized linear models. *The Annals of Statistics 13*(1), 342–368.

Fahrmeir, L. und Tutz, G. (1994). *Multivariate Statistical Modelling based on Generalized Linear Models*. Springer Series in Statistics. New York: Springer.

Fan, Y. und Li, Q. (1996). Consistent model specification tests: Omitted variables and semiparametric functional form. *Econometrica 64*(4), 865–890.

Firth, D., Glosup, J., und Hinkley, D. V. (1991). Model checking with nonparametric curves. *Biometrika 78*(2), 245–252.

Gregory, G. G. (1977). Large sample theory for U-statistics and tests of fit. *The Annals of Statistics 5*(1), 110–123.

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

Härdle, W. und Horowitz, J. L. (1994). Testing a parametric model against a semiparametric alternative. *Econometric theory 10*, 821–848.

Härdle, W. und Mammen, E. (1993). Comparing nonparametric versus parametric regression fits. *The Annals of Statistics 21*(4), 1926–1947.

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

Kozek, A. S. (1991). A nonparametric test of fit of a parametric model. *Journal of Multivariate Analysis 37*, 66–75.

McCullagh, P. und Nelder, J. (1989). *Generalized Linear Models* (Second ed.). Number 37 in Monographs on Statistics and Applied Probability. London: Chapman & Hall.

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

Newey, W. K. (1985). Maximum likelihood specification testing and conditional moment tests. *Econometrica 53*(5), 1047–1070.

Rodrigues-Campos, M. C., Gonzales Manteiga, W., und Cao, R. (1998). Testing the hypothesis of a generalized linear regression model using nonparametric regression estimation. *Journal of Statistical Planning and Inference 67*, 99–122.

Staniswalis, J. G. und Severini, T. A. (1991). Diagnostics for assessing regression models. *Journal of the American Statistical Association 86*(415), 684–692.

Stute, W. (1997). Nonparametric model checks for regression. *The Annals of Statistics 25*(2), 613–641.

Su, J. Q. und Wei, L. J. (1991). A lack-of-fit test for the mean function in a generalized linear model. *Journal of the American Statistical Association 86*(414), 420–426.

Werwatz, A. (1997). A consistent test for misspecification in polychotomous response models. Discussion paper 74, SFB 373, Humboldt Universität zu Berlin.

White, H. (1981). Consequences and detection of misspecified nonlinear regression models. *Journal of the American Statistical Association 76*(374), 419–433.

Wu, C.-F. (1986). Jackknife, bootstrap and other resampling methods in regression analysis (with discussion). *The Annals of Statistics 14*, 1261–1350.

UNIVERSITÄT POTSDAM
Wirtschafts- und Sozialwissenschaftliche Fakultät
STATISTISCHE DISKUSSIONSBEITRÄGE
Herausgeber: Hans Gerhard Strohe
ISSN 0949-068X

Nr. 4 (1996) Berger, Ursula: *Die Landwirtschaft in den drei neuen EU-Mitgliedsstaaten Finnland, Schweden und Österreich - Ein statistischer Überblick -*

Nr. 5 (1996) Betzin, Jörg: *Ein korrespondenzanalytischer Ansatz für Pfadmodelle mit kategorialen Daten*

Nr. 6 (1996) Berger, Ursula: *Die Methoden der EU zur Messung der Einkommenssituation in der Landwirtschaft - Am Beispiel der Bundesrepublik Deutschland -*

Nr. 7 (1997) Strohe, Hans Gerhard / Geppert, Frank: *Algorithmus und Computerprogramm für dynamische Partial Least Squares Modelle*

Nr. 8 (1997) Rambert, Laurence / Strohe, Hans Gerhard: *Statistische Darstellung transformationsbedingter Veränderungen der Wirtschafts- und Beschäftigungsstruktur in Ostdeutschland*

Nr. 9 (1997) Faber, Cathleen: *Die Statistik der Verbraucherpreise in Rußland - Am Beispiel der Erhebung für die Stadt St. Petersburg -*

Nr. 10 (1998) Nosova, Olga: *The Atractiveness of Foreign Direct Investment in Russia and Ukraine - A Statistical Analysis*

Nr. 11 (1999) Gelaschwili, Simon: *Anwendung der Spieltheorie bei der Prognose von Marktprozessen*

Nr. 12 (1999) Strohe, Hans Gerhard / Faber, Cathleen: *Statistik der Transformation - Transformation der Statistik. Preisstatistik in Ostdeutschland und Russland*

Nr. 13 (1999) Müller, Claus: *Kleine und mittelgroße Unternehmen in einer hoch konzentrierten Branche am Beispiel der Elektrotechnik. Eine statistische Langzeitanalyse der Gewerbezählungen seit 1882*

Nr. 14 (1999) Faber, Cathleen: *The Measurement and Development of Georgian Consumer Prices*

Nr. 15 (1999) Geppert, Frank / Hübner, Roland: *Korrelation oder Kointegration - Eignung für Portfoliostrategien am Beispiel verbriefter Immobilenanlagen*

Nr. 16 (2000) Achsani, Noer Azam / Strohe, Hans Gerhard: *Statistischer Überblick über die indonesische Wirtschaft*

Nr. 17 (2000) Bartels, Knut: *Testen der Spezifikation von multinomialen Logit-Modellen*

Bezugsquelle: Universität Potsdam
Lehrstuhl für Statistik und Ökonometrie der
Wirtschafts- und Sozialwissenschaftlichen Fakultät
Postfach 90 03 27, D-15539 Potsdam
Tel. (+49 331) 977-32 25
Fax: (+49 331) 977-32 10
e-mail: stroherz.uni-potsdam.de