Zeitvariierende Effekte in Markenwahl-Modellen

Thomas Kneib

Institut für Statistik
Ludwig-Maximilians-Universität München

gemeinsam mit

Bernhard Baumgartner
Universität Regensburg

Winfried J. Steiner
Technische Universität Clausthal

29.9.2008

Überblick

- Klassische Markenwahlmodelle.
- Semiparametrische Erweiterung auf Modelle mit zeitvariierenden Effekten.
- Inferenz basierend auf gemischten Modellen.
- Anwendung auf Kaffee-Käufe.

Markenwahl-Daten

- Beim Kauf eines Produkts entscheidet sich der Konsument zwischen einer diskreten Auswahl von Alternativen (Substitutionsgüter).

- Ziel der Marketing-Forschung: Identifiziere Faktoren (Kovariablen), die das Entscheidungsverhalten beeinflussen.

- Zwei Typen von Kovariablen:
  - Globale Kovariablen: Unabhängig vom Produkt, z.B. Alter oder Geschlecht des Käufers.
  - Produktspezifische Kovariablen: Abhängig vom Produkt, z.B. Loyalität zu einer Marke, Preis, spezielle Werbung für das Produkt.

- Im Folgenden: Analyse von Käufen der größten Kaffee-Marken.

- Einige Charakteristika des Datensatzes:
  - Scanner-Paneldaten erhoben durch die Gesellschaft für Konsumforschung (GfK) Nürnberg.
  - Fünf Kaffeemarken mit insgesamt 53% Marktanteil.
  - Beobachtungszeitraum: Ein Jahr.
  - 7.439 Haushalte mit insgesamt 32.477 Kaufakten.
  - Aufgeteilt in 16.238 Beobachtungen zur Modellschätzung und 16.239 Beobachtungen zur Modellvalidierung.

- Kovariablen:

Loyalität	Loyalität des Käufers zu einer Marke.
Referenzpreis	Referenzpreis des Käufers, gebildet aus früheren Käufen.
Loss	negative Differenz zwischen Referenzpreis und Preis.
Gain	positive Differenz zwischen Referenzpreis und Preis.
Werbeaktivitäten	Dummy-Variable zu speziellen Werbemaßnahmen.

- Loyalität und Referenzpreis werden durch exponentiell gewichtete Mittelwerte aus früheren Käufen geschätzt.

Multinomiale Logit Modelle

- Idee eines Regressionsmodells zur Beschreibung der Markenwahl: Latente Nutzenvariablen, die mit dem Kauf einer Marke r durch den i-ten Konsumenten verbunden sind:
$$L_i^{(r)}, \quad r = 1, \dots, k.$$

- Die Nutzenvariablen sind nicht beobachtbar (nur die Kaufentscheidung).

- Bei rationalem Verhalten entscheidet sich der Käufer für das Produkt, das den Nutzen maximiert:

$$Y_i = r \quad \Longleftrightarrow \quad L_i^{(r)} = \max_{s=1,\dots,k} L_i^{(s)}.$$

- Beschreibe die Nutzenvariablen in Abhängigkeit von Kovariablen und einem Fehlerterm:

$$L_i^{(r)} = u_i'\alpha^{(r)} + w_i^{(r)}{}'\delta + \varepsilon_i^{(r)}.$$

- Bestandteile des Modells:

$u_i'\alpha^{(r)}$:	produktspezifische Effekte globaler Kovariablen u_i.
$w_i^{(r)}{}'\delta$:	globale Effekte produktspezifischer Kovariablen $w_i^{(r)}$.
$\varepsilon_i^{(r)}$:	Zufälliger Fehler.

- Speziell in der Anwendung:

$$L_i^{(r)} = \alpha^{(r)} + w_i^{(r)}{}'\delta + \varepsilon_i^{(r)}.$$

- Für standardextremwertverteilte Fehler ergibt sich aus

$$Y_i = r \quad \Longleftrightarrow \quad L_i^{(r)} = \max_{s=1,\dots,k} L_i^{(s)}.$$

das multinomiale Logit-Modell:

$$\pi_i^{(r)} = P(Y_i = r) = \frac{\exp(\eta_i^{(r)})}{1 + \sum_{s=1}^{k-1} \exp(\eta_i^{(s)})}, \qquad r = 1, \dots, k-1$$

mit

$$\eta_i^{(r)} = u_i'\alpha^{(r)} + (w_i^{(r)} - w_i^{(k)})'\delta = u_i'\alpha^{(r)} + \bar{w}_i^{(r)\prime}\delta.$$

- Aus Identifizierbarkeitsgründen können nur Kontraste produktspezifischer Kovariablen einbezogen werden.

- Die Kovariableneffekte wirken multiplikativ auf Verhältnis der Wahrscheinlichkeiten $\pi_i^{(r)}/\pi_i^{(k)}$.

- Positive Effekte bewirken eine verstärkte Auswahl der Kategorie r im Verhältnis zur Referenzkategorie k.

- Rein parametrische Markenwahl-Modelle vernachlässigen die zeitliche Dimension der Daten.

- Zeitlich variierende Präferenzen und Effekte ergeben sich beispielsweise durch
  - Veränderungen der ökonomischen Rahmenbedingungen.
  - Promotion-Aktivitäten.
  - Änderungen der Verwendungssituation (z.B. Feiertage).
  - Änderungen der Produkteigenschaften.

$\Rightarrow$ Semiparametrische Erweiterungen des multinomialen Logit-Modells.

Semiparametrische Multinomiale Logit-Modelle

- Erweitere das lineare Modell $L_{it}^{(r)} = \alpha^{(r)} + w_{it}^{(r)\prime}\delta + \varepsilon_{it}^{(r)}$ in zwei Schritten zu einem semiparametrischen Modell.

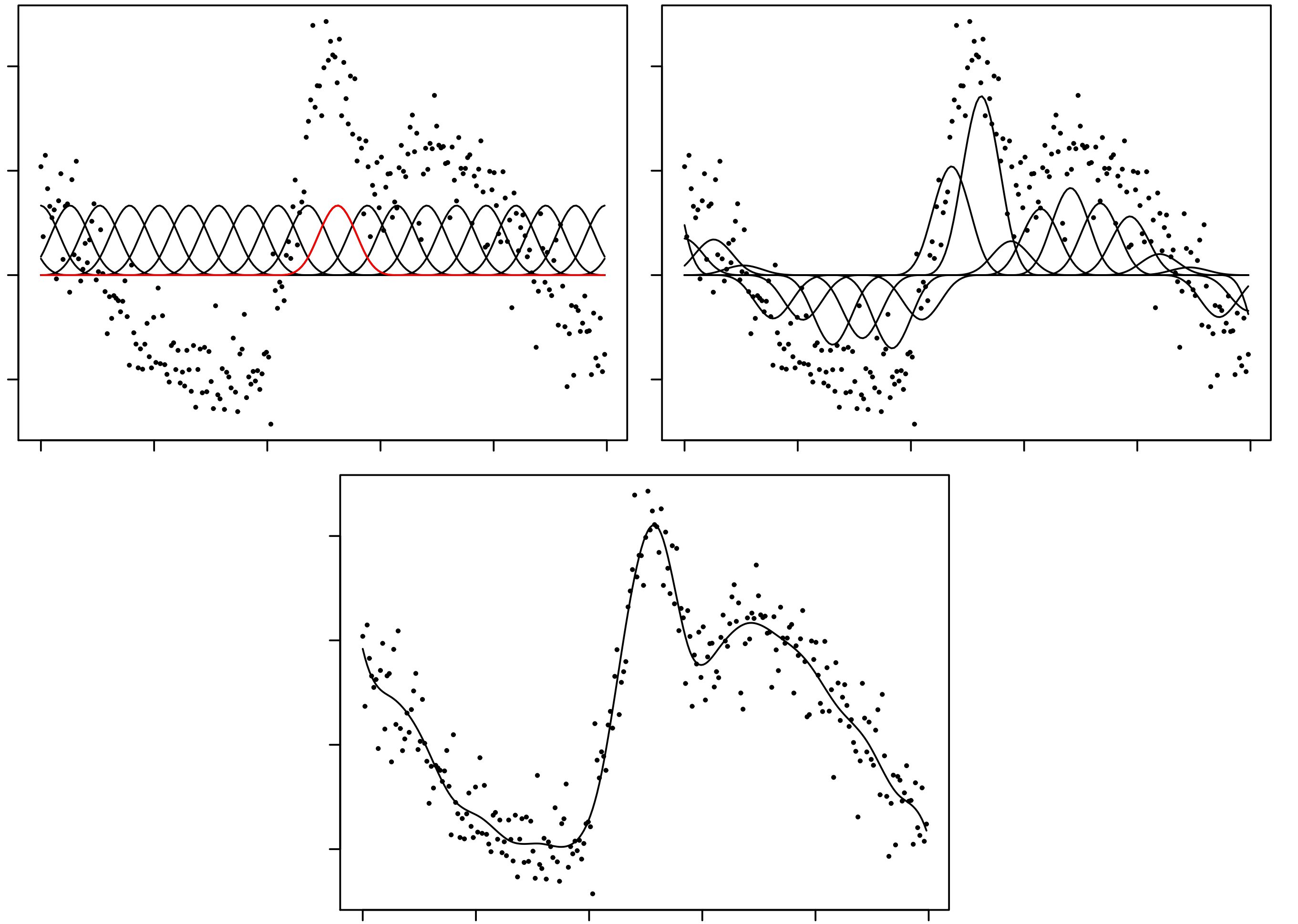
- Zeitvariierende Präferenzen: Ersetze die Konstanten des Modells durch zeitvariierende Präferenzen.
$$L_{it}^{(r)} = f_0^{(r)}(t) + w_{it}^{(r)\prime}\delta + \varepsilon_{it}^{(r)}.$$

- Zeitvariierende Effekte: Ersetze auch die Kovariableneffekte durch zeitvariierende Parameter.
$$L_{it}^{(r)} = f_0^{(r)}(t) + \sum_{j=1}^{J} w_{itj}^{(r)} f_j(t) + \varepsilon_{it}^{(r)}.$$

- Die Funktionen $f_0^{(r)}(t)$ und $f_j(t)$ sollen keine spezifische parametrische Form besitzen und flexibel (basierend auf penalisierten Splines) geschätzt werden.

- Spline-Schätzung: Approximiere die Funktionen $f(t)$ durch B-Spline-Basisfunktionen:

$$f(t) = \sum_{m=1}^{M} \beta_m B_m(t).$$



- Zur Regularisierung des Schätzproblems wird die Likelihood um einen Strafterm ergänzt.

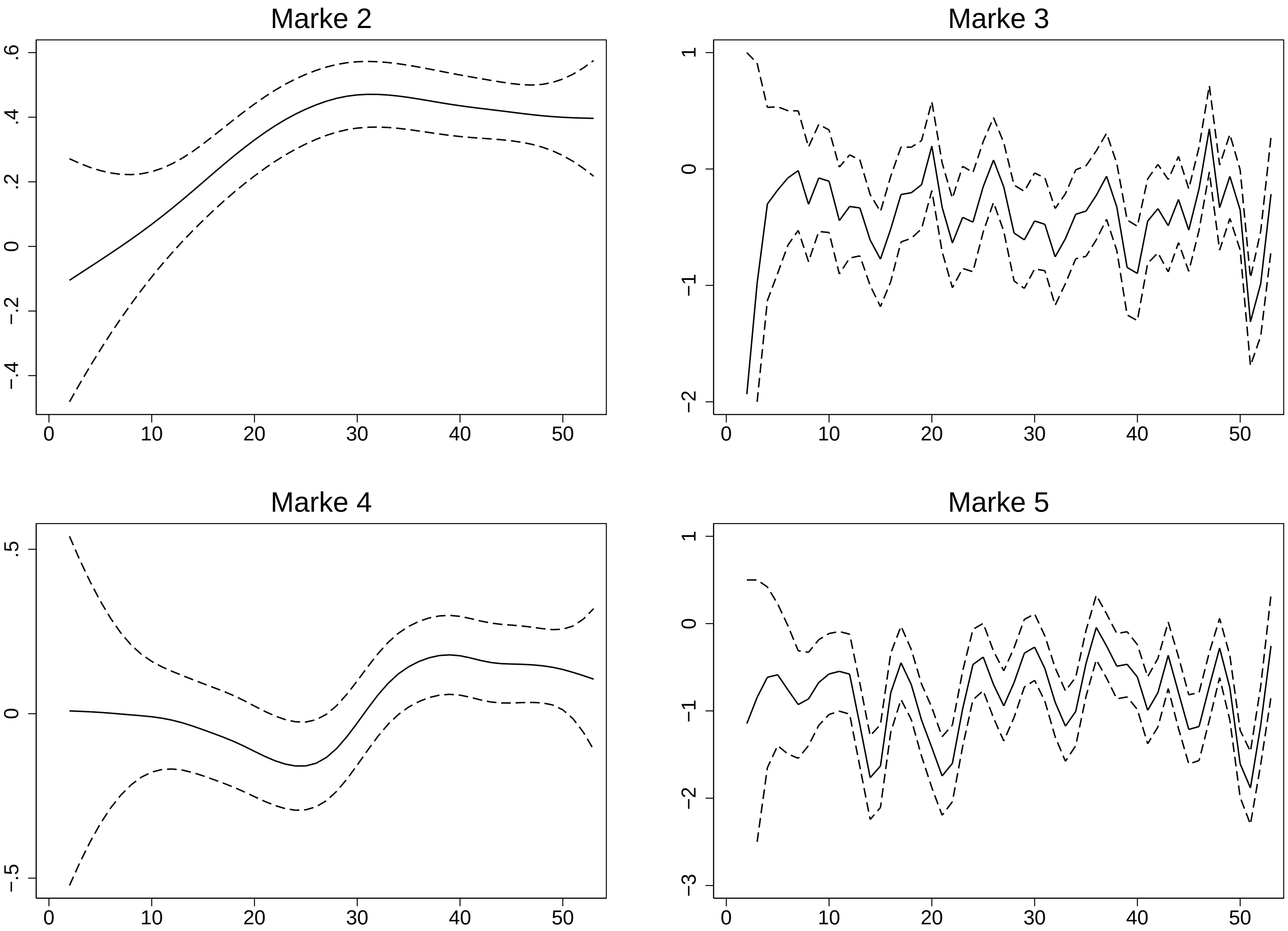
- Strafterme basierend auf Ableitungen lassen sich für B-Splines einfach durch Differenzenstrafterme approximieren:

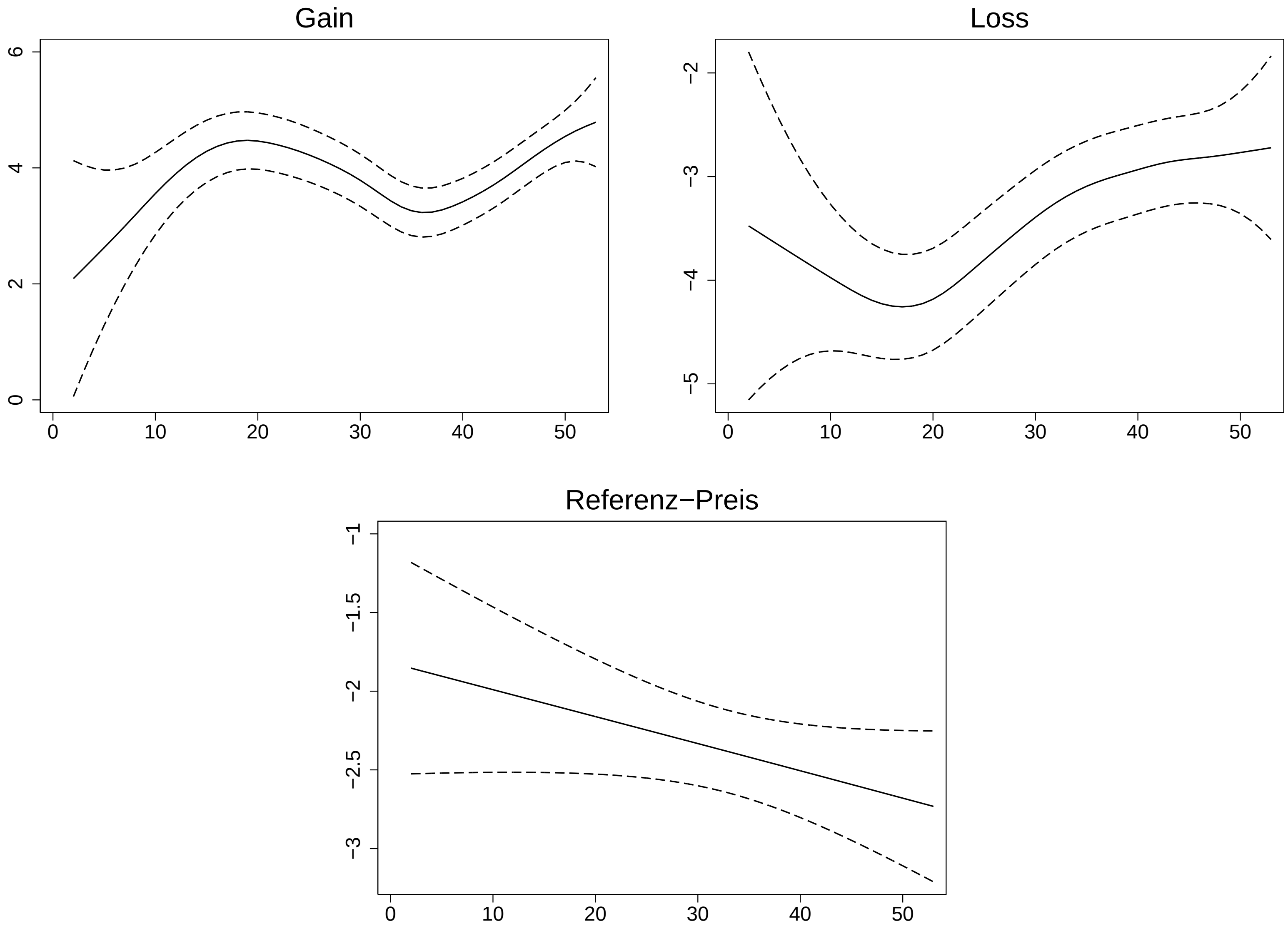
$$\frac{1}{2\tau^2}\sum_{m=2}^{M}(\beta_m - \beta_{m-1})^2 \qquad \text{(Differenzen erster Ordnung)}$$

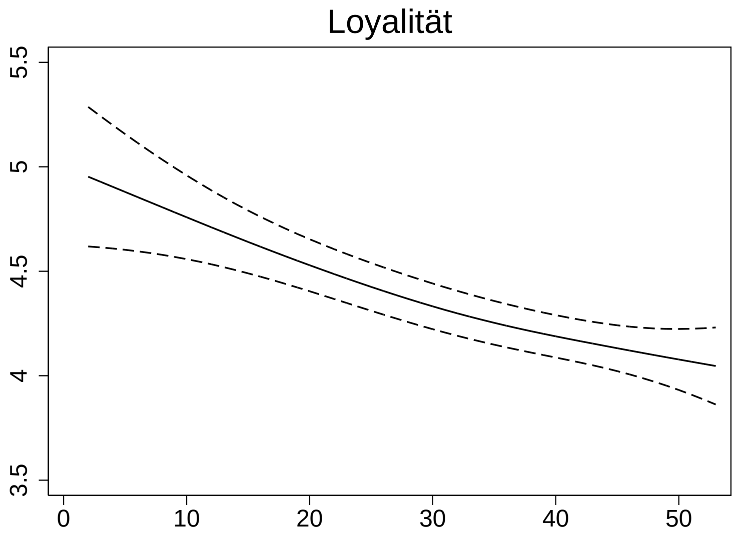
$$\frac{1}{2\tau^2}\sum_{m=3}^{M}(\beta_m - 2\beta_{m-1} + \beta_{m-2})^2 \qquad \text{(Differenzen zweiter Ordnung)}$$

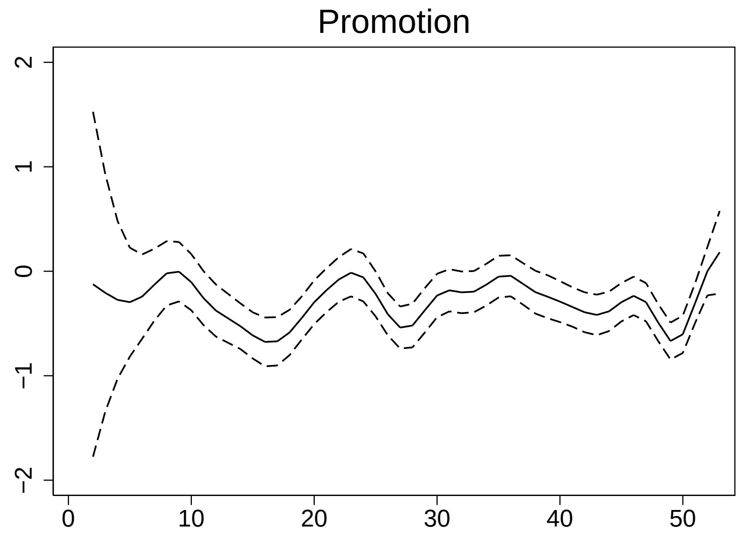
- Der Glättungsparameter τ^2 steuert den Ausgleich zwischen Datenanpassung (τ^2 groß) und Glattheit der Funktionsschätzung (τ^2 klein).

Ergebnisse


Marke 2
−.4
−.2
0
.2
.4
.6
0
10
20
30
40
50
Marke 3
−2
−1
0
1
0
10
20
30
40
50
Marke 4
−.5
0
.5
0
10
20
30
40
50
Marke 5
−3
−2
−1
0
1
0
10
20
30
40
50


Gain
Loss
Referenz−Preis
0
10
20
30
40
50
6
4
2
0
−2
−3
−4
−5
−1
−1.5
−2
−2.5
−3


Loyalität
3.5
4
4.5
5
5.5
0
10
20
30
40
50


Promotion
−2
−1
0
1
2
0
10
20
30
40
50

Exkurs: Inferenz basierend auf gemischten Modellen

- Penalisierte Splines sind ein spezielles Beispiel der penalisierten Likelihood-Schätzung

$$l_{\text{pen}}(\beta) = l(\beta) - \frac{1}{2\tau^2}\beta' K \beta \rightarrow \max_{\beta}.$$

- Die Likelihood $l(\beta)$ bewertet die Modellanpassung, die quadratische Form $\beta' K \beta$ bestraft zu große Modellkomplexität (bestimmt durch die Wahl der Strafmatrix K).

- Der Glättungsparameter τ^2 steuert den Ausgleich zwischen Modellanpassung und Modellkomplexität.

- Beispiele:
  - Penalisierte Splines: $K = D'D$ mit Differenzenmatrix D, z.B. für erste Differenzen

$$D\beta = \begin{pmatrix} -1 & 1 & & & \\ & -1 & 1 & & \\ & & \ddots & \ddots & \\ & & & -1 & 1 \end{pmatrix} \begin{pmatrix} \beta_1 \\ \vdots \\ \beta_M \end{pmatrix} = \begin{pmatrix} \beta_2 - \beta_1 \\ \vdots \\ \beta_M - \beta_{M-1} \end{pmatrix}$$

  - Markov-Zufallsfelder: K entspricht einer Nachbarschaftsmatrix und führt zur Schätzung räumlich glatter Effekte.
  - Zufällige Effekte: Unabhängig und identisch normalverteilte zufällige Effekte führen zu einem Ridge-Strafterm mit $K = I$.

- Für jeden dieser Ansätze lässt sich die penalisierte Likelihood aus verschiedenen Perspektiven betrachten und herleiten.

- Am Beispiel des Random Intercept Modells für Wiederholungsmessungen

$$y_{it} = x'_{it}\gamma + \beta_i + \varepsilon_{it}, \quad i = 1, \dots, n,\ t = 1, \dots, T,$$

mit

$$\begin{aligned} \beta_i &\quad \text{i.i.d.} \quad N(0, \tau^2), \\ \varepsilon_{it} &\quad \text{i.i.d.} \quad N(0, \sigma^2). \end{aligned}$$

- Gemischte Modelle: Die Individuen sind eine Zufallsstichprobe aus einer Population, so dass zwei Quellen der Zufälligkeit zu beachten sind.

- Marginale Sicht: Durch die zufälligen Effekte werden Korrelationen zwischen Beobachtungen eines Individuums induziert.

- Penalisierte Likelihood: Ziel ist die Schätzung individuenspezifischer Effekte. Zur Stabilisierung der Schätzung wird der Ridge-Strafterm

$$\frac{1}{2\tau^2}\sum_{i=1}^{n}\beta_i^2 = \frac{1}{2\tau^2}\beta'\beta$$

  eingeführt.

- Bayesianische Sicht: Die Verteilungsannahme für die zufälligen Effekte drückt Vorwissen über die Parameter β_i aus.

- In ähnlicher Weise lassen sich die vier Perspektiven für die anderen Penalisierungsansätze herleiten.

- Motiviert aus den unterschiedlichen Perspektiven wurden verschiedene Schätzverfahren entwickelt.

- Prinzipiell lässt sich jedes der Schätzverfahren auf jede der Perspektiven übertragen.

- Im Kontext gemischter Modelle am weitesten verbreitet: Restricted Maximum Likelihood Schätzung.

- Diese liefert insbesondere eine Schätzung für den Glättungsparameter und damit eine voll-automatische Schätzung beispielsweise für penalisierte Splines.

- Inferenz in semiparametrischen multinomialen Logit-Modellen basierend auf gemischten Modellen:
  - Schätzung der Spline-Koeffizienten über penalisiertes Fisher-Scoring.
  - Schätzung der Glättungsparameter basierend auf approximativer Restricted Likelihood.

- Im Detail Schwierigkeiten:
  - In Modellen mit nicht normalverteilten Zielvariablen sind die Perspektiven teilweise weniger offensichtlich und schwieriger zu interpretieren.
  - Hat die Strafmatrix K nicht vollen Rang, muss geeignet reparametrisiert werden, um zu einem gemischten Modell zu gelangen.

Modell-Validierung & Scoring-Regeln

- Ist die zusätzliche Modellkomplexität in semiparametrischen Regressionsmodellen tatsächlich notwendig?

- Modell-Validierung basierend auf der Vorhersage-Qualität.

- Wie misst man diese Vorhersage-Qualität und was ist eine Vorhersage?

- Im Folgenden: Vorhersage-Verteilungen

$$\hat{\pi} = (\hat{\pi}^{(1)}, \dots, \hat{\pi}^{(k)})$$

  basierend auf den modellierten Auswahlwahrscheinlichkeiten

$$\pi^{(r)} = P(Y = r).$$

- Eine Scoring-Regel ist eine reellwertige Funktion $S(\hat{\pi}, r)$, die der Beobachtung r einen Wert basierend auf der Vorhersage-Verteilung $\hat{\pi}$ zuweist.

- Ein Score ergibt sich dann durch Summation über Beobachtungen in einem Validierungs-Datensatz

$$S = \sum_{i=1}^{n} S(\hat{\pi}_i, r_i).$$

- Sei π_0 die wahre Verteilung. Dann heißt eine Scoring-Regel
  - Proper falls $\mathbb{E}_{\pi_0}(S(\pi_0, \cdot)) \leq \mathbb{E}_{\pi_0}(S(\hat{\pi}, \cdot))$ für alle $\hat{\pi}$ gilt.
  - Strikt proper falls die Gleichheit genau dann gilt, wenn $\hat{\pi} = \pi_0$.

- Übliche Beispiele:
  - Hit Rate (proper aber nicht strikt proper):

$$S(\hat{\pi}, r_i) = \begin{cases} \frac{1}{n} & \text{falls } \hat{\pi}^{(r_i)} = \max\{\hat{\pi}^{(1)}, \dots, \hat{\pi}^{(k)}\}, \\ 0 & \text{sonst.} \end{cases}$$

– Logarithmischer Score (strikt proper):

$$S(\hat{\pi}, r_i) = \log(\hat{\pi}^{(r_i)}).$$

– Brier-Score (strikt proper):

$$S(\hat{\pi}, r_i) = -\sum_{r=1}^{k} \left(\mathbb{1}(r_i = r) - \hat{\pi}^{(r)} \right)^2$$

– Sphärischer Score (strikt proper):

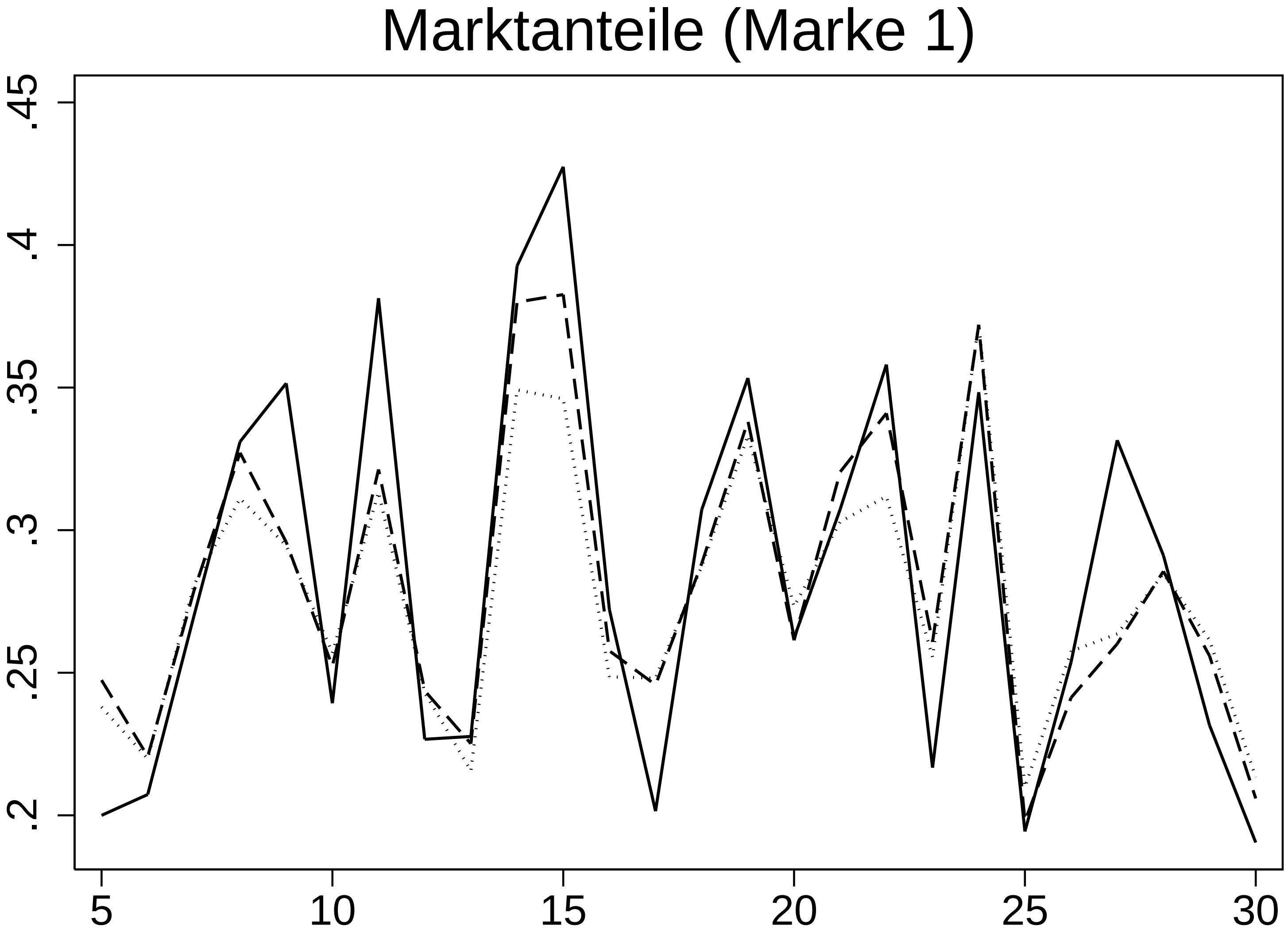
$$S(\hat{\pi}, r_i) = \frac{\hat{\pi}^{(r_i)}}{\sqrt{\sum_{r=1}^{k} (\hat{\pi}^{(r)})^2}}.$$

- Im Beispiel ergibt sich:

	parametrisch	zeitvar. Präf.	zeitvar. Effekte
Hit Rate (Schätzung)	0.6956	0.7011	0.7043
Hit Rate (Vorhersage)	0.6967	0.7007	0.7020
Logarithmisch (Schätzung)	-13816.1104	-13591.8655	-13498.9401
Logarithmisch (Vorhersage)	-13953.8974	-13844.1197	-13801.7610
Brier (Schätzung)	-6913.6315	-6807.7259	-6763.6767
Brier (Vorhersage)	-6930.8855	-6874.8116	-6859.8093
Sphärisch (Schätzung)	12101.5453	12174.2805	12199.5373
Sphärisch (Vorhersage)	12093.4520	12132.6391	12139.2852

- Die Verwendung zeitvariierender Präferenzen und Effekte ergibt eine deutlich verbesserte Vorhersage.

- Diese Verbesserung überträgt sich auch auf die Vorhersage der Marktanteile.



— wahrer Marktanteil, - - - zeitvariierende Effekte, · · · zeitkonstante Effekte

Software

- Die entwickelte Methodik ist im Software-Paket BayesX umgesetzt.
- Eigenständige Software für additive und geoadditive Regressionsmodelle.
- Unterstützt werden univariate Exponentialfamilien, kategoriale Regressionsmodelle und Modelle der zeitstetigen Verweildaueranalyse.
- BayesX ist kostenlos erhältlich unter

`http://www.stat.uni-muenchen.de/~bayesx`

Zusammenfassung

- Semiparametrische Erweiterung des multinomialen Logit-Modells.
- Automatisierte Schätzung aller Modellparameter (inklusive der Glättungsparameter).
- Model-Validierung basierend auf Scoring-Regeln.
- Referenzen:

  - Kneib, T., Baumgartner, B. & Steiner, W. J. (2007). Semiparametric Multinomial Logit Models for Analysing Consumer Choice Behaviour. *AStA Advances in Statistical Analysis*, **91**, 225–244.
  - Kneib, T., Baumgartner, B. & Steiner, W. J. (2008). Time-Varying Coefficients in Brand Choice Models (in Vorbereitung).

- Weitere Forschungsinteressen:
  - Boosting-Verfahren zur Modellwahl und Variablenselektion in geoadditiven Regressionsmodellen.
  - Bayesianische Regularisierungs-Prioris für hochdimensionale Regressionsmodelle.
  - Geoadditive Hazardraten-Modellierung für Verweildauern und Mehrstadien-Modelle.
  - Quantilregression.

- Mehr Informationen unter

`http://www.stat.uni-muenchen.de/~kneib`