Time-varying Coefficients in Brand Choice Modelling

Thomas Kneib

Department of Statistics
Ludwig-Maximilians-University Munich

joint work with

Bernhard Baumgartner
University of Regensburg

Winfried J. Steiner
Clausthal University of Technology

17.9.2008

Overview

- Brand Choice Modelling and the Multinomial Logit Model
- Semiparametric Multinomial Logit Models with Time-varying effects
- Application to Coffee Data.

Brand Choice Data

- When purchasing a product, the consumer is faced with a discrete set of alternatives (substitute goods).
- Aim of marketing research: Identify factors (covariates) guiding the decision.
- Two types of covariates:
  - Global covariates: Independent of the brand, e.g. age or gender of the consumer.
  - Product-specific covariates: Depending on the brand, e.g. loyalty to a brand, price, or advertising.

- In the following: Data on purchases of the coffee brands with largest market share.

- Some characteristics of the data set:
  - Scanner panel data collected by the Gesellschaft für Konsumforschung (GfK).
  - Five coffee brands with 53% market share in total.
  - Data collected over one year.
  - 7.439 households with 32.477 purchase acts in total.
  - Divided in 16.238 observations for model estimation and 16.239 observations for model validation.

- Covariates:

loyalty	loyalty of the consumer to a specific brand.
reference price	internal reference price built through experience.
loss	negative difference between reference price and price.
gain	positive difference between reference price and price.
promotional activity	dummy-variables for the presence of special promotion.

- Loyalty and reference price are estimated based on an exponentially weighted average of former purchases.

Multinomial Logit Models

- General idea of regression models describing brand choice: Latent utility associated with consumer i's decision to buy brand r:

$$L_i^{(r)}, \quad r = 1, \ldots, k.$$

- Note: We do not observe the utilities but only the brand choice decisions.

- Rational behaviour: The consumer chooses the product that maximises her/his utility:

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

- Express the utilities in terms of covariates and random error:

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

- Ingredients of the model:

$u_i'\alpha^{(r)}$:	product-specific effects of global covariates u_i.
$w_i^{(r)\prime}\delta$:	global effects of product-specific covariates $w_i^{(r)}$.
$\varepsilon_i^{(r)}$:	random error.

- In our application:

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

- If the error terms are i.i.d. standard extreme value distributed, the principle of maximum utility

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

yields the multinomial logit model

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

with

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

- Only contrasts of product-specific covariates can be included due to identifiability restrictions.

- The covariate effects act multiplicatively on the ratios of choice probabilities $\pi_i^{(r)}/\pi_i^{(k)}$.

- Positive effects cause an increase in the preference for brand r as compared to the reference category k.

- Purely parametric models neglect the temporal dimension of the data.
- Time-varying preferences and time-varying effects result, for example,
  - if the economic conditions change.
  - due to promotional activities.
  - due to specific usage situations or special events (such as holidays).
  - changes of properties of the product.

$\Rightarrow$ Semiparametric extensions of the multinomial logit model.

Semiparametric Multinomial Logit Models

- Extend the linear model $L_{it}^{(r)} = \alpha^{(r)} + w_{it}^{(r)\prime}\delta + \varepsilon_{it}^{(r)}$ in two steps.

- Time-varying preferences: Replace the constant intercepts with time-varying functions.

$$L_{it}^{(r)} = f_0^{(r)}(t) + w_{it}^{(r)\prime}\delta + \varepsilon_{it}^{(r)}.$$

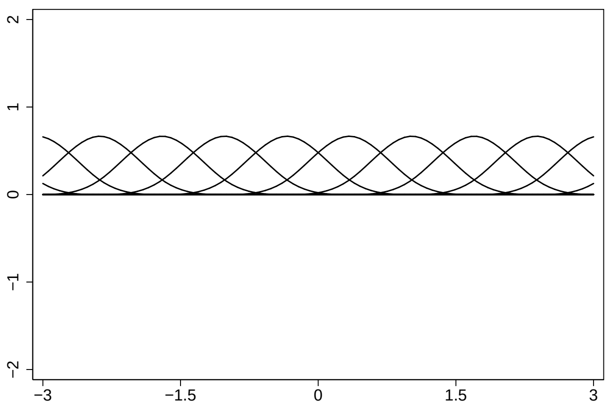
- Time-varying effects: Replace the regression coefficients with time-varying parameters.

$$L_{it}^{(r)} = f_0^{(r)}(t) + \sum_{j=1}^{J} w_{itj}^{(r)} f_j(t) + \varepsilon_{it}^{(r)}.$$

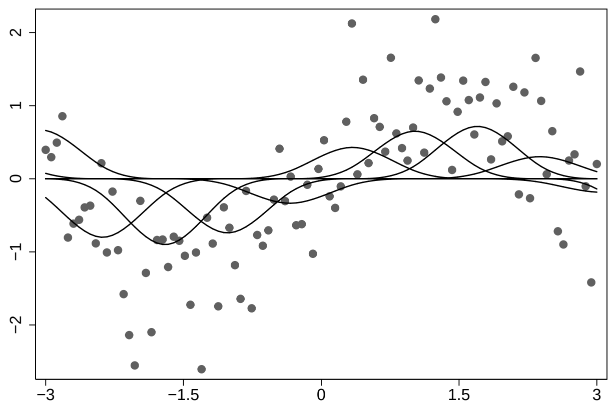
- The functional form of the functions $f_0^{(r)}(t)$ and $f_j(t)$ shall remain unspecified and will be estimated based on penalised splines.

- Spline smoothing: Approximate a function $f(t)$ in terms of a B-spline basis:

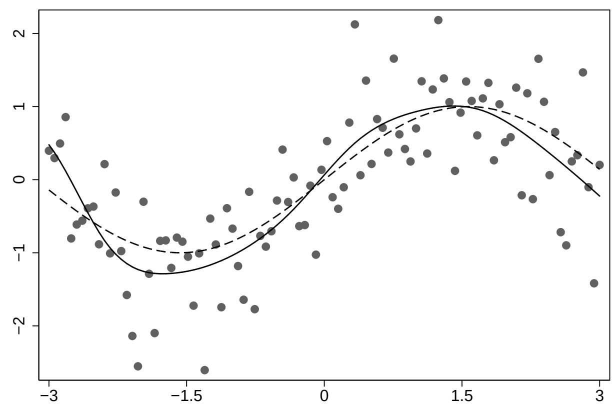
$$f(t) = \sum_{m=1}^{M} \beta_m B_m(t).$$



B-spline basis



Scaled B-splines



Resulting estimate

- To regularise estimation, add a smoothness penalty to the likelihood.

- Integral penalties can be approximated with difference penalties:

$$\frac{1}{2\tau^2}\sum_{m=2}^{M}(\beta_m - \beta_{m-1})^2 \qquad \text{(first order differences)}$$

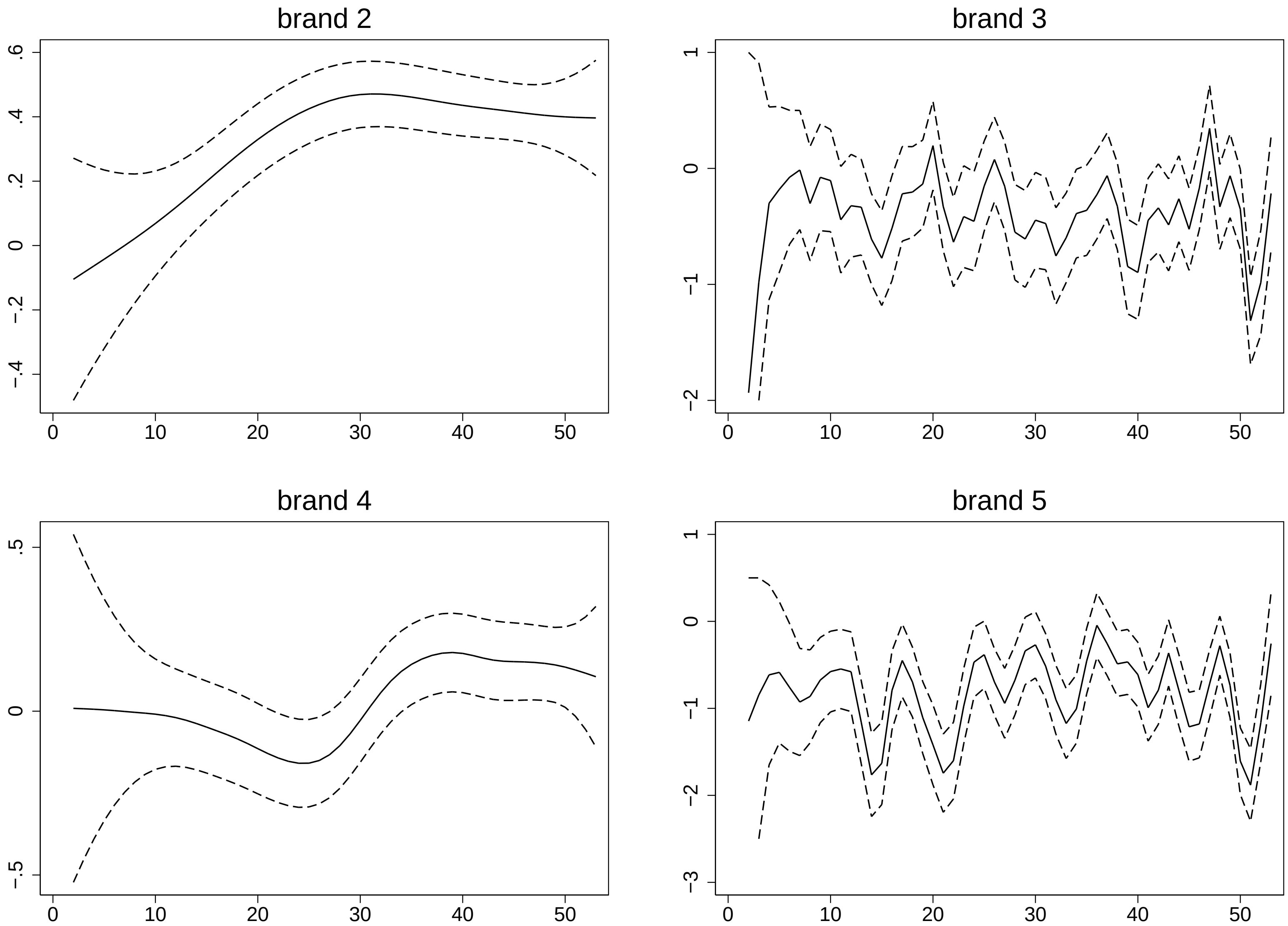
$$\frac{1}{2\tau^2}\sum_{m=3}^{M}(\beta_m - 2\beta_{m-1} + \beta_{m-2})^2 \qquad \text{(second order differences)}$$

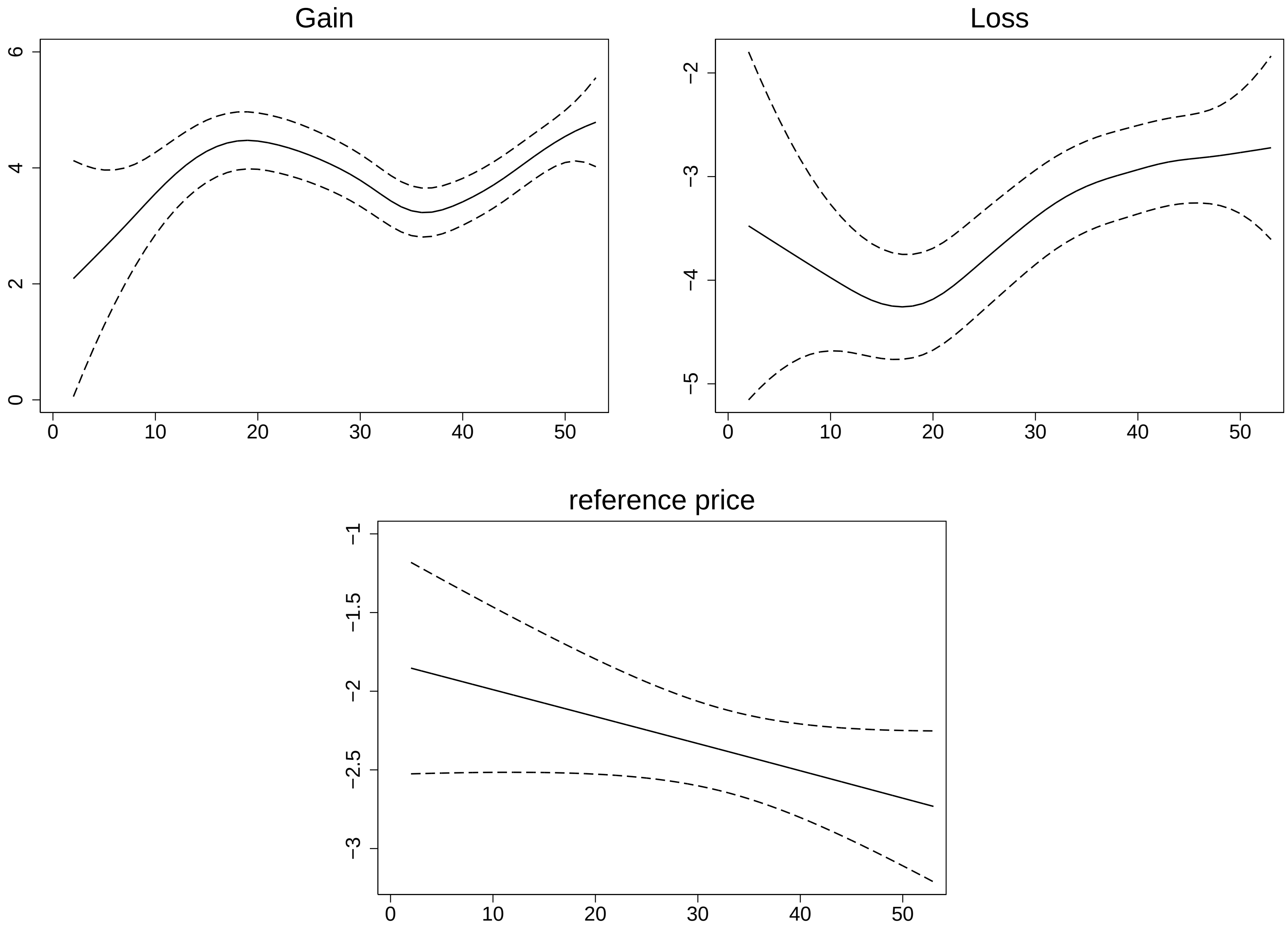
- The smoothing parameter τ^2 governs the trade-off between fidelity to the data (τ^2 large) and smoothness of the function estimate (τ^2 small).

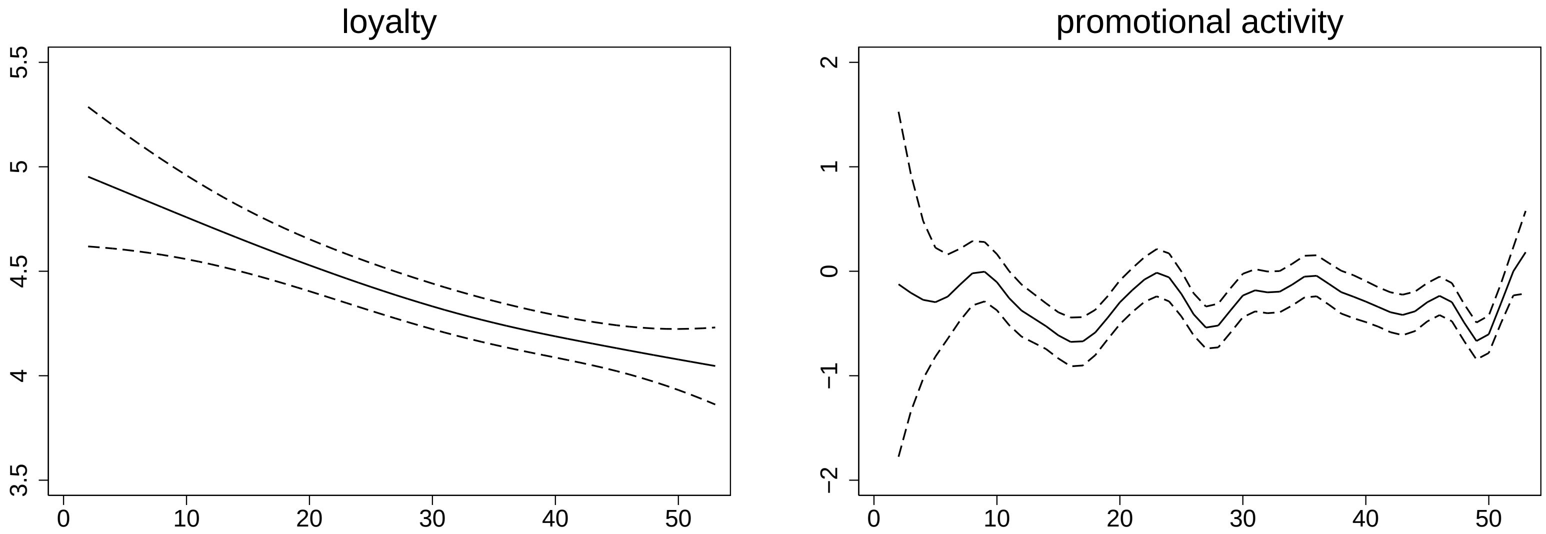
Statistical Inference

- The model contains two types of parameters:
  - Regression coefficients describing parametric or nonparametric effects and
  - smoothing parameters.
- Penalised likelihood estimation for the regression coefficients based on a modified Fisher scoring algorithm.
- Estimation of smoothing parameters based on an approximate marginal likelihood.

Results


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


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


loyalty
5.5
5
4.5
4
3.5
0
10
20
30
40
50
promotional activity
2
1
0
−1
−2
0
10
20
30
40
50

Model Validation & Proper Scoring Rules

- Is the increased model complexity in semiparametric regression models really required for our data?
- Model validation based on the prediction performance.
- What are suitable measures of predictive performance? What is a prediction?
- We consider predictive distributions

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

  based on the estimated choice probabilities

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

- A scoring rule is a real-valued function $S(\hat{\pi}, r)$ that assigns a value to the event that category r is observed when $\hat{\pi}$ is the predictive distribution.

- Score: Sum over individuals in a validation data set

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

- Let π_0 denote the true distribution. Then a scoring rules is called
  - Proper if $\mathbb{E}_{\pi_0}(S(\pi_0, \cdot)) \leq \mathbb{E}_{\pi_0}(S(\hat{\pi}, \cdot))$ for all $\hat{\pi}$.
  - Strictly proper if equality holds only if $\hat{\pi} = \pi_0$.

- Some common examples:
  - Hit rate (proper but not strictly proper):

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

- Logarithmic score (strictly proper):

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

- Brier score (strictly proper):

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

- Spherical score (strictly proper):

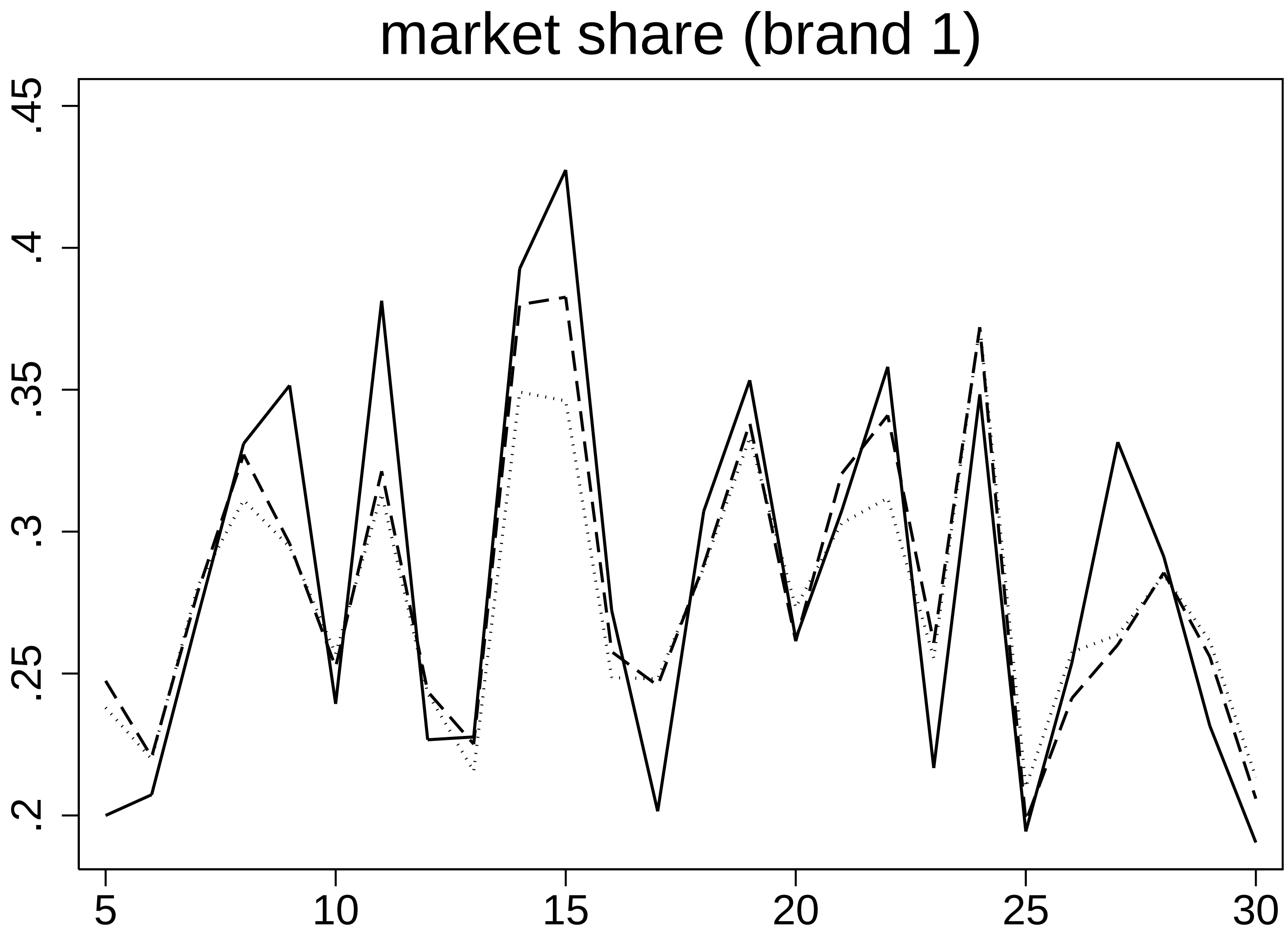
$$S(\hat{\pi}, r_i) = \frac{\hat{\pi}^{(r_i)}}{\sqrt{\sum_{r=1}^{k} (\hat{\pi}^{(r)})^2}}.$$

- In our data sets:

	parametric	time-varying preferences	time-varying effects
hit rate (estimation)	0.6956	0.7011	0.7043
hit rate (prediction)	0.6967	0.7007	0.7020
logarithmic (estimation)	-13816.1104	-13591.8655	-13498.9401
logarithmic (prediction)	-13953.8974	-13844.1197	-13801.7610
Brier (estimation)	-6913.6315	-6807.7259	-6763.6767
Brier (prediction)	-6930.8855	-6874.8116	-6859.8093
spherical (estimation)	12101.5453	12174.2805	12199.5373
spherical (prediction)	12093.4520	12132.6391	12139.2852

- The inclusion of time-varying preferences and effects results in an increased prediction performance.

- This is also reflected in the prediction of market shares.



— true market share, - - - time-varying effects, · · · time-constant effects

Software

- Implemented in the free software package BayesX.
- Stand-alone software for additive and geoadditive regression.
- Supports univariate exponential families, categorical regression and time-continuous duration time models.
- Available from

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

Summary

- Semiparametric extensions of the multinomial logit model.
- Fully automatic estimation of all model parameters (incl. the smoothing parameters).
- Model validation based on scoring rules.
- References:

  - 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 preparation).

- A place called home:

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