Semiparametric Multinomial Logit Models for the Analysis of Brand Choice Behaviour

Thomas Kneib

Department of Statistics
Ludwig-Maximilians-University Munich

joint work with
Bernhard Baumgartner & Winfried J. Steiner
University of Regensburg

28.3.2007

Brand Choice Data

- When purchasing a specific brand, the consumer is faced with a discrete set of alternatives.
- One aim of marketing analyses: Identify the influence of covariates on brand choice behaviour.
- Two types of covariates:
  - Global covariates: Fixed for all categories, e.g. age, gender of the consumer.
  - Brand-specific covariates: Depending on the category, e.g. loyalty to a product, price, presence of special advertisement.
- We will consider data on purchases of the most frequently bought brands of coffee, ketchup and yogurt.

- Main characteristics of the data sets:

	Coffee	Ketchup	Yogurt
Number of brands	five	three	five
Market share	53%	87%	74%
Sample size	49.083	26.820	66.679

- Covariates:

Loyalty	Loyalty of the consumer to a specific brand.
Reference price	Internal reference price built through experience.
Difference between reference price and price	Deviation of the actual price from the reference 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.

- Model the decision using latent utilities associated with buying a specific brand r:

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

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

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

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

- Express the utilities in terms of covariates and an error term:

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

- If the error term is standard extreme value distributed, we obtain the multinomial logit model.

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

with

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

- Some marketing theories suggest the possibility of nonlinear influences of some of the covariates.

- Example: Adaptation level theory.
  - Consumers compare prices to internal reference prices build through experience.
  - Around the reference point (price equals reference price) there may be a region of indifference.
  - Suggests a sigmoid-shaped form of the covariate-effect.

$\Rightarrow$ Semiparametric extensions of the multinomial logit model to validate such hypotheses.

Semiparametric Multinomial Logit Models

- Extend the linear predictor to a semiparametric predictor

$$\eta_i^{(r)} = u_i'\alpha^{(r)} + \bar{w}_i^{(r)\prime}\delta + \sum_{j=1}^{l} f_j^{(r)}(x_{ij}) + \sum_{j=l+1}^{p} \bar{f}_j(x_{ij}^{(r)})$$

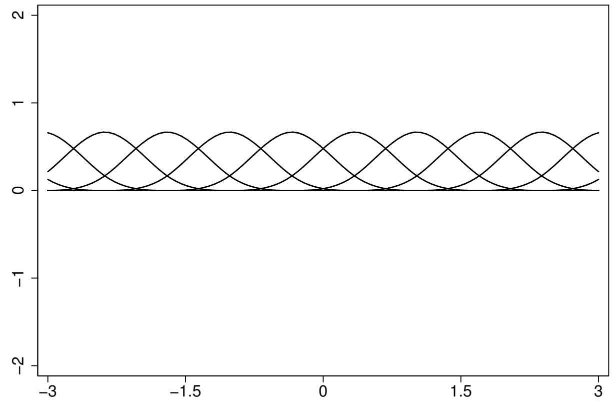
where

$$\bar{f}_j(x_{ij}^{(r)}) = f_j(x_{ij}^{(r)}) - f_j(x_{ij}^{(k)}).$$

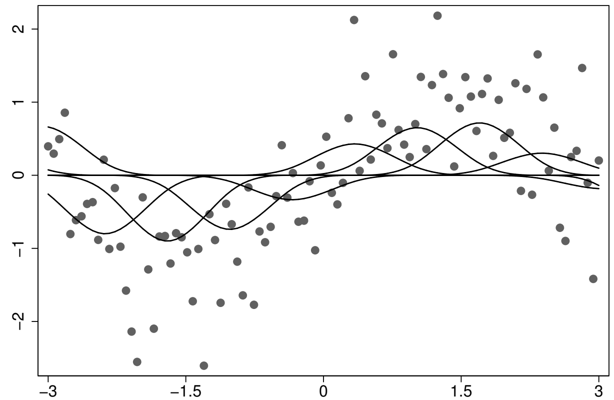
- The functions $f_j^{(r)}$ and f_j are modelled using penalised splines.

- Represent a function $f(x)$ as a linear combination of B-spline basis functions:

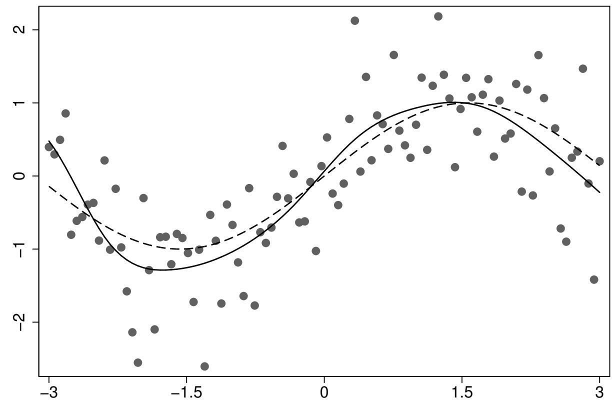
$$\sum_{m=1}^{M} \beta_m B_m(x).$$



B-spline basis



Scaled B-splines



B-spline fit

- Use a large number of basis functions to guarantee enough flexibility but augment a penalty term to the likelihood to ensure smoothness.

- Approximate derivative penalties are obtained by difference penalties, e.g.

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

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

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

- Penalty terms in matrix notation:

$$\frac{1}{2\tau^2}\beta' K \beta$$

with penalty matrix $K = D'D$ and appropriate difference matrices D.

Inference

- Two different types of parameters in the model:
  - Regression coefficients describing either parametric or semiparametric effects, and
  - Smoothing parameters.

- Penalised likelihood for the regression coefficients:

$$l_{\text{pen}}(\alpha, \delta, \beta) = l(\alpha, \delta, \beta) - \sum_{r=1}^{k-1} \sum_{j=1}^{q} \frac{1}{2(\tau_j^{(r)})^2} \beta_j^{(r)\prime} K_j \beta_j^{(r)} - \sum_{j=q+1}^{p} \frac{1}{2\tau_j^2} \beta_j' K_j \beta_j.$$

- $l(\alpha, \delta, \beta)$ is the usual likelihood of a multinomial logit model.

- Maximisation can be achieved by a slight modification of Fisher scoring.

- Estimate smoothing parameters based on marginal likelihood:

$$L(\tau^2) = \int L_{\mathsf{pen}}(\alpha, \delta, \beta, \tau^2) d\alpha \, d\delta \, d\beta \to \max_{\tau^2} .$$

- Laplace approximation to the integral yields a working Gaussian model.

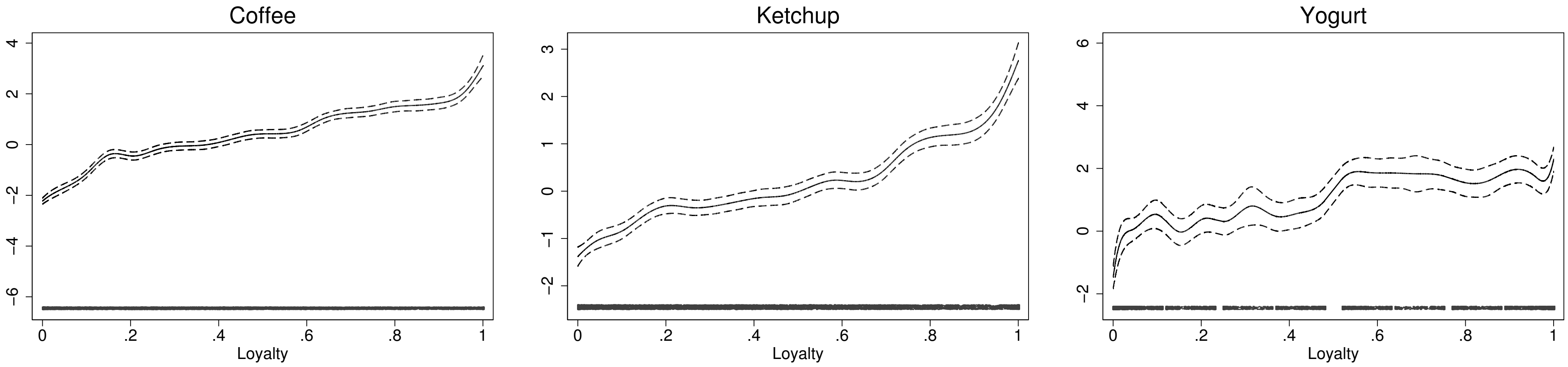
  $\Rightarrow$ Integral becomes tractable.

- Fisher scoring algorithm in the working model.

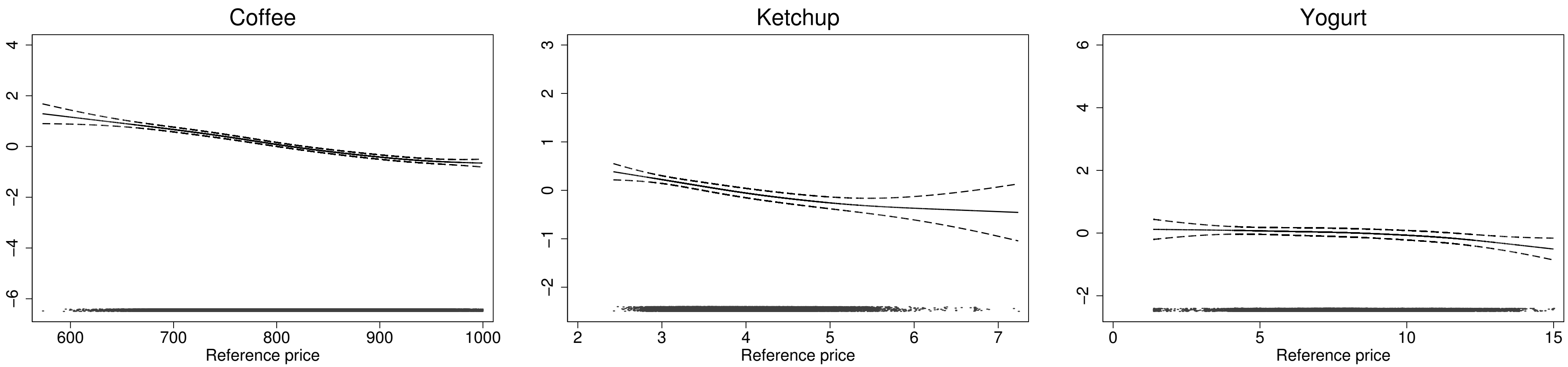
- Marginal likelihood corresponds to restricted maximum likelihood estimation in Gaussian regression models.

Results

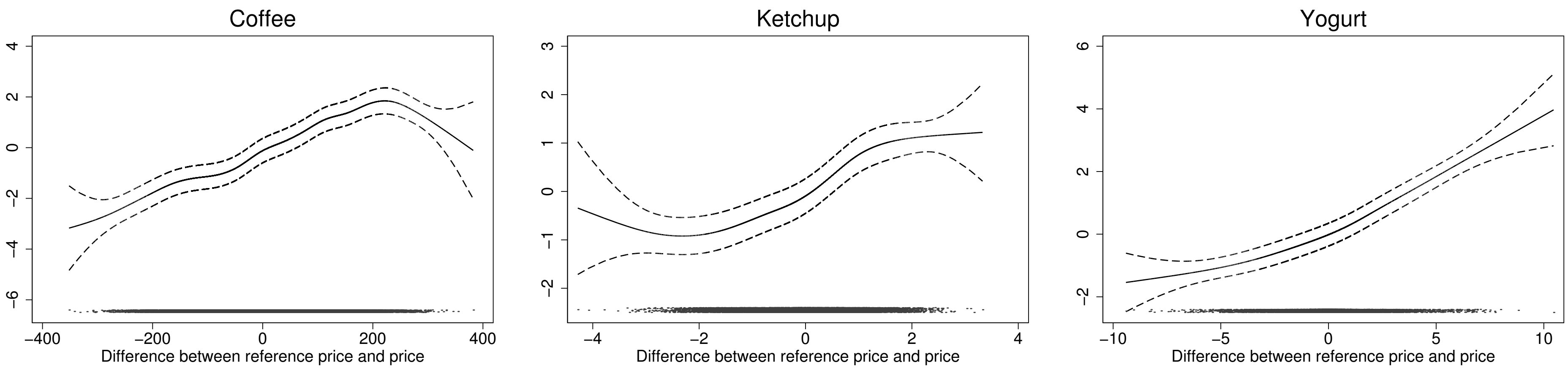
- Loyalty:



- Reference price:



- Difference between reference price and price:



Model Evaluation & Proper Scoring Rules

- We propose to use a more complicated model. Is the increased model complexity necessary?

- Validate the model based on its predictive performance.

- What are suitable measures of predictive performance? What is a prediction?

- We consider predictive distributions

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

  with the model 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 $S(\pi_0, \pi_0) \leq S(\hat{\pi}, \pi_0)$ for all π.
  - Strictly proper if equality holds only if $\hat{\pi} = \pi_0$.

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

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

– 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:

	Coffee		Ketchup		Yogurt	
	parametric	semipar.	parametric	semipar.	parametric	semipar.
Hit rate (est.)	0.70	0.70	0.79	0.79	0.82	0.83
Hit rate (pred.)	0.70	0.66	0.78	0.78	0.83	0.81
Logarithmic (est.)	-13816.90	-13491.45	-5146.61	-5024.40	-8502.60	-7923.95
Logarithmic (pred.)	-13955.80	-15682.87	-5297.58	-5225.32	-24061.49	-26588.13
Brier (est.)	-6912.34	-6789.38	-5192.60	-5222.72	-4261.52	-4044.11
Brier (pred.)	-6930.30	-7646.83	-2990.25	-2962.46	-12919.96	-12416.39
Spherical (est.)	12102.09	12181.02	6455.08	6450.31	12678.38	12798.71
Spherical (pred.)	12093.57	11588.11	7688.11	7701.58	37965.96	38231.85

- Coffee data: Parametric model seems sufficient.
- Ketchup data: Improved performance with semiparametric model.
- Yogurt data: Some indication of a need for semiparametric extensions but no definite answer.

Software

- Proposed methodology is implemented in the software package BayesX.

- Stand-alone software for additive and geoadditive regression models.

- Supports exponential family regression, categorical regression and hazard regression for continuous time survival analysis.

- The current version is Windows-only but a Linux version and a connection to R are work in progress.

- Available from

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

Summary

- Semiparametric extension of the well-known multinomial logit model.
- Fully automated fit (including smoothing parameters).
- Model validation based on proper scoring rules.
- Reference: Kneib, T., Baumgartner, B. & Steiner, W. J. (2007). Semiparametric Multinomial Logit Models for Analysing Consumer Choice Behaviour. Under revision for *AStA Advances in Statistical Analysis*.
- A place called home:

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