

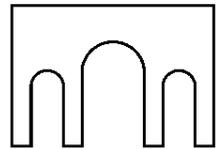
Mixed Model basierte Inferenz in komplexen Regressionsproblemen

Thomas Kneib
Institut für Statistik
Ludwig-Maximilians-Universität München

in Zusammenarbeit mit
Bernhard Baumgartner, Andrea Hennerfeind & Ludwig Fahrmeir

LMU

24.04.2006



Überblick

- Discrete Choice Modelle zur Markenwahl-Analyse
- P-Splines und gemischte Modelle
- Kindersterblichkeit in Nigeria
- Multi-State Modelle zur Beschreibung des menschlichen Schlafverlaufs

Discrete Choice Modelle zur Markenwahl-Analyse

- Ziel: Beschreibung des Entscheidungsprozesses von Kunden beim Kauf eines bestimmten Produkts.
- Häufig verwendetes Verfahren: **Multinomiale Logit-Modelle**.
- Herleitung über **latente Nutzenmodelle**.
- Nutzen beim Kauf eines Produkts $r \in \{1, \dots, k\}$ wird beschrieben durch latente (unbeobachtbare) Größen
$$L_i^{(r)}, \quad r = 1, \dots, k.$$
- Beobachtet wird nur der Kauf eines bestimmten Produkts.

- Annahme: Kaufentscheidung für das Produkt mit **maximalem Nutzen**:

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

- Übliche Parametrisierung des latenten Nutzens:

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

- Für standardextremwertverteilte Fehler ergibt sich

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

mit

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

- Annahme linearer Kovariableneffekte ist in der Praxis selten gerechtfertigt.

⇒ Erweitere das Modell durch einen **semiparametrischen Prädiktor** der Form

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

wobei

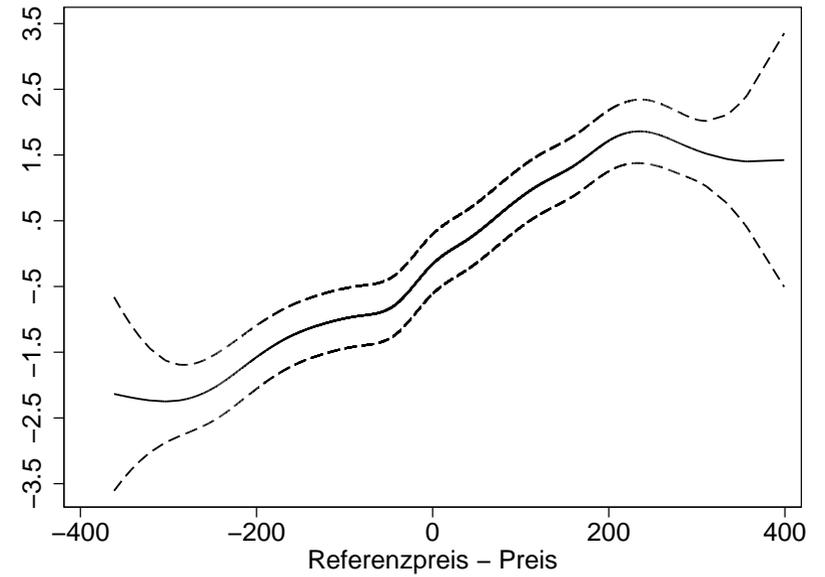
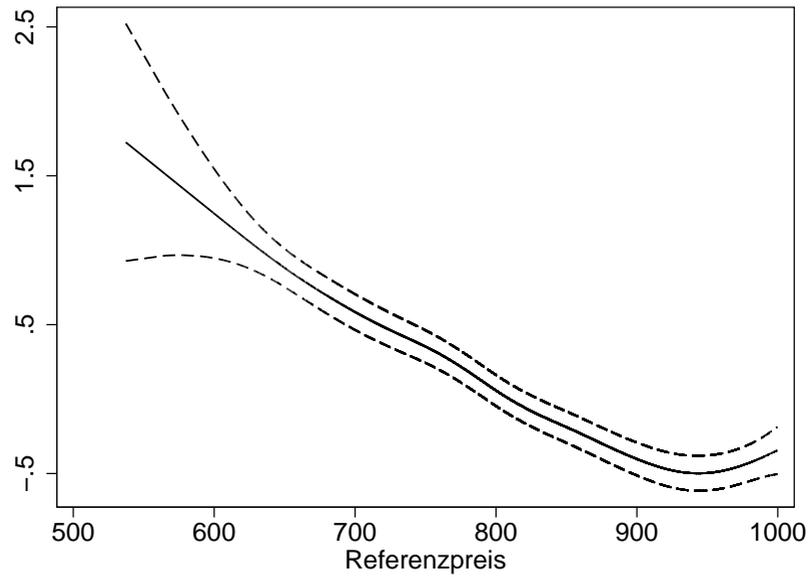
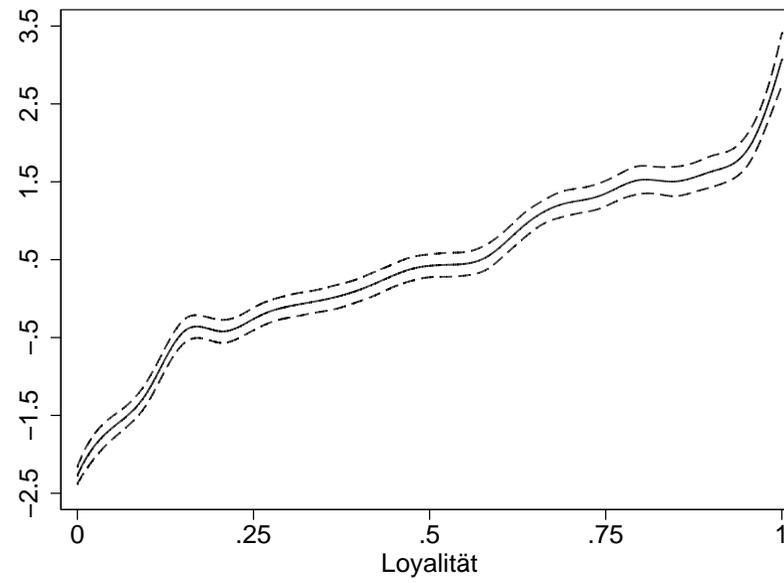
$$\bar{f}_j(x_{ij}^{(r)}) = f_j(x_{ij}^{(r)}) - f_j(x_{ij}^{(k)}).$$

Beispiel: Kaffee

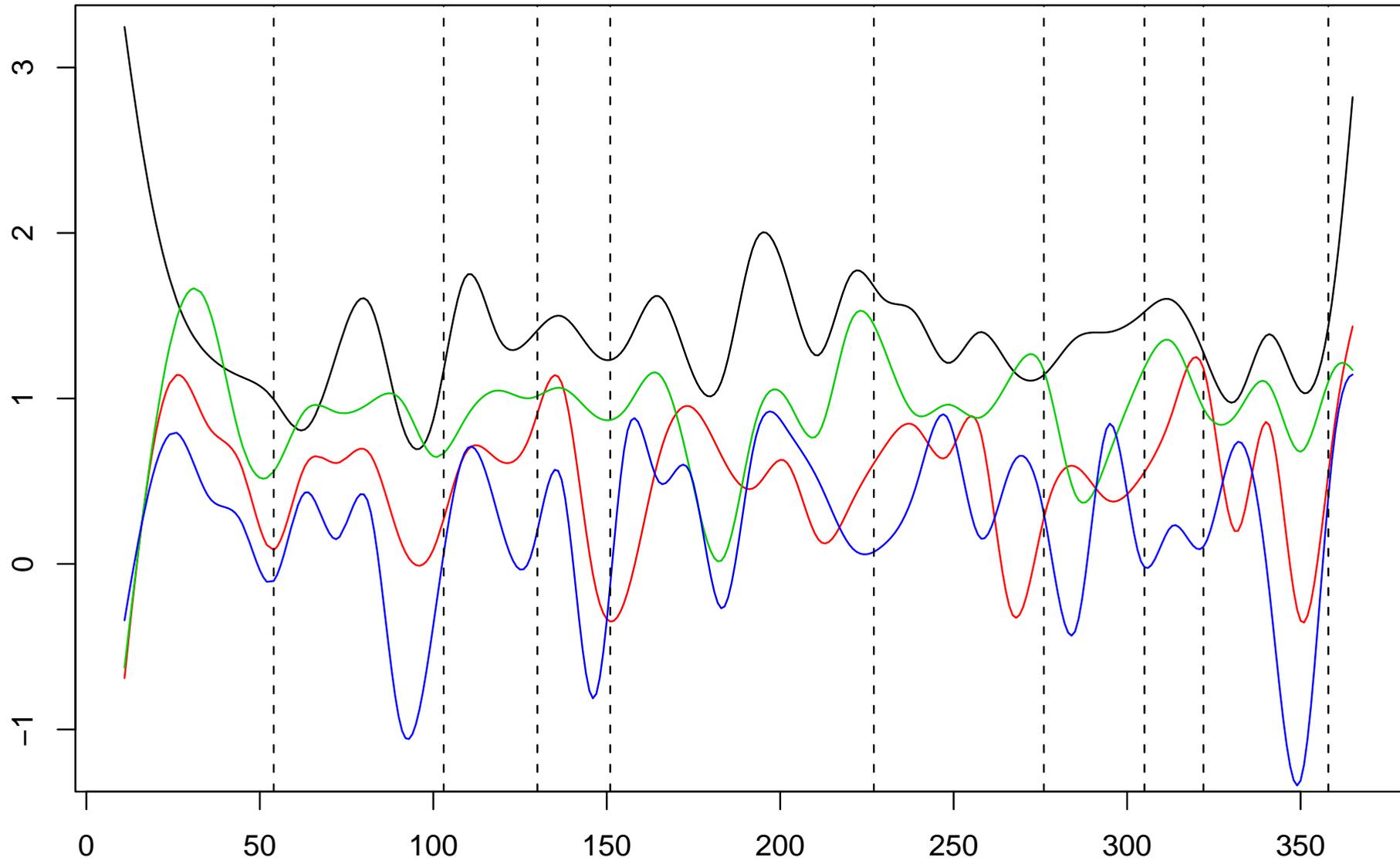
- Daten zu 47.962 Kaufakten.
- Betrachtet werden 5 verschiedene Kaffeemarken.
- Modell für den latenten Nutzen:

$$L_i^{(r)} = \beta \text{promo}_i^{(r)} + f_0^{(r)}(t_i) + f_1(\text{loy}_i^{(r)}) + f_2(\text{refpreis}_i^{(r)}) + f_3(\text{diffpreis}_i^{(r)}) + \varepsilon_i^{(r)}.$$

- Zusätzliches Problem: Bei manchen Käufen sind einzelne Marken nicht verfügbar
 - ⇒ Variierende Zahl von Kategorien.
 - ⇒ Summe in der multinomialen Logit-Formel geht nicht über alle Kategorien.



Zeitvariierende Vorlieben

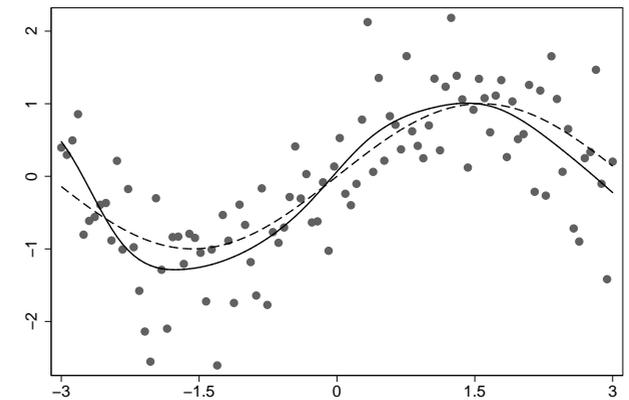
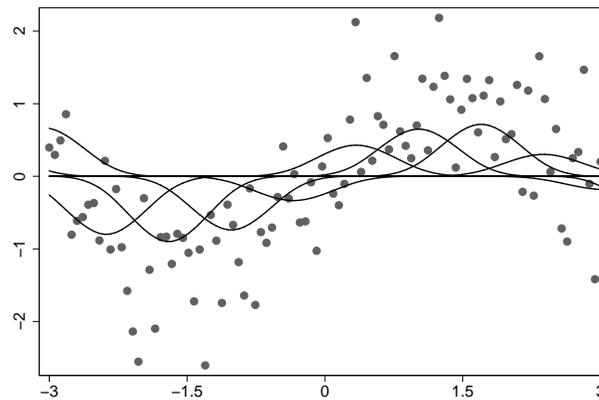
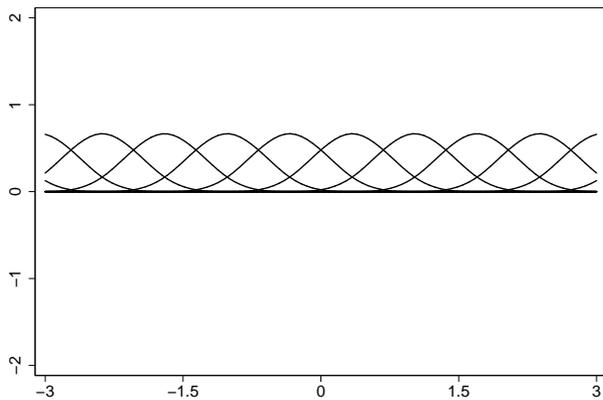


P-Splines

- **Basisfunktionsansatz** zur Beschreibung glatter Funktionen:

$$f(x) = \sum_{j=1}^p \xi_j B_j(x).$$

- Spezielle Basisfunktionen: **B(asic)-Splines**.



- P-Splines: Führe zusätzlich einen **Strafterm** ein, um einen geeigneten Kompromiss zwischen Datentreue und Glattheit zu erhalten.

- Strafterme:

$$Pen(\xi|\tau^2) = \frac{1}{2\tau^2} \sum_{j=2}^p (\xi_j - \xi_{j-1})^2 \quad \text{erste Differenzen,}$$

$$Pen(\xi|\tau^2) = \frac{1}{2\tau^2} \sum_{j=3}^p (\xi_j - 2\xi_{j-1} + \xi_{j-2})^2 \quad \text{zweite Differenzen.}$$

- Darstellbar als quadratische Formen:

$$Pen(\xi|\tau^2) = \frac{1}{2\tau^2} \xi' K \xi.$$

- Penalisiertes Maximum Likelihood-Kriterium:

$$l(\xi) - \frac{1}{2\tau^2} \xi' K \xi \rightarrow \max_{\xi}.$$

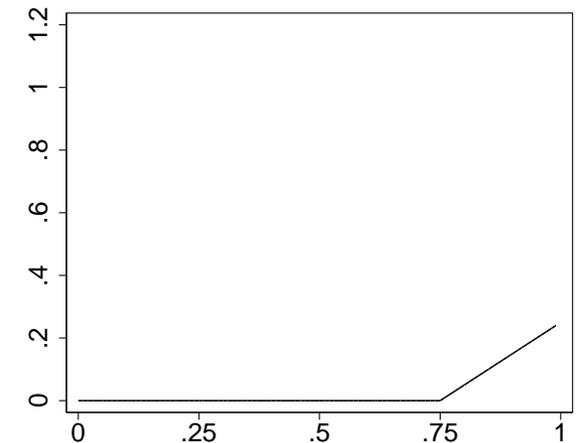
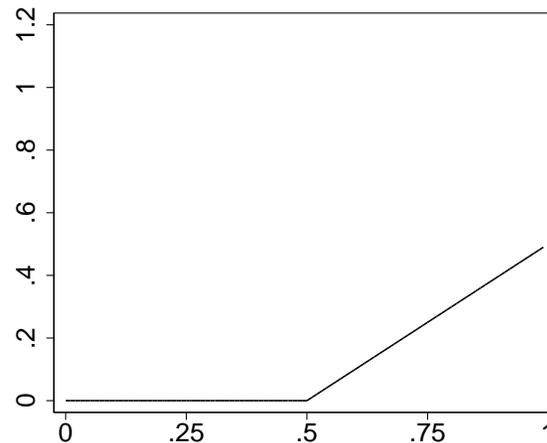
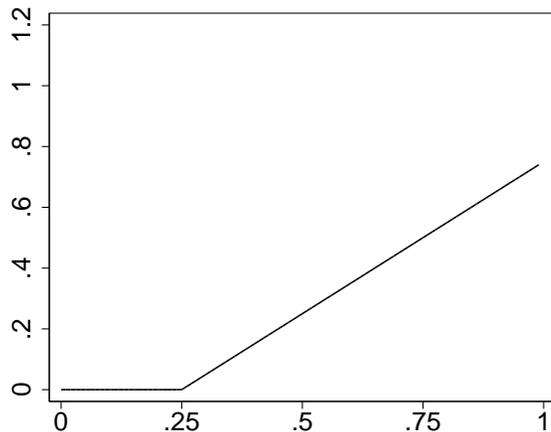
- τ^2 wirkt als **Glättungsparameter**.
- Bei k -ten Differenzen werden Abweichungen von Polynom der Ordnung $k - 1$ bestraft.

- Alternative Definition von P-Splines: **Truncated Power Series Basis**

$$f(x) = \beta_0 + \beta_1 x + \dots + \beta_l x^l + b_1(x - \kappa_1)_+^l \dots + b_p(x - \kappa_p)_+^l$$

mit **Ridge Penalty**

$$\frac{1}{2\tau^2} \sum_{j=1}^p b_j^2.$$



- B-Spline-Basis mit Differenzen-Strafterm erlaubt die **separate Wahl von Spline-Grad und Ordnung der Bestrafung**

⇒ Größere Flexibilität bei der Modellformulierung.

Darstellung als gemischtes Modell

- Penalisierte ML-Schätzung:

$$l(\xi) - \frac{1}{2\tau^2} \xi' K \xi \rightarrow \max_{\xi}.$$

- Vergleich mit gemischtem Modell:

$$\eta_i = x_i' \beta + z_i' b, \quad b \sim N(0, Q)$$

$$l(\beta, b) - \frac{1}{2} b' Q^{-1} b \rightarrow \max_{\beta, b}.$$

⇒ ξ entspricht einem zufälligen Effekt mit Erwartungswert 0 und Präzisionsmatrix $\frac{1}{\tau^2} K$.

- Problem: K hat (bei B-Splines mit Differenzen-Penalty) **nicht vollen Rang**.
- Die Verteilung von ξ ist **teilweise uneigentlich** (nicht normierbar).

- Dichte der (teilweise uneigentlichen) Verteilung von ξ :

$$p(\xi) \propto \exp\left(-\frac{1}{2\tau^2}\xi'K\xi\right).$$

$\Rightarrow \xi$ lässt sich zerlegen in

$$\xi = \tilde{X}\beta + \tilde{Z}b,$$

so dass

$$\underbrace{p(\xi)}_{\text{teilweise uneigentlich}} = \underbrace{p(\beta)}_{\text{uneigentlich}} \cdot \underbrace{p(b)}_{\text{eigentlich}} \propto p(b).$$

- β = unpenalisierter Anteil von ξ mit $\dim(\beta) = \dim(\xi) - \text{rang}(K)$,
- b = penalisierter Anteil von ξ mit $\dim(b) = \text{rang}(K)$.

- Einsetzen der Zerlegung in die P-Spline-Formel ergibt (bei geeigneter Wahl von \tilde{X} und \tilde{Z})

$$\begin{aligned} B\xi &= B(\tilde{X}\beta + \tilde{Z}b) \\ &= X\beta + Zb \end{aligned}$$

wobei

$$p(\beta) \propto \text{const} \quad \text{und} \quad b \sim N(0, \tau^2 I).$$

- β entspricht einem Vektor fester Effekte.
- b entspricht einem Vektor von i.i.d. zufälligen Effekten.

⇒ Darstellung eines P-Splines als gemischtes Modell.

- Vorteil: Zusätzlich zu den Regressionsparametern können auch die Varianz- bzw. Glättungsparameter geschätzt werden!

- Die gleiche Reparametrisierung lässt sich bei anderen Regressionsansätzen einsetzen, die auf Penalisationen der Form

$$Pen(\xi|\tau^2) = \frac{1}{2\tau^2}\xi'K\xi.$$

beruhen.

- Beispiele:
 - **Markov-Zufallsfelder**,
 - Interaktionsoberflächen basierend auf 2d P-Splines,
 - Flexible Saisonkomponenten,
 - Variierende Koeffizienten,
 - Glättungssplines,
 - (Stationäre Gauß-Felder),
 - (zufällige Effekte).

Inferenz

- **Penalisierte Log-Likelihood** in Mixed Model Formulierung:

$$l_p(\beta, b) = l(\beta, b) - \frac{1}{2}b'Q^{-1}b.$$

- Schätzung der Regressionskoeffizienten: Newton-Raphson-Algorithmus bzw. Fisher-Scoring bei gegebenen Varianzen τ^2 anwendbar.

- Score-Funktion und Fisher-Information:

$$s_p(\beta, b) = \begin{pmatrix} \frac{\partial l}{\partial \beta} \\ \frac{\partial l}{\partial b} - Q^{-1}b \end{pmatrix} \quad F_p(\beta, b) = \begin{pmatrix} \frac{\partial^2 l}{\partial \beta \partial \beta'} & \frac{\partial^2 l}{\partial \beta \partial b'} \\ \frac{\partial^2 l}{\partial b \partial \beta'} & \frac{\partial^2 l}{\partial b \partial b'} - Q^{-1} \end{pmatrix}.$$

- **Marginale Likelihood-Schätzung** der Varianzparameter:

$$L(Q) = \int L_p(\beta, b, Q) d\beta db \rightarrow \max_Q.$$

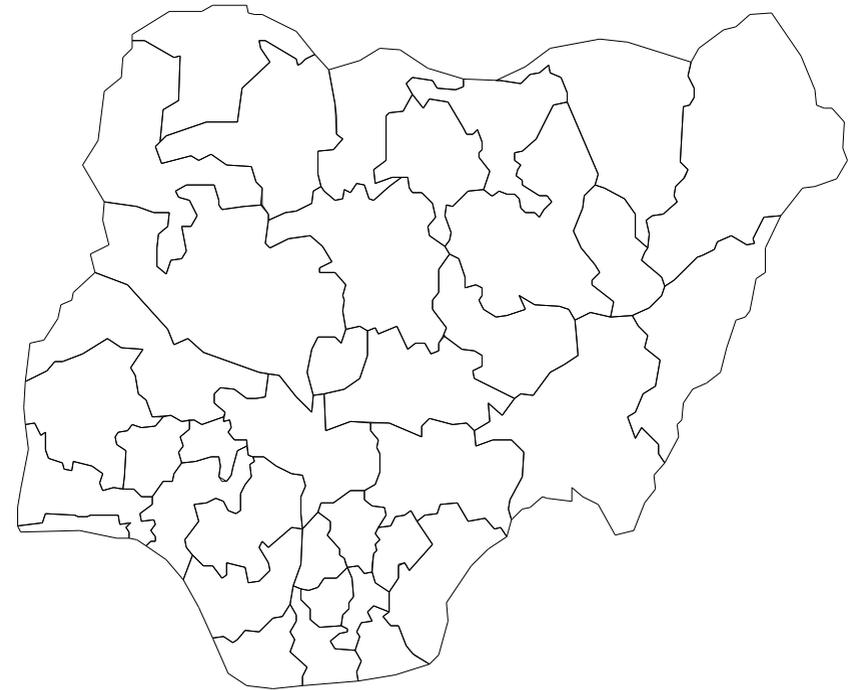
- Hochdimensionales Integral \Rightarrow im Allgemeinen weder analytisch noch numerisch bestimmbar.
- Approximatives Vorgehen:
 - **Laplace-Approximation** an die Likelihood (quadratische Taylor-Entwicklung um den Modus).
 - Ergibt Normalverteilungs-Likelihood \Rightarrow Integral explizit lösbar.
 - Die resultierende marginale Likelihood entspricht der **Restricted Likelihood** des approximativen Normalverteilungs-Modells.
- Für Exponentialfamilien entspricht die Laplace-Approximation der Normalverteilungsapproximation der **iterativ gewichteten KQ-Schätzung**.

- Mixed model basierte Inferenz kann als **empirische Bayes-Schätzung** aufgefasst werden:
 - Unterscheide zwischen primär interessierenden Parametern (den Regressionskoeffizienten) und Hyperparametern (den Varianzparametern).
 - Schätze die Hyperparameter vorab aus ihrer **marginalen Posteriori-Verteilung**.
 - Setze die Schätzer in die Posteriori ein und maximiere bezüglich der primär interessierenden Parameter (**Posteriori-Modus-Schätzer**).
- Volle Bayes-Inferenz: Auch die Hyperparameter werden mit Prioris versehen und simultan mitgeschätzt.

Kindersterblichkeit in Nigeria

- Daten aus dem 2003 Demographic and Health Survey (DHS) in Nigeria.
- **Retrospektive Befragung** zum Gesundheitszustand von Frauen im gebärfähigen Alter und deren Kindern.
- Überlebenszeiten zu $n = 5323$ Kindern.
- Zahlreiche Kovariablen inklusive **räumlicher Information**.
- Analyse basierend auf dem **Cox-Modell**:

$$\lambda(t; u) = \lambda_0(t) \exp(u' \gamma).$$



- **Einschränkungen** des klassischen Cox-Modells:
 - Beschränkt auf rechtszensierte Beobachtungen,
 - Nachträgliche Schätzung der Baseline Hazardrate,
 - Annahme proportionaler Hazardraten,
 - Parametrische Form des Prädiktors,
 - Keine räumlichen Korrelationen.
- Erweiterungen in Bezug auf einzelne Probleme existieren, aber es fehlt eine Möglichkeit **simultan allen Problemen** zu begegnen.

- Flexibles, **semiparametrisches Modell** für die Hazardrate:

$$\lambda(t; \cdot) = \exp \left[g_0(t) + \sum_{j=1}^q g_j(t) z_j + \sum_{k=1}^p f_k(x_k) + f_{spat}(s) + u(t)' \gamma \right]$$

wobei

$g_0(t) = \log(\lambda_0(t))$ logarithmierte **Baseline-Hazardrate**,

g_j **zeitvariierende Effekte** der Kovariablen z_j ,

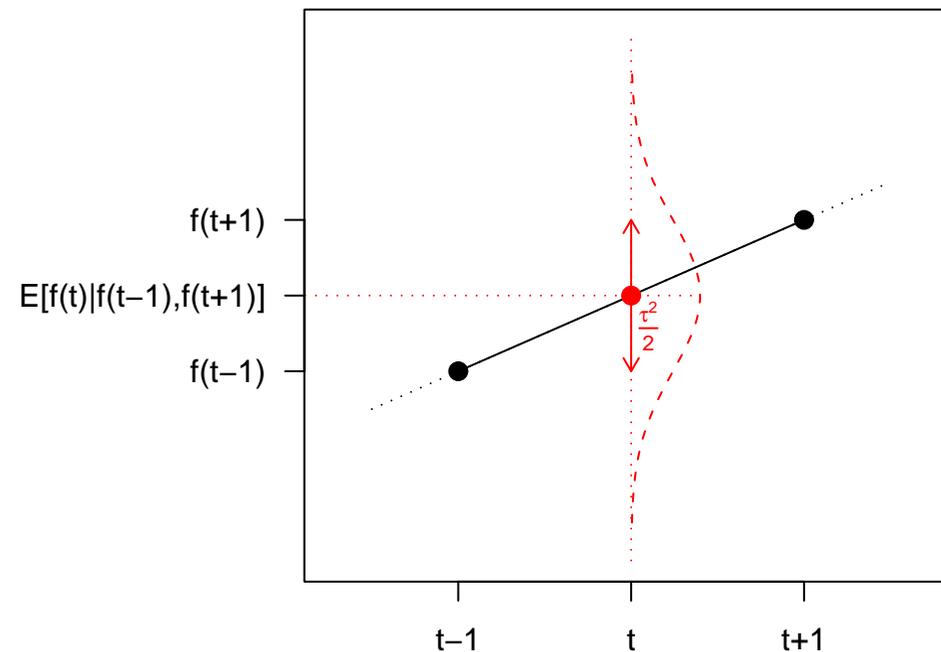
f_k **nonparametrische** Funktionen der stetigen Kovariablen x_k ,

f_{spat} **räumliche** Funktion,

$u(t)$ **zeitvariierende Kovariablen**.

Modellierung räumlicher Effekte

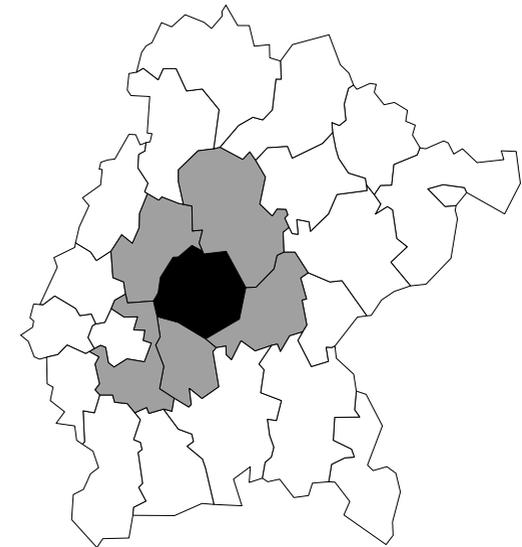
- Speziell: Modellierung für **Regionendaten** $s \in \{1, \dots, S\}$.
- $\xi_s = f_{spat}(s)$ gebe den räumlichen Effekt an einer Lokalisation $s \in \{1, \dots, S\}$ an.
- Verallgemeinere zeitlichen **Random Walk** auf den räumlichen Fall.



- Einfachstes **Markov-Zufallsfeld**:

$$\xi_s | \xi_{s'}, s' \neq s, \tau^2 \sim N \left(\frac{1}{N_s} \sum_{s' \in \partial_s} \xi_{s'}, \frac{\tau^2}{N_s} \right),$$

$\delta_s =$ Menge der Nachbarn von s , $N_s = |\delta_s|$.



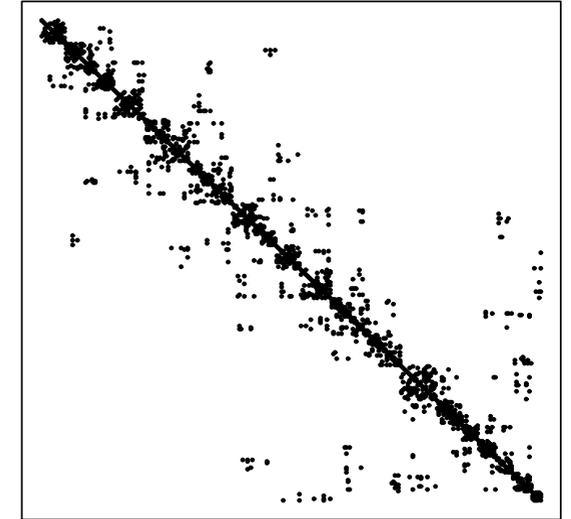
- Auch hier lässt sich ein **penalisiertes ML-Kriterium** aufstellen:

$$l(\xi) - \frac{1}{2\tau^2} \xi' K \xi \rightarrow \max_{\xi}.$$

- K ist eine Nachbarschaftsmatrix.

$$k_{ss'} = \begin{cases} -1 & \text{falls } s \text{ und } s' \text{ benachbart,} \\ 0 & \text{sonst,} \end{cases}$$
$$k_{ss} = N_s.$$

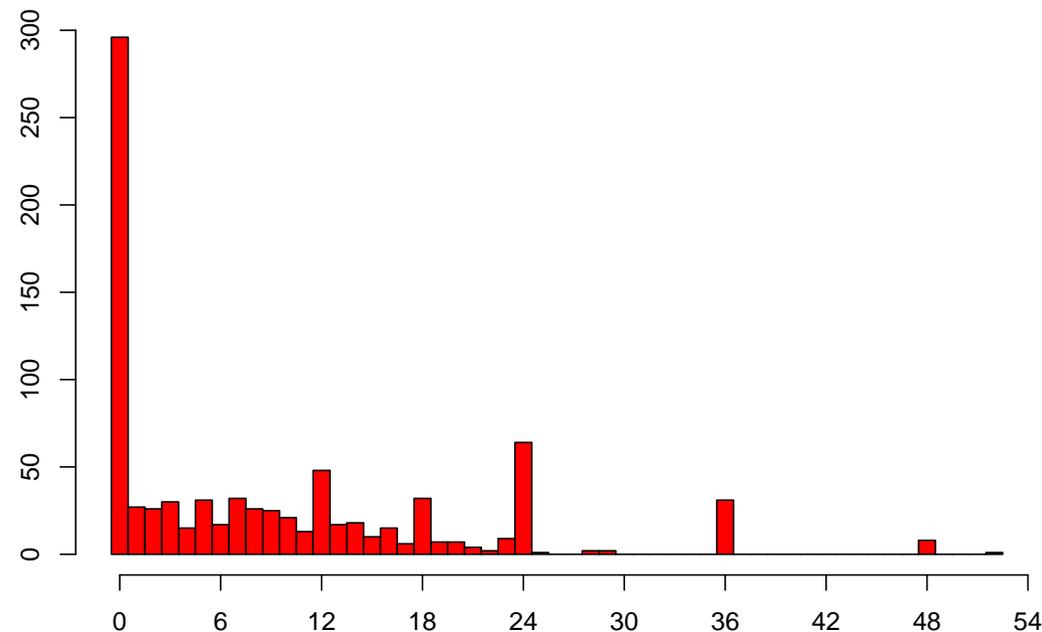
$$\Rightarrow \text{rang}(K) = \dim(\xi) - 1.$$



- Reparametrisierung und Methoden für gemischte Modelle auch hier anwendbar.

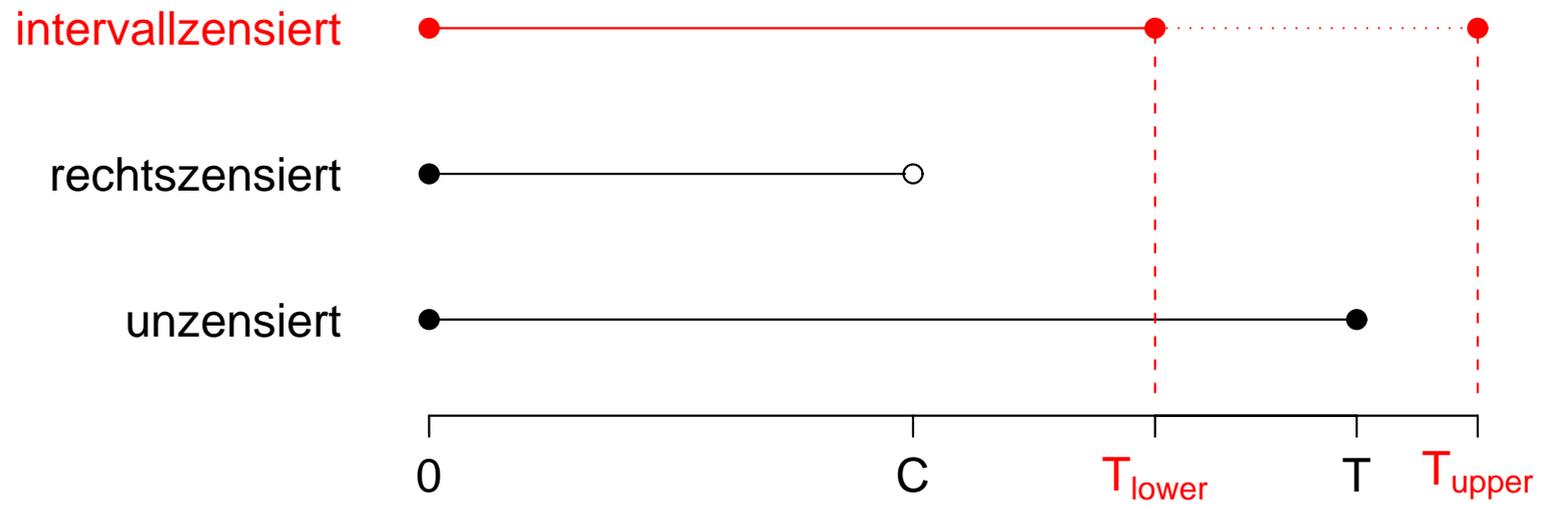
Intervallzensierte Überlebenszeiten

- Theoretisch wird die Überlebenszeit in Tagen erhoben.
- Retrospektives Design \Rightarrow die meisten unzensierten Überlebenszeiten sind gerundet.



- Zensierte Überlebenszeiten sind dagegen exakt erhältlich.

\Rightarrow Betrachte Überlebenszeiten als **intervallzensiert**.



- Likelihood-Beiträge:

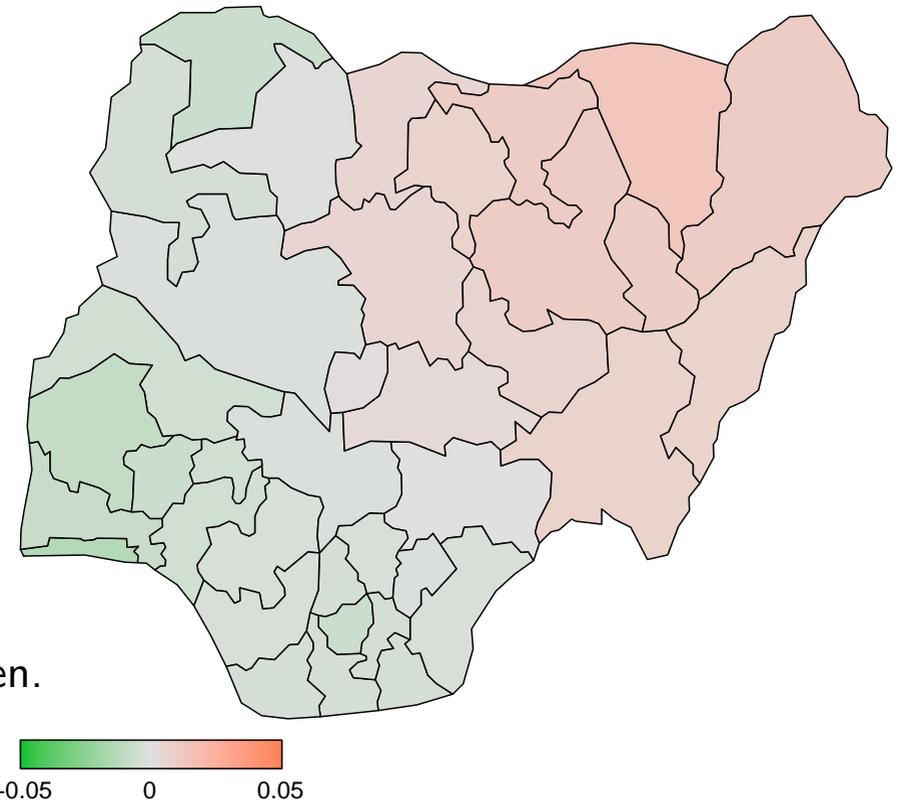
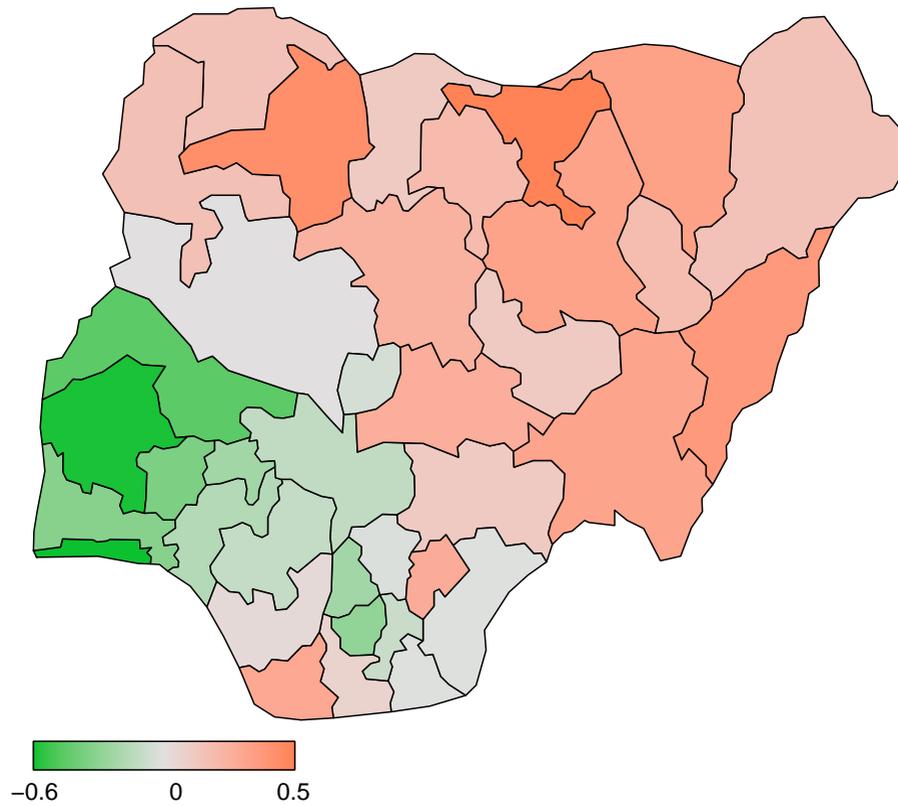
$$\begin{aligned} P(T > C) &= S(C) \\ &= \exp \left[- \int_0^C \lambda(t) dt \right]. \end{aligned}$$

$$\begin{aligned} P(T \in [T_{lower}, T_{upper}]) &= S(T_{lower}) - S(T_{upper}) \\ &= \exp \left[- \int_0^{T_{lower}} \lambda(t) dt \right] - \exp \left[- \int_0^{T_{upper}} \lambda(t) dt \right]. \end{aligned}$$

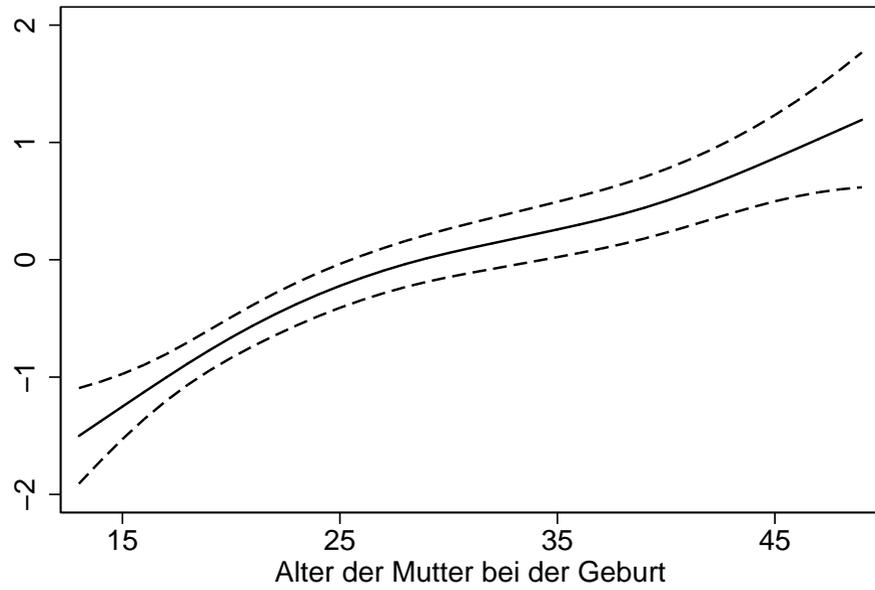
- Ableitungen der Log-Likelihood werden wesentlich komplizierter.
- **Numerische Integrationsverfahren** müssen verwendet werden.
- Stückweise konstante, **zeitvariierende Kovariablen** und **Linkstrunkierung** können einfach berücksichtigt werden.

Ergebnisse

Räumlicher Effekt ohne Berücksichtigung der Kovariablen.

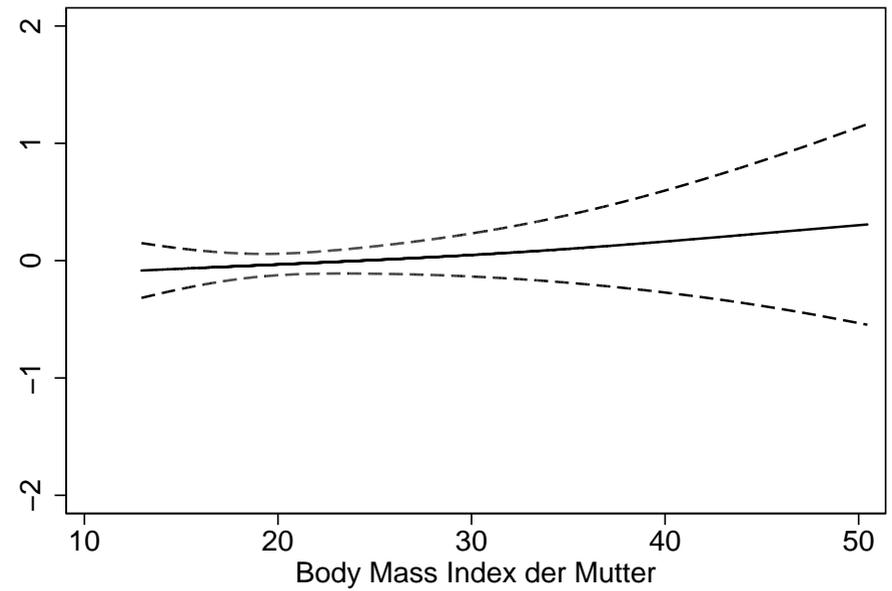


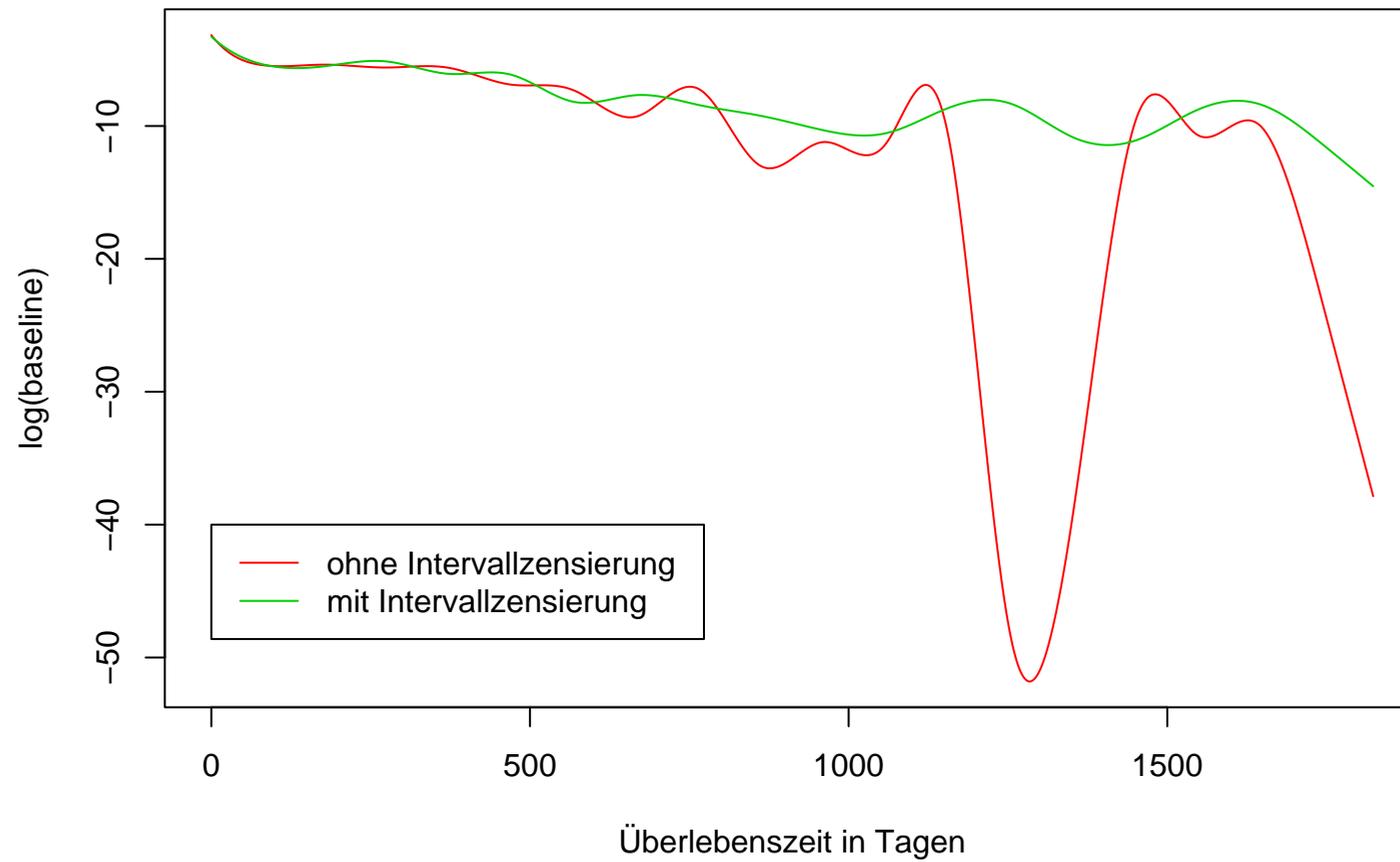
Räumlicher Effekt bei Berücksichtigung der Kovariablen.



Alter der Mutter bei der Geburt.

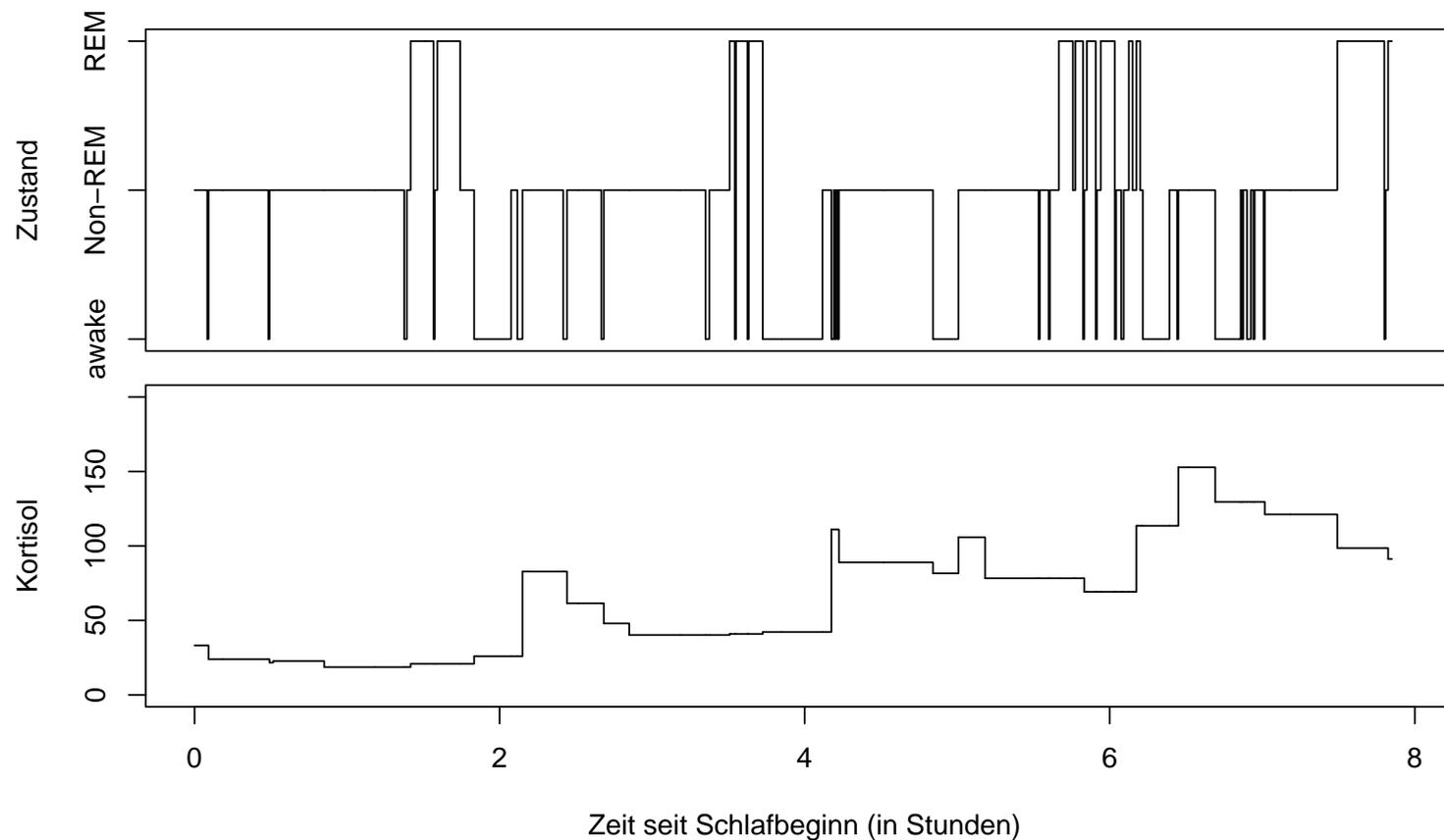
Body Mass Index der Mutter.





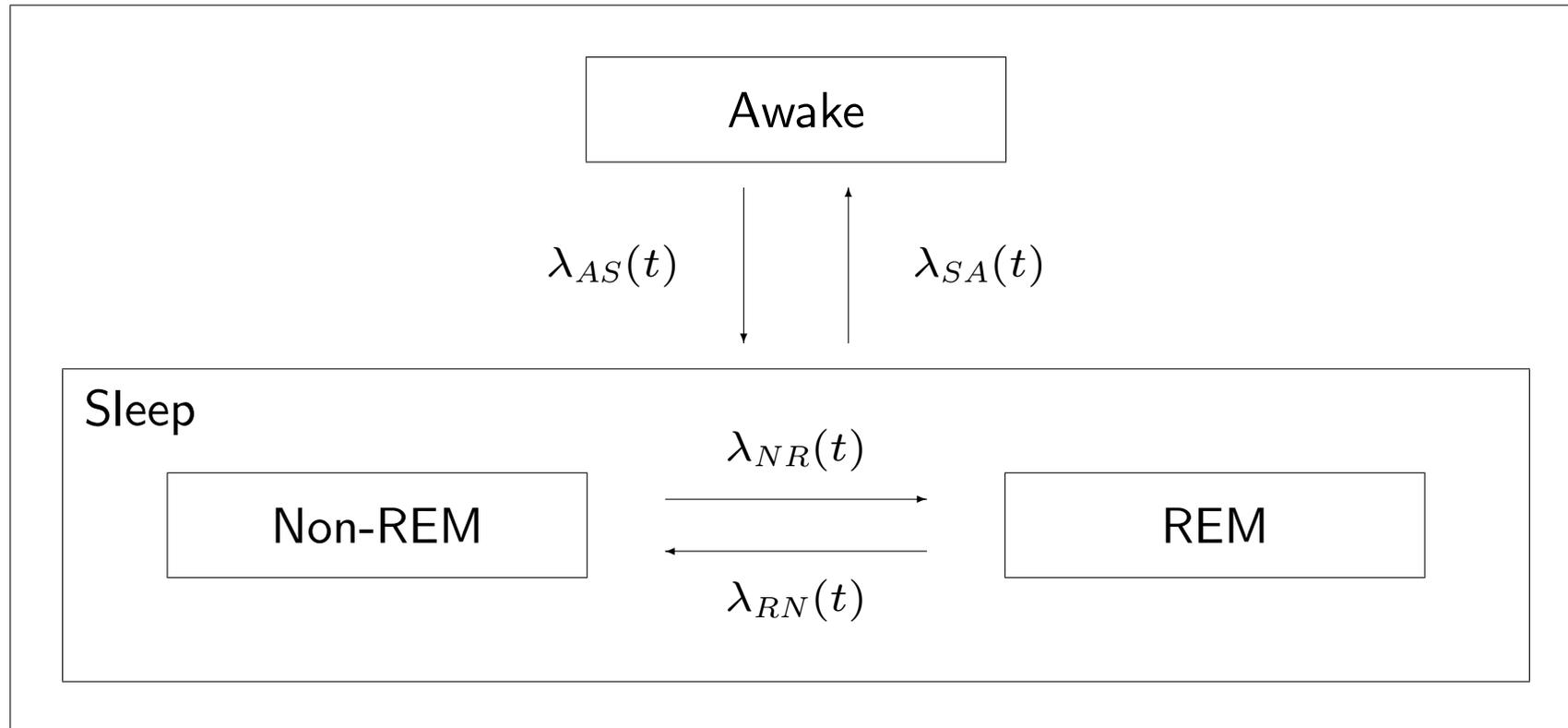
Multi-State Modelle zur Beschreibung des menschlichen Schlafverlaufs

- Betrachte den Schlafverlauf als Realisation eines **zeitstetigen** stochastischen Prozesses.



- Diskreter Zustandsraum: {awake, REM, non-REM}.
 - Wesentliche Größe: **Intensitäten für Übergänge** zwischen den Zuständen.
 - Einfache Modelle: Markov- oder Semi-Markov-Prozesse.
 - Probleme:
 - Schlafverhalten ändert sich während der Nacht.
 - Individuelle Unterschiede im Schlafverhalten.
 - Nur wenige Kovariablen sind direkt erhältlich.
- ⇒ Beschreibe Übergangintensitäten analog zum Cox-Modell.

- Einfache Struktur für die möglichen Übergänge:



- Modell für die Übergänge:

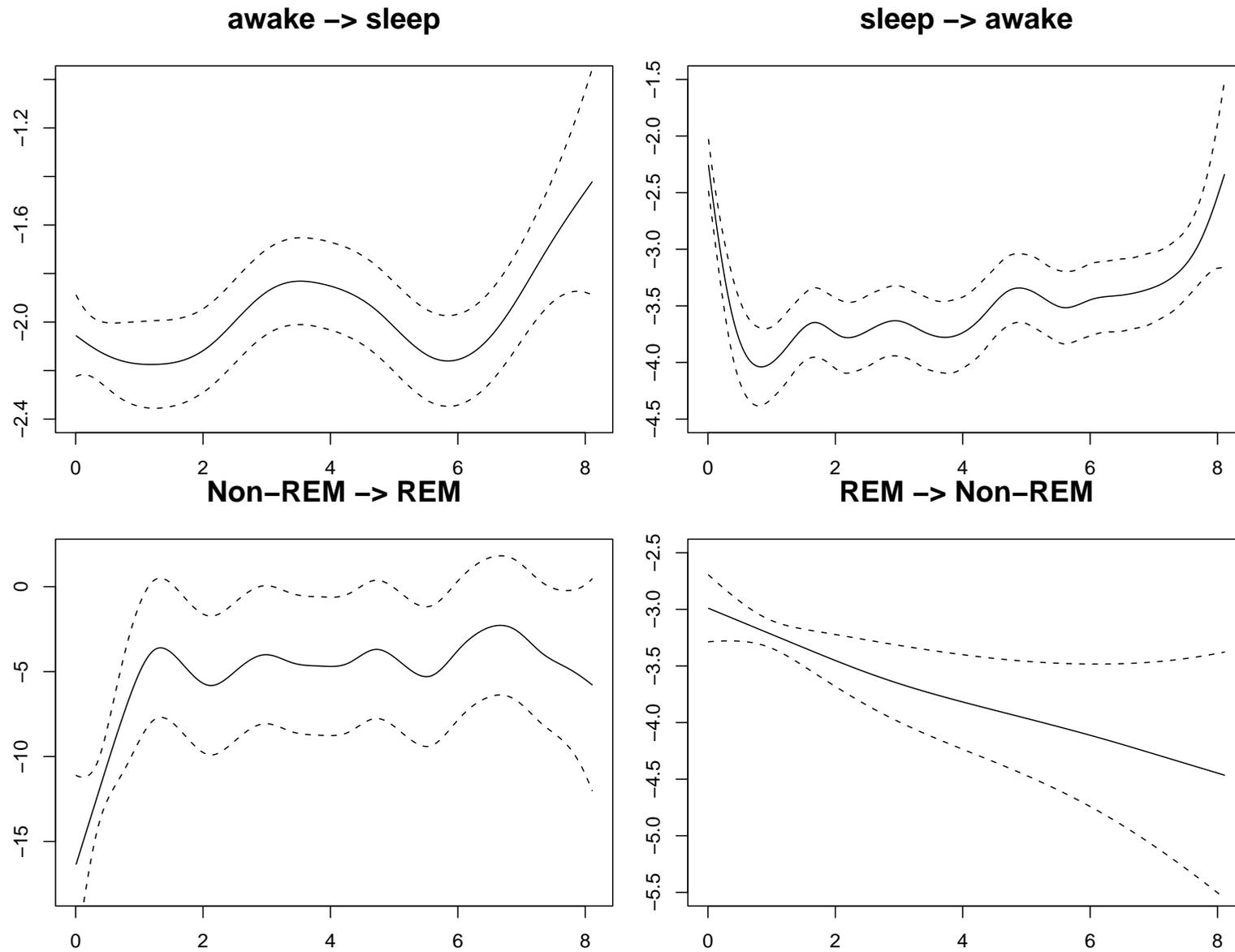
$$\begin{aligned}\lambda_{AS,i}(t) &= \exp \left[\gamma_0^{(AS)}(t) + s_i \beta^{(AS)} + b_i^{(AS)} \right] \\ \lambda_{SA,i}(t) &= \exp \left[\gamma_0^{(SA)}(t) + s_i \beta^{(SA)} + b_i^{(SA)} \right] \\ \lambda_{NR,i}(t) &= \exp \left[\gamma_0^{(NR)}(t) + c_i(t) \gamma_1^{(NR)}(t) + s_i \beta^{(NR)} + b_i^{(NR)} \right] \\ \lambda_{RN,i}(t) &= \exp \left[\gamma_0^{(RN)}(t) + c_i(t) \gamma_1^{(RN)}(t) + s_i \beta^{(RN)} + b_i^{(RN)} \right]\end{aligned}$$

wobei

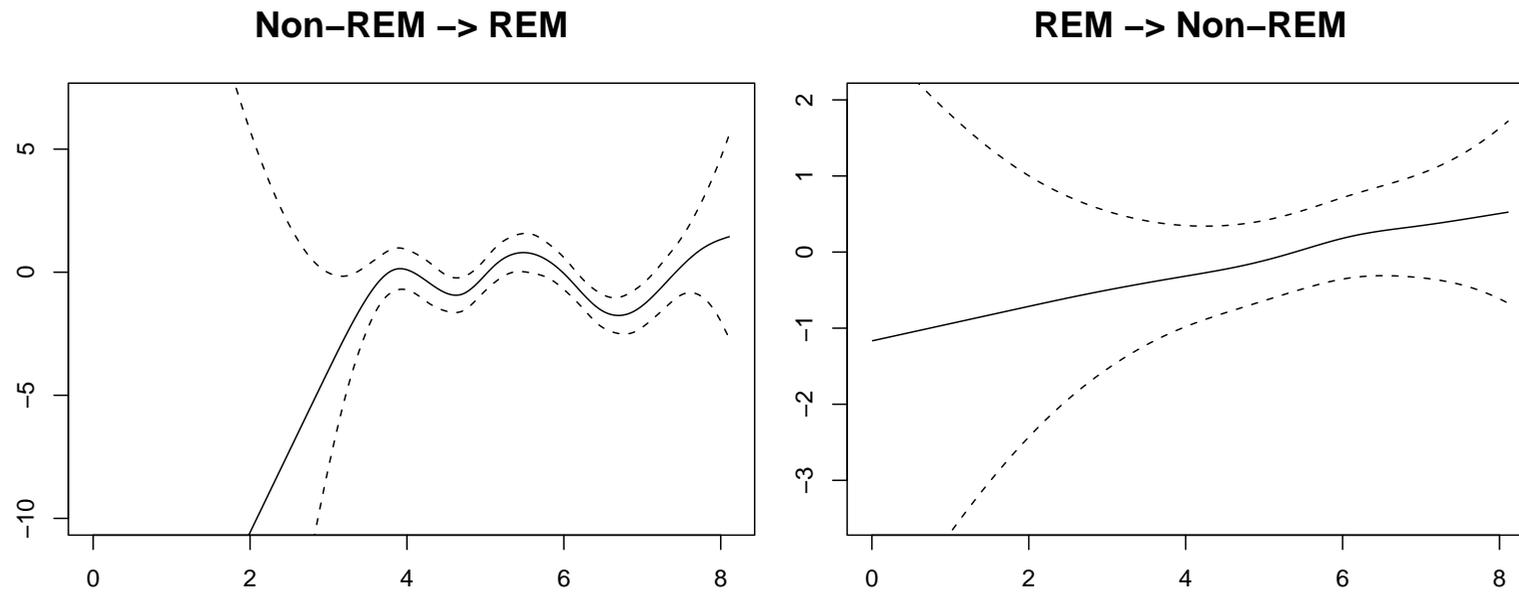
$$\begin{aligned}c_i(t) &= \begin{cases} 1 & \text{Kortisol} > 60 \text{ n mol/l zum Zeitpunkt } t \\ 0 & \text{Kortisol} \leq 60 \text{ n mol/l zum Zeitpunkt } t, \end{cases} \\ s_i &= \begin{cases} 1 & \text{für Männer} \\ 0 & \text{für Frauen,} \end{cases} \\ b_i^{(j)} &= \text{Übergangsspezifische zufällige Effekte.}\end{aligned}$$

- P-Splines zur Modellierung der zeit-variierenden Effekte.
- Zufällige Effekte zur Modellierung unbeobachteter Heterogenität.
- Likelihood-Konstruktion über **Zählprozess-Formulierung** möglich.
- Damit sind die üblichen Konzepte (prinzipiell) wieder anwendbar.
- Insbesondere:
 - **Penalisierte ML-Schätzung** für die Regressionskoeffizienten.
 - **Marginale Likelihood-Schätzung** der Varianzkomponenten.

● Baseline-Effekte:



- Zeitvariierende Effekte bei hohem Kortisol:



Software

- Alle vorgestellten Modelle sind implementiert im Software Paket BayesX.



- Erhältlich unter

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

Gemischte Modelle vs. MCMC

- Nachteile:
 - Nicht modular aufgebaut. Der numerische Aufwand wächst schneller an.
 - Kreditabilitätsintervalle nur asymptotisch.
 - Laplace-Approximation.
 - Nur Plug-in Schätzungen für Funktionale der Parameter.
 - Schwieriger zu erweitern (?).
- Vorteile:
 - Keine Fragen nach Konvergenz und Mixing.
 - Keine Sensitivität in Bezug auf Hyperprioris.
 - In einfachen Modellen wesentlich schneller.
 - Anwendern einfacher zu vermitteln (?)

Zusammenfassung

- Inferenz basierend auf Mixed Model Repräsentationen von additiven und geoadditiven Modellen.
- Anwendbar in äußerst vielfältigen Regressionsmodellen:
 - Univariate und multivariate Exponentialfamilien,
 - Hazardraten-Modelle,
 - Multi-State-Modelle.
- Vermeidung von MCMC-Verfahren.