

Generalisierte Lineare Modelle mit zufälligen Effekten und variierenden Koeffizienten

vorgelegt von
Diplom-Wirtschaftsmathematikerin
Claudia Beate Funck-Hüsges

Von der Fakultät IV - Elektrotechnik und Informatik
der Technischen Universität Berlin
zur Erlangung des akademischen Grades
Doktorin der Naturwissenschaften
- Dr.rer.nat. -

genehmigte Dissertation

Promotionsausschuss:

Vorsitzender: Prof. Dr.-Ing. Adam Wolisz

Berichter: Prof. Dr. Gerhard Tutz

Berichter: Prof. Dr. Ulrich Kockelkorn

Tag der wissenschaftlichen Aussprache: 03.12.2001

Berlin 2001

D 83

Inhaltsverzeichnis

Einleitung	1
1 Das Generalisierte Lineare Modell	3
1.1 Grundlagen des Generalisierten Linearen Modells	4
1.1.1 Modelldefinitionen	4
1.1.2 Schätzung der Parameter	11
2 Modellerweiterungen im GLM	13
2.1 Modellansätze mit zufälligen Effekten	16
2.1.1 EM-Algorithmus mit Gauß-Hermite Integrationstechnik	19
2.1.2 EM-Type Algorithmus	25
2.1.3 Modellansatz der nichtparametrischen Maximum Likelihood	28
2.2 Modellansätze mit (zeit-)variierenden Koeffizienten	33
2.2.1 Parametrische und semi-parametrische Ansätze	34
2.2.2 Lokaler Likelihoodansatz	36
3 Zufällige Effekte und variierende Koeffizienten	40
3.1 Modellansätze und Parameterschätzungen	42
3.1.1 Volles Modell, parametrische und semi- parametrische Ansätze	42
3.1.2 Lokaler Likelihoodansatz mit zufälligen Effekten	43

3.2	Varianz der Parameterschätzungen	48
3.2.1	Varianzschätzungen bei Modellen mit zufälligen Effekten	48
4	Simulationsbeispiele	54
4.1	Schätzung der festen Parameter	56
4.1.1	Das volle Modell mit zufälligen Effekten	56
4.1.2	Lokaler Likelihoodansatz mit zufälligen Effekten und Gauß-Hermite Integrationstechnik	60
4.1.3	Lokaler Likelihoodansatz mit zufälligen Effekten und EM-Type Algorithmus	64
4.2	Bayes-Schätzung der zufälligen Effekte	67
5	Anwendungsbeispiele	70
5.1	Poissonmodelle	71
5.1.1	Emesis Datensatz	71
5.1.2	Epileptic Datensatz	77
5.2	Verweildauermodelle	80
5.2.1	Dauer der Arbeitslosigkeit	80
5.2.2	Magenkrebsstudie	82
6	Anhang	87
6.1	EM-Algorithmus mit direkter Gauß-Hermite Integrationstechnik	87
6.1.1	Numerische Integration mit Gauß-Hermite	87
6.1.2	Schätzung der festen Parameter	89
6.2	EM-Type Algorithmus	97
6.2.1	Blockdiagonalgestalt der Fishermatrix	97
	Zusammenfassung	99
	Literaturverzeichnis	102

Einleitung

Generalisierte Lineare Modelle (GLM) sind ein vielverwendetes Instrumentarium für die Durchführung von Zusammenhangsanalysen, die beispielsweise bei den Auswertungen von medizinischen oder sozialökonomischen Studien verwendet werden.

Ihre große Verbreitung liegt zum einen daran, daß in dieser Modellklasse sowohl das Lineare Modell als auch Spezialfälle für z. B. diskrete, kategoriale oder poissonverteilte Daten enthalten sind und somit sehr unterschiedliche Datenmodelle analysiert werden können.

Andererseits sind Generalisierte Lineare Modelle in den verschiedensten Statistik-Software Programmen implementiert und sind dadurch sehr einfach vom Anwender abrufbar.

Das grundlegende Generalisierte Lineare Modell geht von unabhängigen Beobachtungseinheiten aus. Jedoch liegen in vielen großangelegten Studien Gruppenstrukturen vor, so daß spezielle Modellerweiterungen angewendet werden müssen. Diese zu berücksichtigenden Gruppenstrukturen können durch Meßwiederholungen an den gleichen Beobachtungseinheiten oder auch durch eine familiäre, regionale oder administrative Zugehörigkeit der Beobachtungseinheiten gegeben sein.

Zusätzlich zu der Gruppenstruktur muß häufig auch noch ein möglicher zeitlicher Einfluß berücksichtigt werden, da die Individuen oder Beobachtungs-

einheiten in diesen Studien meist über einen längeren Zeitraum beobachtet werden.

Das Ziel dieser Arbeit ist es, die gruppen- und/oder individuenspezifische Datenstruktur durch die Aufnahme von zufälligen Effekten und gleichzeitig den zeitlichen Einflusses durch Betrachtung von zeitvariierenden Koeffizienten zu berücksichtigen.

Nach der Einleitung in die Thematik dieser Arbeit wird im ersten Kapitel das grundlegende Generalisierte Lineare Modell vorgestellt. Auf die Modellerweiterungen mit zeitvariierenden Koeffizienten und mit zufälligen Effekten wird im zweiten Kapitel eingegangen. Diese beiden Modellansätze werden im dritten Kapitel in gemeinsame Modelle aufgenommen. Im vierten Kapitel wird in den Simulationsbeispielen neben der Güte für die Parameterschätzungen selbst auch die Güte für die Schätzung der Varianzen der Parameter betrachtet. Aus der Vielzahl der Anwendungsmöglichkeiten werden im fünften Kapitel vier Anwendungsbeispiele herausgestellt. Das sechste Kapitel enthält einige technische Details für die im zweiten Kapitel beschriebenen Verfahren. In der nachfolgenden Zusammenfassung erfolgt eine, diese Arbeit abschließende, Beurteilung der vorgestellten Verfahren.

Kapitel 1

Das Generalisierte Lineare Modell

Das Generalisierte Lineare Modell (GLM), welches von Nelder & Wedderburn (1972) vorgeschlagen wurde, ist eine Erweiterung des klassischen Regressionsansatzes im Linearen Modell.

Die im folgenden vorgestellten Modellannahmen im Linearen Modell sind für viele metrische Zielgrößen gut motiviert, falls ein linearer Zusammenhang zwischen der Responsevariablen und den erklärenden Variablen, die auch als Regressoren bzw. Kovariablen bezeichnet werden, besteht.

Liegt die Zielvariable diskret, beispielsweise in Form von binomialen oder multinomialen Zielgrößen oder als Zähldaten vor, so ist das klassische lineare Modell meist nicht mehr passend und es sollte besser die Erweiterung durch das Generalisierte Lineare Modelle angewendet werden.

1.1 Grundlagen des Generalisierten Linearen Modells

Da das Generalisierte Lineare Modell eine Erweiterung des Linearen Modells ist, werden wir zunächst kurz auf diesen Regressionsansatz eingehen und im Anschluß die Modelldefinition des Generalisierten Linearen Modells vorstellen.

1.1.1 Modelldefinitionen

Das Lineare Modell

Für den univariaten Fall der linearen Regression wird vorausgesetzt, daß Beobachtungen (y_i, x_i) mit $i = 1, \dots, N$ vorliegen. Bei dem Modellansatz wird zwischen der eindimensionalen Responsevariablen y_i und dem Vektor $x_i' = (x_{i1}, \dots, x_{im})$ der erklärenden Variablen ein linearer Zusammenhang der Form $y_i = z_i' \beta + \epsilon_i$ für $i = 1, \dots, N$ angenommen.

Dabei ist $\beta' = (\beta_1, \dots, \beta_p)$ der Vektor des p -dimensionalen unbekanntem Parameter und z_i ein p -dimensionaler Designvektor, der durch eine geeignete Funktion aus dem Kovariablenvektor x_i gebildet wird; beispielsweise durch $z_i' = (1, x_i')$. Weiter ist ϵ_i eine Fehler- bzw. Störvariable mit dem Erwartungswert $E(\epsilon_i) = 0$. Es wird vorausgesetzt, daß ϵ_i normalverteilt mit $\epsilon_i \sim \mathcal{N}(0, \sigma^2)$ ist und y_i bei gegebenem x_i bedingt unabhängig ist. Damit folgt bei gegebenem x_i , daß $y_i \sim \mathcal{N}(\mu_i, \sigma^2)$ mit $\mu_i = E(y_i|x_i)$ gilt.

Im linearen Modell ist $\mu_i = z_i' \beta$ für $i = 1, \dots, N$, das bedeutet, der bedingte Erwartungswert ist gleich dem linearen Prädiktor $z_i' \beta$.

Das Generalisierte Lineare Modell

Im Generalisierten Linearen Modell wird im Gegensatz zum Linearen Modell die Verteilungsannahme für die Zielgröße verallgemeinert indem vorausgesetzt wird, daß die Dichte von y_i aus einer Exponentialfamilie ist. Beispiele für die Exponentialfamilie sind die Normal-, Binomial-, Multinomial-, Poisson- und Gammaverteilung. Die Zielgröße y_i kann mehrdimensional sein, d. h. $y_i' = (y_{i1}, \dots, y_{iq})$. Der Zusammenhang zwischen dem Erwartungswert $E(y_i|x_i) = \mu_i = (\mu_{i1}, \dots, \mu_{iq})'$ und dem linearen Prädiktor $Z_i\beta$ mit der Designmatrix Z_i und dem unbekanntem Parametervektor $\beta' = (\beta_1, \dots, \beta_p)$ wird durch eine mehrdimensionale Responsefunktion h mit $\mu_i = h(Z_i\beta)$ hergestellt.

Das Generalisierte Lineare Modell, das beispielsweise in McCullagh & Nelder (1989) und Fahrmeir & Tutz (1994) vorgestellt ist, wird durch die Verteilungsannahme, die Strukturannahme und dem Aufbau der Designmatrix der Kovariablen definiert.

Bei der Verteilungsannahme wird vorausgesetzt, daß die y_i , gegeben x_i , bedingt unabhängig sind und ihre Verteilung aus einer Exponentialfamilie

$$f(y_i|\theta_i, \phi, w_i) = \exp \left\{ \frac{(y_i'\theta_i - b(\theta_i))}{\phi} w_i \right\} c(y_i, \phi, w_i).$$

stammt.

Durch die Strukturannahme wird der Erwartungswert $\mu_i = (\mu_{i1}, \dots, \mu_{iq})' = E(y_i|x_i)$ mit dem linearen Prädiktor $\eta_i = (\eta_{i1}, \dots, \eta_{iq})' = Z_i\beta$ durch eine invertierbare, zweimal stetig differenzierbare Funktion $h : \mathbb{R}^q \rightarrow \mathbb{R}^q$ mit

$$\mu_i = h(\eta_i) = h(Z_i\beta)$$

verbunden.

Die Designmatrix $Z_i \in \mathbb{R}^{q \times p}$ wird durch eine geeignete Abbildung aus den Kovariablen $x_i = (x_{i1}, \dots, x_{im})'$ gebildet. Ist die Ausprägung der Kovariablen kategorial, so wird eine Kodierung der Variablen vorgenommen.

Bemerkungen zur Verteilungsannahme

- Die Funktionen $b_i(\cdot)$ und $c_i(\cdot)$ bestimmen den genauen Typ der Exponentialfamilie, d. h. ob beispielsweise die Normal-, Binomial-, Multinomial-, Poisson- oder Gammaverteilung vorliegt.
- Der natürliche Parameter θ_i ist eine Funktion des Erwartungswertes μ_i , d. h. $\theta_i = \theta(\mu_i)$.
- Der Erwartungswert ist durch die Wahl der Exponentialfamilie mit $\mu = \frac{\partial b(\theta)}{\partial \theta}$ und die Kovarianz mit $cov(y_i|x_i) = \frac{\partial^2 b(\theta)}{\partial \theta \partial \theta'} \frac{\phi}{w_i}$ bestimmt.
- Der Dispersionsparameter ϕ ist von i unabhängig und in der Regel unbekannt. Im Linearen Modell gilt $\phi = \sigma^2$. Soweit nichts anderes angegeben ist, betrachten wir im folgenden ϕ als fest und bekannt mit $\phi \equiv 1$.
- Liegen die Daten in gruppierter Form vor und ist N_i die Anzahl der Beobachtungen in Gruppe i , so sind die Gewichte $w_i = N_i$ für $i = 1, \dots, N$ falls die Zielvariable y_i den Gruppendurchschnitt beschreibt, und es ist $w_i = \frac{1}{N_i}$, falls y_i die Summe der Ausprägungen in den Beobachtungsgruppen ist. Sind die Daten nicht gruppiert, so gilt $w_i \equiv 1$.

Bemerkungen zur Strukturannahme

- Der Vektor $\beta = (\beta_1, \dots, \beta_p)'$ ist der unbekannte zu schätzende Parametervektor.

- Die Funktion $h(\eta_i) = (h_1(\eta_{i1}), \dots, h_q(\eta_{iq}))'$ heißt Responsefunktion. Die zu h inverse Funktion $g = h^{-1}$ heißt Linkfunktion mit $g(\mu_i) = \eta_i = Z_i\beta$.
- Der Wertebereich der Responsefunktion h wurde in der Definition allgemein mit \mathbb{R}^q definiert. Durch die Wahl der Linkfunktion wird der Wertebereich aber meist geeignet eingeschränkt, z. B. für die Multinomialverteilung auf $(0, 1) \times \dots \times (0, 1)$.

Wählt man bei der Verteilungsannahme die Normalverteilung und für die Linkfunktion die identische Abbildung, so erhält man das Lineare Modell als Spezialfall eines Generalisierten Linearen Modells.

Spezielle Datenmodelle

Für jede Exponentialfamilie gibt es eine natürliche Linkfunktion, mit $g(\mu) \equiv \theta(\mu)$, d. h. es gilt $\theta \equiv Z\beta$ oder in Worten ausgedrückt, der natürliche Parameter ist gleich dem linearen Prädiktor. In dieser Arbeit werden wir Modellansätze mit der natürlichen Linkfunktion verwenden, die wir im folgenden kurz vorstellen werden.

Poissonmodell Im Log-linearen Poissonmodell wird $y \sim \mathcal{P}(\lambda)$ angenommen und für den unbekannt Parameter λ wird die Modellannahme $\lambda = E(y|x) = \exp(z'\beta)$ getroffen, d. h. die Responsefunktion in diesem Modell ist die Exponentialfunktion. Der Designvektor z enthält den Regressorenvektor x und gegebenenfalls auch die 1 für den Interzept. Im Poissonmodell gilt für die Varianz $var(y|x) = \lambda$.

Binomiales Logit-Modell Ist die Responsevariable binär kodiert und Bernoulli verteilt, d. h. $y \sim \mathcal{B}(1, \pi)$, so wird im Logit-Modell für den Erwartungswert $E(y|x) = P(y = 1|x) = \pi$ als Responsefunktion die logistische Verteilungsfunktion gewählt. Somit folgt mit dem linearen Prädiktor $\eta = z'\beta$ der Modellansatz $\pi = h(\eta) = \frac{\exp(\eta)}{1+\exp(\eta)}$.

Im ungruppierten Fall liegen N Beobachtungen (y_i, x_i) für $i = 1, \dots, N$ vor. Sind die Daten gruppiert, d. h. zu einem Kovariablenvektor liegen mehrere unabhängige Beobachtungen vor, so liegen die Daten in der Form (y_i, x_i, N_i) vor. Dabei ist y_i der Response-, x_i der Kovariablenvektor und N_i die Anzahl der Beobachtungen in Gruppe i . Insgesamt liegen in N Gruppen $N^* := \sum_{i=1}^N N_i$ Beobachtungen vor.

Damit der Modellansatz für gruppierte und für ungruppierte Daten mit der gleichen Notation verwendet werden kann, setzen wir voraus, daß die Responsevariable y_i die relative Häufigkeit in Gruppe i bezeichnet und definieren $w_i = 1$ für ungruppierte Daten und $w_i = N_i$ für gruppierte Daten. Für die Varianz gilt $\text{var}(y_i|x_i) = \frac{\pi_i(1-\pi_i)}{w_i}$, mit $\pi_i = E(y_i|x_i)$.

Multikategoriales und kumulatives Logit-Modell Diese Modelle sind multivariate Erweiterungen des oben beschriebenen Binomialen Logit-Modells. Wir gehen zunächst von ungruppierten Daten aus, d. h. es ist $y \sim \mathcal{M}(1, \pi)$ mit $\pi = (\pi_1, \dots, \pi_q)$. Die Ausprägungen der zugehörigen Responsevariablen $y_i = (y_{i1}, \dots, y_{iq})$ liegen in $K := q + 1$ Kategorien vor und es gilt:

$$y_{ir} = \begin{cases} 1 & \text{falls } Y_i = r \\ 0 & \text{sonst.} \end{cases}$$

Für gruppierte Daten setzen wir voraus, daß y_i für die einzelnen Kategorien die Gruppendurchschnitte enthält und definieren die Gewichte w_i für jede

Beobachtungsgruppe i analog zum Binomialen Modell.

Im multikategorialen Logit-Modell wird der linearen Prädiktor mit

$$\eta_i = \begin{pmatrix} \eta_{i1} \\ \vdots \\ \eta_{iq} \end{pmatrix} = Z_i \beta = \begin{bmatrix} 1 & x'_i & 0 & 0 & \cdots & \cdots & 0 \\ 0 & 0 & 1 & x'_i & \ddots & & \vdots \\ \vdots & & \ddots & \ddots & \ddots & \ddots & \vdots \\ \vdots & & & \ddots & \ddots & \ddots & 0 \\ 0 & \cdots & \cdots & \cdots & 0 & 1 & x'_i \end{bmatrix} \begin{pmatrix} \beta_{10} \\ \beta_1 \\ \beta_{20} \\ \beta_2 \\ \vdots \\ \beta_{q0} \\ \beta_q \end{pmatrix}$$

und die Responsefunktion $h = (h_1, \dots, h_q)$ mit

$$h_r(\eta_{i1}, \dots, \eta_{iq}) = \frac{\exp(\eta_{ir})}{1 + \sum_{s=1}^q \exp(\eta_{is})}$$

für $r = 1, \dots, q$ gewählt.

Sind die Kategorien von $Y \in \{1, \dots, q + 1 =: K\}$ ordinal, so kann die Ordnung der Responsekategorien ausgenutzt werden.

Wir beschränken uns auf die Definition für das kumulative Logit-Modell. Weitere Modelle, welche die Ordnung der Responsekategorien berücksichtigen, sind bei Agresti (1990), McCullagh (1980) und Fahrmeir & Tutz (1994) beschrieben.

Motiviert wird das kumulative Logit-Modell, das auch als Schwellenwertmodell bezeichnet wird, durch die Existenz einer latenten stetigen Variablen $V = -x'\gamma + \epsilon$, so daß mit den Schwellen $-\infty = \theta_0 < \theta_1 < \dots < \theta_q < \theta_K = \infty$

$$Y = r \quad \Leftrightarrow \quad \theta_{r-1} < V \leq \theta_r \quad \text{für } r=1, \dots, K$$

gilt.

Mit dem Modellansatz $P(y \leq r|x) = F(\theta_r + x'\gamma)$, wobei F die logistische Verteilungsfunktion, d. h. $F(u) = \frac{\exp(u)}{1+\exp(u)}$ ist, sowie dem linearen Prädiktor

$$\eta = \begin{pmatrix} \eta_1 \\ \vdots \\ \eta_q \end{pmatrix} = Z\beta = \begin{bmatrix} 1 & 0 & \dots & \dots & 0 & x' \\ 0 & 1 & 0 & \dots & 0 & x' \\ \vdots & \ddots & \ddots & \ddots & \vdots & \vdots \\ 0 & \dots & 0 & 1 & 0 & x' \\ 0 & \dots & \dots & 0 & 1 & x' \end{bmatrix} \begin{pmatrix} \theta_1 \\ \vdots \\ \theta_q \\ \gamma \end{pmatrix}$$

folgt

$$P(Y = 1|x) = F(\theta_1 + x'\gamma)$$

$$P(Y = r|x) = F(\theta_r + x'\gamma) - F(\theta_{r-1} + x'\gamma) \quad \text{für } r = 2, \dots, q.$$

Da für die Schwellen $\theta_1 \leq \dots \leq \theta_q$ erfüllt sein muß, wird für die Schätzung der Parameter meist eine Parametrisierung der Schwellen mit $\alpha_1 := \theta_1$ und $\alpha_r := \ln(\theta_r - \theta_{r-1})$ für $r = 2, \dots, q$ durchgeführt. Das heißt es wird der Modellansatz $\eta = Z\beta$ mit

$$Z = \begin{pmatrix} 1 & 0 & \dots & 0 & x' \\ 0 & 1 & & & 0 \\ & & \ddots & & \vdots \\ 0 & & & 1 & 0 \end{pmatrix} \quad \text{und } \beta' = (\alpha_1, \dots, \alpha_q, \gamma).$$

verwendet.

Die Reparametrisierung der Schwellen erfolgt mit $\theta_1 = \alpha_1$ und $\theta_r = \alpha_1 + \sum_{s=2}^r \exp(\alpha_s)$ für $r = 2, \dots, q$.

In beiden Modellen gilt für die Kovarianz $cov(y_i) = \frac{1}{w_i}(diag(\pi_i) - \pi_i\pi_i')$, wobei w_i und π_i analog zum binomialen Logit-Modell definiert sind.

1.1.2 Schätzung der Parameter

Der unbekannte Parameter $\beta = (\beta_1, \dots, \beta_p)$ wird aus dem Modellansatz $E(y_i|x_i) = \mu_i = h(\eta_i)$ mit $\eta_i = Z_i\beta$ bestimmt. Die Schätzung des Parameters kann durch Maximierung der Log-Likelihood $l(\beta)$ durchgeführt werden. Dabei wird die Nullstelle der Scorefunktion $s(\beta) = \frac{\partial l}{\partial \beta}$ berechnet, wobei zu beachten ist, daß die beobachtete Fishermatrix $F_{obs}(\beta) = -\frac{\partial^2 l(\beta)}{\partial \beta \partial \beta'}$ positiv definit ist.

Da vorausgesetzt wurde, daß die y_i unabhängig, identisch verteilt sind und die Verteilung aus der Exponentialfamilie stammt, gilt (bis auf einen konstanten Term):

$$l(\beta) = \sum_{i=1}^N l_i(\beta) = \sum_{i=1}^N \ln f(y_i|\theta_i, \phi, w_i) = \sum_{i=1}^N \frac{y_i\theta_i - b(\theta_i)}{\phi} w_i$$

und für die Scorefunktion folgt mit $\phi \equiv 1$:

$$s(\beta) = \frac{\partial l}{\partial \beta} = \sum_{i=1}^N s_i(\beta) = \sum_{i=1}^N Z_i' D_i(\beta) \Sigma_i^{-1}(\beta) [y_i - \mu_i(\beta)]$$

mit $D_i(\beta) = \frac{\partial h(\eta_i)}{\partial \eta}$ und $\Sigma_i(\beta) = cov(y_i)$.

Die numerische Berechnung des Maximum Likelihood Schätzers kann unter Verwendung des Fisher-Scoring Algorithmus bestimmt werden. Ist $F(\beta) = cov(s(\beta)) = \sum_i F_i(\beta)$ die Fisherinformationsmatrix, so gilt mit $F(\beta) = E(F_{obs}(\beta))$:

$$F_i(\beta) = Z_i' D_i(\beta) \Sigma_i^{-1}(\beta) D_i(\beta)' Z_i .$$

Nach der Initialisierung des Startwertes $\hat{\beta}^{(0)}$ wird die Iteration

$$\hat{\beta}^{(k+1)} = \hat{\beta}^{(k)} + F^{-1}\left(\hat{\beta}^{(k)}\right) s\left(\hat{\beta}^{(k)}\right)$$

für $k = 0, 1, 2, \dots$ so lange durchgeführt bis das Abbruchkriterium

$$\frac{\|\hat{\beta}^{(k+1)} - \hat{\beta}^{(k)}\|}{\|\hat{\beta}^{(k)}\|} < \epsilon$$

für ein vorgegebenes ϵ erfüllt ist.

Kapitel 2

Modellerweiterungen im Generalisierten Linearen Modell

Im vorherigen Kapitel wurde das grundlegende Generalisierte Lineare Modell vorgestellt. Bei den Modellannahmen wurde von bedingt unabhängigen Beobachtungen ausgegangen.

Sind die zugrundeliegenden Daten in einer Studie erhoben worden, bei der einzelne Individuen mehrfach untersucht wurden, so kann keinesfalls davon ausgegangen werden, daß die Daten (y_i, x_i) für $i = 1, \dots, N$ bedingt unabhängig sind. Aber auch wenn die einzelnen Untersuchungseinheiten zu Gruppen zusammengefaßt sind, muß diese Gruppenstruktur durch das Modell berücksichtigt werden. Dies ist beispielsweise dann der Fall, wenn sich die Untersuchungseinheiten in gemeinsame Familien oder sonstige soziale, regionale oder administrative Gruppen gliedern lassen.

Wir werden diese Gruppenstruktur durch die Modellierung von zufälligen Effekten berücksichtigen. Durch diesen Ansatz berücksichtigen wir nicht nur

die Gruppenstruktur, sondern es können zusätzlich die Einflüsse von nicht erhobenen Kovariablen untersucht werden.

Diese gruppen- und/oder individuenspezifischen Effekte können auftreten, wenn die Erhebung aller Daten aus technischen bzw. wirtschaftliche Gründen nicht möglich war oder wichtige Variablen aufgrund von fehlerhaften Versuchsplänen übersehen wurden.

Wurden die Daten einer Studie in einem Zeitverlauf gewonnen, so muß zusätzlich diese Zeitabhängigkeit berücksichtigt werden. Wir betrachten hierzu Modelle, die eine Variation der Modellparameter über die Zeit ermöglichen. In diesem Kapitel werden wir daher im folgenden Modellansätze, die entweder zeitvariierende Koeffizienten oder zufällige Effekte berücksichtigen, vorstellen.

Die Modellansätze für variierende Koeffizienten werden in der Literatur in sehr unterschiedlichen Ansätzen betrachtet. Für die Betrachtung der Variation der Modellparametern im Generalisierten Linearen Modell als Glättungsfunktion von anderen Variablen werden in Hastie & Tibshirani (1993) verschiedene Modelle vorgestellt. Aus dieser Modellklasse werden in Tutz & Kauermann (1997) lokal gewichtete Schätzungen betrachtet.

In Friedman (1991) und Stone, Hansen, Kooperberg & Truong (1997) werden Schätzverfahren mit Glättung durch Splines vorgestellt, wobei insbesondere die Wahl der Knoten für die Splines berücksichtigt wird. Allgemeine Ansätze mit Spline-Funktionen werden in Eubank (1988) sowie in Eilers & Marx (1996) für B-splines untersucht.

In Silverman (1984) werden sowohl Glättungsverfahren mit Spline-Funktionen als auch Kernglättungen mit flexibler Bandbreite vorgestellt. Hastie & Loader (1993) schlagen eine lokale Regression mit Kern-Glättung vor. Parametrische und semi-parametrische Glättungsansätze in Verweildauermodellen

werden in Efron (1988) betrachtet. Die Zeitabhängigkeit bei Meßwiederholungen, die beispielsweise bei der Analyse von klinischen Studien auftritt, untersucht Davis (1991) durch semi-parametrische und nichtparametrische Ansätze.

Die in dieser Arbeit verwendeten Verfahren für die Schätzung der Parameter mit normalverteilten zufälligen Effekten werden in Fahrmeir & Tutz (1994), Hennevogl (1991) und Tutz & Hennevogl (1996) vorgestellt.

Generalisierte Lineare Mischmodelle mit unterschiedlichen Schätzmethoden werden in McCulloch (2000) vorgestellt. In Geyer & Thompson (1992) werden Approximierungstechniken behandelt, bei denen die Likelihoodfunktion direkt maximiert wird. In McCulloch (1994), McCulloch (1997) und Booth & Hobert (1999) werden Verfahren vorgestellt, die einen EM-Algorithmus mit Gibbs-Sampling bzw. Metropolis-Hastings Algorithmus verwenden. Ein anderer Ansatz wird in Ruppert, Reish & Carroll (1984) und Quintana, Lui & Pino (1999) gewählt, es wird ein stochastischer Approximationsalgorithmus verwendet und die Likelihood-Gleichung wird als Regressions Problem betrachtet.

Da in Studien meist sowohl gruppenspezifische Effekte als auch Einflüsse über die Zeit berücksichtigt werden müssen, werden wir im nächsten Kapitel beide Modellansätze in ein gemeinsames Modell aufnehmen. Dieser Ansatz wird in der Literatur, bis auf Tutz (1999), nicht weiter betrachtet.

2.1 Modellansätze mit zufälligen Effekten

Liegen die Beobachtungen von verschiedenen Gruppen bzw. Kategorien, die als Cluster bezeichnet werden, vor, kann der Einfluß zwischen dem Response und den erklärenden Variablen von Cluster zu Cluster variieren. Deshalb muß die Gruppenstruktur der Daten im Modellansatz berücksichtigt werden. Insbesondere bei Meßwiederholungen an mehreren Untersuchungseinheiten in einem Zeitverlauf, die im nächsten Abschnitt betrachtet werden, ist mit dieser sogenannten “Extra Variation innerhalb der Daten”, die auch als unbeobachtete Populationsheterogenität bezeichnet wird, zu rechnen.

Da die Beobachtungen nicht in unabhängigen elementaren Beobachtungseinheiten vorliegen, ist die Modellannahme im Generalisierten Linearen Modell, welche von gleicher Varianzstruktur für alle Beobachtungen ausgeht, verletzt, und es können Modelle, die eine endliche Mischungsdichte zulassen, angewendet werden. Diese Modelle werden zum Beispiel in Everitt & Hand (1981), Laird & Louis (1982), Little & Rubin (1987), McLachlan & Basford (1988), Meng & van Dyk (1997b) und Titterington, Smith & Makov (1985) vorgestellt.

Ist die Anzahl der Cluster sehr klein, so ist es möglich mit Hilfe von ANOVA (analysis of variances) und ANCOVA (analysis of covariances) Modellen die Variation zwischen den Clustern zu untersuchen. Man erhält Aussagen über die Variation in den konkret vorliegenden Clustern.

Wenn die Auswahl der beobachteten Cluster als zufällig angesehen wird, ist man primär nicht an einer Aussage über die konkret vorliegenden Cluster, sondern über die zugrunde liegende Population interessiert. Es können Modelle mit zufälligen Effekten, die beispielsweise in Fahrmeir & Tutz (1994), Jansen (1993), Longford (1993), Meng & van Dyk (1997a), Meng (1997) und

Hennevogl (1991) vorgestellt werden, angewendet werden.

Im Modellansatz mit zufälligen Effekten wird davon ausgegangen, daß die Unterschiede, die in der Zusammenhangsanalyse in den Clustern auftreten, durch Variablen erklärt werden könnten, die jedoch nicht beobachtet wurden. Diese unbeobachteten Variablen können beispielsweise auftreten, wenn ihre Erhebung aus technischen bzw. wirtschaftlichen Gründen nicht möglich war oder diese relevante Variablen aufgrund von fehlerhaften Versuchsplänen übersehen wurden.

Im Generalisierten Linearen Modell mit zufälligen Effekten werden die nicht beobachteten Variablen durch zusätzliche Parameter im linearen Prädiktor berücksichtigt, die als Realisation einer Zufallsvariable aufgefaßt werden.

In der Definition für Generalisierte Lineare Modelle mit zufälligen Effekten setzen wir voraus, daß für $i = 1, \dots, N$ Beobachtungseinheiten jeweils n_i Meßwiederholungen vorliegen. Somit liegen insgesamt $N^{**} = \sum_{i=1}^N n_i$ Beobachtungen vor, die sich jedoch auf $i = 1, \dots, N$ Gruppen, bzw. allgemeiner ausgedrückt auf Beobachtungseinheiten verteilen. Wir setzen damit die Datenstruktur (y_{it}, x_{it}) mit $i = 1, \dots, N$ und $t \in \{t_{i1}, \dots, t_{in_i}\}$ voraus.

Im anschließenden Kapitel werden wir den Ansatz der zeitvariierenden Koeffizienten mit der Modellierung von zufälligen Effekten verbinden. Dort werden wir voraussetzen, daß die Beobachtungen (y_{it}, x_{it}) der i -ten Beobachtungseinheit zum Zeitpunkt t erhoben wurden.

Im folgenden betrachten wir Modelle bei denen für jede Beobachtungseinheit i ein zufälliger Effekt b_i in den linearen Prädiktor aufgenommen wird. Das heißt, statt dem linearen Prädiktor $\eta_{it} = Z_{it}\beta$ wird der lineare Prädiktor $\eta_{it} = Z_{it}\beta + W_{it}b_i$ betrachtet. Die Designmatrizen Z_{it} und W_{it} werden durch geeignete Abbildungen aus den Kovariablen gebildet.

Eine übliche Annahme für die zufälligen Effekte b_1, \dots, b_N ist, daß diese unabhängig normalverteilt sind. Beispiele für diesen Modellansatz werden wir in den Abschnitten 2.1.1 und 2.1.2 betrachten. Bei dem in Abschnitt 2.1.3 vorgestellten Verfahren der nichtparametrischen Maximum Likelihood wird keine parametrische Voraussetzung für die zufälligen Effekte getroffen.

Modelldefinition

Die Modelldefinition ist zweistufig und teilt sich in das Beobachtungs- und das Wahrscheinlichkeitsmodell auf.

Das Beobachtungsmodell setzt voraus, daß die Responsewerte y_{i1}, \dots, y_{in_i} für $i = 1, \dots, N$ bedingt unabhängig sind, gegeben x_{i1}, \dots, x_{in_i} und b_i sowie die Dichte der y_{it} zu der Exponentialfamilie gehört. Der Erwartungswert μ_{it} ist über die Linkfunktion h mit dem linearen Prädiktor durch

$$\mu_{it} = E(y_{it}|b_i, x_{it}) = h(\eta_{it}) \text{ mit } \eta_{it} = Z_{it}\beta + W_{it}b_i$$

verbunden.

Das Wahrscheinlichkeitsmodell bestimmt die Verteilung der zufälligen Effekte b_1, \dots, b_N . In den Abschnitten 2.1.1 und 2.1.2 setzten wir voraus, daß diese für $i = 1, \dots, N$ unabhängig normalverteilt sind mit

$$b_i \stackrel{iid}{\sim} \mathcal{N}(0, Q).$$

Die zu schätzenden Modellparameter in diesem Modellansatz sind der Parametervektor β , der den globalen Einfluß der Regressoren beschreibt, die Kovarianzmatrix Q und die zufälligen Effekte b_i , welche die clusterspezifischen Abweichungen vom globalen Einfluß β beschreiben.

Für die Schätzung der Parameter mit normalverteilten zufälligen Effekten geben wir im folgenden zwei unterschiedliche Verfahren an. Das erste Verfahren ist der EM-Algorithmus mit Gauß-Hermite Integrationstechnik und das zweite Verfahren der EM-Type. Beide Verfahren werden in Fahrmeir & Tutz (1994), Hennevogl (1991) und Tutz & Hennevogl (1996) vorgestellt.

2.1.1 EM-Algorithmus mit Gauß-Hermite Integrations- technik

Die Schätzung der Parameter in diesem Algorithmus erfolgt in zwei Schritten. Zuerst werden die festen Parameter β und Q geschätzt und anschließend die zufälligen Effekte b_i .

Schätzung der festen Parameter

Um die Notation übersichtlicher zu gestalten, beschreiben wir kurz das verwendete Verfahren für einen eindimensionalen zufälligen Effekt. Im Anhang im Abschnitt 6.1.2 gehen wir auf die Details des Verfahrens ein und betrachten dabei den allgemeinen Fall bei dem die zufälligen Effekte b_i mehrdimensional sein können. Im folgenden eindimensionalen Fall schreiben wir statt der Kovarianzmatrix Q einfach σ^2 .

Der lineare Prädiktor im univariaten Generalisierten Linearen Modell mit einem eindimensionalen zufälligen Effekt

$$\eta_{it} = x'_{it} \beta + b_i \quad \text{mit } b_i \stackrel{iid}{\sim} \mathcal{N}(0, \sigma^2)$$

läßt sich mit $b_i = a_i \sigma$ umparametrisieren zu

$$\eta_{it} = x'_{it} \beta + a_i \sigma \quad \text{mit } a_i \stackrel{iid}{\sim} \mathcal{N}(0, 1).$$

Mit der üblichen Voraussetzung der bedingten Unabhängigkeit zwischen den einzelnen Beobachtungseinheiten bzw. innerhalb der Meßwiederholungen zu einer Beobachtungseinheit gilt für die Likelihood

$$L(\beta, \sigma) = \prod_{i=1}^N \int f(y_i|a_i; \beta, \sigma) p(a_i) da_i \quad \text{mit} \quad f(y_i|a_i; \beta, \sigma) = \prod_{t=1}^{n_i} f(y_{it}|a_i; \beta, \sigma).$$

Wobei $f(\cdot)$ die bedingte Dichte, die aus einer Exponentialfamilie stammt und $p(\cdot)$ die Dichte der Standardnormalverteilung ist.

Diese Mischungsdichte kann nur für einige Spezialfälle analytisch gelöst werden, so daß für die Lösung des Integrals numerische Verfahren angewendet werden.

Bei der Gauß-Hermite Integrationstechnik wird das Integral durch eine endliche Summe über M Quadratur-Stützstellen d_j , ($j = 1, \dots, M$) mit bekannten Quadratur-Gewichten v_j , $j = 1, \dots, M$ approximiert, so daß

$$L(\beta, \sigma) \approx \prod_{i=1}^N \sum_{j=1}^M v_j f(y_i|d_j; \beta, \sigma)$$

folgt. Die Gewichte v_j und Stützstellen d_j für die Gauß-Hermite Integration liegen in Tabellen (Abramowitz & Stegun, 1972) vor.

Das bedeutet, daß die Likelihood durch eine endliche Mischung von Dichten aus einer Exponentialfamilie mit bekannten Mischungsgewichten v_j und bekannten Maßpunkten d_j approximiert wird. Ist $l(\beta, \sigma)$ die Log-Likelihood, d. h.

$$l(\beta, \sigma) = \sum_{i=1}^N \ln \sum_{j=1}^M v_j f_{ij} \quad \text{mit} \quad f_{ij} = f(y_i|d_j; \beta, \sigma)$$

so folgt (Hinde, 1982; Aitkin, 1996)

$$\frac{\partial l}{\partial \beta} = \sum_{i=1}^N \sum_{j=1}^M \frac{v_j f_{ij} \frac{\partial \ln f_{ij}}{\partial \beta}}{\sum_{l=1}^M v_l f_{il}} = \sum_{i=1}^N \sum_{j=1}^M c_{ij} s_{ij}(\beta)$$

$$\frac{\partial l}{\partial \sigma} = \sum_{i=1}^N \sum_{j=1}^M \frac{v_j f_{ij} \frac{\partial \ln f_{ij}}{\partial \sigma}}{\sum_{l=1}^M v_l f_{il}} = \sum_{i=1}^N \sum_{j=1}^M c_{ij} s_{ij}(\sigma).$$

Dabei ist c_{ij} die posteriori Wahrscheinlichkeit, daß die Beobachtungseinheit i aus der Mischungskomponente j stammt, mit

$$c_{ij} = \frac{v_j f_{ij}}{\sum_{s=1}^M v_s f_{is}}.$$

Die Komponenten $s_{ij}(\beta)$ und $s_{ij}(\sigma)$ sind die β - bzw. σ -Komponenten der Scorefunktion für die i -te Beobachtung in der j -ten Mischungskomponente. Da die Stützstellen d_j in Tabellen vorliegen, sind diese Größen bekannt. Die Parameter β und σ sind die zugehörigen Regressionskoeffizienten.

Beim Maximieren der Log-Likelihood wird die Scorefunktion gleich Null gesetzt, und man erhält die gleichen gewichteten Summen, die als Gleichungen eines gewöhnlichen Generalisierten Linearen Modells mit dem linearen Prädiktor

$$\eta_{itj} = x'_{it} \beta + d_j \sigma = [x'_{it}, d_j] \begin{pmatrix} \beta \\ \sigma \end{pmatrix} \quad \text{mit} \quad \begin{cases} i = 1, \dots, N \\ t = 1, \dots, n_i \\ j = 1, \dots, M \end{cases}$$

entstehen, wenn die Beobachtungen mit c_{ij} gewichtet werden und für den zugehörigen Response $y_{itj} \equiv y_{it}$ gilt.

Das Lösen der Gleichungen für gegebene Gewichte c_{ij} und die Berechnung der Gewichte mit der letzten Parameterschätzung für $\alpha' = (\beta', \sigma)$ führt Fahrmeir & Tutz (1994) zu folgendem EM-Algorithmus:

- Die Anzahl M der Stützstellen für die Gauß-Hermite Integration wird schrittweise erhöht.
- Die Daten des Regressionsmodells werden mit der Anzahl der Stützstellen vervielfacht, indem jede Zeile $y_{it} = x'_{it} \beta$ des Modellansatzes durch

$$\begin{pmatrix} y_{it} \\ \vdots \\ y_{it} \end{pmatrix} = \begin{bmatrix} x'_{it} & d_1 \\ \vdots & \vdots \\ x'_{it} & d_M \end{bmatrix} \begin{pmatrix} \alpha \end{pmatrix} \quad \text{mit } \alpha = \begin{pmatrix} \beta \\ \sigma \end{pmatrix}$$

ersetzt wird.

- Der Startwert des Schätzers für die Iteration ist $\hat{\alpha}^{(0)} = (\hat{\beta}^{(0)}, \hat{\sigma}^{(0)})'$, wobei $\hat{\beta}^{(0)}$ der Schätzer des Generalisierten Linearen Modells ohne zufällige Effekte mit dem linearen Prädiktor $\eta_{it} = x'_{it}\beta$ ist. Für $\hat{\sigma}^{(0)}$ wird eine Konstante beispielsweise $\hat{\sigma}^{(0)} = 2$ gewählt.
- Jeder Iterationsschritt $k = 0, 1, 2, \dots$ besteht aus den folgenden Schritten:

– Berechnung der Gewichte mit:

$$c_{ij}(\alpha^{(k)}) = \frac{v_j \cdot f(y_i | d_j; \alpha^{(k)})}{\sum_{s=1}^M v_s \cdot f(y_i | d_s; \alpha^{(k)})}$$

– Schätzung von $\hat{\alpha}^{(k+1)}$ durch das mit $c_{ij}(\alpha^{(k)})$ gewichtete Regressionsmodell.

- Ist das Abbruchkriterium:

$$\frac{\|\hat{\alpha}^{(k+1)} - \hat{\alpha}^{(k)}\|}{\|\hat{\alpha}^{(k)}\|} < \epsilon$$

für ein vorgegebenes ϵ erfüllt, so wird die Iteration beendet.

Durch die Verwendung des EM-Algorithmus ist sichergestellt, daß sich in jedem Iterationsschritt der Wert der Likelihood erhöht (Wu, 1983). Es ist jedoch nicht gewährleistet, daß das globale Maximum gefunden wird.

Der Startwert $\sigma^{(0)} = 0$ sollte deshalb vermieden werden. Denn bei $\alpha' = (\beta', 0)$, d. h. beim Modellansatz ohne Berücksichtigung von zufälligen Effekten, liegt meist ein lokales Maximum vor. In den durchgeführten Simulationen hat sich gezeigt, daß $\sigma^{(0)} = 2$ eine gute Wahl für den Startwert ist.

Um die optimale Anzahl der gewählten Stützstellenanzahl M bei der Gauß-Hermite Integration zu finden, muß der Algorithmus für $M = 2, 4, 6, \dots$ solange durchgeführt werden, bis sich der Wert der Likelihood nicht mehr erhöht.

Schätzung der zufälligen Effekte

Nachdem die festen Parameter β und σ^2 , bzw. allgemein die Kovarianzmatrix Q , geschätzt wurden, kann die Bayes-Schätzung der zufälligen Effekte b_i erfolgen.

Der unbekannte Parameter b_i wurde als Zufallsgröße mit $b_i \sim \mathcal{N}(0, Q)$ vorausgesetzt, so daß die Priori-Dichte mit

$$p(b_i) = \frac{1}{(2\pi)^{q/2} |Q|^{1/2}} \exp \left\{ -\frac{1}{2} b_i' Q^{-1} b_i \right\}$$

als bekannt angenommen wird. Wir schätzen b_i durch Maximierung der Posteriori Dichte $p(b_i|Y)$, d. h. wir maximieren $l_p(b_i|Y) := \ln p(b_i|Y)$ mit:

$$\begin{aligned}
l_p(b_i|Y) &= \ln L(b_i|Y) + \ln p(b_i) \\
&= \ln L(b_i|Y) - \frac{1}{2} b_i' Q^{-1} b_i,
\end{aligned}$$

wobei $L(b_i|Y)$ die Likelihoodfunktion, abhängig von b_i , bei gegebenen Daten Y bezeichnet und konstante Terme vernachlässigt wurden.

Die Bestimmung von b_i kann durch Anwendung des Fisher-Scoring Algorithmus durchgeführt werden.

Mit dem Startwert $b_i^{(0)} = 0$ wird die Integration

$$b_i^{(k+1)} = b_i^{(k)} + F^{-1}(b_i^{(k)}) s(b_i^{(k)})$$

für $k = 1, 2, \dots$ durchgeführt bis ein Abbruchkriterium erfüllt ist, wobei $F(\cdot)$ die Fishermatrix und $s(\cdot)$ die Scorefunktion bezeichnen.

Als Anwendungsbeispiel betrachten wir im folgenden das Binäre Logit-Modell bei dem die natürliche Linkfunktion verwendet wird. Liegt ein eindimensionaler zufälliger Effekt vor und ist $W_{it} \equiv 1$, d.h.

$$l_p(b_i|Y) = \ln L(b_i|Y) - \frac{1}{2} Q^{-1} b_i^2$$

so folgt für Scorefunktion

$$s(b_i) = \sum_{t=1}^{n_i} n_{it} [y_{it} - \mu_{it}(b_i)] - Q^{-1} b_i$$

sowie für die Fishermatrix

$$F(b_i) = \sum_{t=1}^{n_i} n_{it} \mu_{it}(b_i) (1 - \mu_{it}(b_i)) + Q^{-1}$$

mit $\mu_{it}(b_i) = h(Z_{it} \beta + b_i)$, wobei für β und Q die entsprechenden Schätzer verwendet werden.

2.1.2 EM-Type Algorithmus

Beim EM-Type Algorithmus wird eine simultane Schätzung der Parameter β , Q und b_i , $i = 1, \dots, N$ mit Hilfe des EM-Algorithmus durchgeführt. Das grundlegende Prinzip des EM-Type Algorithmus wird auch beim EM-Algorithmus mit Gauß-Hermite Integrationstechnik verwendet und im Anhang in Abschnitt 6.1.2 vorgestellt.

Kurz zusammengefaßt kann der EM-Type Algorithmus wie folgt beschrieben werden. Die Parameter $(\beta', b'_1, \dots, b'_N) = \delta'$ werden als unbekannte Daten betrachtet und Q ist der zu schätzende Parameter.

Mit diesem Ansatz können im $(k + 1)$ -ten Iterationsschritt im M-Schritt die Schätzer für

$$\delta'^{\langle k+1 \rangle} = (\beta'^{\langle k+1 \rangle}, b'_1{}^{\langle k+1 \rangle}, \dots, b'_N{}^{\langle k+1 \rangle}), \text{ gegeben } Q^{\langle k \rangle}$$

bestimmt werden. Im E-Schritt wird dann Q mit dem Ansatz

$$M(Q|Q^{\langle k \rangle}) = E(\ln f(Y, \delta, Q)|Y, Q^{\langle k \rangle})$$

geschätzt.

Auf die Einzelheiten des EM-Type Algorithmus wird, unterteilt in den E- bzw. M-Schritt, im folgenden eingegangen.

M-Schritt

Mit gegebenem $Q^{\langle k \rangle}$ gilt für die posteriori Log-Likelihood $l(\delta)$

$$\begin{aligned} l(\delta) &= \sum_{i=1}^N \ln f(y_i|b_i, \beta) - \frac{1}{2} \sum_{i=1}^N b'_i (Q^{\langle k \rangle})^{-1} b_i \\ &= \sum_{i=1}^N \sum_{t=1}^{n_i} \ln f(y_{it}|b_i, \beta) - \frac{1}{2} \sum_{i=1}^N b'_i (Q^{\langle k \rangle})^{-1} b_i, \end{aligned}$$

so daß für die Scorefunktion

$$\begin{aligned}\frac{\partial l(\delta)}{\partial \beta} &= \sum_{i=1}^N \sum_{t=1}^{n_i} Z'_{it} D_{it} \Sigma_{it}^{-1} (y_{it} - \mu_{it}) \\ \frac{\partial l(\delta)}{\partial b_i} &= \sum_{t=1}^T W'_{it} D_{it} \Sigma_{it}^{-1} (y_{it} - \mu_{it}) - (Q^{(k)})^{-1} b_i\end{aligned}$$

mit $D_{it} = D_{it}(\beta, b_i) = \frac{\partial h(\eta_{it})}{\eta}$, $\Sigma_{it}^{-1} = \Sigma_{it}^{-1}(\beta, b_i) = (\text{cov}(y_{it}|\beta, b_i))^{-1}$ und $\mu_{it} = \mu_{it}(\beta, b_i) = h(\eta_{it})$ folgt. Für die Fishermatrix gilt

$$F(\delta) = \begin{bmatrix} F_{\beta\beta} & F_{\beta b_1} & \cdots & F_{\beta b_N} \\ F_{b_1\beta} & F_{b_1 b_1} & & \textcircled{0} \\ \vdots & & \ddots & \\ F_{b_N\beta} & \textcircled{0} & & F_{b_n b_n} \end{bmatrix}$$

mit

$$\begin{aligned}F_{\beta\beta} &= \sum_{i=1}^N \sum_{t=1}^{n_i} Z'_{it} D_{it} \Sigma_{it}^{-1} D'_{it} Z_{it} \\ F_{\beta b_i} &= \sum_{t=1}^{n_i} Z'_{it} D_{it} \Sigma_{it}^{-1} D'_{it} W_{it} \\ F_{b_i b_i} &= \sum_{t=1}^{n_i} W'_{it} D_{it} \Sigma_{it}^{-1} D'_{it} + (Q^{(k)})^{-1} \\ F_{b_i \beta} &= F'_{\beta b_i}.\end{aligned}$$

Die Berechnung von $\delta^{(k+1)}$ kann durch den Fisher-Scoring Algorithmus mit

$$\delta^{(k+1)} = \delta^{(k)} + F^{-1}(\delta^{(k)}) s(\delta^{(k)})$$

durchgeführt werden.

Im E-Schritt werden die Schätzer \hat{b}_i sowie ihre posteriori Kovarianzen V_{ii} für $i = 1, \dots, n$ benötigt. Letztere ergeben sich aus

$$F^{-1}(\delta) = \begin{bmatrix} V_{\beta\beta} & V_{\beta 1} & \cdots & V_{\beta n} \\ V_{1\beta} & V_{11} & \cdots & V_{1n} \\ \vdots & & \ddots & \\ V_{n\beta} & V_{n1} & \cdots & V_{nn} \end{bmatrix}$$

mit

$$V_{\beta\beta} = (F_{\beta\beta} - \sum_{i=1}^n F_{\beta i} F_{ii}^{-1} F_{i\beta})^{-1}$$

$$V_{ii} = F_{ii}^{-1} + F_{ii}^{-1} F_{i\beta} V_{\beta\beta} F_{\beta i} F_{ii}^{-1}$$

$$V_{\beta i} = V_{i\beta}' = -V_{\beta\beta} F_{\beta i} F_{ii}^{-1}$$

$$V_{ij} = V_{ji}' = F_{ii}^{-1} F_{i\beta} V_{\beta\beta} F_{\beta j} F_{jj}^{-1}, i \neq j$$

Im Anhang in Abschnitt 6.2.1 wird gezeigt, daß die Blockdiagonalgestalt der Fishermatrix ausgenutzt werden kann, so daß für die Berechnung von $\delta^{(k+1)}$ nicht die gesammte Fishermatrix $F(\delta^{(k)})$ invertiert werden muß.

E-Schritt

Da für die gemeinsame Dichte

$$f(Y, \delta, Q) = \prod_{i=1}^N f(Y, \delta) p(b_i, Q)$$

gilt und nur der letzte Faktor in der Dichte von Q abhängt, läßt sich der E-Schritt im $\langle k + 1 \rangle$ -ten Iterationsschritt reduzieren zu:

$$\tilde{M}(Q|Q^{\langle k \rangle}) = E(\ln p(\delta, Q)|Y, Q^{\langle k \rangle})$$

Da $b_i \sim \mathcal{N}(0, Q)$ vorausgesetzt wurde, erhält man bis auf konstante Terme

$$\tilde{M}(Q|Q^{\langle k \rangle}) = -\frac{N}{2} \ln |Q| - \sum_{i=1}^N \text{tr}(Q^{-1} E(b_i b_i' | Y, Q^{\langle k \rangle}))$$

und damit kann $Q^{\langle k+1 \rangle}$ bestimmt werden mit

$$Q^{\langle k+1 \rangle} = \frac{1}{N} \sum_{i=1}^N \text{cov}(b_i | y_i, Q^{\langle k \rangle}) + E(b_i | y_i, Q^{\langle k \rangle}) E(b_i | y_i, Q^{\langle k \rangle})' .$$

In Tutz & Hennevogl (1996) wird vorgeschlagen, wegen des großen numerischen Aufwands bei der Berechnung der Erwartungswerte, den posteriori Erwartungswert von b_i durch den posteriori Mode $\hat{b}_i^{\langle k \rangle}$ zu ersetzen und die posteriori Kovarianzen durch \hat{V}_{ii} . Damit folgt

$$Q^{\langle k \rangle} = \frac{1}{N} \sum_{i=1}^N \hat{V}_{ii}^{\langle k \rangle} + \hat{b}_i^{\langle k \rangle} (\hat{b}_i^{\langle k \rangle})' .$$

2.1.3 Modellansatz der nichtparametrischen Maximum Likelihood

Im vorherigen Abschnitt wurde vorausgesetzt, daß die eindimensionalen zufälligen Effekte unabhängig normalverteilt mit $b_i \stackrel{iid}{\sim} \mathcal{N}(0, \sigma^2)$ sind. Das heißt, es wird eine parametrische Form vorausgesetzt. Da die Mischungsdichte nicht bekannt ist und die Parameterschätzungen empfindlich von der Spezifizierung der parametrischen Form abhängen kann, bietet ein nichtparametrischer Ansatz, wie die nichtparametrische Maximum Likelihood-Schätzung Vorteile.

Bei diesem Verfahren, das vorgestellt wird in Aitkin (1995), Aitkin (1996), Aitkin & Aitkin (1996), Aitkin & Francis (1995) und Aitkin & Wilson (1980), wird die Mischverteilung zusammen mit den Modellparametern des Generalisierten Linearen Modells geschätzt. Das primäre Ziel ist jedoch nicht die Mischverteilung zu schätzen, sondern es sollen falsche Schlußfolgerungen aufgrund fehlerhafter Modellvoraussetzungen vermieden werden.

Bei der Parameterschätzung mit der Gauß-Hermite Integration sind die Stützstellen d_j und die zugehörigen Gewichte v_j für eine gewählte Anzahl von Stützstellen bei der gesamten Iteration konstant. Es wird ein gewichtetes Generalisiertes Lineares Modell mit dem linearen Prädiktor $\eta_{itj} = x'_{it} \beta + d_j \sigma$ betrachtet.

An Stelle von $d_j \sigma$ werden bei der nichtparametrischen Maximum Likelihood Schätzung die Stützstellen $\alpha_1, \dots, \alpha_M$ als Parameter in den linearen Prädiktor mit

$$\eta_{itj} = x'_{it} \beta + \alpha_j$$

aufgenommen, so daß α_j der Parameter für den Interzept der j -ten Komponente der Mischverteilung ist.

Die Parameter $\alpha_1, \dots, \alpha_M$ können geschätzt werden, indem Indikatoren mit M Stufen in das Modell aufgenommen werden. Da Identifizierungsprobleme mit dem Interzept von β und einer Komponente α_j entstehen, wird $\alpha_M \equiv 0$ gesetzt. Alternativ ist es auch möglich den Interzept aus dem Modell zu entfernen.

Die Gewichte zu den entsprechenden Stützstellen p_1, \dots, p_M mit $\sum_{j=1}^M p_j = 1$ müssen in jedem Iterationsschritt bestimmt werden. Durch Differenzieren der

Log-Likelihood

$$l(\beta, \alpha_1, \dots, \alpha_M) = \sum_{i=1}^N \ln \sum_{j=1}^M p_j f_{ij} \quad \text{mit } f_{ij} = f(y_i | \beta, \alpha_1, \dots, \alpha_M).$$

nach p_j folgt

$$\frac{\partial l}{\partial p_k} = \sum_{i=1}^N \frac{f_{ij} - f_{iM}}{\sum_{l=1}^M p_l f_{il}} = \sum_{i=1}^N \left\{ \frac{w_{ik}}{p_k} - \frac{w_{iM}}{p_M} \right\}$$

und mit $\frac{\partial l}{\partial p_k} \stackrel{!}{=} 0$ gilt:

$$\hat{p}_j = \sum_{i=1}^N \frac{w_{ij}}{N}.$$

Die Schätzung der Parameter kann dadurch mit folgendem EM-Algorithmus durchgeführt werden:

- Die Anzahl M der Stützstellen wird fest gewählt und schrittweise erhöht bis das Maximum der Likelihood erreicht ist.
- Die Daten des Regressionsmodells werden mit der Anzahl der Stützstellen vervielfacht, indem jede Zeile $y_{it} = x'_{it} \beta$ des Modellansatzes durch

$$\begin{pmatrix} y_{it} \\ \vdots \\ y_{it} \end{pmatrix} = \begin{bmatrix} x'_{it} & 1 & 0 & \dots & 0 \\ \vdots & & \ddots & & \\ \vdots & & & \ddots & \\ x'_{it} & 0 & \dots & 0 & 1 \\ x'_{it} & 0 & \dots & 0 & 0 \end{bmatrix} \begin{pmatrix} \beta \\ z_1 \\ z_2 \\ \vdots \\ z_{M-1} \end{pmatrix}$$

ersetzt wird. Es werden jeweils M Zeilen, aber nur $M - 1$ neue Parameter in das Regressionsmodell aufgenommen, so daß $\alpha_M \equiv 0$ erfüllt ist und keine Identifizierungsprobleme auftreten.

- Startwert für die Iteration ist die Berechnung der Gewichte $w_{ij}^{(0)}$, wobei $\beta^{(0)}$ der Schätzer des Generalisierten Linearen Modells ohne zufällige Effekte ist und die Stützstellen und Gewichte aus den Verfahren mit Gauß-Hermite Integration übernommen werden, so daß $z_j^{(0)} = d_j$ und $p_j^{(0)} = v_j$ für $j = 1, \dots, M$ gilt, d. h. es ist:

$$w_{ij}^{(0)} = \frac{v_j \cdot f(y_i | \beta^{(0)}, d_j)}{\sum_{l=1}^M v_l \cdot f(y_i | \beta^{(0)}, d_l)} .$$

- Jeder Iterationsschritt $k = 0, 1, 2, \dots$ besteht aus den folgenden Schritten:

- Schätzung von $\hat{\gamma}^{(k+1)} = (\hat{\beta}^{(k+1)}, \hat{z}_1^{(k+1)}, \dots, \hat{z}_{M-1}^{(k+1)})'$ durch das mit $w_{ij}(\gamma^{(k+1)})$ gewichtete Regressionsmodell
- Update der posteriori Wahrscheinlichkeit durch

$$p_j^{(k+1)} = \frac{\sum_{i=1}^N w_{ij}^{(k)}}{N}$$

- Berechnung der Gewichte für Beobachtungseinheit i und Stützstelle j mit

$$w_{ij}(\gamma^{(k+1)}) = \frac{p_j^{(k)} \cdot f(y_i | \beta^{(k)}, z_j^{(k)})}{\sum_{l=1}^M p_l^{(k)} \cdot f(y_i | \beta^{(k)}, z_l^{(k)})} .$$

- Die Iteration wird für eine fest gewählte Anzahl von Stützstellen solange durchgeführt, bis ein lokales Maximum für die Likelihood erreicht wurde. Dann wird die Anzahl der Stützstellen erhöht bis kein größerer Wert für das Maximum der Likelihood gefunden wird.

Damit die Designmatrix des Regressionsansatzes vollen Rang hat, wurde eine Parametrisierung des Modellansatzes $\eta_{itj} = x'_{it} \beta + \alpha_j$ durch:

$$\tilde{\eta}_{it} = \begin{pmatrix} \tilde{\eta}_{it1} \\ \vdots \\ \tilde{\eta}_{itM} \end{pmatrix} = \begin{bmatrix} x'_{it} & 1 & 0 & \cdots & 0 \\ \vdots & & \ddots & & \\ \vdots & & & \ddots & \\ x'_{it} & 0 & \cdots & 0 & 1 \\ x'_{it} & 0 & \cdots & 0 & 0 \end{bmatrix} \begin{pmatrix} \tilde{\beta} \\ \tilde{\alpha}_1 \\ \tilde{\alpha}_2 \\ \vdots \\ \tilde{\alpha}_{M-1} \end{pmatrix}$$

mit $\tilde{\alpha}_M \equiv 0$ gewählt.

Unter der Nebenbedingung $\sum_{j=1}^M \alpha_j = 0$ können die ursprünglichen Parameter $\beta = (\beta_0, \beta_1, \dots, \beta_q)$ und $\alpha = (\alpha_1, \dots, \alpha_M)$ mit:

$$\begin{aligned} \beta_0 &= \tilde{\beta}_0 + \frac{1}{M} \sum_{j=1}^M \tilde{\alpha}_j \\ \beta_j &= \tilde{\beta}_j && \text{für } j = 1, \dots, q \\ \alpha_j &= \tilde{\beta}_0 - \beta_0 + \tilde{\alpha}_j && \text{für } j = 1, \dots, M \end{aligned}$$

reparametrisiert werden.

2.2 Modellansätze mit (zeit-)variierenden Koeffizienten

Beim Modellansatz mit (zeit-)variierenden Koeffizienten betrachten wir den zeitabhängigen Regressionsparameter $\beta(t)$ und setzen voraus, daß die Daten in der Form (y_i, x_i, t) vorliegen. Das heißt, daß neben dem Response- und Kovariablenvektor auch noch der Beobachtungszeitpunkt t für jede Beobachtung bekannt ist. Im allgemeinen gehen wir davon aus, daß für die i -te Beobachtungseinheit mehrere Beobachtungszeitpunkte bzw. Meßwiederholungen zu verschiedenen Zeitpunkten vorliegen.

Weiter wird angenommen, daß die Menge der beobachteten Zeitpunkte diskret und endlich ist, so daß $t \in \{1, \dots, T\}$ vorausgesetzt werden kann.

Wir verwenden die Daten (y_i, x_i, t) in der Notation (y_{it}, x_{it}) mit $i = 1, \dots, N$ und $t \in \{1, \dots, T\}$, wobei N die Anzahl der Beobachtungseinheiten ist.

Liegt ein binomiales oder multinomiales Modell mit gruppierten Daten vor, so wird vorausgesetzt, daß die Daten in der Form (y_{it}, x_{it}, n_{it}) mit $i = 1, \dots, N$ und $t \in \{1, \dots, T\}$ vorliegen. Dabei bezeichnet N die Anzahl der Gruppen, n_{it} die Anzahl der Beobachtungen, x_{it} den Kovariablenvektor und y_{it} den Gruppendurchschnitt jeweils für die Gruppe i zum Zeitpunkt t . Die Anzahl der Beobachtungen ist $N^* = \sum_i \sum_t n_{it}$.

Im weiteren unterscheiden wir nicht zwischen gruppierten und ungruppierten Daten. Wir setzen $n_{it} = 1$ für ungruppierte Daten und sprechen allgemein von der i -ten Beobachtungseinheit.

Es wurde nicht vorausgesetzt, daß zu einer Beobachtungseinheit i für alle Zeitpunkte $t = 1, \dots, T$ Beobachtungen vorliegen. Trotzdem werden wir die Summation $\sum_{i=1}^N \sum_{t=1}^T$ verwenden, indem wir die entsprechenden Summanden mit 0 definieren, wenn für eine Beobachtungseinheit i zum Zeitpunkt t

keine Beobachtung vorliegt.

Um Zeitabhängigkeit zu modellieren, betrachten wir den Regressionskoeffizienten β als eine Funktion der Zeit t , d. h. $\beta' = \beta'(t) = (\beta_1(t), \dots, \beta_p(t))$. Die in dieser Arbeit verwendeten Ansätze für $\beta_j(t)$ ($j \in 1, \dots, p$) werden im folgenden kurz vorgestellt. Dabei gehen wir immer davon aus, daß Zeitabhängigkeit vorliegt, auch wenn mit den Modellen allgemeinere Abhängigkeitsstrukturen modelliert werden können.

2.2.1 Parametrische und semi-parametrische Ansätze

Polynomiale Regression

Bei der Polynomialen Regression wird die Zeitabhängigkeit des Regressionskoeffizienten $\beta_j(t)$ durch ein Polynom von t mit fest vorgegebenem Grad s^* modelliert:

$$\beta_j(t) = \sum_{r=0}^{s^*} \gamma_{j,r} t^r .$$

Regressions Spline

Bei diesem Ansatz (Hastie & Tibshirani, 1990; Fahrmeir, Hamerle & Tutz, 1994) wird der obige Modellansatz verändert indem der Parameter $\beta_j(t)$ stückweise durch Polynome definiert wird. Der Spline ist an den Bruchstellen, den sogenannten Knoten stetig differenzierbar.

Eine mögliche Anwendung ist der stückweise-kubische Spline Efron (1988). Dabei wird $\beta_j(t)$ aus $s+1$ ($s \in \mathbb{N}$) kubischen Funktionen zusammengesetzt, so daß die Funktion auch an den Knoten ξ_1, \dots, ξ_s , zweimal stetig differenzierbar

ist:

$$\beta_j(t) = \gamma_{j0} + \gamma_{j1} t + \gamma_{j2} t^2 + \gamma_{j3} t^3 + \sum_{r=1}^s \gamma_{jr+3} (t - \xi_r)_+^3,$$

mit $(t - \xi_r)_+ = \max\{0, t - \xi_r\}$.

Liegen die Beobachtungen teilweise spärlich vor, beispielsweise ist dies in Verweildauermodellen gegen Ende des Beobachtungszeitraumes der Fall, so wird häufig ein kubisch-linearer Regressions Spline verwendet:

$$\beta_j(t) = \gamma_{j0} + \gamma_{j1} t + \gamma_{j2} (t - \xi)_-^2 + \gamma_{j3} (t - \xi)_-$$

mit $(t - \xi)_- = \min\{0, t - \xi\}$.

Für $t < \xi$ hat die Funktion $\beta_j(t)$ einen kubischen und für $t > \xi$ einen linearen Verlauf. Die Funktion $\beta_j(t)$ ist für alle t , insbesondere für $t = \xi$ stetig differenzierbar.

Neben diesen Regressions Splines ist es auch möglich, mehrere kubische Funktionen und eine lineare Funktion stückweise zu definieren, so daß die Funktion $\beta_j(t)$ für $t > \xi^*$ aus einer Geraden und für $t < \xi^*$ aus einem stückweise-kubischen Spline besteht, es ist:

$$\beta_j(t) = \gamma_{j0} + \gamma_{j1} t + \gamma_{j2} (t - \xi^*)_-^2 + \gamma_{j3} (t - \xi^*)_-^3 + \sum_{r=1}^s \gamma_{jr+3} (t - \xi_r)_-^3 .$$

Mit den Knoten $\xi_r < \xi_{r-1} < \dots, \xi_1 < \xi^*$ ist $\beta_j(t)$ für $t \geq \xi^*$ einmal und für $t < \xi^*$ zweimal stetig differenzierbar.

Die in diesem Abschnitt vorgeschlagenen Modelle sind sehr einfach zu implementieren. Nach Anpassung der Designmatrix können Standardverfahren zur Schätzung der Koeffizienten γ_{jr} verwendet werden um $\beta_j(t)$ zu bestimmen.

Um die Interpretation der variierenden Koeffizienten zu erleichtern und um numerische Schätzprobleme zu vermeiden, muß darauf geachtet werden, daß der Grad des Polynoms bzw. die Anzahl der Knoten der vorhandenen Datenstruktur gerecht wird. Das heißt die Anzahl der Spalten in der Designmatrix sollte möglichst klein sein, um die Struktur der zeitlichen Abhängigkeit zu erkennen und um eine Überparametrisierung zu vermeiden

2.2.2 Lokaler Likelihoodansatz

Dieser Modellansatz, der auch in Kauermann & Tutz (1995), Kauermann & Tutz (2000), Tutz & Kauermann (1995) und Tutz & Kauermann (1997) behandelt wird, ist ein spezielles variierendes Koeffizientenmodell, wie es in Hastie & Tibshirani (1993) vorgestellt wird. Die Modellparameter variieren in Abhängigkeit einer erhobenen Variablen, dem sogenannten Effektmodifizierer. In unserem Ansatz entspricht der Effektmodifizierer der Zeit.

Zu jedem Zeitpunkt $t \in 1, \dots, T$ wird $\beta(t) = (\beta_1(t), \dots, \beta_p(t))'$ durch ein separates Generalisiertes Lineares Modell geschätzt. Dabei gehen die Beobachtungen (y_{is}, x_{is}) mit $s \in \{1, \dots, T\}$ gewichtet, in Abhängigkeit zu dem beobachteten Zeitpunkt s , in die Schätzung ein. Die dabei verwendete Gewichtsfunktion $\omega_\gamma(t, s)$ ist unter Verwendung einer Kernfunktion $K(u)$ in Abhängigkeit des Glättungsparameters γ durch

$$\omega_\gamma(t, s) = c_t \cdot K\left(\frac{t-s}{\gamma}\right)$$

definiert.

Als Kernfunktionen werden verschiedene unimodale, symmetrische Funk-

tionen verwendet. Beispielsweise können der Epanechnikov-Kern

$$K(u) = \begin{cases} \frac{3}{4}(1 - u^2) & \text{für } |u| \leq 1 \\ 0 & \text{sonst} \end{cases}$$

oder der Gauß-Kern

$$K(u) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}u^2\right).$$

angewendet werden.

Mit der Wahl der Normierungskonstanten $c_t = K(0)^{-1}$ gilt:

$$\begin{aligned} \gamma \rightarrow \infty : \quad \omega_\gamma(t, s) &= 1 && \text{für } s = 1, \dots, T \\ \gamma \rightarrow 0 : \quad \omega_\gamma(t, s) &= \begin{cases} 1 & \text{für } s = t \\ 0 & \text{für } s \neq t \end{cases} \end{aligned}$$

d. h. im Lokalen Likelihoodansatz sind auch zwei Spezialfälle enthalten. Für $\gamma \rightarrow \infty$ erhalten alle Beobachtungen unabhängig von Beobachtungszeitpunkt das Gewicht 1. Die Parameterschätzung mit dem Lokalen Likelihoodansatz entspricht der Schätzung mit dem Modellansatz bei dem keine Zeitabhängigkeit für β berücksichtigt wird. Für $\gamma \rightarrow 0$ gehen in die Parameterschätzung von $\beta(t)$ nur die Beobachtungen zum Zeitpunkt t ein.

Durch die Wahl des Glättungsparameters können somit sehr flexibel die unterschiedlichen Gewichte festgelegt werden. In Abbildung 2.1 werden die Gewichte $\omega_\gamma(t, s)$ für $t = 10$ und $s = (1, 2, \dots, 21)$ in Abhängigkeit des Glättungsparameters γ unter Verwendung des Gauß-Kerns dargestellt.

Liegt ein Verweildauermodell vor, bei dem gegen Ende des Beobachtungszeitraumes nur noch wenige Beobachtungen vorliegen, so kann der Glättungsparameter γ in Abhängigkeit der Anzahl der vorliegenden Beobachtungen beispielsweise mit $\gamma_t = \frac{1}{\sqrt{n_t}}\gamma$ gewählt werden.

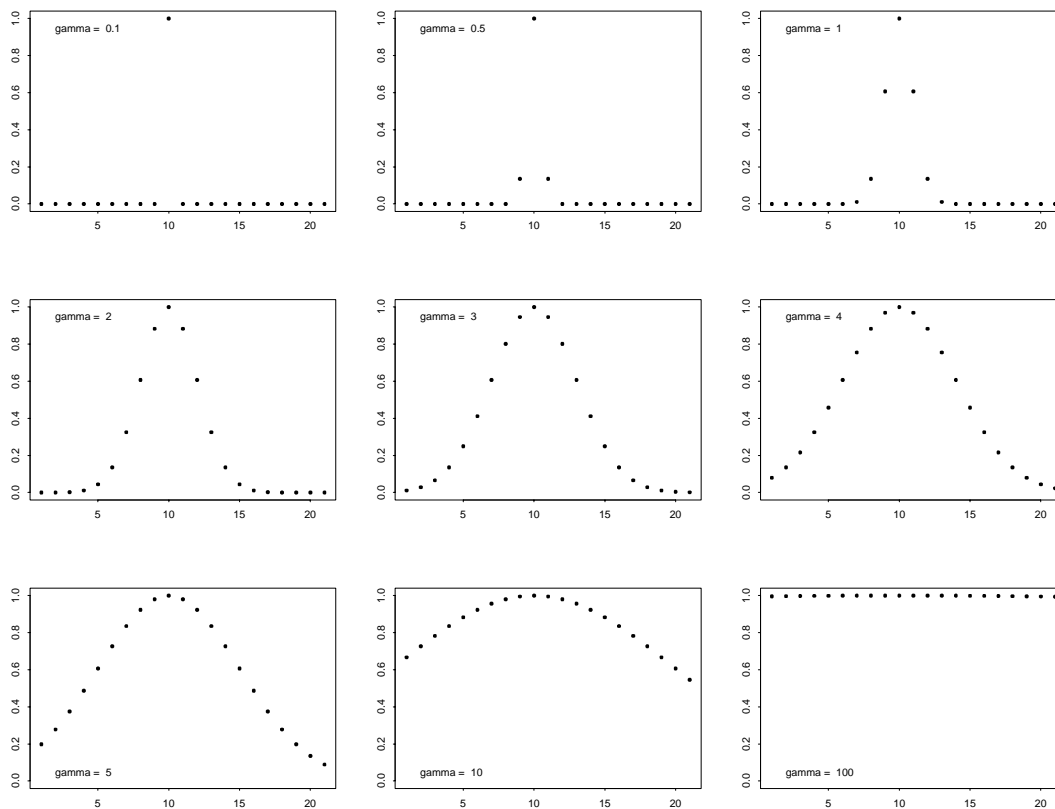


Abbildung 2.1: Gewichte des Gauß-Kerns in Abhängigkeit des Glättungsparameters

Die Wahl des Glättungsparameters γ kann durch Kreuzvalidierung, die beispielsweise in Rice & Silverman (1991), Fan, Hall, Martin & Patil (1996) und Tutz (1989)) behandelt wird, erfolgen. Dabei wird der Wert von γ gewählt bei dem $CV(\gamma)$ mit

$$CV(\gamma) = \frac{1}{N^*} \sum_{i=1}^N \sum_{t=1}^T l(y_{it}, \hat{\mu}_{it}^{-(i,t)})$$

minimal ist.

Dabei ist $\hat{\mu}_{it}^{-(i,t)}$ der geschätzte Response, gegeben x_{it} , der durch den Schätzer

$\widehat{\beta(t)}^{-(i,t)}$ bestimmt wird. Der Schätzer $\widehat{\beta(t)}^{-(i,t)}$ ist derjenige Schätzer zum Zeitpunkt $t \in \{1, \dots, T\}$ des Regressionskoeffizienten, bei dem alle Beobachtungen bis auf die Beobachtung (y_{it}, x_{it}) in das Modell eingehen.

Die Anzahl aller Beobachtungen ist

$$N^* = \sum_{i=1}^N n_i .$$

Weiter ist $l(\cdot)$ eine Verlustfunktion, beispielsweise werden die folgenden Verlustfunktionen verwendet:

- Quadratische Verlustfunktion

$$l(y_{it}, \hat{\mu}_{it}^{-(i,t)}) = (y_{it} - \hat{\mu}_{it}^{-(i,t)})^2$$

- Kullback-Leibler Verlustfunktion

$$l(y_{it}, \hat{\mu}_{it}^{-(i,t)}) = y_{it} \ln \left\{ \frac{y_{it}}{\hat{\mu}_{it}^{-(i,t)}} \right\}$$

- Minimale Pearson-Residuen

$$l(y_{it}, \hat{\mu}_{it}^{-(i,t)}) = \frac{(y_{it} - \hat{\mu}_{it}^{-(i,t)})^2}{v(\hat{\mu}_{it}^{-(i,t)})},$$

wobei $v(\hat{\mu}_{it}^{-(i,t)})$ die Varianzfunktion ist. Im binären Modell gilt beispielsweise $v(\hat{\mu}_{it}^{-(i,t)}) = \hat{\mu}_{it}^{-(i,t)} (1 - \hat{\mu}_{it}^{-(i,t)})$ und im Poissonmodell $v(\hat{\mu}_{it}^{-(i,t)}) = \hat{\mu}_{it}^{-(i,t)}$.

Da in der verwendeten Datenstruktur (y_{it}, x_{it}) korrelierte Daten vorliegen und deshalb keine Unabhängigkeit der Residuen angenommen werden kann, sollte dies bei Kreuzvalidierung berücksichtigt werden. Verfahren für die Bestimmung der Kreuzvalidierung bei abhängigen Beobachtungen werden in van der Linde (1994) vorgestellt.

Kapitel 3

Zufällige Effekte und variierende Koeffizienten im Generalisierten Linearen Modell

Im vorherigen Kapitel wurde die Modellierung mit variierenden Koeffizienten bzw. mit zufälligen Effekten vorgestellt. Beide Ansätze für Modell-erweiterungen im Generalisierten Linearen Modell werden in diesem Kapitel miteinander kombiniert.

Die Ausgangssituation bei der Verwendung von zeitvariierenden Koeffizientenmodellen ist, daß im Zeitverlauf meist die gleichen Beobachtungseinheiten oder -gruppen betrachtet werden. Auf Grund der Meßwiederholungen oder der vorhandenen Gruppenstruktur, welche beispielsweise die familiäre, regionale oder administrative Zugehörigkeit der einzelnen Beobachtungseinheit wiedergibt, ist die Voraussetzung der homogenen Varianzstruktur nicht erfüllt. Werden diesen Modellen zufällige Effekte hinzugefügt,

so wird nicht nur die Gruppenstruktur bzw. die Struktur der Meßwiederholungen berücksichtigt, sondern dieser Modellansatz ermöglicht auch den gruppen- bzw. individuenspezifischen Einfluß der einzelnen Beobachtungseinheit zu quantifizieren.

Werden die Daten für ein Generalisiertes Lineares Modell mit zufälligen Effekten in einem Zeitverlauf erhoben, so kann nicht vorausgesetzt werden, daß der Einfluß der Koeffizienten zeitunabhängig ist. Um den zeitspezifischen Einfluß der Kovariablen zu berücksichtigen, können zeitvariierende Koeffizienten in das Modell aufgenommen werden.

Es ist somit nur natürlich, sowohl den Modellansatz mit zufälligen Effekten durch die Aufnahme von zeitvariierenden Koeffizienten zu erweitern, als auch beim Modellansatz mit zeitvariierenden Koeffizienten zusätzlich zufällige Effekte zu betrachten.

3.1 Modellansätze und Parameterschätzungen

Bei der Betrachtung eines Generalisierten Linearen Modells, das sowohl zeitvariierende Koeffizienten als auch zufällige Effekte berücksichtigt, setzen wir folgende Datenstruktur voraus.

Die Daten liegen in der Form (y_{it}, x_{it}) , für $i = 1, \dots, N$ Beobachtungseinheiten bzw. -gruppen zu den entsprechenden Zeitpunkten $t \in \{1, \dots, T\}$ vor.

Bei der Modellierung wird im folgenden vorausgesetzt, daß der Regressionskoeffizient $\beta = (\beta(1), \dots, \beta(T))$ den zeitvariierenden und die zufälligen Effekte b_1, \dots, b_N den individuen-, gruppen- oder auch umweltspezifischen Einfluß auf die Zielgröße beschreiben.

3.1.1 Volles Modell, parametrische und semi-parametrische Ansätze

Ist die Anzahl der betrachteten Zeitpunkte gering, so kann, wenn keine zufälligen Effekte berücksichtigt werden, separat für jeden betrachteten Zeitpunkt ein Modell geschätzt werden. Dabei werden immer die Beobachtungen betrachtet, die zu dem zugehörigen Zeitpunkt erhoben wurden. Unter Berücksichtigung von zufälligen Effekten kann mit diesem Ansatz das folgende Modell betrachtet werden.

Ist \mathbb{O} die Matrix, deren Koeffizienten alle 0 sind, so lautet der lineare Prädiktor im vollen Modell

$$\eta_{it} = \tilde{Z}_{it}\beta + W_{it}b_i$$

mit $\tilde{Z}_{it} = (\mathbb{O}, \dots, Z_{it}, \dots, \mathbb{O})$ und $\beta = (\beta(1), \dots, \beta(T))$ sowie $b_i \stackrel{iid}{\sim} \mathcal{N}(0, Q)$.

Bei diesem Modellansatz kann die Schätzung der Parameter durch den vorgestellten EM-Algorithmus mit direkter Gauß-Hermite Integration ohne weitere Anpassungen durchgeführt werden.

Die im vorherigen Kapitel vorgestellten parametrischen und semi-parametrischen Modelle lassen sich ebenso einfach erweitern. Der lineare Prädiktor $\eta_{it} = Z_{it}\beta$ wird durch die Aufnahme der zufälligen Effekte durch

$$\eta_{it} = Z_{it}\beta(t) + W_{it}b_i \text{ mit } b_i \stackrel{iid}{\sim} \mathcal{N}(0, Q)$$

ersetzt. Wobei für $\beta(t)$ eine der in 2.2.1 vorgestellten Parametrisierungen gewählt werden kann.

Mit zunehmender Anzahl der zu betrachteten Zeitpunkte wird die Anzahl der zu schätzenden Parameter meist so groß, daß numerische Schätzprobleme bei der Bestimmung der Parameterschätzungen auftreten. Deshalb können die oben vorgestellten Modelle nur sehr eingeschränkt verwendet werden. Der folgende Lokale Likelihoodansatz unter Berücksichtigung von zufälligen Effekten ist flexibler, da durch die Verwendung der Kernfunktion mit der Wahl des Glättungsparameters eine beliebige zeitliche Glättung erfolgen kann.

3.1.2 Lokaler Likelihoodansatz mit zufälligen Effekten

Im folgenden stellen wir zwei Algorithmen vor, die sowohl zufällige Effekte als auch zeitvariierende Koeffizienten, die mit einem Lokalem Likelihoodansatz geschätzt werden, berücksichtigen. Für die Berücksichtigung der zufälligen Effekte wird zuerst der EM-Algorithmus mit Gauß-Hermite Integrationstechnik verwendet und im zweiten vorgestellten Algorithmus der EM-Type verwendet.

Der lineare Prädiktor des Modellansatzes lautet in beiden Varianten:

$$Z_{it}\beta + W_{it}b_i \text{ mit } b_i \stackrel{iid}{\sim} \mathcal{N}(0, Q) \text{ und } \beta = (\beta(1), \dots, \beta(T))$$

Für die zufälligen Effekte b_1, \dots, b_N , die den individuen- oder gruppenspezifischen Einfluß beschreiben, wird keine zeitliche Abhängigkeit vorausgesetzt. Das heißt einerseits, daß die zufälligen Effekte b_1, \dots, b_N unabhängig vom betrachteten Zeitpunkt $t \in 1, \dots, T$ sind. Andererseits muß bei der Schätzung des zeitvariierenden Regessionskoeffizienten $\beta = (\beta(1), \dots, \beta(T))$ der Einfluß der zufälligen Effekte b_1, \dots, b_N für jeden Zeitpunkt $t \in 1, \dots, T$ gleichbleibend berücksichtigt werden.

EM-Algorithmus mit Gauß-Hermite Integrationstechnik

In dem zuerst vorgeschlagenen Algorithmus wird ein iteratives Verfahren verwendet, das den zeitvariierenden Koeffizienten β schätzt, indem die zufälligen Effekte konstant gehalten werden, d. h. $W_{it}b_i$ geht als OFFSET in die Modellgleichung ein. Die Schätzung des Parametervektors $\beta = (\beta(1), \dots, \beta(T))$ kann mit dem Lokalen Likelihoodansatz erfolgen.

Anschließend wird die Kovarianzmatrix Q geschätzt, wobei $Z_{it}\beta$ konstant gehalten wird. Der im vorherigen Kapitel vorgestellte EM-Algorithmus mit Gauß-Hermite Integration kann ohne weitere Anpassungen übernommen werden.

Da nun sowohl β als Q bekannt sind, können die zufälligen Effekte b_1, \dots, b_N bestimmt werden.

Dieses Verfahren muß iterativ angewendet werden und es ergibt sich folgender Ablauf:

- Um Startwerte für die Iteration zu erhalten, wird das Modell mit zufälligen Effekten ohne variierenden Koeffizienten verwendet.

Der lineare Prädiktor dieses Modells hat die Form

$$\eta_{it} = Z_{it} \beta + W_{it} b_i \quad \text{mit} \quad b_i \stackrel{iid}{\sim} \mathcal{N}(0, Q).$$

- Zuerst werden $\hat{\beta}$ und \hat{Q} geschätzt.
- Anschließend erfolgt die Bayes-Schätzung der zufälligen Effekte $\hat{b}_1, \dots, \hat{b}_N$.

- Unter Verwendung der Startwerte $\hat{b}_i^{(0)} := \hat{b}_i$ für $i = 1, \dots, N$ wird die folgende Iteration für $p = 0, 1, \dots$ durchgeführt:

- Die zeitvariierenden Koeffizienten $\hat{\beta}(t)^{\langle p+1 \rangle}$ werden durch den Lokalen Likelihoodansatz geschätzt, bei dem die zufälligen Effekte $W_{it} \hat{b}_i^{\langle p \rangle}$ konstant gehalten werden. Das heißt es gilt

$$\eta_{it} = Z_{it} \beta(t)^{\langle p+1 \rangle} + OFFSET \left(W_{it} \hat{b}_i^{\langle p \rangle} \right).$$

- Die Aktualisierung der zufälligen Effekte $\hat{b}_i^{\langle p+1 \rangle}$ für $i = 1, \dots, N$ erfolgt in zwei Schritten:

Zuerst wird der Schätzer $\hat{Q}^{\langle p+1 \rangle}$ durch ein Modell mit zufälligen Effekten, bei dem die zeitvariierenden Koeffizienten $\hat{\beta}(t)^{\langle p+1 \rangle}$ Teil des OFFSETS sind, bestimmt. Es wird das Modell

$$\eta_{it} = OFFSET \left(Z_{it} \hat{\beta}(t)^{\langle p+1 \rangle} \right) + W_{it} \hat{b}_i^{\langle p \rangle}$$

mit $b_i \sim \mathcal{N}(0, Q^{\langle p+1 \rangle})$ betrachtet und anschließend erfolgt die Aktualisierung der zufälligen Effekte $\hat{b}_i^{\langle p+1 \rangle}$ für $i = 1, \dots, N$.

- Die Iteration wird beendet, falls $\frac{\|\hat{\beta}(t)^{\langle p+1 \rangle} - \hat{\beta}(t)^{\langle p \rangle}\|}{\|\hat{\beta}(t)^{\langle p \rangle}\|} < \epsilon$ gilt.

EM-Type Algorithmus

Da beim EM-Type Algorithmus eine simultane Schätzung des Parametervektors $\beta = (\beta(1), \dots, \beta(T))$ und der zufälligen Effekte b_i , $i = 1, \dots, N$ mit $b_i \stackrel{iid}{\sim} \mathcal{N}(0, Q)$ durchgeführt wird, können zeitvariierenden Koeffizienten problemlos berücksichtigt werden.

Ist $Q^{(p)}$ aus den vorherigen Iterationsschritt bekannt, so gilt für die Likelihood mit $\delta' = (\beta'(1), \dots, \beta'(T), b_1, \dots, b_N)$:

$$\begin{aligned} l(\delta) &= \sum_{i=1}^N \ln f(y_i | b_i, \beta) - \frac{1}{2} \sum_{i=1}^N b_i' (Q^{(p)})^{-1} b_i \\ &= \sum_{i=1}^N \sum_{s=1}^T w_\gamma(t, s) \ln f(y_{is} | b_i, \beta(t)) - \frac{1}{2} \sum_{i=1}^N b_i' (Q^{(p)})^{-1} b_i \end{aligned}$$

und für die Scorefunktion folgt:

$$\begin{aligned} \frac{\partial l(\delta)}{\partial \beta(t)} &= \sum_{i=1}^N \sum_{s=1}^T w_\gamma(t, s) Z'_{is} D_{is}(\beta(t), b_i) \Sigma_{is}^{-1}(\beta(t), b_i) (y_{is} - \mu_{is}(\beta(t), b_i)) \\ \frac{\partial l(\delta)}{\partial b_i} &= \sum_{t=1}^T W'_{it} D_{it}(\beta(t), b_i) \Sigma_{it}^{-1}(\beta(t), b_i) (y_{it} - \mu_{it}(\beta(t), b_i)) - (Q^{(p)})^{-1} b_i \end{aligned}$$

Damit hat die Fishermatrix die folgende partitionierte Darstellung:

$$F(\delta)^{\langle p \rangle} = \begin{bmatrix} F_{\beta^1 \beta^1} & & \mathbb{O} & F_{\beta^1 b_1} & \cdots & \cdots & F_{\beta^1 b_N} \\ & \ddots & & \vdots & & & \vdots \\ \mathbb{O} & & F_{\beta^T \beta^T} & F_{\beta^T b_1} & \cdots & \cdots & F_{\beta^T b_N} \\ F_{b_1 \beta^1} & \cdots & F_{b_1 \beta^T} & F_{b_1 b_1} & & & \\ \vdots & & \vdots & & \ddots & & \mathbb{O} \\ \vdots & & \vdots & \mathbb{O} & & \ddots & \\ F_{b_N \beta^1} & \cdots & F_{b_N \beta^T} & & & & F_{b_n b_n} \end{bmatrix}$$

mit:

$$\begin{aligned} F_{\beta^t \beta^t} &= \sum_{i=1}^N \sum_{s=1}^T w_\gamma(t, s) Z'_{is} D_{is}(\beta(t), b_i) \Sigma_{is}^{-1}(\beta(t), b_i) D'_{is}(\beta(t), b_i) Z_{is} \\ F_{\beta^t b_i} &= \sum_{s=1}^T w_\gamma(t, s) Z'_{is} D_{is}(\beta(t), b_i) \Sigma_{is}^{-1}(\beta(t), b_i) D'_{is}(\beta(t), b_i) W_{is} \\ F_{b_i \beta^t} &= F'_{\beta^t b_i} \\ F_{b_i b_i} &= \sum_{t=1}^T W'_{it} D_{it}(\beta(t), b_i) \Sigma_{it}^{-1}(\beta(t), b_i) D'_{it}(\beta(t), b_i) + (Q^{\langle p \rangle})^{-1} \end{aligned}$$

Diese Blockgestalt der Fishermatrix kann, wie im Anhang in Abschnitt 6.1.2 gezeigt wird, für eine Vereinfachung bei der Berechnung der Schätzer ausgenutzt werden.

3.2 Varianz der Parameterschätzungen

Um inhaltlich die geschätzten Parameterwerte interpretieren zu können, muß neben der Schätzung der Parameterwerte vor allem die Varianz dieser Schätzwerte berücksichtigt werden. Im folgenden betrachten wir die Schätzungen der Varianzen der Parameterschätzungen für die oben vorgeschlagenen Modelle.

Im Generalisierten Linearen Modell ohne zufällige Effekte und ohne variierende Koeffizienten gilt für den geschätzten Parametervektor $\hat{\beta}$:

$$\hat{\beta} \sim \mathcal{N}(\beta, F^{-1}(\hat{\beta})).$$

Das heißt, die Schätzung der Kovarianz kann durch $\widehat{cov}(\hat{\beta}) = F^{-1}(\hat{\beta})$ erfolgen, wobei $F^{-1}(\hat{\beta})$ die inverse Fishermatrix bezeichnet. Bei diesem Modellansatz liegen Beobachtungen (y_i, x_i) mit $i = 1, \dots, N$ vor und der lineare Prädiktor lautet $\eta_i = Z_i\beta$, so daß für die Fishermatrix

$$F(\beta) = \sum_{i=1}^N F_i(\beta) = \sum_{i=1}^N Z_i' D_i \Sigma_i^{-1} D_i' Z_i,$$

mit $D_i = D_i(\beta) = \frac{\partial h(\eta_i)}{\partial \eta}$ und $\Sigma_i = \Sigma_i(\beta) = cov(y_i|\beta)$ gilt.

3.2.1 Varianzschätzungen bei Modellen mit zufälligen Effekten

Im Generalisierten Linearen Modell mit zufälligen Effekten sind im linearen Prädiktor neben den geschätzten festen Parametern β und Q auch die zufälligen Effekte b_i für $i = 1, \dots, N$ enthalten.

EM-Algorithmus mit Gauß-Hermite Integrationstechnik

Beim EM-Algorithmus mit Gauß-Hermite Integrationstechnik kann wegen der verwendeten Mischungsdichte für die Varianzschätzung nicht die Fisher-matrix aus dem letzten Iterationsschritt verwendet werden. In Fahrmeir & Tutz (1994) und in Gourieroux & Montfort (1989) wird vorgeschlagen die Fisher-matrix mit Hilfe der Scorefunktion zu approximieren.

Im Anhang in Abschnitt 6.1.2 wird auf die Parameterschätzung für den Modellansatz mit dem linearen Prädiktor

$$\eta_{it} = Z_{it}\beta + W_{it}b_i \quad \text{mit} \quad b_i \stackrel{iid}{\sim} \mathcal{N}(0, Q)$$

eingegangen. Dieser Modellansatz wird mit $b_i = Q^{1/2}a_i$ und $\theta = \text{vec}(Q^{1/2})$ sowie $\alpha' = (\beta', \theta')$ mit $\beta = (\beta(1), \dots, \beta(T))$ unter Verwendung des Kronecker Produktes \otimes und der Cholesky-Zerlegung von Q mit $Q = Q^{1/2}Q^{T/2}$ umparametrisiert zu

$$\eta_{it} = [Z_{it} \quad , \quad a_i' \otimes W_{it}] \alpha \quad \text{mit} \quad a_i \stackrel{iid}{\sim} \mathcal{N}(0, I).$$

Ist M die Anzahl der Stützstellen und $c_{ij}^{GH} = c_{ij}^{GH}(\alpha)$ die Gewichte der i -ten Beobachtungseinheit für Stützstelle j der Gauß-Hermite Integration, so die Scorefunktion der i -ten Beobachtungseinheit mit

$$s_i(\alpha) \approx \sum_{j=1}^M c_{ij}^{GH} \frac{\partial \ln f(y_i | d_j, \alpha)}{\partial \alpha} = \sum_{t=1}^T \sum_{j=1}^M c_{ij}^{GH} \frac{\partial \ln f(y_{it} | d_j, \alpha)}{\partial \alpha}$$

approximiert werden. Die Kovarianz des Parametervektors $\hat{\alpha}' = (\hat{\beta}', \hat{\theta}')$ kann somit durch

$$\widehat{\text{cov}}(\hat{\alpha}) \approx \widehat{\text{cov}}_{SC}(\hat{\alpha}) := (F^{SC}(\hat{\alpha}))^{-1} \quad \text{mit} \quad F^{SC}(\hat{\alpha}) = \sum_{i=1}^N s_i(\hat{\alpha}) \cdot s_i'(\hat{\alpha})$$

geschätzt werden.

Mit diesem Ansatz läßt sich die Scorefunktion unter Verwendung des linearen Prädiktors $\eta_{itj} = [Z_{it}, d'_j \otimes W_{it}] \alpha$ durch

$$s_i^{GH}(\alpha) = \sum_{t=1}^T \sum_{j=1}^M c_{ij}^{GH}(\alpha) [Z_{it}, d'_j \otimes W_{it}]' D_{itj} \Sigma_{itj}^{-1} (y_{it} - \mu_{itj})$$

mit $D_{itj} = \frac{\partial h(\eta_{itj})}{\partial \eta}$, $\mu_{itj} = h(\eta_{itj})$ und $\Sigma_{itj} = cov(y_{it}|d_j, \alpha)$ berechnen.

Volles Modell Für das volle Modell kann der Ansatz ohne Veränderungen übernommen werden. Es wurde für die die Schätzung von $\alpha' = (\beta', \theta')$ mit $\beta = (\beta(1), \dots, \beta(T))$ ein Generalisiertes Lineares Modell mit zufälligen Effekten und dem folgenden linearem Prädiktor

$$\mu_{it} = h(\tilde{Z}_{it} \beta + b_i) \text{ mit } \tilde{Z}_{it} = (\mathbb{O}, \dots, Z_{it}, \dots, \mathbb{O}),$$

verwendet. Die Scorefunktion ist somit approximiert durch

$$\begin{aligned} s_i(\alpha) &\approx s_i^{GH}(\alpha) = \sum_{j=1}^M c_{ij}^{GH}(\alpha) \frac{\partial \ln f(y_i|d_j; \alpha)}{\partial \alpha} \\ &= \sum_{t=1}^T \sum_{j=1}^M c_{ij}^{GH}(\alpha) [\tilde{Z}_{it}, d'_j \otimes W_{it}]' D_{itj} \Sigma_{itj}^{-1} (y_{it} - \mu_{itj}) \end{aligned}$$

und dabei ist

$$\begin{aligned} D_{itj} &= \frac{\partial h(\eta_{itj})}{\partial \eta} \\ \Sigma_{itj} &= cov(y_{it}|d_j, \alpha) \\ \mu_{itj} &= h(\eta_{itj}) \text{ mit } \eta_{itj} = [\tilde{Z}_{it}, d'_j \otimes W_{it}] \begin{pmatrix} \beta(1) \\ \vdots \\ \beta(T) \\ \theta \end{pmatrix}. \end{aligned}$$

Delta-Methode Die Delta-Methode wird verwendet, um die Schätzung der Varianzen der Parameterschätzungen im Modellen mit parametrischen und semi-parametrischen zeitvariierenden Koeffizienten zu berechnen.

Wurde beispielsweise $\beta_j(t)$ parametrisiert als ein Polynom von t mit Grad s^* (vgl. Abschnitt 2.2.1), so gilt $\beta_j(t) = \sum_{r=0}^{s^*} \gamma_{jr} t^r$ und für die Kovarianz von $cov(\beta_j(t))$ folgt

$$cov(\beta_j(t)) = t cov(\gamma_j) t'.$$

Zudem wird die Delta-Methode auch verwendet, um die Konfidenzbänder für die Schätzung der Varianz $\widehat{var}(\hat{\mu})$ mit $\hat{\mu} = h(\hat{\eta}) = h(Z \hat{\beta})$ zu bestimmen. Mit $D(\hat{\beta}) = \frac{\partial h(\hat{\eta})}{\partial \hat{\eta}}$ folgt

$$\widehat{var}(\hat{\mu}) = D(\hat{\beta})' Z \widehat{cov}(\hat{\beta}) Z' D(\hat{\beta}).$$

Lokaler Likelihoodansatz Bei der Parameterschätzung durch den Lokalen Likelihoodansatz bei dem die zufälligen Effekte durch Gauß-Hermite Integrationstechnik geschätzt werden, erfolgt die Schätzung von $\beta(t)$ lokal für jeden betrachteten Zeitpunkt $t \in 1, \dots, T$. Dagegen wird die Schätzung von θ global, d. h. unabhängig von t durchgeführt.

Beim Verfahren für die Berechnungen der Varianzen der Parameterschätzungen muß deshalb folgende Unterscheidung getroffen werden:

- Der Schätzung von $\beta(t)$ liegt ein Generalisiertes Lineares Modell mit zufälligen Effekten zugrunde. Das heißt, die Approximation der lokalen Scorefunktion erfolgt mit

$$\begin{aligned} s_i(\beta(t), \theta) &\approx \sum_{s=1}^T \sum_{j=1}^M c_{ij}^{GH}(\beta(t), \theta) w_\gamma(s, t) \frac{\partial \ln f(y_{is} | d_j, \beta(t), \theta)}{\partial \beta(t)} \\ &= \sum_{s=1}^T \sum_{j=1}^M c_{ij}^{GH}(\beta(t), \theta) w_\gamma(s, t) [Z_{is}, d_j' \otimes W_{is}]' D_{isj} \Sigma_{isj}^{-1} (y_{is} - \mu_{isj}), \end{aligned}$$

wobei $w_\gamma(s, t)$ die Gewichte der gewählten Kernfunktion bezeichnen und

$$\begin{aligned} D_{isj} &= \frac{\partial h(\eta_{isj})}{\partial \eta} \\ \Sigma_{isj} &= \text{cov}(y_{is} | d_j, \beta(t), \theta) \\ \mu_{isj} &= h(\eta_{isj}) \text{ mit } \eta_{isj} = [Z_{is}, d'_j \otimes W_{is}] \begin{pmatrix} \beta(t) \\ \theta \end{pmatrix} \end{aligned}$$

gilt.

Damit wird die Kovarianzmatrix geschätzt mit:

$$\widehat{\text{cov}}(\beta(t)', \theta') = \left(\sum_{i=1}^N s_i(\beta(t), \theta) s_i'(\beta(t), \theta) \right)^{-1}.$$

- Die Schätzung von θ erfolgt global, so daß der folgende marginale Ansatz gewählt werden kann. Mit $\alpha = (\beta(1), \dots, \beta(T), \theta)$ und $\tilde{Z}_{it} = (\mathbb{O}, \dots, Z_{it}, \dots, \mathbb{O})$ gilt für die Scorefunktion

$$s_i(\alpha) \approx \sum_{t=1}^T \sum_{j=1}^M c_{ij}^{GH}(\alpha) \left[\tilde{Z}_{it}, d'_j \otimes W_{it} \right]' D_{itj} \Sigma_{itj}^{-1} (y_{it} - \mu_{itj})$$

mit

$$\begin{aligned} D_{itj} &= \frac{\partial h(\eta_{itj})}{\partial \eta} \\ \Sigma_{itj} &= \text{cov}(y_{it} | d_j, \beta(t), \theta) \\ \mu_{itj} &= h(\eta_{itj}) \text{ mit } \eta_{itj} = \left[\tilde{Z}_{it}, d'_j \otimes W_{it} \right] \begin{pmatrix} \beta(1) \\ \vdots \\ \beta(T) \\ \theta \end{pmatrix} \end{aligned}$$

und somit folgt für die Kovarianzmatrix

$$\text{cov}(\alpha) = \left(\sum_{i=1}^N s_i(\alpha) s_i'(\alpha) \right)^{-1}.$$

Die Kovarianzen $\text{cov}(\beta(t))$ ergeben sich dann aus den entsprechenden Teilmatrizen von $\text{cov}(\beta(t)', \theta')$ und die Schätzung der Kovarianz von θ aus der entsprechenden Teilmatrix von $\text{cov}(\alpha) = \text{cov}(\beta(1)', \dots, \beta(T)', \theta')$.

Kapitel 4

Simulationsbeispiele

Im vorherigen Kapitel 3 wurden verschiedene Modellerweiterungen des Generalisierten Linearen Modells vorgestellt, die gleichzeitig eine zeitliche Variation der Parameter ermöglichen und gruppen- bzw. individuenspezifische Effekte durch die Aufnahme von zufälligen Effekten im Modellansatz berücksichtigen. Es wurden verschiedene Schätzverfahren für die Bestimmung der Parameterschätzungen und deren Varianzen angegeben. Die Güte der vorgestellten Verfahren wird in diesem Kapitel durch Simulationsbeispiele verifiziert.

Zunächst werden die vorgestellten Modellansätze für die Schätzung der festen Parameter angewendet und anschließend wird auf die Bayes-Schätzung der zufälligen Effekte eingegangen.

Bei der Schätzung der festen Parameter werden der Lokale Likelihoodansatz und der Ansatz des vollen Modells miteinander verglichen. Dabei zeigt sich, daß der Lokale Likelihoodansatz, bei dem für die Schätzung der zufälligen Effekte die Gauß-Hermite Integrationstechnik verwendet wird, sehr gute Resultate liefert. Wird dagegen der EM-Type Algorithmus angewendet, so liefert dieser nur zufriedenstellende Ergebnisse, wenn die Streuung der zufälligen Ef-

fekte sehr gering ist. Damit ist eine Aussage über die Schätzung der zufälligen Effekte beim EM-Type Algorithmus nur bedingt aussagekräftig.

4.1 Schätzung der festen Parameter

In den folgenden Simulationstudien werden die Ergebnisse aus dem vollen Modell mit denen des Lokalen Likelihoodansatzes mit Gauß-Hermit Integrationstechnik bzw. mit EM-Type Algorithmus jeweils unter Berücksichtigung von zufälligen Effekten vorgestellt.

Aufbau der Simulationsstudie

In den Simulationsstudien wird das binäre logistische Modell

$$\mu_{it} = h(\eta_{it}) = \frac{\exp(\eta)}{1 + \exp(\eta)}$$

mit $\eta_{it} = \beta_0(t) + \beta_1(t)x_{it} + b_i$ und $b_i \sim \mathcal{N}(0, Q)$ mit 100 Beobachtungseinheiten für 10 Zeitpunkte gewählt.

Der Regressor x_{it} ist mit $x_{it} = x_i$ für $t = 1, \dots, 10$ und $x_i \sim \mathcal{N}(0, 1)$ für $i = 1, \dots, 100$ gewählt worden. Die Parametervorgaben für β sind

$$\begin{aligned}\beta_0 &= (\beta_0(1), \beta_0(2), \dots, \beta_0(10))' = (1, 0.9, \dots, 0.1)' \\ \beta_1 &= (\beta_1(1), \beta_1(2), \dots, \beta_1(10))' = (0.1, 0.2, \dots, 1)' .\end{aligned}$$

Für θ wird beim vollen Modell und beim Lokalen Likelihoodansatz mit Gauß-Hermite Integrationstechnik die Parametervorgabe mit $\theta = \sqrt{Q} = 2$, dagegen beim Lokalen Likelihoodansatz unter Verwendung des EM-Type Algorithmus mit $\theta = \sqrt{Q} = \sqrt{(0.1)}$ gewählt.

4.1.1 Das volle Modell mit zufälligen Effekten

Das volle Modell unter Berücksichtigung von zufälligen Effekten wurde im Abschnitt 3.1.1 vorgestellt. Die Schätzung der Parameter θ und $\beta = (\beta_0(1),$

$\beta_1(1), \dots, \beta_0(10), \beta_1(10)$) erfolgt durch den Modellansatz eines Generalisier-
ten Linearen Modells mit zufälligen Effekten und dem linearem Prädiktor

$$\mu_{it} = h(\tilde{Z}_{it}\beta + b_i) \text{ mit } \tilde{Z}_{it} = (\mathbb{O}, \dots, Z_{it}, \dots, \mathbb{O}) \text{ mit } Z_{it} = (1, x_{it}).$$

Schätzung der Parametervarianzen

Die Schätzung der Varianz wird, wie in Abschnitt 3.2.1 vorgestellt, durch die
Approximation der inversen Fishermatrix mit der Scorefunktion, d. h.

$$\widehat{cov}(\hat{\alpha}) \approx (F^{SC})^{-1} \text{ mit } F^{SC} = \sum_{i=1}^n s_i(\hat{\alpha}) s_i'(\hat{\alpha})$$

durchgeführt. Ist M die Anzahl der Stützstellen für die numerische Integra-
tion so gilt:

$$\begin{aligned} s_i(\alpha) &\approx s_i^{GH}(\alpha) = \sum_{j=1}^M c_{ij}^{GH}(\alpha) \frac{\partial \ln f(y_i|d_j; \alpha)}{\partial \alpha} \\ &= \sum_{t=1}^T \sum_{j=1}^M c_{it}^{GH}(\alpha) \left(\tilde{Z}_{it}, d_j \right)' D_{itj}(\eta_{itj}) \Sigma_{it}^{-1}(\eta_{itj}) (y_{it} - \mu_{it}(\eta_{itj})) \\ \text{mit } \eta_{itj} &= \left(\tilde{Z}_{it}, d_j \right) \alpha \end{aligned}$$

In den folgenden Abbildungen werden die Ergebnisse der Simulationsstudie
graphisch dargestellt. Dabei wird deutlich, daß die empirischen Standardab-
weichungen

$$\begin{aligned} St_{emp}(\beta(t)) &:= \sqrt{\frac{1}{U-1} \sum_{u=1}^U \left(\hat{\beta}^u(t) - \bar{\beta}(t) \right)^2} \quad \text{bzw.} \\ St_{emp}(\theta) &:= \sqrt{\frac{1}{U-1} \sum_{u=1}^U \left(\hat{\theta}^u - \bar{\theta} \right)^2} \end{aligned}$$

sehr gut mit den entsprechenden geschätzten Standardabweichungen

$$St_{SC}(\beta(t)) := \frac{1}{U} \sum_{u=1}^U \sqrt{\widehat{var}_{FSC}(\hat{\beta}^u(t))} \quad \text{bzw.}$$

$$St_{SC}(\theta) := \frac{1}{U} \sum_{u=1}^U \sqrt{\widehat{var}_{FSC}(\hat{\theta}^u)}$$

übereinstimmen. Mit $\bar{\beta}(t) = \frac{1}{U} \sum_{u=1}^U \hat{\beta}^u(t)$ und $\bar{\theta}(t) = \frac{1}{U} \sum_{u=1}^U \hat{\theta}^u(t)$ werden die durchschnittlichen Parameterschätzungen bezeichnet.

Ergebnisse der Simulationsstudie

In den Abbildungen 4.1 und 4.2 werden für die Parameter β_0 und β_1 zu jedem betrachteten Zeitpunkt $t = 1, 2, \dots, 10$ alle Parameterschätzungen in Form eines Box-Plots eingetragen. Zusätzlich werden die Modellvorgaben sowie die empirischen und geschätzten Standardabweichungen mit

-----	Modellvorgabe
○	+– empirische Standardabweichung
×	+– geschätzte Standardabweichung

gekennzeichnet.

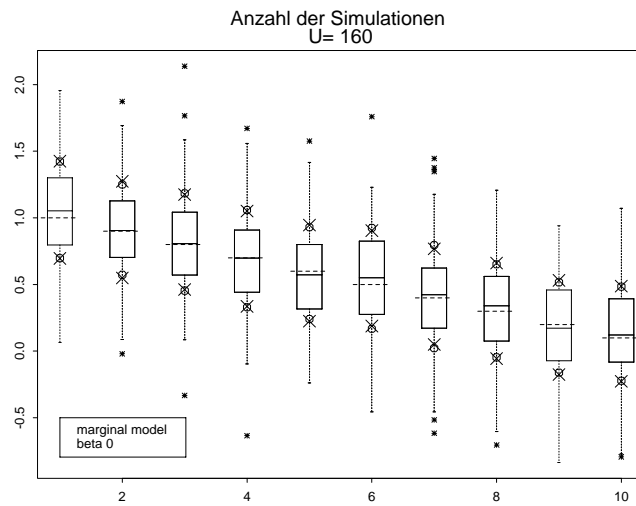


Abbildung 4.1: Simulationsergebnisse für $\beta_0(1), \beta_0(2), \dots, \beta_0(10)$ im vollen Modell

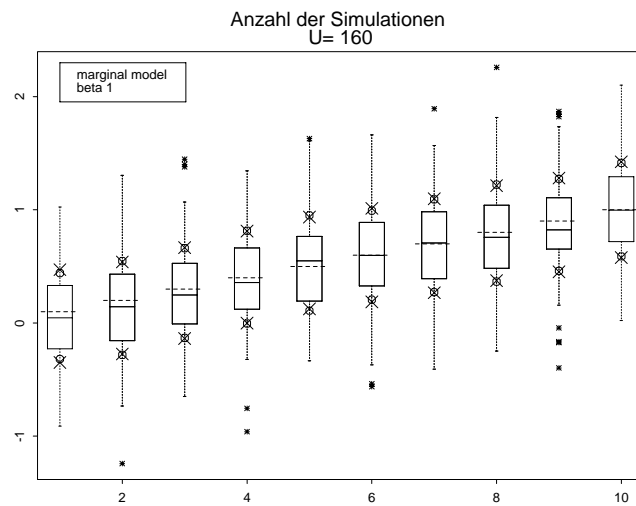


Abbildung 4.2: Simulationsergebnisse für $\beta_1(1), \beta_1(2), \dots, \beta_1(10)$ im vollen Modell

Für die zugehörigen Parameterschätzungen von θ wurden die folgenden durchschnittlichen Parameter- bzw. Varianzwerte berechnet.

θ	$\bar{\theta} = \frac{1}{U} \sum_{u=1}^U \hat{\theta}^u$	$St_{emp}(\theta)$	$St_{SC}(\theta)$
2	2.024	0.227	0.214

4.1.2 Lokaler Likelihoodansatz mit zufälligen Effekten und Gauß-Hermite Integrationstechnik

In diesem Abschnitt werden die Ergebnisse der Parameterschätzungen beim Lokalen Likelihoodansatz unter Berücksichtigung von zufälligen Effekten vorgestellt, wobei die Gauß-Hermite Integrationstechnik verwendet wird.

Die Schätzung von $\alpha' := (\beta', \theta)$ und $\beta = (\beta_0(1), \beta_1(1), \dots, \beta_0(10), \beta_1(10))$ erfolgt somit durch ein Generalisiertes Lineares Modell mit zufälligen Effekten und (zeit-)variierenden Koeffizienten unter Verwendung des Lokalen Likelihood Ansatzes mit dem linearen Prädiktor:

$$\mu_{it} = h(Z_{it}\beta(t) + b_i) \text{ mit } Z_{it} = (1, x_{it}), \beta(t) = (\beta_0(t), \beta_1(t)) \text{ und } b_i \sim \mathcal{N}(0, Q).$$

Es wird das in Abschnitt 3.1.2 vorgestellte Verfahren verwendet, wobei für die Kernfunktion der Gauß-Kern mit $\gamma = 1$ und als Startwert $\theta^{(0)} = 2$ sowie für das Abbruchkriterium der Iteration $\epsilon = 0.05$ gewählt wurden.

Schätzung der Parametervarianzen

Bei der Parameterschätzung wird der Koeffizienten $\beta(t)$ mit Hilfe der lokalen Likelihood und den Gewichten des Gaußkerns berechnet, dagegen erfolgte die Parameterschätzung von θ global, d. h. ohne lokale Gewichtung. Deshalb muß bei der Schätzung der Varianzen für die Parameterschätzungen für

$var(\beta(t))$ und $var(\theta)$ unterschiedliche Verfahren angewendet werden, die in Abschnitt 3.2.1 vorgestellt wurden.

- Die Lokale Schätzung von $var(\beta(t))$ erfolgt mit:

$$\begin{aligned}
cov(\beta(t), \theta) &= \left(\sum_{i=1}^N s_i(\beta(t), \theta) s_i'(\beta(t), \theta) \right)^{-1} \quad \text{und} \\
s_i(\beta(t), \theta) &\approx \sum_{s=1}^T \sum_{j=1}^M c_{ij}^{GH}(\beta(t), \theta) w_\gamma(s, t) \frac{\partial \ln f(y_{is}|d_j, \beta(t), \theta)}{\partial \beta(t)} \\
&= \sum_{s=1}^T \sum_{j=1}^M c_{ij}^{GH}(\beta(t), \theta) w_\gamma(s, t) [Z_{is}, d_j' \otimes W_{is}]' D_{isj} \Sigma_{isj}^{-1} (y_{is} - \mu_{isj}) .
\end{aligned}$$

- Für die globale Schätzung von θ mit $\alpha = (\beta(1), \dots, \beta(T), \theta)$ und $\tilde{Z}_{it} = (\mathbb{O}, \dots, Z_{it}, \dots, \mathbb{O})$ gilt:

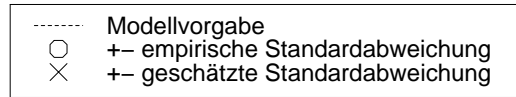
$$\begin{aligned}
cov(\alpha) &= \left(\sum_{i=1}^N s_i(\alpha) s_i'(\alpha) \right)^{-1} \quad \text{mit} \\
s_i(\alpha) &\approx \sum_{t=1}^T \sum_{j=1}^M c_{ij}^{GH}(\alpha) [\tilde{Z}_{it}, d_j' \otimes W_{it}]' D_{itj} \Sigma_{itj}^{-1} (y_{it} - \mu_{itj}) .
\end{aligned}$$

Für die Schätzung der Kovarianz von $\beta(t)$ bzw. θ müssen die entsprechenden Teilmatrizen von $cov(\alpha)$ betrachtet werden.

Ergebnisse der Simulationsstudie

Die Ergebnisse der Simulation werden für jeden betrachteten Zeitpunkt $t = 1, 2, \dots, 10$ in Abbildung 4.3 und 4.4 in Form eines Box-Plots eingetragen.

Mit



werden zusätzlich die Modellvorgaben sowie die empirischen und geschätzten Standardabweichungen gekennzeichnet.

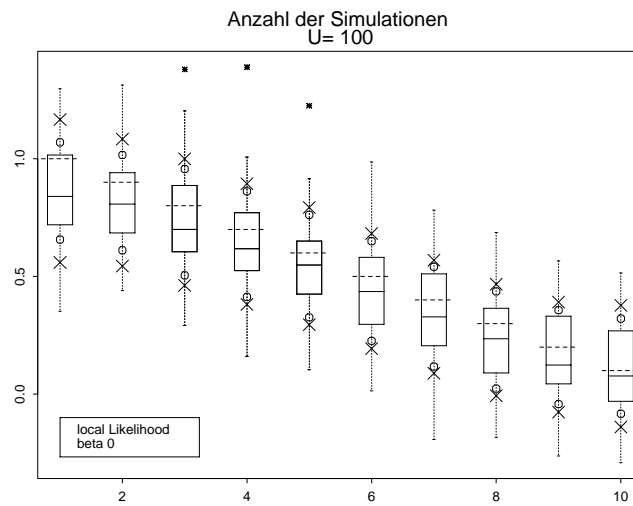


Abbildung 4.3: Simulationsergebnisse für $\beta_0(1), \beta_0(2), \dots, \beta_0(10)$ beim Lokalen Likelihoodansatz mit zufälligen Effekten und Gauß-Hermite Integrations-technik und $\gamma = 1$

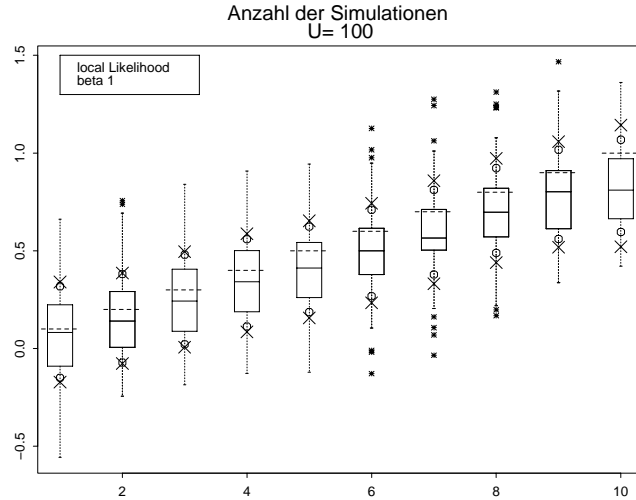


Abbildung 4.4: Simulationsergebnisse für $\beta_1(1), \beta_1(2), \dots, \beta_1(10)$ beim Lokalen Likelihoodansatz mit zufälligen Effekten und Gauß-Hermite Integrations-technik und $\gamma = 1$

Die Ergebnisse der Parameterschätzungen von θ werden in der folgenden Tabelle zusammengestellt:

θ	$\bar{\theta} = \frac{1}{U} \sum_{u=1}^U \hat{\theta}^u$	$\sqrt{\frac{1}{U-1} \sum_{u=1}^U (\hat{\theta}^u - \bar{\theta})^2}$	$\frac{1}{U} \sum_{u=1}^U \sqrt{\widehat{\text{var}}_{FSC}(\hat{\theta}^u)}$
2	1.890	0.199	0.202

Die Struktur der zeitvariierenden Koeffizienten wird bei der Schätzung der Parameterwerte exakt wiedergegeben. Daß die Modellvorgaben im Mittel nicht genau getroffen werden, ist nicht verwunderlich, da beim verwendeten Algorithmus alle Beobachtungen gewichtet in Abhängigkeit der verwendeten Kernfunktion eingehen. Insbesondere sind deshalb zu Beginn (d. h. für

$t = 1$ und $t = 2$) und zum Ende (d. h. für $t = 9$ und $t = 10$) die geschätzten Parametervarianzen größer als in der Mitte des Beobachtungszeitraumes für $t = 4$, $t = 5$ und $t = 6$, bei denen die geschätzten und empirischen Varianzen sehr gut übereinstimmen.

4.1.3 Lokaler Likelihoodansatz mit zufälligen Effekten und EM-Type Algorithmus

Der Lokale Likelihoodansatz unter Berücksichtigung von zufälligen Effekten mit Anwendung des EM-Type Algorithmus wurde in Abschnitt 2.1.2 vorgestellt. Bei den durchgeführten Simulationsstudien hat sich gezeigt, daß der EM-Type Algorithmus ein sehr schnelles Verfahren ist, und recht gute Ergebnisse liefert, wenn die Standardabweichung der normalverteilten zufälligen Effekte sehr klein ist (vgl. Hennevogl, 1991).

Es werden jedoch keine guten Schätzergebnisse erzielt, wenn die Standardabweichung der zufälligen Effekte nicht sehr klein ist. Deshalb wurden die Parametervorgaben für β_0 und β_1 mit

$$\begin{aligned}\beta_0 &= (\beta_0(1), \beta_0(2), \dots, \beta_0(10))' = (1, 0.9, \dots, 0.1)' \\ \beta_1 &= (\beta_1(1), \beta_1(2), \dots, \beta_1(10))' = (0.1, 0.2, \dots, 1)'\end{aligned}$$

identisch zu den Vorgaben der letzten beiden Simulationsstudien, jedoch die Standardabweichung der normalverteilten zufälligen Effekte mit $\theta = \sqrt{Q} = \sqrt{0.1}$, statt mit $\theta = 2$ gewählt.

Die Schätzung von $\alpha' := (\beta', \theta)$ und $\beta = (\beta_0(1), \beta_1(1), \dots, \beta_0(10), \beta_1(10))$ erfolgt durch ein Generalisiertes Lineares Modell mit zufälligen Effekten und (zeit-)variierenden Koeffizienten unter Verwendung der lokalen Likelihood und dem linearen Prädiktor:

$$\mu_{it} = h(Z_{it}\beta(t) + b_i) \text{ mit } Z_{it} = (1, x_{it}), \beta(t) = (\beta_0(t), \beta_1(t)) \text{ und } b_i \sim \mathcal{N}(0, Q)$$

Ergebnisse der Simulationsstudie

Die Schätzungen der Parameter $\beta_0(1), \beta_0(2), \dots, \beta_0(10)$ und $\beta_1(1), \beta_1(2), \dots, \beta_1(10)$ werden in den Abbildungen 4.5 und 4.6 mit Box-Plots dargestellt. So wie in Abbildung 4.7 sind die Schätzungen für θ dargestellt. Die Schätzungen für θ sind nicht zufriedenstellend geschätzt worden. Ist die Varianz der zufälligen Effekte größer, so werden die Schätzungen für den Modellparameter β sehr ungenau. Deshalb werden wir dieses Verfahren nicht bei den Anwendungsbeispielen im anschließenden Kapitel 5 verwenden.

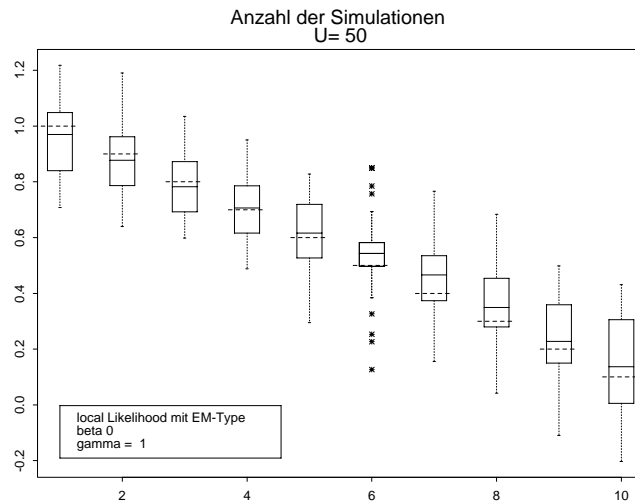


Abbildung 4.5: Simulationsergebnisse für $\beta_0(1), \beta_0(2), \dots, \beta_0(10)$ beim Lokalen Likelihoodansatz mit zufälligen Effekten und EM-Type Algorithmus

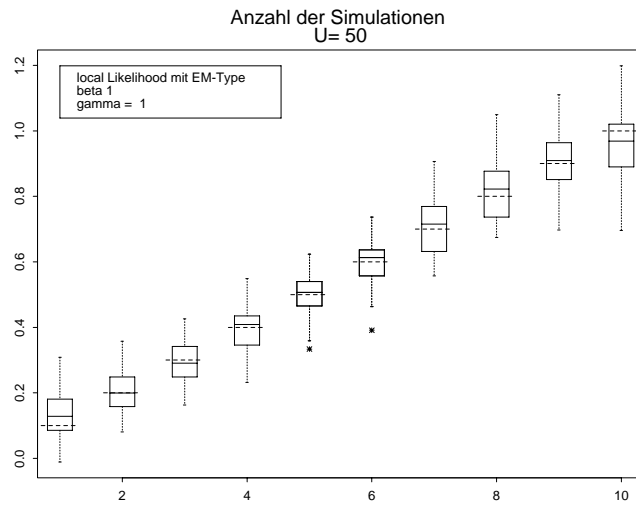


Abbildung 4.6: Simulationsergebnisse für $\beta_1(1), \beta_1(2), \dots, \beta_1(10)$ beim Lokalen Likelihoodansatz mit zufälligen Effekten und EM-Type Algorithmus

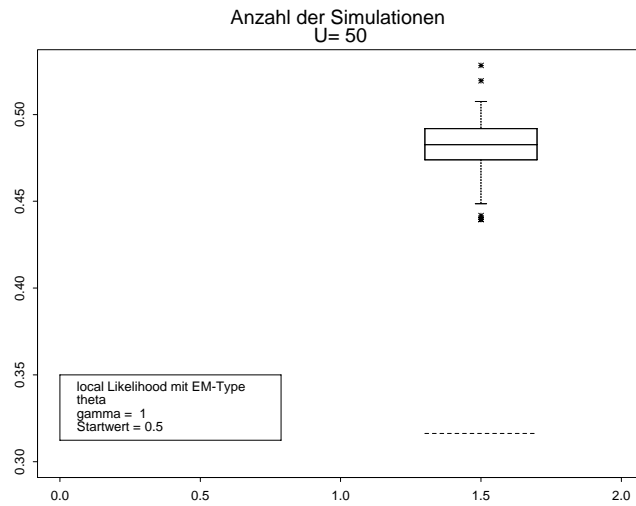


Abbildung 4.7: Simulationsergebnisse für θ beim Lokalen Likelihoodansatz mit zufälligen Effekten und EM-Type Algorithmus

4.2 Bayes-Schätzung der zufälligen Effekte

Die Bayes-Schätzung der zufälligen Effekte kann mit dem in Abschnitt 2.1.1 vorgestellten Verfahren durchgeführt werden.

In Abbildung 4.8 sind die Simulationsergebnisse für die Bayes-Schätzung der zufälligen Effekte b_i dargestellt. Dabei ist ein binäres logistisches Modell mit zufälligen Effekten und dem Modellansatz

$$E(y_{it}|x_{it}) = \frac{\exp(\eta_{it})}{1 + \exp(\eta_{it})} \text{ und } \eta_{it} = \beta_0 + x_{it}^{(1)}\beta_1 + x_{it}^{(2)}\beta_2 + b_i .$$

gewählt worden.

Bei der Simulation ist die Anzahl der Untersuchungseinheiten $N = 100$. Die Anzahl der Meßwiederholungen ist für alle Untersuchungseinheiten konstant mit $n_i = 4$ festgelegt worden.

Der Regressor $x_i^{(1)}$ ist binär mit $x_i^{(1)} \in \{0, 1\}$ und $x_i^{(2)}$ ist metrisch durch $x_i^{(2)} \sim \mathcal{N}(0, 10)$ gewählt worden. Für alle Meßwiederholungen $t = 1, \dots, 4$ sind diese Regressoren nicht verändert worden. Damit gilt

$$(1, x_i^{(1)}, x_i^{(2)}) := (1, x_{it}^{(1)}, x_{it}^{(2)}) .$$

Weiter gilt für die zufälligen Effekte $b_i \sim \mathcal{N}(0, Q)$ mit $Q = \sigma^2 = 4$.

Die Schätzungen dieser Effekte für 50 Simulationsdurchläufe sind in der obigen Abbildung 4.8 dargestellt. In vier Fällen konnte selbst nach 100 Iterationsschritten kein b_i gefunden werden, welches das Abbruchkriterium $\frac{|b_i^{(neu)} - b_i^{(alt)}|}{|b_i^{(neu)}|} \leq 0.001$ erfüllt. In allen anderen Fällen wurde der Schätzwert nach maximal 8 Iterationen gefunden.

Es ist auffallend, daß sehr große bzw. sehr kleine Simulationsvorgaben von b_i nicht ausreichend geschätzt werden. Der QQ-Plot der Simulationsvorgaben für b_i gegen die Quantile der Normalverteilung $\mathcal{N}(0, 4)$ in Abbildung 4.9

Schätzung der Randeffekte

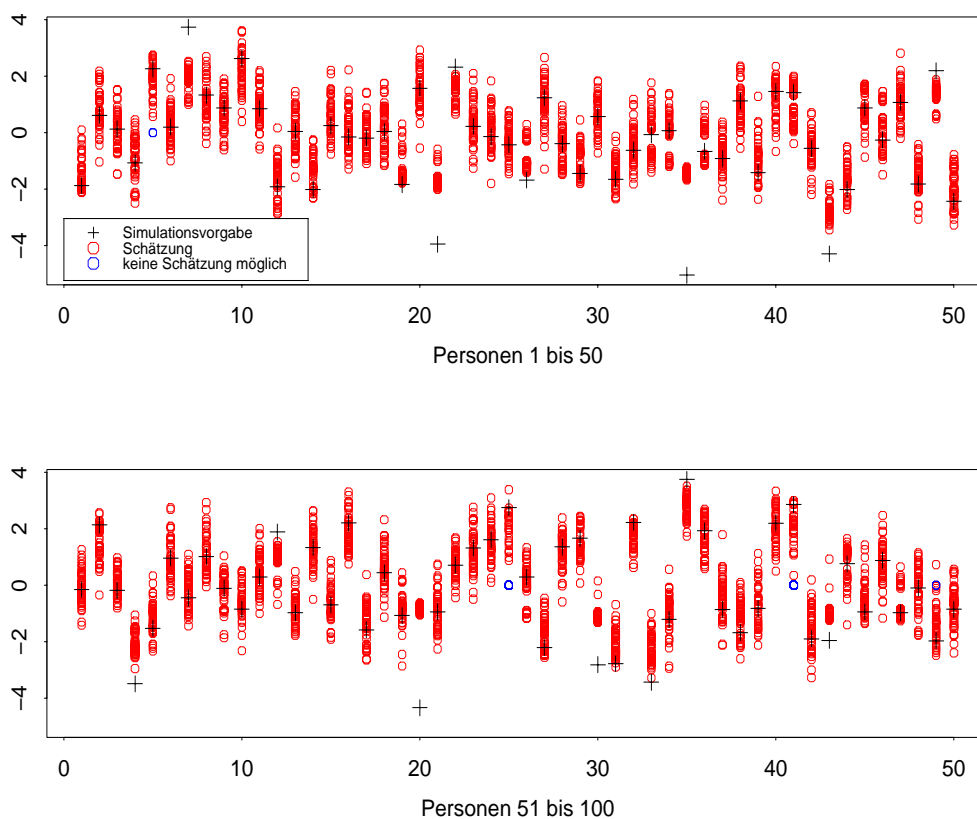


Abbildung 4.8: Simulationsergebnisse für θ

zeigt, daß die Modellvoraussetzung $b_i \sim \mathcal{N}(0, 4)$ in der konkret vorliegenden Stichprobe nicht exakt gegeben ist.

Dies ist auch der Grund weshalb die Schätzung der festen Parameter nicht exakt mit den Modellvorgaben übereinstimmen. Dies wird in Abbildung 4.10 verdeutlicht, in dem die Kerndichteschätzungen für die festen Parameter β_0 , β_1 und β_2 sowie die aufsteigend sortierten Parameterschätzungen für σ dargestellt werden.

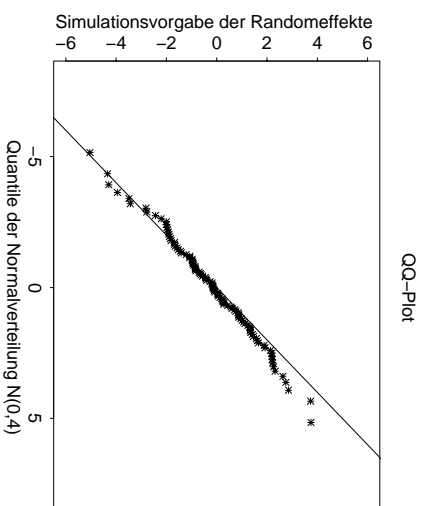


Abbildung 4.9: QQ-Plot der Simulationsvorgaben

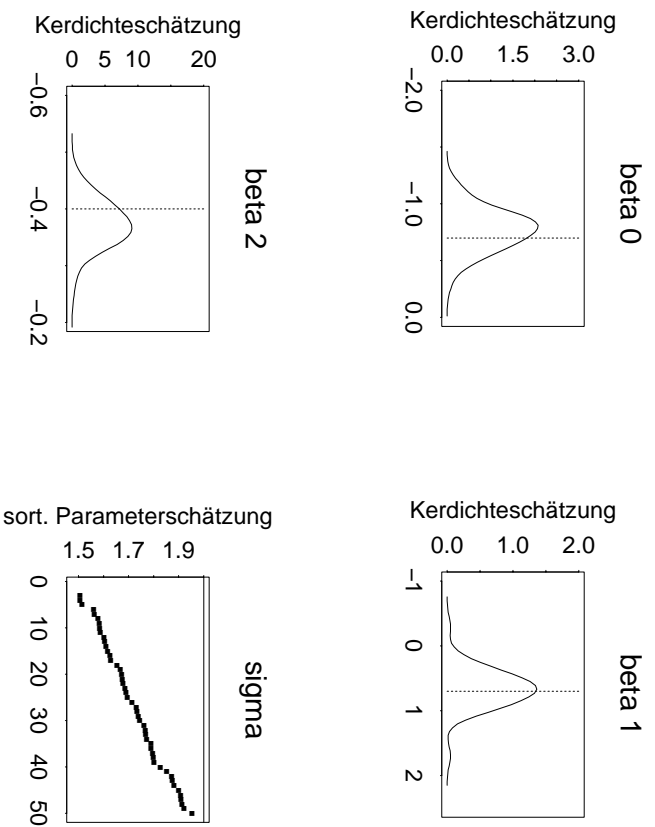


Abbildung 4.10: Schätzungen der festen Parameter

Kapitel 5

Anwendungsbeispiele

Es werden zunächst zwei aus dem medizinischen Umfeld stammende Anwendungsbeispiele für Poissonmodelle vorgestellt. Anschließend wird auf zwei Beispiele für Verweildauermodelle aus dem sozial-ökonomischen bzw. aus dem medizinischen Bereich eingegangen.

In allen Studien werden die relevanten Merkmale der gleichen Personen zu verschiedenen Zeitpunkten untersucht, so daß es sich aus statistischer Sicht um Meßwiederholungen an den gleichen Beobachtungseinheiten im Zeitverlauf handelt. Durch die Verwendung von variierenden Koeffizienten kann dieser zeitliche Einfluß untersucht werden, und durch die Aufnahme der zufälligen Effekte wird sowohl die Datenstruktur der Meßwiederholungen als auch der individuenspezifische Einfluß berücksichtigt.

5.1 Poissonmodelle

5.1.1 Emesis Datensatz

Datenbeschreibung

Häufige Nebenwirkungen der Chemotherapie sind Übelkeit und Erbrechen (Emesis), die nach Angaben der Patienten subjektiv als die am meisten belastende Nebenwirkung empfunden wird.

Durch eine neue Art von Medikamenten sollen diese Nebenwirkungen verringert werden. Die 122 Patientinnen der vorliegenden Studie (Dietz & Böhning, 1994) mit gynäkologischen Tumoren wurden alle mit einer Chemotherapie in Kombination mit einem anti-emetischen Wirkstoff behandelt. Es wurden dabei zwei unterschiedliche Wirkstoffe verwendet. Mit Medikament 1 (`Med_1`) wurden 70 und mit Medikament 2 (`Med_2`) wurden 52 Patientinnen behandelt.

Das Empfinden der Übelkeit einer Patientin objektiv zu messen ist sehr schwierig. Einfacher ist, es die Anzahl der Erbrechenanfälle pro Tag zu zählen, wobei wiederholtes Erbrechen in einer kurzen Zeitspanne als ein Anfall gewertet wird. Der Zusammenhang zwischen der Anzahl der Erbrechenanfälle und dem subjektiven Empfinden der Übelkeit ist offensichtlich gegeben, so daß in dem hier vorgeschlagenen Modell die Zielgröße die Anzahl der Erbrechenanfälle pro Tag ist.

Es ist sinnvoll, die Patientinnen über mehrere Tage zu beobachten, um sicherzustellen, daß ein neues Anti-Emetikum eventuell nicht nur zu einer zeitlichen Verzögerung der Emesis führt. Die Patientinnen in dieser Studie wurden nach der Anzahl der Erbrechenanfälle pro Tag am Therapietag selbst ($t = 1$) und an den nächsten fünf folgenden Tagen ($t = 2, \dots, t = 6$) befragt.

Als weitere prognostische Faktoren wurde das Alter der Patientinnen (`age`) und die Information, ob dies die erste Chemotherapie (`Zyk_1 = 1`) oder eine

weitere Chemotherapie ($Zyk_1 = 0$) ist, berücksichtigt.

Modellansatz

Die Zielgröße, d. h. die Anzahl der täglichen Erbrechenanfälle, wurde als eine poissonverteilte Zufallsvariable gewählt. Es wurde der Modellansatz

$$E(y_{it} | x_{it}^{Med_1}, x_{it}^{Med_2}, x_{it}^{age}, x_{it}^{Zyk_1}) = \exp(x_{it}^{Med_1} + x_{it}^{Med_2} + \ln(x_{it}^{age}) + x_{it}^{Zyk_1} + b_i)$$

mit $b_i \stackrel{iid}{\sim} \mathcal{N}(0, \sigma^2)$ verwendet. Der lineare Prädiktor des Modells enthält keinen Interzept, damit der Einfluß der unterschiedlichen Medikamente, d. h. die Parameterwerte von (Med_1), bzw. (Med_2) direkt interpretiert werden können.

Das verwendete Modell ist ein Generalisiertes Lineares Modell mit zufälligen Effekten und zeitvariierenden Koeffizienten. Als Modellansatz wird der zentrale Ansatz dieser Arbeit, der Lokale Likelihoodansatz unter Berücksichtigung von zufälligen Effekten verwendet, wobei die Berechnung mit Hilfe der Gauß-Hermite Integrationstechnik durchgeführt wird.

In diesem Anwendungsbeispiel werden wir auf die Wahl des Glättungsparameters der Kernfunktion eingehen. Dabei werden verschiedene Optimalitätskriterien bezüglich der Wahl des Glättungsparameters angewendet und für unterschiedliche Glättungsparameters die Ergebnisse gegenübergestellt.

Ergebnisse

Die Aufgabenstellung der Studie ist die Unterschiede der beiden Medikamente hinsichtlich der Emetogenität zu beurteilen. Wie die folgende Darstellung 5.1 zeigt, ist die Anzahl der Erbrechenanfälle im gesamten Zeitverlauf bei der Einnahme von Medikament 2 geringer als bei Medikament 1.

Neben dem Zeitverlauf der Parameterschätzungen, werden auch die Kurven der punkweisen Fehlerbänder gezeigt. Letzter ergeben sich aus der Addi-

tion, bzw. Subtraktion der Parameterschätzwerte und den entsprechenden Varianzen der Parameterschätzungen.

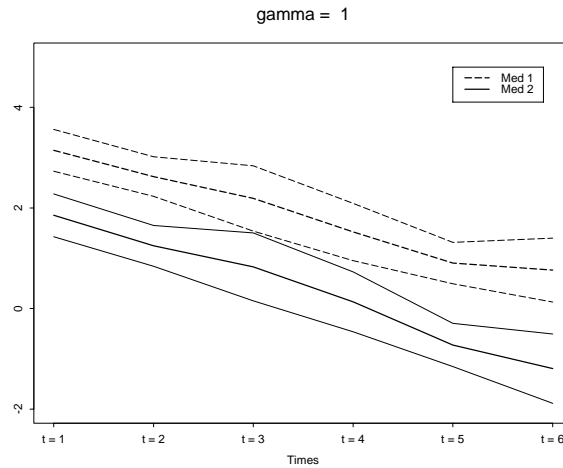


Abbildung 5.1: Vergleich der Parameterschätzungen von Med_1 und Med_2

Es wurde der Glättungsparameter mit $\gamma = 1$, d. h. der optimale Glättungsparameter nach dem Kreuzvalidierungskriterium mit Kullback-Leibler Verlustfunktion, gewählt.

Bevor wir auf den Einfluß des Alters und des Therapiezyklusses eingehen, stellen wir die Ergebnisse der Kreuzvalidierung für die Wahl des Glättungsparameters vor.

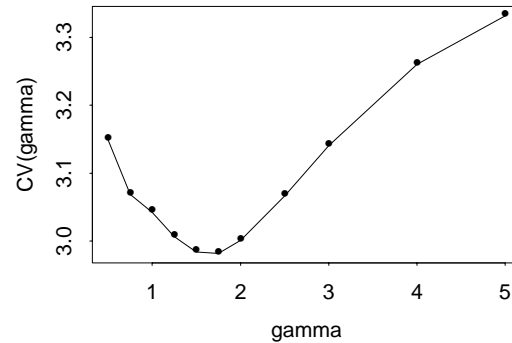
Wahl und Einfluß des Glättungsparameters

Die Wahl des Glättungsparameters wurde mit Kreuzvalidierung, die in Abschnitt 2.2.2 bei der Einführung des Lokalen Likelihoodansatzes vorgestellt wurde, durchgeführt. Der optimale Glättungsparameter ist der Parameter, der im Minimum der Verlustfunktion liegt. Wie in Abbildung 5.2 zu sehen ist, wird der optimale Glättungsparameter durch Kreuzvalidierung bei Verwendung der Quadratischen bzw. der Poisson-Verlustfunktion mit $\gamma = 1.5$, dagegen bei der Kullback-Leibler Verlustfunktion mit $\gamma = 1$ bestimmt.

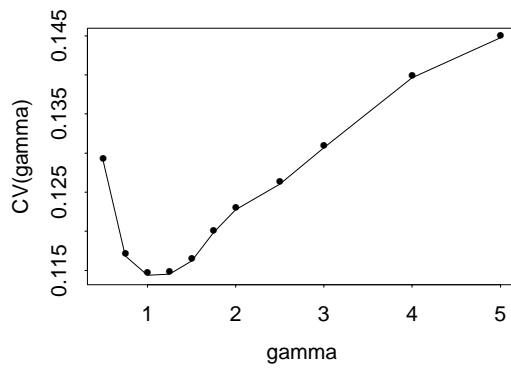
Emesis Datensatz

Regressoren:
Med 1, Med 2, Age, Zyklus 1

Quadratische Verlustfunktion



Kullback-Leibler Verlustfunktion



Pearson Verlustfunktion

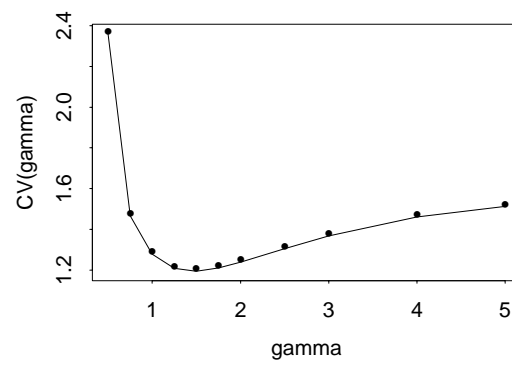


Abbildung 5.2: Wahl des Glättungsparameters durch Kreuzvalidierung

Durch die Bestimmung des optimalen Glättungsparameters mit Kreuzvalidierung durch eine der oben genannten Verlustfunktionen wird sichergestellt, daß die Struktur der zeitabhängigen Koeffizienten erkennbar ist. Es wird jedoch über starke zeitliche Variation in den Daten geglättet, damit der zeitliche Trend optimal zu erkennen ist.

Wird ein Parameterwert für den Glättungsparameter gewählt, der deutlich größer als der optimale Glättungsparameter ist, ist die relevante zeitabhängige Struktur nicht mehr erkennbar. Wählt man dagegen einen Parameterwert für den Glättungsparameter, der kleiner als der optimale Glättungsparameter

ist, läßt sich die den Daten zugrundeliegende Struktur genauer analysieren. Aus diesem Grund werden im folgenden die Parameterschätzungen mit dem Glättungsparameter $\gamma = 0.5$ und $\gamma = 1.5$ in den Abbildung 5.3 und 5.4 dargestellt.

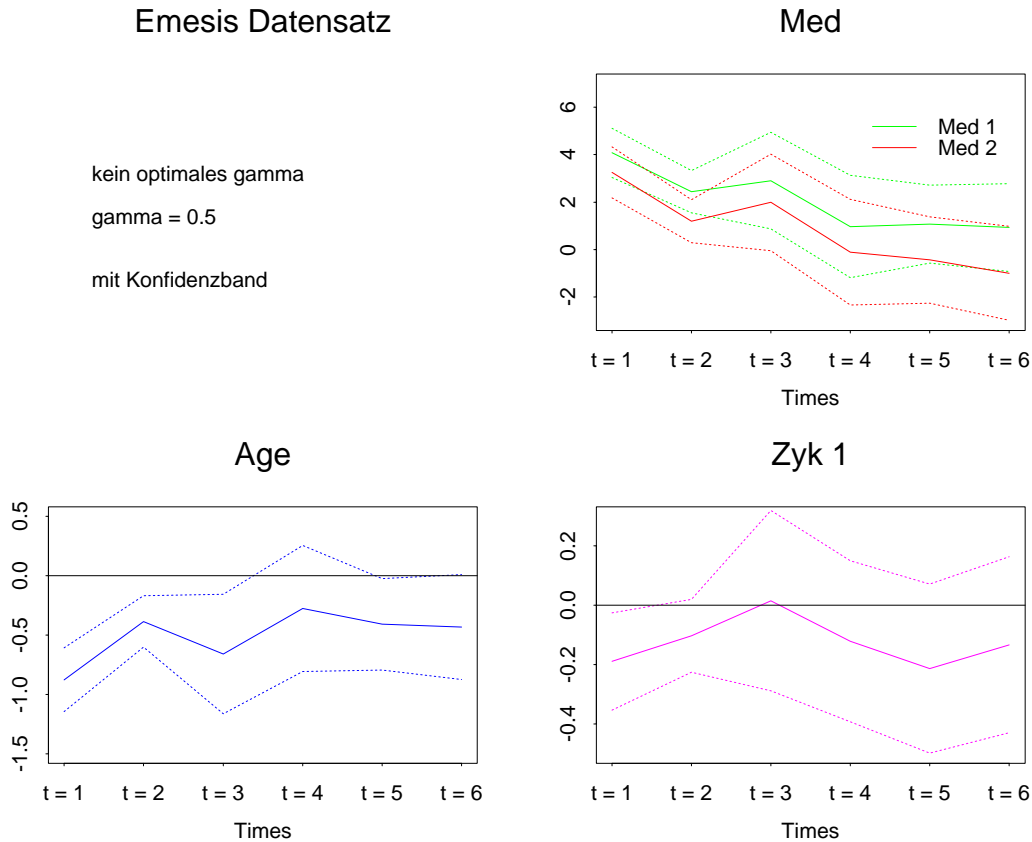


Abbildung 5.3: Parameterschätzungen mit $\gamma = 0.5$

In dieser Studie hatten Patientinnen mit der ersten Chemotherapiebehandlung weniger Erbrechenanfälle als Patientinnen mit zwei oder mehr Behandlungen.

Am zweiten Tag nach der Behandlung ($t = 3$) war die Anzahl der Anfälle jedoch gleich. Da die obere Grenze des Konfidenzbandes auch an den ande-

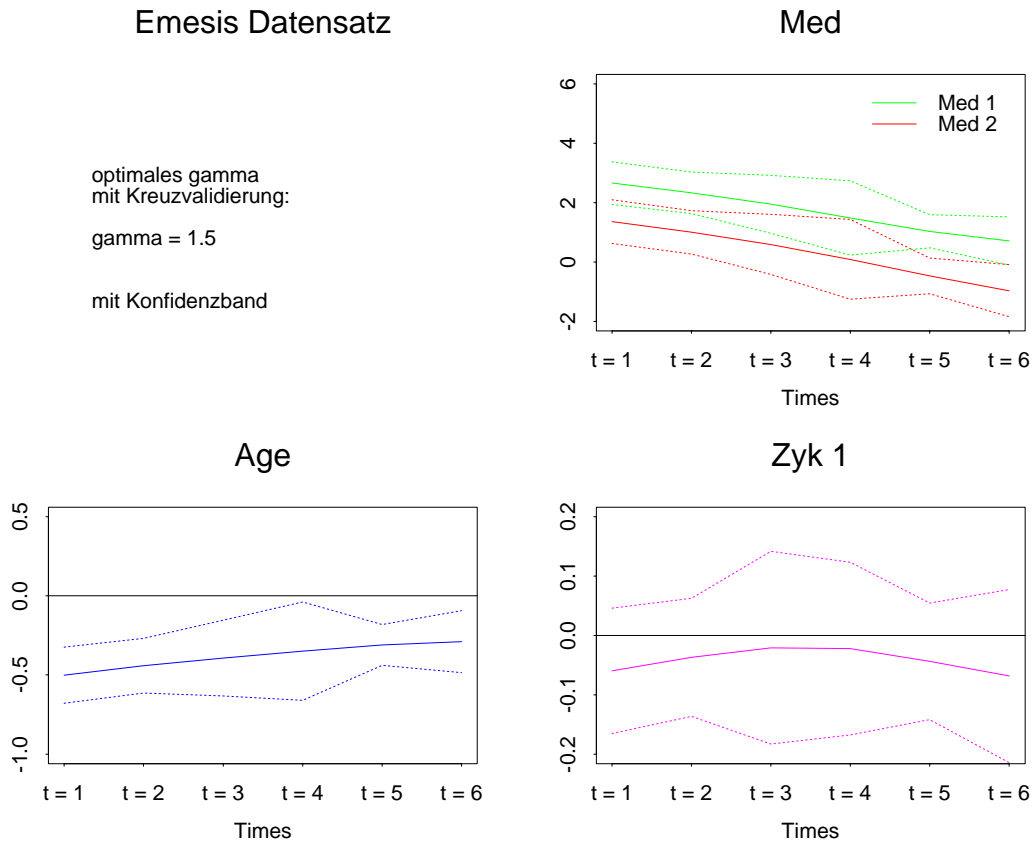


Abbildung 5.4: Parameterschätzungen mit $\gamma = 1.5$

ren Tagen oberhalb der Null-Linie liegt, ist an allen Tagen der Einfluß des Therapiezyklusses nicht signifikant von Null verschieden.

Bei jüngeren Patientinnen waren die Erbrechenanfänge häufiger. Auch dieser Effekt ist nicht signifikant, da bei dem optimalen Gamma, d. h. $\gamma = 1.5$ die Null-Linie überhalb des Konfidenzbandes liegt.

Die punktwisen Konfidenzintervalle für $\gamma = 1.5$ sind zum Teil an den Rändern untypisch schmaler, als in der Mitte des Beobachtungszeitraumes.

Betrachtet man die Ergebnisse der Parameterschätzungen mit $\gamma = 0.5$, so

erkennt man, daß über eine große “zackenartige” Variation der Parameterschätzungen geglättet wurde und deshalb die Konfidenzintervalle sehr breit sind.

5.1.2 Epileptic Datensatz

Datenbeschreibung

In Thall & Vail (1990) und Breslow & Clayton (1993) sowie in Diggle, Liang & Zeger (1994) werden die Ergebnisse aus einer klinischen Studie vorgestellt, bei der 59 Patienten, die an epileptischen Anfällen leiden, zusätzlich zur Standardbehandlung entweder ein neues Medikament oder ein Placebo verabreicht wurde. Die Patienten haben sich ab Beginn der Behandlung viermal, jeweils nach zwei Wochen, in der Klinik vorgestellt.

Die Daten liegen in der Form (y_{it}, x_{it}) mit $i = 1, \dots, 59$ (Personen) $t = 1, \dots, 4$ (Meßwiederholungen) vor. Dabei ist y_{it} die Anzahl der epileptischen Anfälle der i -ten Person in einem zweiwöchigen Intervall vor dem t -ten Arztbesuch. Für den Regressorenvektor $x_{it} = x_i = (1, x_i^{Treat}, x_i^{Base}, x_i^{Age})$ der i -ten Person gilt $x_i^{Treat} = 0$, wenn keine zusätzliche Behandlung (Placebo) bzw. $x_i^{Treat} = 1$, wenn die Behandlung mit dem neuen Medikament erfolgt ist. Der BASELINE x_i^{Base} stellt die patientenspezifische Anzahl der epileptischen Anfälle dar und ist die logarithmierte Anzahl der epileptischen Anfälle vor Beginn der neuen Behandlungsmethode innerhalb eines Zeitraumes von 8 Wochen.

Modellansätze

Im folgenden werden die Ergebnisse verschiedener Modellansätze vorgestellt, wobei jeweils ein Poissonmodell verwendet wurde. In den Tabellen werden

neben den Parameterschätzungen auch die zugehörigen Standardabweichungen angegeben.

Beim ersten Modellansatz werden keine variierenden Koeffizienten und keine zufällige Effekte verwendet.

β^0	β^{Treat}	β^{Base}	β^{Age}
-3.85 (0.39)	-0.04 (0.05)	1.22 (0.03)	0.52 (0.1)

Der obige Modellansatz kann durch Berücksichtigung von zufälligen Effekten erweitert werden und man erhält die folgenden Parameterschätzungen.

β^0	β^{Treat}	β^{Base}	β^{Age}	σ
-4.54 (0.39)	-0.25 (0.04)	1.28 (0.03)	0.74 (0.09)	0.6735 (0.031)

Im nächsten Schritt werden neben der Aufnahme von zufälligen Effekten auch zeitvariierende Koeffizienten betrachtet. Dabei werden drei verschiedene Ansätze für die zeitvariierenden Koeffizienten verwendet. Beim ersten Ansatz wird der Marginale Likelihoodansatz verwendet, beim zweiten die Polynomiale Regression mit einem Polynom 3. Grades. Bei diesem Ansatz können keine Standardabweichungen für die Parameterschätzungen angegeben werden, da die approximierte Fishermatrix aus numerischen Gründen nicht invertiert werden kann. Beim zuletzt betrachteten Modell wird der Lokale Likelihoodansatz mit $\gamma = 0.1$ verwendet.

Obwohl die drei Modellansätze aufgrund der unterschiedlichen Berücksichtigung des zeitlichen Einflusses zu verschiedenen Parameterschätzungen für die zeitabhängigen Koeffizienten führen, lassen sich die gleichen Tendenzen erkennen. Damit ist die qualitative Aussage der Parameterschätzungen in

allen drei Modellen gleich. Insbesondere wird der zeitunabhängige individuenspezifische Einfluß der zufälligen Effekte in allen Modellen gleich geschätzt.

Volles Modell					
	β_t^0	β_t^{Treat}	β_t^{Base}	β_t^{Age}	σ
t=1	-6.13 (0.66)	-0.23 (0.1)	1.40 (0.05)	1.1 (0.16)	0.67 (0.04)
t=2	-2.79 (1.19)	-0.19 (0.11)	1.13 (0.09)	0.38 (0.34)	
t=3	-4.99 (0.63)	-0.25 (0.09)	1.33 (0.10)	0.83 (0.15)	
t=4	-4.03 (0.94)	-0.36 (0.16)	1.25 (0.08)	0.61 (0.26)	
Polynomiale Regression					
	β_t^0	β_t^{Treat}	β_t^{Base}	β_t^{Age}	σ
t=1	-6.13	-0.23	1.40	1.1	0.67
t=2	-2.79	-0.19	1.13	0.38	
t=3	-4.99	-0.25	1.33	0.83	
t=4	-4.03	-0.36	1.25	0.61	
Lokaler Likelihoodansatz					
	β_t^0	β_t^{Treat}	β_t^{Base}	β_t^{Age}	σ
t=1	-4.03 (0.55)	-0.32 (0.06)	1.18 (0.04)	0.64 (0.14)	0.61 (0.04)
t=2	-0.79 (0.94)	-0.25 (0.09)	0.91 (0.06)	-0.05 (0.28)	
t=3	-2.94 (0.55)	-0.33 (0.05)	1.11 (0.06)	0.38 (0.13)	
t=4	-2.04 (0.69)	-0.43 (0.10)	1.03 (0.06)	0.17 (0.19)	

5.2 Verweildauermodelle

5.2.1 Dauer der Arbeitslosigkeit

Datensatzbeschreibung

Bei diesem Anwendungsbeispiel werden Daten des SOEP (Sozio-ökonomisches Panel (Hanefeld, 1987)) analysiert. Betrachtet werden insgesamt 1.188 Personen mit deutscher Staatsangehörigkeit aus den alten Bundesländern, die in den Jahren 1983-1992 arbeitslos gemeldet waren. Als Einflußgrößen wird das Alter der Person zu Beginn der Arbeitslosigkeit sowie das Geschlecht aufgenommen.

Modellansatz

Die Dauer der Arbeitslosigkeit (in Monaten) wird durch ein Verweildauermodell mit der Hazardrate

$$\lambda(t|x_{it}) = P(T = t|T \leq t; x_{it}) = \frac{\exp(\eta_{it})}{1 + \exp(\eta_{it})}$$

und dem linearen Prädiktor

$$\eta_{it} = x'_{it}\beta(t) + b_i$$

modelliert. Der eindimensionale zufällige Effekt b_i modelliert den personenspezifischen Einfluß. Der Regressorenvektor x enthält neben dem Interzept als weitere Einflußgrößen das Alter und das Geschlecht, d. h. es ist $x'_{it} = (1, x_i^{SEX}, x_i^{AGE})$. Das Geschlecht ist binär kodiert mit $x^{SEX} = 1$ für Frauen und $x^{SEX} = 0$ für Männer. Die metrische Größe des Alter zu Beginn der Arbeitslosigkeit geht mit $x^{AGE} = (AGE - 35)$ in das Modell ein.

Wir verwenden den Modellansatz für Regressionsplines mit kubisch-linearem Verlauf. Durch den kubischen Verlauf kann die Variabilität am Anfang der Arbeitslosigkeit sehr gut modelliert werden. Da insbesondere für $t > 20$ nur

noch sehr wenige Beobachtungen vorliegen, können durch die Verwendung des linearen Funktionsansatzes Schätzprobleme vermieden werden.

Ergebnisse

In der folgenden Abbildung 5.5 werden die Ergebnisse des kubisch-linearen Regressionsansatzes mit der Knotenwahl $\xi = 18$ dem Lokalen Likelihoodansatzes gegenübergestellt.

In beiden Ansätzen werden ungefähr die gleichen zeitabhängigen Tendenzen aufgezeigt, solange die Arbeitslosigkeit nicht mehr als 18 Monate andauert. Da nach diesem Zeitpunkt die Datenbasis sehr dünn ist, sollen die Parameterschätzung für $t > 20$ nicht interpretiert werden.

Es wird jedoch auch deutlich, welchen Vorteil der flexible Ansatz der Lokalen Likelihood im Gegensatz zur kubisch-linearen Regressions bietet. Zum Beispiel kann der zeitabhängige Verlauf des Koeffizienten *SEX* nicht adäquat durch eine kubisch-lineare Funktion ausgedrückt werden.

Das Minimum des kubischen Verlaufes zum Zeitpunkt $t = 6$ mit dem Parameterschätzwert -2 , entspricht nicht dem zeitabhängigen Verhalten des Koeffizienten *SEX*. Denn betrachtet man den Verlauf dieses Koeffizienten beim Lokalen Likelihood Ansatz, so werden hier zwei lokale Minima bei $t = 6$ und $t = 14$ und vor allem ein globales Minimum für $t = 24$ aufgezeigt. Diese zeitabhängige Struktur kann nicht durch einen kubisch-linearen Verlauf dargestellt werden, so daß das Minimum des Koeffizienten *SEX* beim kubisch-linearen Regressionsansatz nicht durch die Datenstruktur, sondern durch den parametrischen Modellansatz bedingt wird.

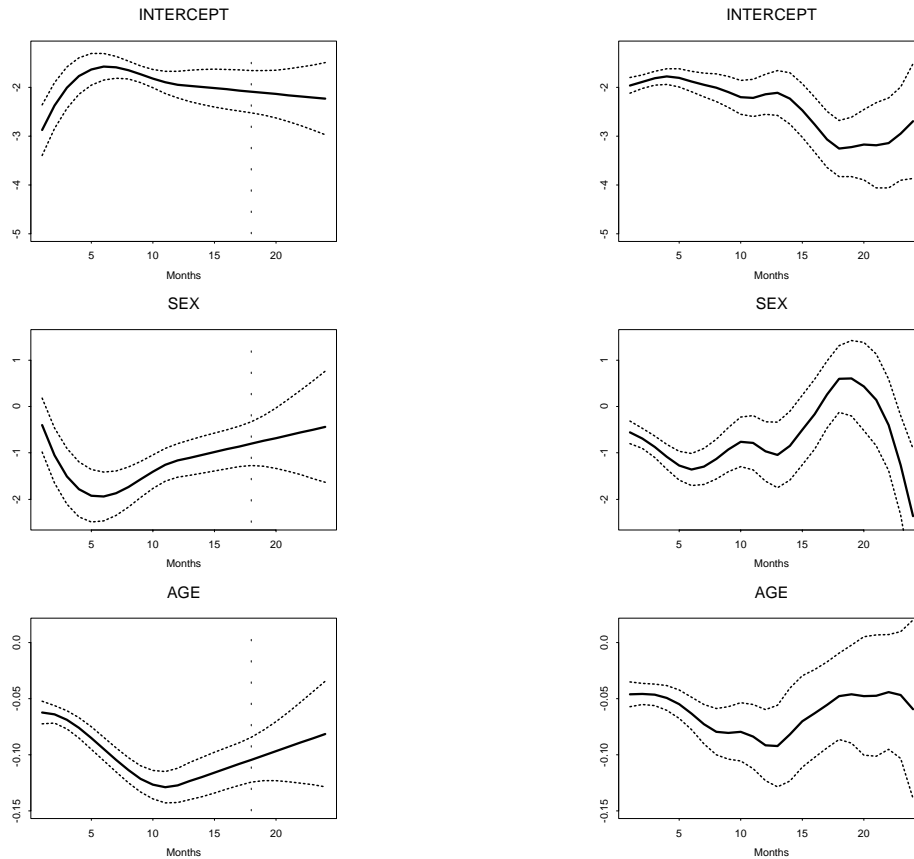


Abbildung 5.5: Parameterschätzungen mit kubisch-linearem Regressionsansatz (links) und Lokalem Likelihoodansatz (rechts)

5.2.2 Magenkrebsstudie

Datensatzbeschreibung

Der Datensatz einer Magenkrebsstudie (Fahrmeir, 1994) enthält die Überlebenszeiten (in Monaten) von 90 Patienten einer Magenkrebsstudie. Die Patienten wurden mit einer Chemotherapie bzw. mit einer kombinierten Therapieform behandelt. Durch die Studie soll der Einfluß der Therapieformen im Zeitverlauf untersucht werden.

Modellansatz

Wie bereits im vorherigen Beispiel betrachten wir ein Verweildauermodell mit der Hazardrate

$$\lambda(t|x_{it}) = P(T = t|T \leq t; x_{it}) = \frac{\exp(\eta_{it})}{1 + \exp(\eta_{it})}$$

und dem linearen Prädiktor

$$\eta_{it} = \beta_0(t) + x'_{it}\beta_{\text{TREATMENT}}(t) + b_i .$$

Der personenspezifischen Einfluß wird durch den eindimensionale zufällige Effekt b_i berücksichtigt. Neben der Therapieform werden keine weiteren Einflußgrößen betrachtet. Der Einfluß der Therapie ist mit $x = 0$ für die Chemotherapie und mit $x = 1$ für die kombinierte Therapieform codiert.

Da in diesem Beispiel die Überlebensraten der Patienten untersucht werden, liegen gegen Ende des Beobachtungszeitraumes nur noch wenige Datensätze vor. Wir betrachten im folgenden Modelle, die diese Struktur berücksichtigen. Beim Lokalen Likelihood Ansatz verwenden wir den Gaußkern und wählen den Glättungsparameter in Abhängigkeit der Anzahl, der unter Risiko stehenden Personen, mit

$$\gamma_t = \frac{1}{\sqrt{n_t}} \gamma .$$

Dadurch wird gegen Ende des Beobachtungszeitraumes, in dem nur noch wenige Beobachtungen vorhanden sind, eine größere Fensterbreite gewählt als zu Beginn.

Wir vergleichen die Ergebnisse mit dem Modellansatz durch kubisch-lineare Regression. Dabei betrachten wir sowohl eine kubische Funktion, die linear fortgesetzt wird, als auch den Fall wenn zwei oder drei stückweise kubische Funktionen linear fortgesetzt werden.

In den folgenden Abbildungen sind die Hazardraten mit den Bändern der punktwweisen Standardabweichungen für die beiden Behandlungsarten eingezeichnet. Die Hazardraten der Patienten, die mit der kombinierten Therapieform behandelt wurden (durchgezogene Linie), sind zunächst höher, jedoch nach 2 Jahren ist ein gegenteiliger Effekt zu beobachten. Beim Ansatz der Regressionssplines wurden ein, zwei und drei stückweise kubische Funktionen mit einer linearen Funktion fortgesetzt. Die Stellen der Knoten sind durch senkrechte Linien gekennzeichnet.

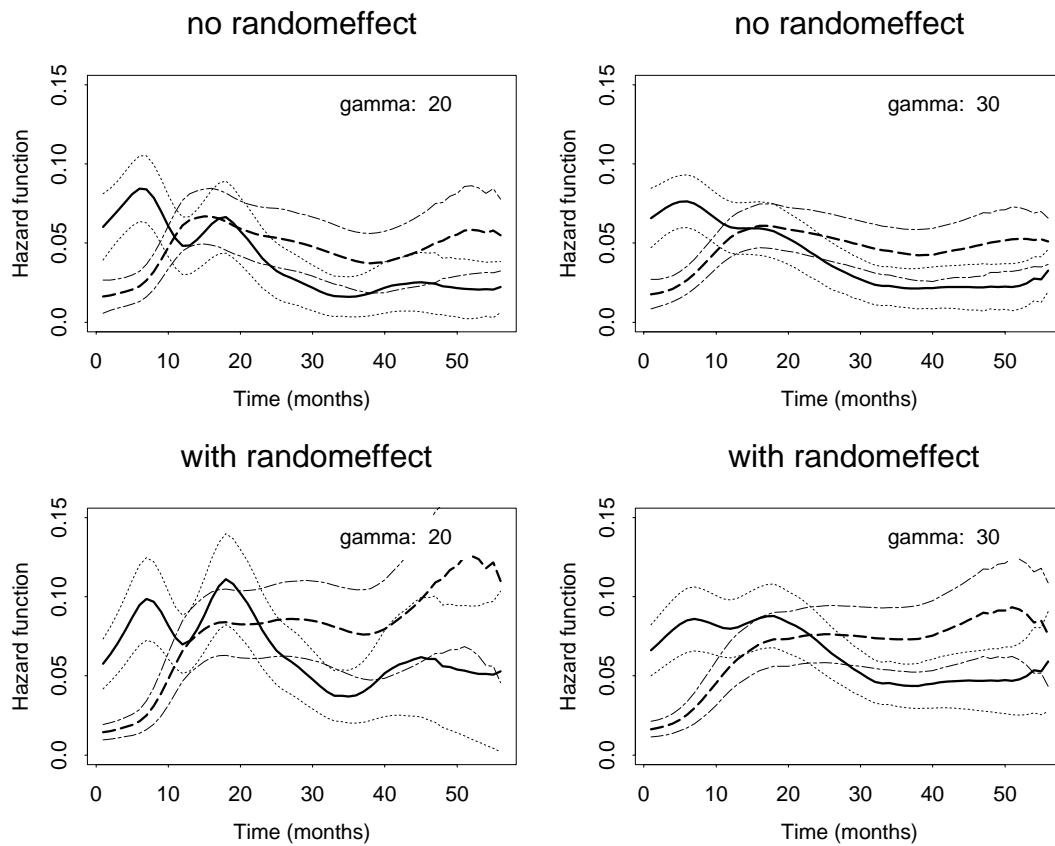


Abbildung 5.6: Hazardraten beim Lokalen Likelihood Ansatz

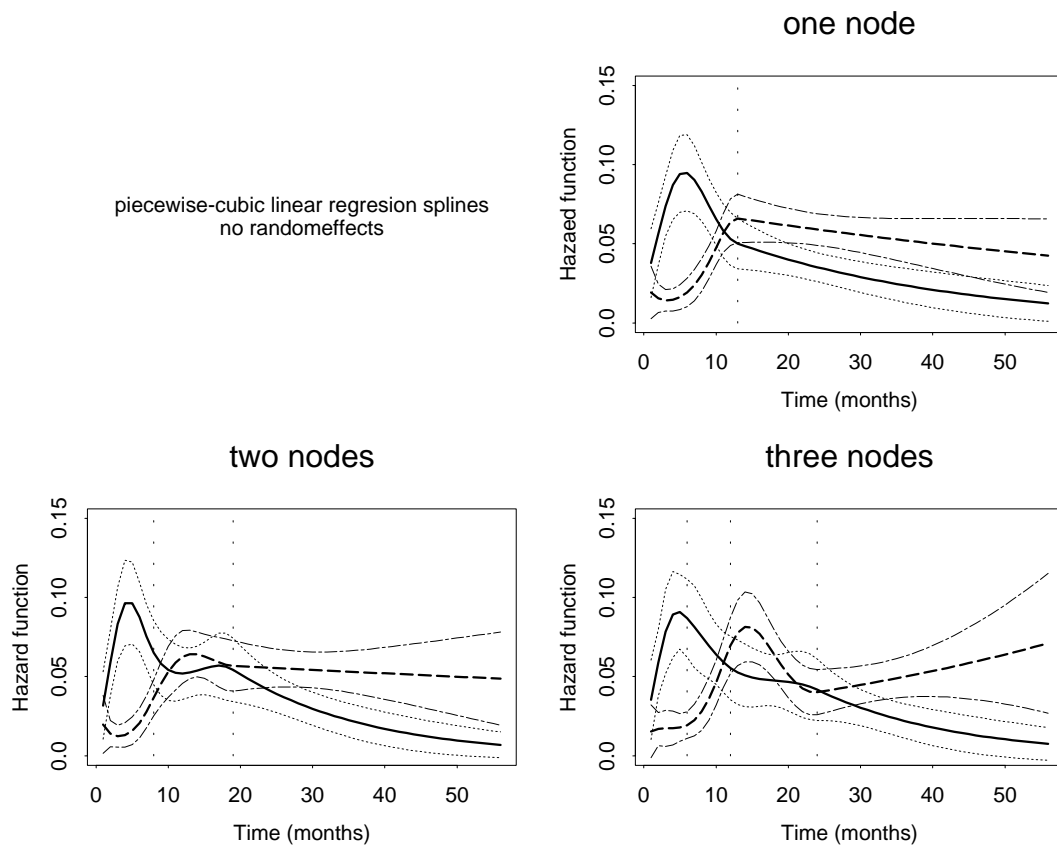


Abbildung 5.7: Hazardraten beim Regressionsspline ohne Berücksichtigung von zufälligen Effekten

piecewise-cubic linear regression splines
with randomeffects

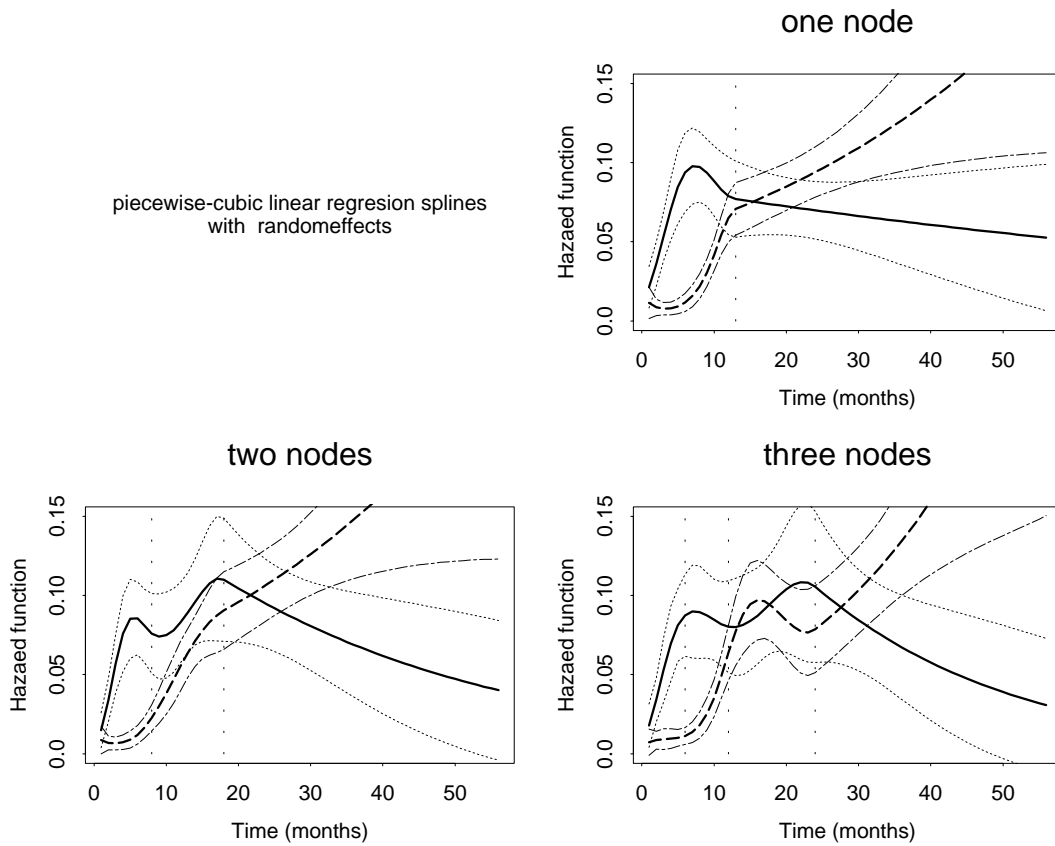


Abbildung 5.8: Hazardraten beim Regressionsspline mit Berücksichtigung von zufälligen Effekten

Kapitel 6

Anhang

6.1 EM-Algorithmus mit direkter Gauß-Hermite Integrationstechnik

In diesem Abschnitt geben wir einige technische Details für das verwendete Verfahren des EM-Algorithmus mit direkter Gauß-Hermite Integrationstechnik an.

6.1.1 Numerische Integration mit Gauß-Hermite

Ist ein Integral der Form

$$\int h(x) dx, \text{ mit } h(x) = f(x) \cdot g(x)$$

nicht analytisch lösbar, so kann es durch numerische Integration (Crouch & Spiegelman, 1990) approximiert werden. Wir beschränken uns im folgenden zunächst auf die Darstellung eines eindimensionalen Integrales, um die Notation übersichtlicher zu gestalten.

Bei der numerischen Integration wird das Integral durch eine gewichtete Summe approximiert:

$$\int_{\mathbb{R}} h(x) dx = \int_{\mathbb{R}} f(x) \cdot g(x) dx \approx \sum_{j=1}^M f(x_j) \cdot g_j .$$

Dabei bezeichnen g_j die Gewichte, x_j die Stützstellen sowie M die Anzahl der Stützstellen für die numerische Integration.

In unserem Fall entspricht $g(x)$ der Standardnormalverteilung, d. h. die betrachtete Funktion hat die Form

$$h^*(x) = f(x) \cdot \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2}\right) .$$

Für die Anwendung der Gauß-Hermite Integration (Abramowitz & Stegun, 1972), die Funktionen der Form

$$h_{gh}(x) = f(x) \cdot \exp\{-x^2\}$$

behandelt, müssen die Gewichte und Stützstellen transformiert werden. Deshalb substituieren wir $z = \frac{x}{\sqrt{2}}$ bzw. $dx = \sqrt{2} dz$ und erhalten

$$\begin{aligned} \int_{\mathbb{R}} h^*(x) dx &= \int_{\mathbb{R}} f(x) \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2}\right) dx \\ &= \int_{\mathbb{R}} \frac{1}{\sqrt{2\pi}} \sqrt{2} \exp(-z^2) f(\sqrt{2}z) dz \\ &\approx \sum_{j=1}^M \frac{1}{\sqrt{\pi}} g_j f(\sqrt{2}x_j) . \end{aligned}$$

Die Gewichte g_j und die Stützstellen x_j der Gauß-Hermite Integration, welche tabellarisiert vorliegen, werden mit $v_j = \frac{1}{\sqrt{\pi}} g_j$ und $d_j = \sqrt{2} x_j$ transformiert,

so daß das Integral mit

$$\int_{\mathbb{R}} h^*(x) dx \approx \sum_{j=1}^M v_j f(d_j)$$

approximiert wird.

Die Übertragung auf den mehrdimensionalen Fall erfolgt analog und wir erhalten

$$\int_{\mathbb{R}^q} h^*(x) dx \approx \sum_{j=1}^M v_j f(d_j) \quad \text{mit } x = (x_1, \dots, x_q)'$$

Dabei ist $j \in \{j_1, \dots, j_q\}$ ein Multiindex. Es wird insgesamt über $M = M_1 \cdot \dots \cdot M_q$ Stützstellen approximiert. Die Gewichte $v_j = (v_{j_1}, \dots, v_{j_q})$ und die zugehörigen Stützstellen $d_j = (d_{j_1}, \dots, d_{j_q})$ sind Elemente aus den Kartesischen Produkten

$$v_j \in V_1 \times \dots \times V_q \quad \text{bzw.} \quad d_j \in D_1 \times \dots \times D_q .$$

Dabei ist $V_l = \{v_1, \dots, v_{M_l}\}$ die Menge der transformierten Gewichte und $D_l = \{d_1, \dots, d_{M_l}\}$ die Menge der transformierten Stützstellen bei der Wahl von M_l Stützstellen für $l = 1, \dots, q$.

Das bedeutet, daß die Anzahl der Stützstellen bei mehrdimensionalen Integralen sehr groß wird. Soll z. B. ein dreidimensionales Integral approximiert werden, wobei in jeder Dimension 10 Stützstellen verwendet werden, so werden insgesamt $10 \cdot 10 \cdot 10 = 1000$ Stützstellen verwendet.

6.1.2 Schätzung der festen Parameter

Die Schätzung der festen Parameter erfolgt durch den EM-Algorithmus unter Verwendung der oben beschriebenen Gauß-Hermite Integrationstechnik.

Dichte

Die Beobachtungseinheiten $y_i = (y_{i1}, \dots, y_{in_i})$ sind für $i = 1, \dots, N$, gegeben β , Q und b_1, \dots, b_N , bedingt unabhängig. Damit gilt für die Dichte von $Y = (y_1, \dots, y_N)'$

$$f(Y|b_1, \dots, b_N; \beta) = \prod_{i=1}^N f(y_i|b_i; \beta) \text{ mit } f(y_i|b_i; \beta) = \prod_{t=1}^{n_i} f(y_{it}|b_i; \beta).$$

Die Schätzung der Parameter β und Q kann durch Maximierung der Marginalen Log-Likelihood

$$l(\beta, Q) = \sum_{i=1}^N \ln L_i(\beta, Q) \text{ mit } L_i(\beta, Q) = \int f(y_i|b_i; \beta) p(b_i, Q) db_i$$

durchgeführt werden. Dabei bezeichnet $p(b_i, Q)$ die Dichte von b_i , die als normalverteilt vorausgesetzt wurde.

Das Integral über diese Mischungsdichte kann nur für einige Spezialfälle analytisch gelöst werden, so daß für die Lösung des Integrales numerische Verfahren verwendet werden.

Bevor diese Verfahren angewendet werden können, wird der Modellansatz für den linearen Prädiktor

$$\eta_{it} = Z_{it}\beta + W_{it}b_i$$

umparametrisiert.

Umparametrisierung des Linearen Prädiktors

Für die zufälligen Effekte b_i hatten wir vorausgesetzt, daß $b_i \stackrel{iid}{\sim} \mathcal{N}(0, Q)$ gilt. Da Q eine Kovarianzmatrix ist, also positiv definit ist, läßt sich Q durch die Cholesky-Zerlegung in $Q = Q^{1/2}Q^{T/2}$ zerlegen, wobei $Q^{1/2}$ eine untere

Dreiecksmatrix ist und $Q^{T/2}$ die dazu transponierte obere Dreiecksmatrix. Die obere Dreiecksmatrix $Q^{1/2}$ läßt sich vektorisieren mit:

$$\theta = (q_{11}, \dots, q_{1r}, q_{22}, \dots, q_{2r}, \dots, q_{rr})' \text{ mit } \begin{pmatrix} q_{11} & q_{12} & \dots & q_{1r} \\ 0 & q_{22} & \dots & q_{2r} \\ \vdots & \dots & \ddots & \vdots \\ \vdots & \dots & \dots & q_{rr} \end{pmatrix} = Q^{1/2} .$$

Der Vektor $\theta = \text{vec}(Q^{1/2})$ enthält nur die Koeffizienten aus $Q^{1/2}$, welche nicht nach Konstruktion von $Q^{1/2}$ gleich 0 sind. Ist θ bekannt, so ist auch die Kovarianzmatrix Q eindeutig bestimmt.

Mit der Umparametrisierung $b_i = Q^{1/2} a_i$ und $a_i \stackrel{iid}{\sim} \mathcal{N}(0, I)$, erhalten wir

$$W_{it} b_i = W_{it} Q^{1/2} a_i = [a_i' \otimes W_{it}] \theta$$

unter Verwendung des Kroneckerproduktes \otimes .

Damit folgt für den linearen Prädiktor

$$\eta_{it} = [Z_{it} \quad , \quad a_i' \otimes W_{it}] \alpha$$

mit $\alpha = \begin{pmatrix} \beta \\ \theta \end{pmatrix}$ und $a_i \stackrel{iid}{\sim} \mathcal{N}(0, I)$.

Diesen Modellansatz für den linearen Prädiktor werden wir im folgenden verwenden, denn er hat zwei entscheidende Vorteile. Zum einen wurde der lineare Prädiktor in die übliche Form eines linearen Modells überführt. Durch eine Schätzung von α erhält man Schätzer für die beiden festen Parameter β und Q . Der zweite Vorteil ist, daß in der Mischungsdichte von Y die Dichte $p(b_i, Q)$, welche eine normalverteilte Dichte mit unbekannter Kovarianz ist, durch eine standardnormalverteilte Dichte, d. h. durch $p(a_i, I)$ ersetzt werden kann.

Schätzung von β und Q

Das Modell $\eta_{it} = Z_{it}\beta + W_{it}b_i$ mit $b_i \sim \mathcal{N}(0, Q)$ liegt nach der Parametrisierung in der Form:

$$\eta_{it} = [Z_{it}, a_i' \otimes W_{it}] \alpha \quad \text{mit } a_i \stackrel{iid}{\sim} \mathcal{N}(0, I)$$

mit $\alpha = \begin{pmatrix} \beta \\ \theta \end{pmatrix}$ und $\theta = \text{vec}(Q^{1/2})$ mit $Q = Q^{1/2}Q^{T/2}$ vor.

Falls die a_i bekannt sind, ist das Problem ein gewöhnliches Generalisiertes Lineares Modell. Da dies aber nicht der Fall ist, wenden wir den EM-Algorithmus an.

Der EM-Algorithmus, der auch Grundlage des EM-Type Algorithmus ist, der in Abschnitt 2.1.2 auf Seite 25 vorgestellt wurde ist ein iteratives Verfahren (Dempster, Laird & Rubin, 1977; Little & Rubin, 1987; McLachlan & Krishnan, 1997; Sundberg, 1974) für die Bestimmung eines Maximum Likelihood Schätzers in unvollständigen Datensituationen. Der Algorithmus besteht in jedem Iterationsschritt aus einem *Expectation-Schritt* (E-Schritt) und einem *Maximizing-Schritt* (M-Schritt).

E-Schritt: (Berechnung des bedingten Erwartungswertes)

$$M(\alpha | \alpha^{(k)}) = E \{ \ln f(Y, A | \alpha) \mid Y, \alpha^{(k)} \}$$

$$\text{mit } Y = \begin{pmatrix} y_1 \\ \vdots \\ y_N \end{pmatrix} \text{ und } A = \begin{pmatrix} a_1 \\ \vdots \\ a_N \end{pmatrix}.$$

Anschließend wird diese Log-Likelihood im M-Schritt maximiert, um einen neuen Schätzer $\alpha^{(p+1)}$ für α zu erhalten.

M-Schritt: (Maximierung von $M(\alpha, \alpha^{(k)})$)

$$\frac{\partial M(\alpha, \alpha^{(k)})}{\partial \alpha} = 0$$

Im EM-Algorithmus wird im E-Schritt der bedingte Erwartungswert

$$M(\alpha | \alpha^{(k)}) = E \{ \ln f(Y, A | \alpha) | Y, \alpha^{(k)} \}$$

mit $Y = \begin{pmatrix} y_1 \\ \vdots \\ y_N \end{pmatrix}$ und $A = \begin{pmatrix} a_1 \\ \vdots \\ a_N \end{pmatrix}$ berechnet.

Da wir im M-Schritt diesen Erwartungswert bezüglich α maximieren wollen, sind für uns konstante Terme oder Terme, die von α unabhängig sind, ohne Bedeutung.

Zunächst ist

$$\begin{aligned} M(\alpha | \alpha^{(k)}) &= E \{ \ln f(Y, A | \alpha) | Y, \alpha^{(k)} \} \\ &= \int \ln f(Y, A | \alpha) f(A | Y, \alpha^{(k)}) dA. \end{aligned}$$

Der erste Term des obigen Integrals lautet

$$\begin{aligned}
\ln f(Y, A|\alpha) &= \ln (f(Y|A, \alpha) g(A)) \\
&= \sum_{i=1}^N \ln f(y_i|a_i, \alpha) + \sum_{i=1}^N \ln g(a_i),
\end{aligned}$$

wobei $\sum_{i=1}^N \ln g(a_i)$ unabhängig von α ist.

Für den zweiten Term ist

$$\begin{aligned}
f(A|Y, \alpha^{(k)}) &= \frac{f(Y, A|\alpha^{(k)})}{f(Y|\alpha^{(k)})} \\
&= \frac{f(Y|A, \alpha^{(k)}) g(A)}{\int f(Y|A, \alpha^{(k)}) g(A) dA} \\
&= \frac{\prod_{i=1}^N f(y_i|a_i, \alpha^{(k)}) \prod_{i=1}^N g(a_i)}{\prod_{i=1}^N \int f(y_i|a_i, \alpha^{(k)}) g(a_i) da_i}.
\end{aligned}$$

Der Nenner, der von α abhängt, ist eine Konstante, die für die Maximierung nicht berücksichtigt werden muß.

Die zu maximierende Funktion läßt sich nun vereinfachen zu

$$\tilde{M}(\alpha|\alpha^{(k)}) = \sum_{i=1}^N \int \ln f(y_i|a_i, \alpha) f(y_i|a_i, \alpha^{(k)}) g(a_i) da_i.$$

Durch Gauß-Hermite Integration wird das Integral durch folgende Summe approximiert

$$\tilde{M}(\alpha, \alpha^{(k)}) \approx M^{GH}(\alpha, \alpha^{(k)}) = \sum_{i=1}^N \sum_{j=1}^M \ln f(y_i|d_j, \alpha) \frac{v_j f(y_i|d_j, \alpha^{(k)})}{\sum_{s=1}^M v_s f(y_i|d_s, \alpha^{(k)})}$$

Die Gewichte v_j und die Stützstellen d_j sind die transformierten Gewichte bzw. Stützstellen der Gauß-Hermite Integration, wie sie im vorherigen Abschnitt definiert wurden.

Wir definieren

$$c_{ij} = \frac{v_j f(y_i|d_j, \alpha^{(k)})}{\sum_{s=1}^M v_s f(y_i|d_s, \alpha^{(k)})}$$

und maximieren im M-Schritt $M^{GH}(\alpha|\alpha^{(k)})$, durch Nullsetzen der Ableitung $M^{GH}(\alpha|\alpha^{(k)})$, d. h.

$$\frac{\partial M^{GH}(\alpha|\alpha^{(k)})}{\partial \alpha} = \sum_{i=1}^N \sum_{j=1}^M c_{ij} \frac{\partial \ln f(y_i|d_j, \alpha)}{\partial \alpha} \stackrel{!}{=} 0$$

Da $\alpha = (\beta, \theta)'$ ist, folgt $\frac{\partial \ln f(y_i|d_j, \alpha)}{\partial \alpha} = \begin{pmatrix} \frac{\partial \ln f(y_i|d_j, \beta)}{\partial \beta} \\ \frac{\partial \ln f(y_i|d_j, \theta)}{\partial \theta} \end{pmatrix}$ und damit entspricht $\frac{\partial M^{GH}(\alpha|\alpha^{(k)})}{\partial \alpha}$ der Scorefunktion dem mit c_{ij} gewichtete GLM:

$$\eta_{itj} = [Z_{it}, d_j' \otimes W_{it}] \begin{pmatrix} \beta \\ \theta \end{pmatrix}$$

mit $h(\eta_{itj}) = E(y_{it})$ für $j = 1, \dots, M$.

Die Parameter β und θ werden mit einem gewichteten GLM geschätzt, wobei die Designmatrix $[Z_{it}, d_j' \otimes W_{it}]$ neben den Regressoren auch noch die Stützstellen für die Gauß-Hermite Integration enthält.

Die Responsematrix wird mit $y_{itj} = y_{it}$ für $j = 1, \dots, M$ entsprechend angepaßt. Jede Beobachtung y_{itj} wird für $t = 1, \dots, n_i$ bei der Schätzung des

GLM's mit

$$c_{ij} = c_{ij}(\alpha^{(k)}) = \frac{v_j f(y_i|d_j, \alpha^{(k)})}{\sum_{s=1}^M v_s f(y_i|d_s, \alpha^{(k)})}$$

gewichtet.

6.2 EM-Type Algorithmus

6.2.1 Blockdiagonalgestalt der Fishermatrix

Da beim EM-Type Algorithmus die Fishermatrix in Blockdiagonalgestalt vorliegt, kann diese Struktur für die Parameterschätzung mit dem Fisher-Scoring Algorithmus ausgenutzt werden.

Mit

$$F_{\beta\beta} = \begin{bmatrix} F_{\beta^1\beta^1} & & \mathbb{O} \\ & \ddots & \\ \mathbb{O} & & F_{\beta^T\beta^T} \end{bmatrix} \quad \text{und} \quad F_{\beta b_i} = \begin{bmatrix} F_{\beta^1 b_i} \\ \vdots \\ F_{\beta^T b_i} \end{bmatrix} = F'_{b_i\beta}$$

folgt:

$$F(\delta)^{\langle p \rangle} = \begin{bmatrix} F_{\beta\beta} & F_{\beta b_1} & \cdots & F_{\beta b_N} \\ F_{b_1\beta} & F_{b_1 b_1} & & \mathbb{O} \\ \vdots & & \ddots & \\ F_{b_N\beta} & \mathbb{O} & & F_{b_N b_N} \end{bmatrix}$$

Ist $\delta^{\langle p+1 \rangle} = \delta^{\langle p \rangle} + \Delta\delta^{\langle p \rangle}$, so folgt beim Fisher-Scoring Algorithmus:

$$\delta^{\langle p+1 \rangle} = \delta^{\langle p \rangle} + F^{-1}(\delta^{\langle p \rangle}) s(\delta^{\langle p \rangle}),$$

daß $s(\delta^{\langle p \rangle}) = F(\delta^{\langle p \rangle}) \Delta\delta^{\langle p \rangle}$ gilt.

Da die Fishermatrix Blockdiagonalgestalt hat, gilt

$$s(\beta^{\langle p \rangle}) = F_{\beta\beta} \Delta\beta^{\langle p \rangle} + \sum_{i=1}^N (F_{\beta b_i} \Delta b_i^{\langle p \rangle})$$

$$s(b_i^{\langle p \rangle}) = F_{b_i\beta} \Delta\beta^{\langle p \rangle} + F_{b_i b_i} \Delta b_i^{\langle p \rangle}$$

Auflösen der 2. Gleichung nach $\Delta b_i^{(p)}$ führt zu

$$\Delta b_i^{(p)} = F_{b_i b_i}^{-1} \left\{ s \left(b_i^{(p)} \right) - F_{b_i \beta} \Delta \beta^{(p)} \right\}$$

und mit Einsetzen von $\Delta b_i^{(p)}$ in die 1. Gleichung folgt

$$\begin{aligned} s \left(\beta^{(p)} \right) &= F_{\beta \beta} \Delta \beta^{(p)} + \sum_{i=1}^N \left(F_{\beta b_i} F_{b_i b_i}^{-1} \left\{ s \left(b_i^{(p)} \right) - F_{b_i \beta} \Delta \beta^{(p)} \right\} \right) \\ &= F_{\beta \beta} \Delta \beta^{(p)} + \sum_{i=1}^N \left(F_{\beta b_i} F_{b_i b_i}^{-1} s \left(b_i^{(p)} \right) \right) - \Delta \beta^{(p)} \sum_{i=1}^N \left(F_{\beta b_i} F_{b_i b_i}^{-1} F_{b_i \beta} \right). \end{aligned}$$

So daß für $\Delta \beta^{(p)}$ gilt:

$$\Delta \beta^{(p)} = \left\{ F_{\beta \beta} - \sum_{i=1}^N \left(F_{\beta b_i} F_{b_i b_i}^{-1} F_{b_i \beta} \right) \right\}^{-1} \left\{ s \left(\beta^{(p)} \right) - \sum_{i=1}^N \left(F_{\beta b_i} F_{b_i b_i}^{-1} s \left(b_i^{(p)} \right) \right) \right\}.$$

Zusammenfassung

In vielen Studien werden die Beobachtungseinheiten über einen längeren Zeitraum betrachtet. Es liegt somit einerseits eine gruppen- und/oder individuenspezifische Datenstruktur vor und andererseits muß auch ein möglicher zeitlicher Einfluß berücksichtigt werden. In dieser Arbeit wurden verschiedene Modellansätze vorgestellt, die den gruppen- bzw. individuenspezifischen Einfluß durch die Aufnahme von zufälligen Effekten berücksichtigen und gleichzeitig einen zeitlichen Einfluß durch die Betrachtung von zeitvariierenden Koeffizienten ermöglichen.

In den Simulationsstudien und Anwendungsbeispielen hat sich gezeigt, daß der Lokale Likelihoodansatz, der durch die Aufnahme von zufälligen Effekten erweitert wurde, von allen vorgestellten Modellansätzen am besten geeignet ist.

Bei diesem Ansatz werden durch eine Kernfunktion die einzelnen Beobachtungen in Abhängigkeit des betrachteten Beobachtungszeitpunktes unterschiedlich gewichtet, so daß durch die Wahl des Glättungsparameters der Kernfunktion ein sehr flexibler Glättungsansatz möglich ist. Der optimale Glättungsparameter kann durch Kreuzvalidierung, die jedoch mit großem Rechenaufwand verbunden ist, bestimmt werden.

Selbstverständlich müssen die Varianzen der Parameterschätzungen für deren Interpretation immer berücksichtigt werden. Jedoch können durch die Be-

trachtung der punktweisen Konfidenzintervalle alle Parameterschätzungen, die in Abhängigkeit des Glättungsparameters erzielt wurden, inhaltlich interpretiert werden, so daß auch Parameterschätzungen interpretiert werden können, die nicht mit dem optimalen Glättungsparameter bestimmt wurden. Damit ist es durch den Lokalen Likelihoodansatz möglich, einerseits durch eine geringe Glättung die einzelnen betrachteten Zeitpunkte separat zu analysieren und andererseits durch eine starke Glättung einen globalen Trendüberblick über den gesamten Zeitraum zu erhalten.

Beispiele für die Auswirkungen unterschiedlicher Glättungsparameter wurden im Emesis Datensatz in Abschnitt 5.1.1 gezeigt.

Beim Lokalen Likelihoodansatz mit zufälligen Effekten wurde im Abschnitt 3.1.2 zwei Verfahren, der EM-Algorithmus mit Gauß-Hermite Integrations-technik und der EM-Type Algorithmus, für die Bestimmung der Parameterschätzwerte angegeben. Der EM-Algorithmus mit Gauß-Hermite Integrationstechnik sollte trotz seines großen Rechenaufwands aufgrund der betrachteten Integrationsstützstellen, bevorzugt angewendet werden, da bei diesem Verfahren auch bei einer deutlichen Ausprägung der Varianz der gruppen- bzw. individuenspezifischen Einflüsse die Parameterschätzungen durchgeführt werden können. Im Simulationsbeispiel 4.1.3 wurde gezeigt, daß der EM-Type Algorithmus in diesem Fall nicht geeignet ist, da die Schätzungen der zufälligen Effekte unzureichend sind.

Für den EM-Algorithmus mit Gauß-Hermite Integrationstechnik, konnte im Simulationsbeispiel 4.1.2 als auch in den Anwendungsbeispielen in Kapitel 5 gezeigt werden, daß dieses Verfahren gut geeignet ist, die Parameterschätzungen selbst und die Bestimmung ihrer Varianzen durchzuführen.

Für den in Abschnitt 2.1.3 vorgestellten Modellansatz der Nicht Parametrischen Maximum Likelihood konnte keine zufriedenstellende Erweiterung

durch die Berücksichtigung von zeitvariierenden Koeffizienten entwickelt werden.

Ein weiterer sehr einfacher Ansatz für die Berücksichtigung von zeitvariierenden Koeffizienten und zufälligen Effekten ist der Marginale Likelihoodansatz. Bei diesem Ansatz kann durch den modellbedingten Aufbau der Designmatrix nur eine geringe Anzahl von betrachteten Zeitpunkten berücksichtigt werden, da sonst in der Regel numerische Probleme bei der Bestimmung der Parameterschätzungen entstehen. Deshalb kann dieser Modellansatz in der Praxis nur sehr beschränkt eingesetzt werden.

Der Vorteil des flexiblen Lokalen Likelihoodansatzes gegenüber der parametrischen und semi-parametrischen Ansätze wurde in den Anwendungsbeispielen für die Dauer der Arbeitslosigkeit in Abschnitt 5.2.1 und der Magenkrebsstudie in Abschnitt 5.2.2 dargestellt werden. Wird keine dem Modell entsprechende parametrische Form gewählt, so können sehr leicht falsche Rückschlüsse getroffen werden.

Zusätzlich wurde in Abschnitt 5.1.2 am Beispiel der epileptischen Anfälle deutlich, daß die Bestimmung der Varianzen der Parameterschätzungen mit der Delta-Methode bei der Polynomialen Regression aus numerischen Gründen nicht immer möglich ist.

Literatur

- Abramowitz, M. und Stegun, I. (1972). *Handbook of Mathematical Functions*. New York: Dover.
- Agresti, A. (1990). *Categorical Data Analysis*. New York: Wiley.
- Aitkin, M. (1995). NPML estimation of the mixing distribution in general statistical models with unobserved random effects. (*working-paper*).
- Aitkin, M. (1996). A general maximum likelihood analysis of overdispersion in generalized linear models. *Statistics and Computing* 6, 251–262.
- Aitkin, M. und Aitkin, I. (1996). A hybrid EM/Gauss-Newton algorithm for maximum likelihood in mixture distributions. (*working-paper*) .
- Aitkin, M. und Francis, B. (1995). Fitting overdispersed generalized linear models by nonparametric maximum likelihood. *GLIM Newsletters* 25, 37–45.
- Aitkin, M. und Wilson, G. (1980). Mixture models, outliers, and the EM algorithm. *Technometrics* 22(3), 325–331.
- Booth, J. und Hobert, J. (1999). Maximizing generalized linear mixed models with an automated monte carlo em alorithm. *Journal of the Royal Statistical Society Ser. B* 61, 265–285.
- Breslow, N. und Clayton, D. (1993). Approximate inference in generalized linear mixed models. *Journal of the American Statistical Associati-*

on 88(421), 9–25.

- Crouch, E. und Spiegelman, D. (1990). The evaluation of integrals of the form $\int_{-\infty}^{+\infty} f(t)\exp(-t^2)dt$: Application to logistic-normal models. *Journal of the American Statistical Association* 85(410), 464–469.
- Davis, C. (1991). Semi-parametric and non-parametric methods for the analysis of repeated measurements with applications to clinical trials. *Statistics in Medicine* 10, 1959–1980.
- Dempster, A., Laird, N., und Rubin, D. (1977). Maximum likelihood from incomplete data via the EM-algorithm. *Journal of the Royal Statistical Society B* 39, 1–38.
- Dietz, E. und Böhning, D. (1994). Analysis of longitudinal data using finite mixture models. *Statistische Hefte* 35, 203–210.
- Diggle, P., Liang, K., und Zeger, S. (1994). *Analysis of Longitudinal Data*. Oxford University Press.
- Efron, B. (1988). Logistic regression, survival analysis, and the Kaplan-Meier-curve. *Journal of the American Statistical Association* 83, 414–425.
- Eilers, P. und Marx, B. (1996). Flexible smoothing with B-splines and penalties. *Statistical Science* 11 (2), 89–121.
- Eubank, R. (1988). *Spline Smoothing and Nonparametric Regression*. New York: Marcel Dekker.
- Everitt, B. und Hand, D. (1981). *Finite Mixture Distributions*. London: Chapman and Hall.
- Fahrmeir, L. (1994). Dynamic modelling and penalized likelihood estimation for discrete time survival data. *Biometrika* 81 (2).
- Fahrmeir, L., Hamerle, A., und Tutz, G. (1994). *Multivariate statistische Verfahren*. Berlin: de Gruyter.

- Fahrmeir, L. und Tutz, G. (1994). *Multivariate Statistical Modelling Based on Generalized Linear Models*. New York: Springer Verlag.
- Fan, J., Hall, P., Martin, M., und Patil, P. (1996). On local smoothing of nonparametric curve estimators. *Journal of the American Statistical Association* 91(433).
- Friedman, J. (1991). Multivariate adaptive regression splines. *The Annals of Statistics* 19, 1–141.
- Geyer, C. und Thompson, E. (1992). Constrained monte carlo maximum likelihood for depend data. *Journal of the Royal Statistical Society Ser. B* 54, 657–683.
- Gourieroux, C. und Montfort, A. (1989). Simulation based inference in models with heterogeneity. *Document de Travail INSEE/ENSAE No. 8902*.
- Hanefeld, U. (1987). *Das sozio-ökonomische Panel*. Frankfurt: Campus.
- Hastie, T. und Loader, C. (1993). Local regression: Automatic kernel carpentry. *Statistical Science* 8(2), 120–143.
- Hastie, T. und Tibshirani, R. (1990). *Generalized Additive Models*. London: Chapman and Hall.
- Hastie, T. und Tibshirani, R. (1993). Varying-coefficient models. *Journal of the Royal Statistical Society B* 55(4), 757–796.
- Hennevogl, W. (1991). *Schätzung generalisierter Regressions- und Zeitreihenmodelle mit varriierenden Parametern*. Dissertation, Universität Regensburg.
- Hinde, J. (1982). Compound poisson regression models. In R. Gilchrist (Hrsg.), *GLIM'82 Internat. Conf. Generalized Linear Models*, New York, S. 109–121. Springer.

- Jansen, R. (1993). Maximum likelihood in a generalized linear finite mixture model by using the EM algorithm. *Biometrics* 49, 227–231.
- Kauermann, G. und Tutz, G. (1995). Local likelihood estimation and bias reduction in varying coefficient models. Forschungsbericht, Technische Universität Berlin. 95-9.
- Kauermann, G. und Tutz, G. (2000). Local likelihood estimation in varying-coefficient models including additive bias correction. *Journal of Nonparametric Statistics* 12, 343–371.
- Laird, N. M. und Louis, T. A. (1982). Approximate posterior distributions for incomplete data problems. *Journal of the Royal Statistical Society B* 44(2), 190–200.
- Little, R. und Rubin, D. (1987). *Statistical Analysis with Missing Data*. New York: John Wiley & Sons.
- Longford, N. (1993). *Random Coefficient Models*. Oxford: Clarendon Press.
- McCullagh, P. (1980). Regression model for ordinal data (with discussion). *Journal of the Royal Statistical Society B* 42, 109–127.
- McCullagh, P. und Nelder, J. (1989). *Generalized Linear Models* (2 Aufl.). London, New York: Chapman and Hall.
- McCulloch, C. (1994). Maximum likelihood variance components estimation for binary data. *Journal of the American Statistical Association* 89(425), 330–335.
- McCulloch, C. (1997). Maximum likelihood algorithms for generalized linear mixed models. *Journal of the American Statistical Association* 92, 162–170.
- McCulloch, C. (2000). An introduction to generalized linear mixed models. (*working-paper*).

- McLachlan, G. und Basford, K. (1988). *Mixture Models. Inference and Applications to Clustering*. New York: Marcel Dekker.
- McLachlan, G. und Krishnan, T. (1997). *The EM Algorithm and Extensions*. New York: Wiley.
- Meng, X.-L. (1997). The EM algorithm and medical studies: a historical link. *Statistical Methods in Medical Research* 6, 3–23.
- Meng, X.-L. und van Dyk, D. (1997a). The EM algorithm - an old folk-song sung to a fast new time. *B* 59(3), 511–567.
- Meng, X.-L. und van Dyk, D. (1997b). Fast EM-type implementations for mixed-effects models. *Journal of the Royal Statistical Society*. (Submitted).
- Nelder, J. A. und Wedderburn, R. W. M. (1972). Generalized linear models. *Journal of the Royal Statistical Society A* 135, 370–384.
- Quintana, R., Lui, J., und Pino, G. (1999). Monte carlo em with importance reweighting and its application in random effects models. *Computational Statistics and Data Analysis* 29, 429–444.
- Rice, J. und Silverman, B. (1991). Estimating the mean and covariance structure nonparametrically when the data are curves. *Journal of the Royal Statistical Society B* 53(1), 233–243.
- Ruppert, D., Reish, R., und Carroll, R. (1984). Optimization using stochastic approximation and monte carlo simulation (with application to harvesting of atlantic menhaden. *Biometrics* 40, 535–545.
- Silverman, B. (1984). Spline smoothing: The equivalent variable kernel method. *The Annals of Statistics* 12(3), 898–916.
- Stone, C., Hansen, M., Kooperberg, C., und Truong, Y. (1997). Polynomial splines and their tensor products in extended linear models. *The Annals of Statistics* 25, 1371–1470.

- Sundberg, R. (1974). Maximum likelihood theory for incomplete data from an exponential family. *Scand. J. Statist.* 1, 49–58.
- Thall, P. und Vail, S. (1990). Some covariance models for longitudinal count data with overdispersion. *Biometrics* 46, 657–671.
- Titterington, D., Smith, A., und Makov, U. (1985). *Statistical Analysis of Finite Mixture Distributions*. New York: Wiley.
- Tutz, G. (1989). On cross-validation for discrete kernel estimates in discrimination. *Communications in Statistics, Theory and Methods* 11, 4145–4162.
- Tutz, G. (1999). Varying coefficients in generalized linear random effects models: A local likelihood approach. *Discussion Paper, Ludwig-Maximilian-Universität München, Sonderforschungsbereich 386* 171.
- Tutz, G. und Hennevogl, W. (1996). Random effects in ordinal regression models. *Computational Statistics and Data Analysis* 22, 537–557.
- Tutz, G. und Kauermann, G. (1995). Varying coefficients in multivariate generalized linear models: a local likelihood approach. *Forschungsberichte des Fachbereiches Informatik, TU-Berlin*.
- Tutz, G. und Kauermann, G. (1997). Local estimators in multivariate generalized linear models with varying coefficients. *Computational Statistics* 12, 193–208.
- van der Linde, A. (1994). On cross-validation for smoothing splines in the case of dependent observations. *36*(1), 67–73.
- Wu, C. (1983). On the convergence properties of the EM algorithm. *The Annals of Statistics* 11, 95–103.