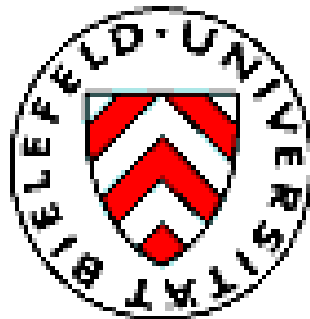


Flexible Modelle für die Verweilzeiten:  
P-Spline basierte Methoden

Dissertation

zur Erlangung des akademischen Grades: Dr. rer. pol.  
der Fakultät für Wirtschaftswissenschaften  
der Universität Bielefeld

Dipl.-Stat. Pavel Khomski



Gutachter: Prof. Dr. G. Kauermann  
Prof. Dr. J. Frohn

Bielefeld, im April 2007

*Data and models are two sources of information in a statistical analysis. Data carry noise but are "unbiased", whereas models, effectively a set of constraints, help reduce noise but are responsible for "biases".*

*(Chong Gu: Smoothing Spline ANOVA Models, 2002, Springer)*

# Inhaltsverzeichnis

<b>1</b>	<b>Einleitung</b>	<b>1</b>
<b>2</b>	<b>Theoretische Grundlagen</b>	<b>3</b>
2.1	Kurzer Abriss des klassischen Cox-Modells . . . . .	3
2.2	Flexibilisierung des Proportional-Hazard Modells . . . . .	7
2.3	Spline-gestützte Modellierung . . . . .	8
2.4	Idee der penalisierten Likelihood . . . . .	12
2.5	Flexible Modellierung mit P-Splines . . . . .	14
2.6	Splinebasierte Konzepte in der Analyse von Verweildauern: Literatur- überblick . . . . .	18
<b>3</b>	<b>Modellierungsstrategie</b>	<b>21</b>
3.1	Gestalt der Likelihoodfunktion . . . . .	21
3.2	Approximation der Likelihood; Poisson -Modell . . . . .	23
3.3	Wahl der Penalisierungsmatrix . . . . .	26
3.4	GLMM-Darstellung des Modells . . . . .	27
3.5	Regulierung des Smoothing-Effektes; Freiheitsgrade der Modell- komponenten . . . . .	29
3.6	Varianz der Modellparameter . . . . .	30
<b>4</b>	<b>Anwendungsmodelle: Überblick</b>	<b>31</b>
<b>5</b>	<b>Two-Way Additives Hazard-Modell mit Kalendereffekten</b>	<b>33</b>
5.1	Modellansatz . . . . .	34
5.2	Schätzalgorithmus . . . . .	35
5.3	Berechnung der Penalisierungsparameter . . . . .	38
5.4	Bestimmung von Varianzen und Freiheitsgraden . . . . .	39
5.5	Simulationsstudie . . . . .	40
5.6	Anwendungsbeispiel . . . . .	43
5.7	Diskussion zur Weiterentwicklung . . . . .	45
<b>6</b>	<b>Two-Way Additives Hazard-Modell mit saisonaler Zeitkom- ponente</b>	<b>48</b>
6.1	Modellansatz und Schätzstrategie . . . . .	49
6.2	Simulationsstudie . . . . .	53
6.3	Auswertung des Datenmaterials . . . . .	54

<b>7</b>	<b>Two-Way-Interaction Hazard-Modell</b>	<b>57</b>
7.1	Modell-Likelihood . . . . .	57
7.2	Berechnung der Smoothing-Parameter . . . . .	59
7.3	Anwendung an die Mortalitätsdaten . . . . .	61
<b>8</b>	<b>Competing-Risks-with-Frailties Modell</b>	<b>63</b>
8.1	Allgemeiner Überblick . . . . .	63
8.2	Identifizierbarkeit; Spezifizierung der Likelihood . . . . .	65
8.3	Frailty als unbeobachtete Heterogenität . . . . .	67
8.4	Wahl der glatten Komponenten; Approximation der Likelihood . . . . .	69
8.5	Spezifizierung der Frailty-Verteilung . . . . .	71
8.6	Schätzroutine; EM-Algorithmus . . . . .	73
8.6.1	Wahl der Penalisierungsmatrix . . . . .	77
8.7	GLMM-Darstellung; Bestimmung der Penalisierungsparameter . . . . .	79
8.8	Kovarianzen der Modellparameter . . . . .	81
8.9	Simulationen . . . . .	83
8.10	Anwendungsbeispiel . . . . .	86
<b>9</b>	<b>Multi-Zustandsmodell</b>	<b>91</b>
<b>10</b>	<b>Zusammenfassung und Ausblick</b>	<b>95</b>
<b>11</b>	<b>Datenmaterial</b>	<b>97</b>
11.1	Arbeitslosigkeitsdaten des SOEP . . . . .	97
11.2	Kindersterblichkeiten . . . . .	99
<b>12</b>	<b>R-package zur Schätzung der glatten Komponenten</b>	<b>103</b>
<b>A</b>	<b>Mathematische und statistische Basics</b>	<b>106</b>
A.1	Grundlegende Begriffe der Analyse von Verweildauern . . . . .	106
A.2	Mathematische Funktionen . . . . .	107
A.3	Vektor- und Matrixableitungen . . . . .	108
A.4	Verteilungen . . . . .	109
A.5	Laplace-Approximation . . . . .	110
A.6	Laplace-Approximation mit nichtinvertierbarer Penalisierungsmatrix . . . . .	110
A.7	I. und II. Ableitungen zur Bestimmung des Smoothing-Parameters . . . . .	112
A.8	Grundriss der Bayes-Inferenz . . . . .	113
A.9	Grundriss des EM-Algorithmus . . . . .	114
A.10	Trapezoide Approximation des Integrals . . . . .	115
A.11	Grundriss der Optimierungsstrategie für mixture-Gewichte in Competing Risks . . . . .	116
<b>B</b>	<b>Simulationsergebnisse</b>	<b>118</b>
B.1	Simulation mit truncated polynomials . . . . .	118
B.2	Simulation mit B-splines . . . . .	128
B.3	Simulation für das Modell der Competing Risks . . . . .	138
B.3.1	Simulation auf Monatsbasis . . . . .	138



B.3.2 Simulation auf Tagesbasis . . . . .	151
<b>Literatur</b>	<b>162</b>

# Kapitel 1

## Einleitung

Diese Untersuchung konzentriert sich auf die flexible Modellierung von Überlebenszeiten (im Angelsächsischen bekannt als "Survival Modelling") und speziell von Kovariableneffekten. Als Ausgangspunkt der Modellierung dient der so genannte "Cox-Proportional-Hazard"-Ansatz. In diesem Modell wird der proportionale Effekt der Kovariablen auf die Hazardrate postuliert. Diese Annahme ist aber zu restriktiv, da sie in vielen verschiedenen Anwendungen nicht erfüllt ist. Damit bietet sich eine Reihe von möglichen Flexibilisierungen, vgl. hierzu z. B. Therneau & Grambsch (2000) mit zahlreichen Referenzen darin.

Zur flexiblen Modellierung der Kovariableneffekte werden die so genannten P-Splines benutzt, welche in dieser Notation von Eilers & Marx (1996) eingeführt wurden und sich in letzter Zeit zunehmender Beliebtheit erfreuen. Ihre grundlegenden Eigenschaften sind bekannt, und in vielen Fachartikeln wird über gute Erfahrungen im Modellieren damit berichtet. (Eine komparative Studie bezüglich der Wahl der jeweiligen P-Splines geben Eilers & Marx, 2005.) Der Entwicklungsstand der Modellierung mit P-Splines wirft aber noch viele offene Fragen auf. Wir geben eine zusammenfassende Übersicht und einige Literaturhinweise im Zusammenhang mit der flexiblen Modellierung mit Splines (und insbesondere mit P-Splines) in Kapitel 2. Kapitel 3 erläutert die theoretischen Grundlagen für die spätere Entwicklung und stützt sich in seiner Ausführung auf die Arbeit von Kauermann (2005b).

Die in dieser Arbeit zu präsentierenden Modelle sind aus der Forschung am Lehrstuhl für Statistik, Universität Bielefeld, entstanden. Jedes Modell wird in einzelnen Kapiteln dargestellt und ausführlich diskutiert. Es sind dies unter anderem das Modell der "Two-Way" additiven Hazards (Kapitel 5 und 6) und der "Competing Risks" (Kapitel 8). Eine Weiterentwicklung des erstgenannten Modells ist das Modell mit Wechselwirkungen, das in Kapitel 7 geschildert wird. Auf der Vorgehensweise des Modells der Competing Risks baut das Multi-Zustandsmodell, das in Kapitel 9 beschrieben wird. Dieses Modell befindet sich aber immer noch in einer Entwicklungsphase, so dass es nur skizzenhaft darzustellen ist. Darüber hinaus sind einige Erweiterungen der Grundmodelle möglich, welche, neben einer allgemeinen Diskussion, im

abschließenden Kapitel 10 erläutert werden.

In Kapitel 11 ist das zur Auswertung herangezogene Datenmaterial beschrieben. Dies sind die Daten über die Zeiten von Arbeitslosigkeiten von Personen, zur Verfügung gestellt vom Deutschen Institut für Wirtschaft (DIW). In Kooperation mit dem Institut für Tropenhygiene, Universität zu Heidelberg, wurden zur Analyse außerdem die Mortalitätsdaten von Kindern im Westafrika (Burkina Faso) herangezogen. Kapitel 12 präsentiert die für das Modellieren entwickelte Software in R-Umgebung (siehe R Development Core Team, 2006).

In zunehmendem Maß gilt das Forschungsinteresse dem Einsatz von Hazard-Modellen in den Wirtschaftsbereichen, wie CRM ("customer relationship management" bezüglich der "customer life time" und dem Abwanderungsverhalten von Kunden, siehe z. B. Krafft, 2002, und Hüppelshäuser, Krafft & Rüger, 2006) und in Banken ("default"-Modelle beim Kreditausfall, siehe Stepanova & Thomas, 2002, sowie Shumway, 2001). Im erstgenannten Bereich wird der Autor dieser Dissertation seine mehr auf die Praxis ausgerichteten Aktivitäten fortsetzen und hofft dabei auf das Interesse eines Unternehmens an der Erprobung und Anwendung der Modelle dieses breiten und herausfordernden wissenschaftlichen Gebiets. "Last but not least" gilt mein persönlicher Dank meinem Betreuer, Herrn Prof. Dr. Kauermann, ohne dessen ständige Hilfe und Unterstützung diese Arbeit niemals zustande gekommen wäre.

# Kapitel 2

## Theoretische Grundlagen

### 2.1 Kurzer Abriss des klassischen Cox-Modells

Der wohl am weitesten verbreitete (semi-)parametrische Ansatz in der Analyse von Überlebenszeiten geht auf Cox (1972) zurück und hat in seiner Popularität bis heute kaum etwas verloren, auch wenn das zugrunde gelegte Modell mit einigen erheblichen restriktiven Annahmen einhergeht, da die Hazardrate  $h(t; \mathbf{x})$  in der folgenden Form spezifiziert wird (grundlegende Definitionen siehe Anhang A.1):

$$h(t; \mathbf{x}) = h_0(t) \exp\{\mathbf{x}^T \boldsymbol{\beta}\}. \quad (2.1)$$

Hier ist  $t$  die Überlebenszeit,  $\mathbf{x}$  der Kovariablenvektor, fix oder zeitabhängig, mit dem zugehörigen fixen (d. h. von der Zeit  $t$  unabhängigen) Effektenvektor  $\boldsymbol{\beta}$ . Mit  $h_0$  ist die Baseline-Hazardrate bezeichnet, welche dem Fall ohne Kovariablen entspricht, also dem Intercept des Modells. Wegen seiner Spezifikation wird das Modell auch "Cox-Proportional-Model" genannt, was man unmittelbar aus der Form von (2.1) sehen kann. Bildet man hier das Verhältnis zweier Hazards, auch relatives Risiko genannt, jeweils auf zwei verschiedenen Levels des Kovariablenvektors, und nutzt die Unabhängigkeitsannahme der Effekte  $\boldsymbol{\beta}$  von der Zeit  $t$ , ergibt sich, wegen der Kürzung der zeitabhängigen Baseline-Hazards, die Proportionalität der Hazards. Den Einfluss der Kovariablen und deren Interpretation kann man auch aus der logarithmierten Form des Modells,

$$\log\{h(t; \mathbf{x})\} = \log\{h_0(t)\} + \mathbf{x}^T \boldsymbol{\beta}, \quad (2.2)$$

erkennen, die in ihrer Struktur einem linearen Modell ähnlich ist. Das Modell und die damit verbundenen statistischen Probleme, wie etwa Spezifikation der Likelihood, Schätzung der Parameter, Modelldiagnostik und Hypothesenüberprüfung, werden in zahlreichen Standardwerken ausführlich behandelt, als Auswahl seien z. B. Cox & Oakes (1984), Klein & Moeschberger (1997), Kalbfleisch & Prentice (2002) und Lawless (2003) erwähnt. Die Schätzmethode

(für rechtszensierte Daten) erfolgt im Kontext der "partiellen Likelihood". Das allgemeine Konzept hierfür ist von Cox (1975) entwickelt worden (siehe auch Wong, 1986, zur Weiterentwicklung dieses Prinzips). Die Konstruktion einer solchen partiellen Likelihood erlaubt im Modell der (rechtszensierten) Überlebenszeiten eine völlig unspezifizierte Form des Intercept-Terms, also der Baseline-Hazard. Im Rahmen der Schätzmethode spielt diese die Rolle eines Nuisance-Parameters. Das Interesse gilt vor allem dem Parameter  $\beta$ . Mit den Bezeichnungen  $\{(T_j, \delta_j; \mathbf{x}_j)\}_{j=1, \dots, N}$  für die beobachtete Überlebenszeit  $T_j$ , den Zensierungsindikator  $\delta_j$  und den (fixen) Kovariablenvektor  $\mathbf{x}_j$  der  $j$ -ten Beobachtung resultiert zunächst im Fall ohne Bindungen für exakt  $K$  Ereignisse (oder Events) die partielle Likelihood als

$$\mathcal{L}(\beta) = \prod_{i=1}^K \frac{\exp\{\beta^T \mathbf{x}_i\}}{\sum_{j \in R(t_i)} \exp\{\beta^T \mathbf{x}_j\}}, \quad (2.3)$$

wobei  $R(t_i)$  die Risikomenge zum Zeitpunkt  $t_i$  bezeichnet, also alle Beobachtungen, die bis unmittelbar vor dem Zeitpunkt  $t_i$  noch kein Event erfahren haben und nicht aus der Untersuchung ausgeschieden sind. Daraus sind die Ableitungen der *log*-Likelihood leicht zu bestimmen, um iterativ (z. B. nach Newton-Raphson) die Lösung  $\hat{\beta}$  zu berechnen.

Die Form der resultierenden partiellen Likelihood kann für Überlebenszeiten auf mehrere Arten hergeleitet werden. Breslow (1972) schätzt simultan den Parameter  $\beta$  und die Baseline  $h_0(t)$ , indem die Funktion  $h_0(t)$  als stückweise konstant zwischen den Events angenommen wird. In Johansen (1983) resultiert sie als profile-Likelihood, indem, ausgehend von der Likelihood der Daten mit zunächst fixem Parameter  $\beta$ , über die Baseline-Hazard maximiert wird, und der so gewonnene Schätzer für die kumulative Hazard in die Likelihoodfunktion eingesetzt wird. Diese profile-Likelihood ist dann bezüglich  $\beta$  proportional zur partiellen Likelihood. Im Fall der Bindungen (die bei stetiger Zeit streng genommen ausgeschlossen sind und nur wegen des Messniveaus vorkommen können), also mehrerer, zu einem bestimmten Zeitpunkt beobachteter Events, kennt man alternative Ansätze, z. B. von Breslow (1974) und von Efron (1977), zur Berechnung der Schätzer.

Es existieren viele Techniken zum Überprüfen der Proportionalitätsannahme. Eine visuelle Möglichkeit basiert auf der grafischen Darstellung der logarithmierten geschätzten kumulativen Hazards in unterschiedlichen Gruppen, spezifiziert durch die Wahl des Kovariablenvektors  $\mathbf{x}$ . Sollten die geschätzten Linien in etwa parallel verlaufen, deutet dies auf die Richtigkeit der Annahme hin. Tatsächlich gilt im Modell (2.1) für die kumulative Hazard  $H(t, \mathbf{x}) = \int_0^t h(u; \mathbf{x}) du$ :

$$H(t, \mathbf{x}) = H_0(t) \exp\{\beta^T \mathbf{x}\},$$

so dass  $\log\{H(t, \mathbf{x}_1)\} - \log\{H(t, \mathbf{x}_2)\} = \beta^T(\mathbf{x}_1 - \mathbf{x}_2)$  (so genannte "log-log" Transformation der Survivorfunktion). Im univariaten Fall  $x \in \mathbb{R}$  mit zwei

Ausprägungen  $x_1 = 1$ ,  $x_2 = 0$  fällt die Entscheidung zugunsten des konstanten Terms  $\beta$ , falls die Differenz (ungefähr) parallel zur Zeitachse verläuft. Klein & Moeschberger (1997), Kapitel 11, diskutieren weitere grafische Möglichkeiten zur Überprüfung der Proportionalität. Die Abweichungen von dieser Annahme implizieren einen direkten Test auf die Abhängigkeit  $\beta_j(t) = \beta_j$  des  $j$ -ten Koeffizienten von der Zeit. Therneau & Grambsch (2000), Kapitel 6, geben einen Gesamtüberblick über die in der Literatur diskutierten Teststrategien. Im Wesentlichen beruht die Teststatistik auf der Miteinbeziehung von so genannten (skalierten) Schoenfeld-Residuen (Schoenfeld, 1980, und Grambsch & Therneau, 1994) und auf der Darstellung  $\beta_j(t) = \beta_j + \theta_j(g_j(t) - \bar{g}_j)$ , wobei  $g_j(t)$  Transformation der Zeit  $t$  und  $\bar{g}_j$  das arithmetische Mittel der (an der Stelle der Eventzeiten)  $g_j(t_k)$  bezeichnen. Man regressiert also  $\beta_j(t)$  auf  $g_j(t)$ . Die Nullhypothese entspricht dann zu  $\theta_j \equiv 0$ . Die Details sind in Grambsch & Therneau (1994) zu finden und ansatzweise auch in Therneau & Grambsch (2000). Die beiden Autoren haben eine S-Funktion geschrieben, `cox.zph`, mit deren Hilfe man sowohl einen globalen als auch einen elementweisen Test auf die funktionale Abhängigkeit durchführen kann. Eine Spezifikation der Zeit-Transformation ist dabei möglich.

Wir können als Motivation für die weitere Betrachtung die SOEP-Daten der Arbeitslosen (siehe Kapitel 11) zur Veranschaulichung heranziehen, um Abweichungen von der Proportionalitätsannahme festzustellen. Die Plots für die Merkmale (in Dummy-Kodierung) "Nationanders" (also nicht deutsch) und "Alteralt" (Alter des Individuums höher als 50 Jahre) verdeutlichen den mit der Zeit variierenden Einfluss dieser Faktoren. Die geplotteten Punkte sind die skalierten Schoenfeld-Residuen, addiert zu den jeweiligen, mittels der Cox-Regression geschätzten Werten für  $\beta$ 's. Wir geben den R-Code sowie die produzierten Plots in Abbildung 2.1 dazu. (Ein detaillierter Kommentar zum Programmoutput ist in Therneau & Grambsch, 2000, gegeben.)

```
fit.cox<-coxph(Surv(time, status)~Nation+Geschlecht+Alter+Beruf+
               Hochschule, data=arbeitslosen)
               coef exp(coef) se(coef) z p
Nationanders -0.212 0.809 0.0537 -3.95 7.9e-05
Geschlechtfrau -0.695 0.499 0.0503 -13.81 0.0e+00
Alterjung 0.300 1.350 0.0532 5.64 1.7e-08
Alteralt -1.845 0.158 0.1022 -18.06 0.0e+00
Berufnein -0.296 0.744 0.0521 -5.68 1.4e-08
Hochschuleja 0.498 1.646 0.0913 5.46 4.9e-08
```

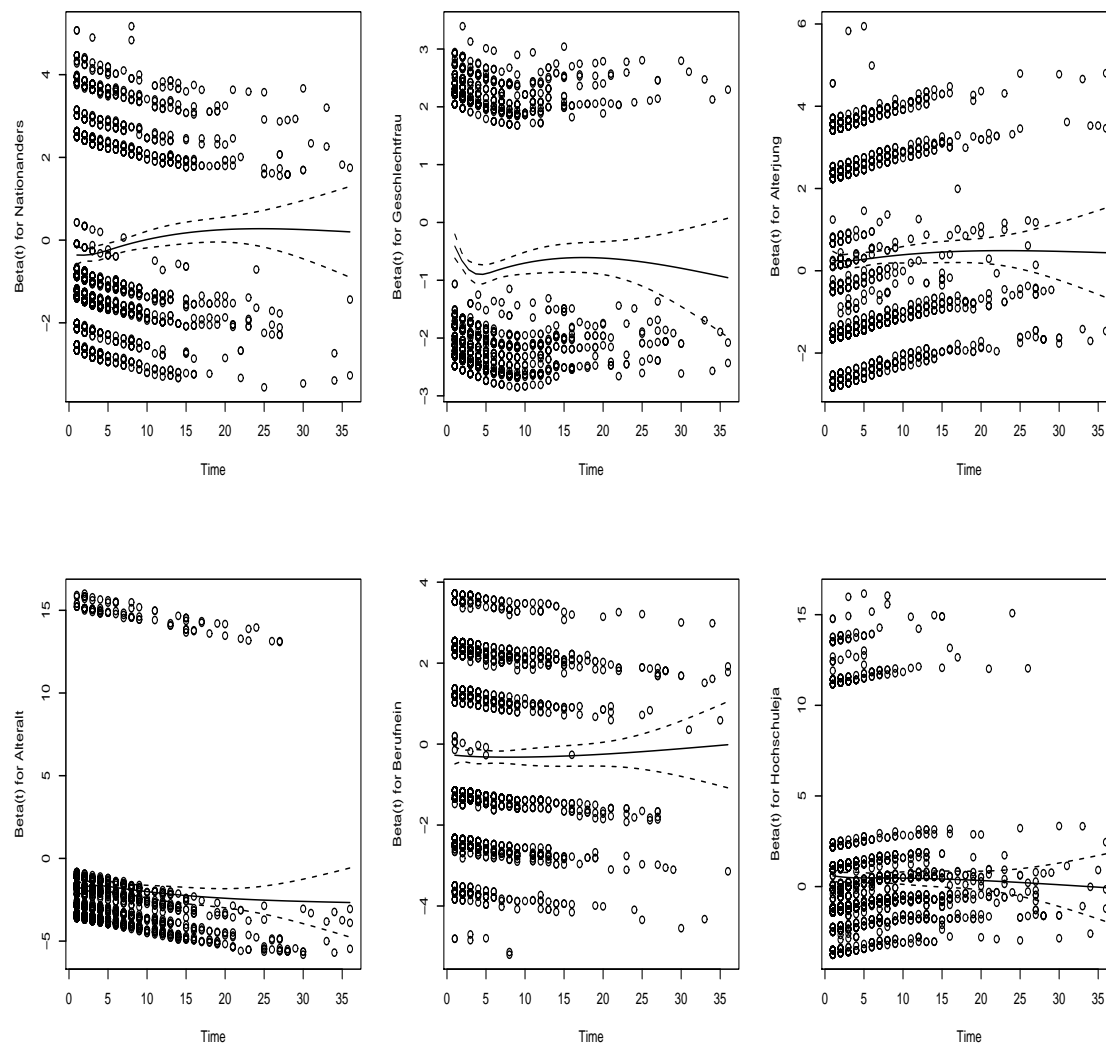


Abbildung 2.1: Scatterplots der skalierten Schoenfeld-Residuen sowie der geglätteten Faktoreffekte. Zeit ist die Verweildauer im Arbeitslosigkeitszustand in Monaten.

```
test.fit<-cox.zph(fit.coxph,transform='identity')
```

	rho	chisq	p
Nationanders	0.08230	11.736	0.000613
Geschlechtfrau	-0.01422	0.366	0.545078
Alterjung	0.03070	1.743	0.186809
Alteralt	-0.04307	3.370	0.066401
Berufnein	0.00589	0.063	0.801838
Hochschuleja	-0.01955	0.690	0.406153
GLOBAL	NA	23.263	0.000713

Ein weiteres Beispiel, in dem die Proportionalität der Hazardraten nicht zutrifft, wird in Kauermann (2005b) analysiert und bezieht sich auf den Erfolg (also Überlebensdauer) neu gegründeter Betriebe basierend auf der "Münchner

Gründerstudie" (siehe Brüderl, Preisendörfer & Ziegler, 1992, und Brüderl, Preisendörfer & Ziegler, 1998).

Wir sehen also, dass die Rahmen des klassischen Cox-Modells wesentlich erweitert werden müssen, um das Datenmaterial mit flexibleren Ansätzen auswerten zu können. Zum oben diskutierten klassischen Cox-Modell existieren zahlreiche Erweiterungen. Einen Überblick über den "State-of-the-art" geben z. B. Therneau & Grambsch (2000) und Ibrahim, Chen & Sinha (2001) im Bayes-Kontext.

## 2.2 Flexibilisierung des Proportional-Hazard Modells

Im Licht der oben diskutierten Problematik in der Anwendung des Cox-Modells konzentrieren wir uns im weiteren Verlauf auf die flexible, d. h. in Kovariableneffekten zeitabhängige, Modellierung der Hazardrate. Zugrunde wird das folgende Modell gelegt:

$$h(t) = h(t; \mathbf{x}) = h_0(t) \exp\{\mathbf{x}^T \boldsymbol{\beta}(t)\} = \exp\{\mathbf{z}^T \boldsymbol{\alpha}(t)\}, \quad (2.4)$$

oder auf der *log*-Skala:

$$\log\{h(t; \mathbf{x})\} = \mathbf{z}^T \boldsymbol{\alpha}(t) = \sum_{k=0}^p \alpha_k(t) z_k, \quad (2.5)$$

mit  $\mathbf{z}^T = [1, \mathbf{x}^T]$  und  $\boldsymbol{\alpha}(t) = [\log\{h_0(t)\}, \boldsymbol{\beta}(t)^T]^T$ . Dabei ist  $\mathbf{x}$  ein fixer, d. h. von der Zeit  $t$  unabhängiger Kovariablenvektor. Mit dieser Erweiterung ist automatisch auch die Baseline-Hazardrate  $h_0(t)$  mit in die im Folgenden noch zu spezifizierende Schätzroutine einbezogen. Diese "direkte" Schätzung der Baseline ist im gewöhnlichen Cox-Modell wegen der zugrunde liegenden Schätzmethode der partiellen Likelihood nicht möglich.

Die (mit der Überlebenszeit  $t$ ) variierenden Effekte  $\alpha_k(t)$ ,  $k = 0, 1, \dots, p$ , sollen glatt modelliert werden. Diese Forderung entspricht dem intuitiven Gedanken, dass die Effekte, mit denen die Kovariablen die Hazardrate beeinflussen, kein unregelmäßiges und ruckartiges Verhalten aufweisen. Außer dieser Forderung legen wir keine Restriktionen auf die Gestalt der Koeffizienten, d. h. die Modellierung erfolgt im datengesteuerten nichtparametrischen Rahmen. Dazu ist lediglich die Darstellungsform für die Modellkomponenten zu wählen. In der Arbeit betrachten wir den Spline-gestützten Ansatz (genauer die so genannten P-Splines, siehe Abschnitt 2.5). Dieser hat sich in vielen Anwendungen als gut funktionierend erwiesen und zeigt zahlreiche grundlegende Eigenschaften, die gerade in der rechenintensiven Computer-gestützten Modellbildung in vollem Umfang ausgenutzt werden. Wir diskutieren Splines



im nächsten Abschnitt 2.3 und stellen zwei Spline-Basen dar, mit denen wir die Modellierung vornehmen werden. Kurz erwähnt sei an dieser Stelle ein weiteres auf den so genannten "reproducing kernels" basierendes Prinzip der flexiblen Modellbildung, welches in der modernen statistischen Literatur auf Wahba (1990) zurückgeht (vgl. auch Gu, 2002, und Anwendungen darin). Eine wichtige Verbindung zwischen den beiden Konzepten der "penalized splines" und "reproducing kernels" wird in Pearce & Wand (2006) untersucht.

Einer der Punkte, welcher im Zusammenhang mit der flexiblen Darstellung auftritt, beruht auf der Zunahme der Freiheitsgrade. Die Erhöhung der zu adjustierenden Größen, welche im Spline-gestützten Modell unvermeidlich ist, wirkt auf die Unstabilität der Likelihoodfunktion. Um diesem Nebeneffekt entgegenzuwirken, soll die Likelihood penalisiert werden. Das Konzept der penalisierten Likelihood betrachten wir in Abschnitt 2.4, geben dort auch die motivierenden Ideen und Formen der so genannten "Roughness Penalty" - (ge)wichtiges Prinzip in der modernen nichtparametrischen Modellbildung - wieder. Den damit einhergehenden Effekt des "Trade-off" zwischen der Varianz des Schätzers und seiner Verzerrung (oder "Bias") diskutieren wir ebenfalls dort.

## 2.3 Spline-gestützte Modellierung

Die Abweichung von der parametrischen Annahme, so wie diese im linearen Regressionsmodell (oder auf deren Basis) postuliert wird, erfordert andere Überlegungen, und zwar dahingehend, wie die Einflüsse der Kovariablen und die Form der Kovariablen selbst modelliert werden sollen. Wenn keine bestimmte Spezifikation vorliegt, spricht man von einer flexiblen Modellierungsform und geht von einem nichtparametrischen Ansatz aus. Natürlich können einige der Kovariablen von vornherein in der parametrischen Form modelliert werden, falls eine bestimmte Sicherheit (oder Expertenwissen) über ihren Einfluss und über ihre Form im Modell gegeben ist. In diesem Fall spricht man auch von einem semiparametrischen Ansatz.

Für die flexible Form der Kovariablen(-einflüsse) gibt es mehrere konkurrierende Ansätze. Zahlreich in Theorie und Anwendung sind die auf Splines basierenden Modelle vertreten, die in der statistischen Literatur insbesondere durch die Arbeiten von Eubank (1988), Wahba (1990), Hastie & Tibshirani (1990, 1993) und Green & Silverman (1994) sowie Gu (2002), Ramsay & Silverman (2005) und Wood (2006) popularisiert wurden. Gemeinsam ist allen Ausführungen die Wahl eines Funktionenraums und dementsprechend einer Basis in diesem Raum, in der die zu modellierende Kovariable oder deren Einfluss dargestellt werden soll. Es geht in diesem Fall dann lediglich um eine funktionale Approximation, deren Qualität nicht zuletzt auch von der späteren Schätzmethodik abhängt. Man wählt also ein bekanntes System  $\Phi = \{\phi_k\}_{k=1, \dots, K}$  von  $K$  linear unabhängigen (im mathematischen Sinne) Funktionen  $\phi_k$ , die die Eigenschaft haben, eine bestimmte Klasse von Funktionen (z. B. stetig oder stetig differen-

zierbar) "gut genug" approximieren zu können, wenn  $K$  groß gewählt wird. Dabei wird die zu approximierende Funktion als lineare Kombination der Basisfunktionen derart dargestellt,

$$x(t) = \sum_{k=1}^K c_k \phi_k(t), \quad (2.6)$$

wobei  $x(\cdot)$  die zu modellierende funktionale Form der Kovariablen bzw. deren Einfluss ist,  $c_k$  die Gewichte oder die Parameter der Darstellung, die im Allgemeinen unbekannt sind und somit im Modell geschätzt werden sollen, und  $t$  das mathematische Argument und inhaltlich das im Modell interessierende Merkmal (oder den Prädiktor) bezeichnet, dessen Wirkung gerade durch  $x(t)$  spezifiziert wird.

In Abhängigkeit von der Fragestellung, dem Expertenwissen oder schlicht der Erfahrung wird die Basis spezifiziert. So ist z. B. die trigonometrische Basis dann eine gute Wahl, wenn die Daten eine Periodizität aufweisen (siehe weitere Diskussion dazu und Anwendungsbeispiele in Ramsay & Silverman, 2005). Zunehmender Beliebtheit beim statistischen Modellieren erfreuen sich die so genannten B-Splines. In einer der Anwendungen werden wir sehen, wie auch mit deren Hilfe periodische Einflüsse modelliert werden können. Als Standardwerk zur umfassenden Betrachtung von B-Splines gilt de Boor (2001), weniger mathematisch herausfordernd sind Dierckx (1993) und Rogers (2001). Die B-Splines gelten als flexibel genug, numerisch stabil, leicht und effizient zu programmieren sowie als schnell zu berechnen. In den angegebenen Werken finden sich auch die zugehörigen rekursiven Formeln (siehe hierzu auch Eilers & Marx, 1996, für eine schnelle Einführung und Konstruktion). Die funktionale Form der B-Splines hängt, neben der Festsetzung des polynomialen Grades der einzelnen Spline-Stücke, von der Wahl der so genannten inneren Knoten und Grenz- (oder Anker-)knoten ab, wobei die Grenzknoten maßgeblich das Verhalten der Spline-Funktion am Rand ihrer Definition beeinflussen. Es gibt zwei standardde Arten von Ankerknoten, die zusammenfallenden ("coincident boundary knots") und die äquidistanten periodischen ("periodic boundary knots") Ankerknoten. Bei der ersten Art verschwinden alle B-Splines außerhalb des Intervalls, über dem die Splines konstruiert wurden. Bei der zweiten Wahl werden die grenzenden Spline-Funktionen auch über den Rand hinaus definiert, aber nur auf dem gegebenen Intervall evaluiert. (Streng betrachtet kann auch eine spezielle Wahl von inneren Knoten eine wichtige Rolle spielen, z. B. dann, wenn die Unstetigkeitsstellen in den zu modellierenden Funktionen berücksichtigt werden sollen; Näheres hierzu diskutieren Ramsay & Silverman, 2005, in einem Anwendungsbeispiel. Bei einer glatten Modellierung der Modellkomponenten ist die Wahl der inneren Knoten nicht von großer Relevanz, die jeweiligen Referenzen sind an den entsprechenden Stellen angeführt.)

Zur Veranschaulichung plotten wir zwei B-Spline Basen mit jeweiliger Wahl der Ankerknoten. Die Splines werden in R im package `splines` implementiert. Außerdem spielt bei der Wahl der Ankerknoten als "coincident boundary

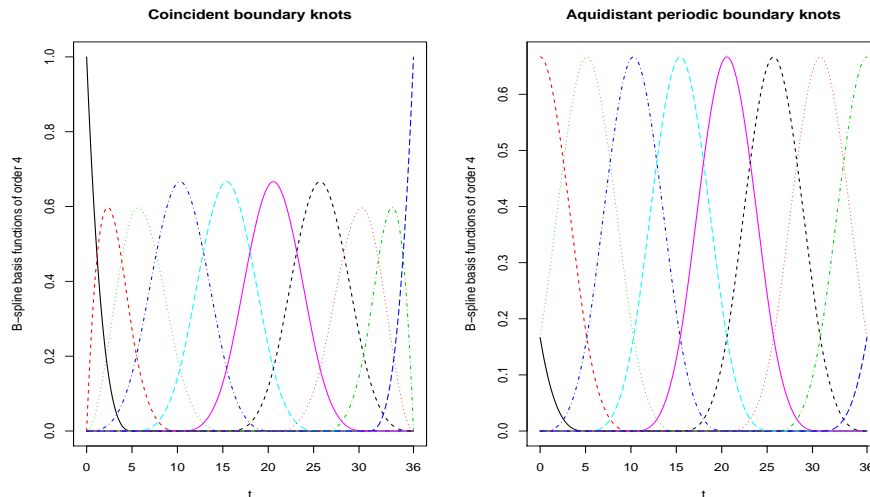


Abbildung 2.2: Zwei B-Spline Basen mit jeweils Spline-Funktionen der 4. Ordnung, also 3. Grades. Der Plot links ist mit "coincident boundary knots" und der Plot rechts ist mit äquidistanten periodischen Ankerknoten.

knots" das so genannte Intercept eine Rolle, da die Splines so implementiert sind, dass sich die Basisfunktionen in jedem Berechnungspunkt zu 1 addieren. Der R-Code zum in Abbildung 2.2 dargestellten rechten Plot ist wie folgt:

```
library(splines)
bs.order<-4
bs.degree<-bs.order-1
knots.t<-seq(0,36,le=8)
intervall.t<-knots.t[length(knots.t)]-knots.t[1]
knots.t.left<-knots.t[(length(knots.t)-bs.degree):(length(knots.t)-1)]-intervall.t
knots.t.right<-knots.t[2:(bs.degree+1)]+intervall.t
Basis.t.periodic<-splineDesign(knots=c(knots.t.left,knots.t,knots.t.right),
                               ord=bs.order, x=seq(0,36,le=1000))
matplot(seq(0,36,le=1000),Basis.t.periodic,type='l',xlab='t',
        ylab='B-spline basis functions of order 4',axes=F,
        main='Äquidistant periodische boundary knots')
axis(1,at=c(0,5,10,15,20,25,30,36))
axis(2,at=c(0,0.1,0.2,0.3,0.4,0.5,0.6,0.7))
box(type='o')
```

Von Ruppert, Wand & Carroll (2003) wird eine andere Splinebasis favorisiert, die der "truncated polynomials", welche zu der entsprechenden B-Spline Basis äquivalent ist, im Sinne einer bestimmten bijektiven Matrix-Transformation. Die Basis zum polynomialen Grad  $p$  besteht aus den Funktionen

$$1, t, \dots, t^p, (t - \tau_1)_+^p, \dots, (t - \tau_K)_+^p,$$

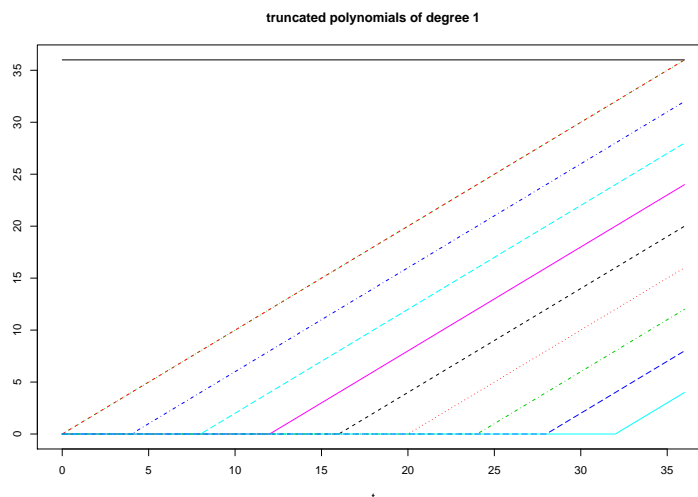


Abbildung 2.3: Truncated polynomial Splinebasis für den Grad  $p = 1$ . Die Knoten  $\tau_k$ ,  $k = 1, \dots, 8$ , sind als Quantile des Intervalls 0 bis 36 gewählt.

so dass die Spline-Darstellung die Form

$$x(t) = \beta_0 + \beta_1 t + \dots + \beta_p t^p + \sum_{k=1}^K u_k (t - \tau_k)_+^p \quad (2.7)$$

hat. Grafisch sieht die Basis aus wie in Abbildung 2.3 für  $p = 1$  und  $K = 8$  dargestellt.

Welche Splinebasis ist also zu wählen? Ruppert, Wand & Carroll (2003) argumentieren wie folgt:

*In principle, a change of the basis does not change the fit - though some bases are more numerically stable and allow computation of a fit with a greater accuracy. Besides numerical stability, reasons for selecting one basis over another are ease of implementation ... and interpretability ...*

Im Zusammenhang mit P-Splines (siehe Abschnitt 2.5) stufen Eilers & Marx (1996) die B-Splines als eine bessere, im Vergleich zu den "truncated polynomials", Wahl ein. Sie diskutieren in einer komparativen Studie (Eilers & Marx, 2005) auch die Vorteile in Bezug auf die Spezifizierung der Spline-Knoten, auch wenn diese (zumindest deren Anzahl) keine erhebliche Rolle beim Modellieren mit P-Splines, basierend auf truncated polynomials, spielt (Ruppert, 2002). In dieser Arbeit kommen beide Basen zum Einsatz. Dabei lassen wir uns von der relativen Leichtigkeit der Implementierung und vom Anwendungsmodell leiten. Die komparative Performance-Studie steht nicht zur Diskussion.

## 2.4 Idee der penalisierten Likelihood: "a roughness-penalty approach"

In diesem Abschnitt betrachten wir einen weiteren Baustein unserer Modellierungsstrategie, das Konzept der penalisierten Likelihoodfunktion, welches gerade auf dem Gebiet des (nichtparametrischen) Smoothing in den letzten Jahren enorm an Bedeutung gewonnen hat. Zugrunde gelegt wird die simple Idee der Bestrafung einer Modellanpassung (oder des Modellfits) an die Daten, sollte diese Anpassung gewisse Abweichungen vom erwünschten (oder optimalen) Verlauf aufweisen. Dabei sind zwei extreme Modellverläufe möglich. Das eine Extrem ist die Dateninterpolation: Die modellierte (glatte) Komponente durchläuft alle Datenpunkte und weist damit eine sehr große Varianz (aber andererseits keine Verzerrung) auf. Das andere Extrem stellt die geschätzte gerade Linie (oder Hyperebene) dar, mit der kleinstmöglichen Varianz, aber mit womöglich großer Verzerrung. Beide Extrem-Modelle sind nicht optimal, das Optimum liegt also "irgendwo dazwischen". Im ersten Fall soll das Modell weniger Überanpassung aufweisen (man spricht vom Overfitting). Im zweiten Fall (Underfitting) soll die geschätzte Komponente mehr Flexibilität aufweisen können, sofern die Daten komplexere als lineare Struktur beinhalten. Eine Möglichkeit, diesem trade-off entgegenzutreten, ist die Einführung einer zusätzlichen Komponente im Modell, die die Glattheit der Anpassung steuert. Der zu "kurvige" bzw. zu "flache" Verlauf einer glatten Funktion wird durch ihre Ableitungen beschrieben. Um den gesamten Verlauf zu beschreiben, integriert man z. B. das Quadrat der 2-ten Ableitung über den gegebenen Wertebereich, oder man nimmt das absolute Maximum der 2-ten Ableitung.

Ursprünglich in der numerischen Literatur erschienen, bezog sich das Problem der Penalisierung auf die Feststellung optimaler Eigenschaften von so genannten Smoothing-Splines. Man kann relativ leicht beweisen (Green & Silverman, 1994, de Boor, 2001), dass eine bestimmte Spline-Funktion (so genanntes "natural cubic spline") die folgende Optimalität besitzt: Unter allen zweimal stetig differenzierbaren Funktionen  $g$  auf dem Intervall  $a < t_1 < \dots < t_n < b$  mit den Punkten  $t_i$  soll für die  $n$  Beobachtungspunkte  $y_1, \dots, y_n$  eine optimale Funktion bestimmt werden, welche das Kriterium

$$\sum_{i=1}^n \{y_i - g(t_i)\}^2 + \lambda \int_a^b \{g''(t)\}^2 dt \quad (2.8)$$

minimiert. Das in der Mathematik bekannte Variationsprinzip liefert als Lösung den natural cubic spline. Dieser unterscheidet sich von der Klasse der B-Splines 3-ten Grades mit  $t_i$  als inneren Knoten nur noch durch den Verlauf am Rande des Intervalls, er ist linear außerhalb des Intervalls  $[a, b]$ . Die Erhöhung der Glattheitsforderung hat Splines höheren Grades zur Folge. Falls nun als innere Knoten die Menge aller unifizierten Beobachtungstellen gewählt wird, spricht man in der statistischen Literatur vom Smoothing-Spline als optimaler Funktion, im Sinne von (2.8).

Betrachten wir detailliert die einzelnen Komponenten in (2.8). Der erste Term ist die RSS-Summe ("residual sum of squares") und stellt im Regressionsmodell (bis auf die multiplikative Konstante) die negative *log*-Likelihood von  $g$ , gegeben normalverteilte Daten  $y_i$  mit Erwartungswert  $g(t_i)$ . Der zweite Term ist die Penalisierung, die die Glattheit des Schätzers steuert und deren Wirkung sich in der optimalen Wahl von  $\lambda$  niederschlägt. Zusammenfassend bekommen wir also:

$$\ell_p(x) = - \sum_{i=1}^n \{y_i - g(t_i)\}^2 - \lambda \int_a^b \{g''(x)\}^2 dx, \quad (2.9)$$

und die Maximierung dieses Ausdrucks ist äquivalent zur Minimierung in (2.8).

Allgemein betrachtet, sei nun eine durch die Splinebasis zu approximierende Funktion  $x(\cdot)$  aus (2.6) zu modellieren. Wir schreiben zunächst  $x(t) = \mathbf{c}^T \mathbf{\Phi}(t)$  mit  $\mathbf{c} = [c_1, \dots, c_K]^T$  und  $\mathbf{\Phi}(t) = [\phi_1(t), \dots, \phi_K(t)]^T$ . Die Ableitungen von  $x(\cdot)$  sind leicht zu berechnen, indem diese von den Funktionen  $\phi_k(\cdot)$  evaluiert werden. Dann kann man das Integral der 2-ten Ableitungen in der Matrixnotation als quadratische Funktion  $\mathbf{c}^T \mathbf{D} \mathbf{c}$  beschreiben mit

$$\mathbf{D} = \int_a^b \mathbf{\Phi}^{(2)} \{\mathbf{\Phi}^{(2)}\}^T dt.$$

Die penalisierte *log*-Likelihood im normalen Modell wird dann zu

$$-(\mathbf{y} - \mathbf{x})^T (\mathbf{y} - \mathbf{x}) - \lambda \mathbf{c}^T \mathbf{D} \mathbf{c} = -\|\mathbf{y} - \mathbf{c}^T \mathbf{\Phi}\|^2 - \lambda \mathbf{c}^T \mathbf{D} \mathbf{c}, \quad (2.10)$$

mit bekannten  $\mathbf{\Phi}$  und  $\mathbf{D}$ , die nur noch bezüglich  $\mathbf{c}$  (beim gegebenen  $\lambda$ ) zu maximieren ist. Das Ergebnis ist der wohl bekannte restringierte Kleinste-Quadrate-Schätzer in der Form

$$\hat{\mathbf{c}} = (\mathbf{\Phi}^T \mathbf{\Phi} + \lambda \mathbf{D})^{-1} \mathbf{\Phi}^T \mathbf{y}, \quad (2.11)$$

mit der Smoothing-Matrix  $\mathbf{S}_\lambda$  im Ausdruck

$$\hat{\mathbf{y}} = \mathbf{S}_\lambda \mathbf{y} = \mathbf{\Phi} (\mathbf{\Phi}^T \mathbf{\Phi} + \lambda \mathbf{D})^{-1} \mathbf{\Phi}^T \mathbf{y}. \quad (2.12)$$

Aus (2.11) wird die Wirkung des Parameters  $\lambda$  unmittelbar deutlich. Für  $\lambda = 0$  ist  $\hat{\mathbf{c}}$  der gewöhnliche Kleinste-Quadrate Schätzer (bezüglich der Basisfunktionen) und stellt eine Interpolation der Daten dar, falls die Anzahl der Funktionen  $\phi_k$  gleich der Anzahl der unterschiedlichen Beobachtungen gewählt ist, also  $K = n$ . Für  $\lambda \rightarrow \infty$  wird der Penalisierungsterm in (2.10) dominierend, so dass die Maximierung des Ausdrucks darin das Ergebnis  $\mathbf{c}^T \mathbf{D} \mathbf{c} = 0$  zur Folge hat. Da die Matrix  $\mathbf{D}$  nichtnegativ definit ist, gilt die Darstellung  $\mathbf{D} = \mathbf{L}^T \mathbf{L}$ ,

so dass  $\mathbf{Lc} = 0$  erfüllt ist. Nun hängt die Lösung  $\mathbf{c}_{\lambda=\infty}$  mit der Wahl der Penalisierungsmatrix zusammen (siehe Diskussion dazu in Abschnitt 3.3). Für die truncated polynomials ist die Standardwahl  $\mathbf{L} = \text{diag}\{\mathbf{0}_{p+1}, \mathbf{1}_K\}$  mit  $p$  und  $K$  wie in (2.7). Damit ergibt sich als extreme Lösung für die geschätzte Funktion das Polynom des Grades  $p$ . Für die B-Splines hängt bei der Wahl der Penalisierungsmatrizen als Differenzenmatrizen der extreme Funktionsverlauf mit der Wahl der Differenzenordnung  $s$  zusammen. So sind z. B. bei der Wahl  $s = 1$  alle  $c_1 = \dots = c_K$ , und die Funktion ist die Konstante  $c_1$  (für die normierten B-Splines). Analog überlegt man sich approximativen polynomialen Verlauf bei anderen Differenzenordnungen (siehe Diskussion dazu in Eilers & Marx, 1996).

Die obige Betrachtung legt daher nahe, das Konzept der penalisierten Likelihood allgemein und nicht nur für normalverteilte Daten einzuführen. Diese Idee ist in der statistischen Literatur erstmalig systematisch von Good & Gaskins (1971) (primär für die Dichteschätzung) untersucht worden. Inzwischen gibt es mehrere motivierende Überlegungen dazu. Einleuchtende Argumente in Bezug auf Smoothing-Splines (Bayessche Sicht) zeigt Silverman (1985) auf. Eine weitere zusammenfassende Übersicht hierzu geben Green & Silverman (1994). Allgemein betrachtet stellt die Methode der penalisierten Likelihood ein Beispiel eines weiteren in der mathematischen Optimierungsliteratur wohl bekannten Regularisierungsprinzips dar, das mindestens auf Tikhonov (1963) zurückgeht und das von O'Sullivan (1986) im Rahmen eines so genannten "Ill-posed Inverse Problem" diskutiert wird.

Eine simple und nützliche Implikation aus dieser Betrachtungsweise ist die folgende. Die unbekannte glatte Modellkomponente wird approximiert durch eine Linearkombination von bekannten Basisfunktionen, relativ groß in der Anzahl, mit dem Ziel eines relativ kleinen Bias. Um die daraus resultierende hohe Flexibilität (zu viele freie Parameter oder Gewichte) des Schätzers einzuschränken, penalisieren wir einen bestimmten Teil der Basisfunktionen bzw. alle Funktionen, indem die  $\log$ -Likelihood des Modells penalisiert wird. Der Penalisierungsterm soll wohl überlegt gewählt werden und die erwünschten Approximationseigenschaften wiedergeben. So berücksichtigen z. B. die auf den Ableitungen basierenden Strafterme die Krümmung der resultierenden Komponenten; zu hohe Krümmung wird durch eine hohe Penalisierung bestraft. Eine detaillierte Diskussion zur Wahl der Penalisierungsterme erfolgt in Abschnitt 3.3.

## 2.5 P-Splines für flexible Modellierung glatter Komponenten

In den letzten beiden Abschnitten wurde die Spline-gestützte Modellbildung allgemein dargestellt. Dabei spielte die Penalisierung der Likelihoodfunktion eine erhebliche Rolle. Zumeist wurden aber (im "klassischen" Sinne) solche

Spline-Basen betrachtet, die für jede (unifizierte) Merkmalsausprägung eine Spline-Funktion haben, für  $n$  Beobachtungen also  $n$  Spline-Funktionen. Ein zu großes  $n$  führt damit insbesondere bei der asymptotischen Betrachtungsweise zu einer großen Smoothing-Matrix  $\mathbf{S}_\lambda$  in (2.12). Daher wären solche Modellkonzepte wichtig, die in etwa gleiche Approximationsgüte aufweisen wie die klassischen Smoothing-Splines und dabei mit weniger Spline-Funktionen auskommen. Die ersten Ideen hierzu sind schon in den 1980er-Jahren erschienen (siehe z. B. Parker & Rice, 1985; O’Sullivan, 1986 und O’Sullivan, 1988a). Green & Silverman (1994), Abschnitt 3.6, betrachten das so genannte "basis functions approach" (mit der im Voraus festgelegten Funktionenbasis) und Penalisierung. Hastie (1996) untersucht die so genannten Pseudosplines, die eine "Low-Rank" Smoothing-Matrix erzeugen und damit, im Vergleich zu den klassischen Smoothing-Splines, schneller numerisch effiziente Schätzer produzieren. Doch systematisch untersucht und popularisiert wurde das Konzept von Eilers & Marx (1996). Von beiden Autoren stammt auch die Bezeichnung "P-Splines". Vor allem bestechen die P-Splines durch ihre numerische Stabilität, ausreichende Flexibilität und durch ihre einfache Implementierungsmöglichkeit.

Die Idee besteht in der Wahl einer fixen in der Anzahl, hochdimensionalen Funktionenbasis, deren Dimension nicht mit dem Stichprobenumfang  $n$  wächst, sondern konstant bleibt. Die zu schätzende Komponente wird durch die Linearkombination von Basisfunktionen, so wie in (2.6), dargestellt. Durch die Miteinbeziehung dieser Basis entsteht dennoch eine beträchtliche Anzahl von Parametern (oder Gewichten), die zu instabilen Schätzern führen. Daher wird auf diese Parameter eine geeignet gewählte Penalisierung gelegt. Der Penalisierungsterm soll in seiner Struktur möglichst einfach für die Implementierung und zugleich effizient und logisch begründet gewählt werden. Im Zusammenhang mit den B-Splines beschreiben Eilers & Marx (1996) das Prinzip wie folgt:

*The basic idea is not to use the integral of a squared higher derivative of the fitted curve in the penalty, but instead to use a simple difference penalty on the coefficients themselves of adjacent B-Splines.*

Auch Differenzen höherer Ordnung können benutzt werden, insbesondere dann, wenn auch die glatten Schätzer für Ableitungen bestimmt werden sollen. Dies kann z. B. bei der Analyse von funktionalen Daten von großer Relevanz sein (siehe Ramsay & Silverman, 2005). Berechtigterweise sei an dieser Stelle dennoch erwähnt, dass die Einfachheit der Implementierung keineswegs die allgemeine Wahl der Penalisierungsmatrix in ihrer ursprünglichen Form als Integral der quadrierten Ableitung einengt, wenn die B-Splines der Ordnung 4, also 3-ten Grades gewählt werden, wie Wand & Ormerod (2007) dies zeigen, mit effizienter Implementierung in R.

"In Konkurrenz" zu den P-Splines, basierend auf B-Splines, steht die Basis der truncated polynomials mit der einfachen Ridge-Penalisierung. Diese Splines werden von Ruppert, Wand & Carroll (2003) untersucht und ebenfalls als P-



Splines bezeichnet. Auch die alternativen Spezifizierungen von Penalisationen werden dort kurz erwähnt. Auf jeden Fall sollte die Wahl der Penalisation im Zusammenhang mit der Wahl der Splinebasis wohl überlegt getroffen werden.

Die komparative Studie von Eilers & Marx (2005) stuft B-Splines als "bessere" Wahl ein. B-Splines sind wegen ihrer numerischen Stabilität beliebt. Andererseits gibt es keinen triftigen Grund, auf das Modellieren mit truncated polynomials zu verzichten, zumal aus den Simulationsstudien (siehe z. B. Ruppert, 2002) als Fazit ein eindeutiges "Pro" gezogen wird. In unserer Anwendung kommen beide Spline-Basen zum Einsatz. Dabei können wir keine eindeutigen Prioritäten zugunsten der einen oder der anderen Basis setzen.

Zwei weitere Probleme bleiben mit der Wahl von P-Splines als Modellbaustein bestehen: die Wahl von Spline-Knoten und die Bestimmung des optimalen Wertes  $\lambda_{\text{opt}}$  zur Regulierung des Smoothing-Effektes. Allgemein existieren in der Literatur beim Modellieren mit Splines zahlreiche Ideen und Vorschläge zur Spezifizierung der Knoten, von heuristischen ad-hoc Überlegungen bis hin zur völlig automatisierten Knotenwahlstrategie, sowohl in der Position als auch in der Anzahl. Eilers & Marx (1996) erwähnen einige Artikel in diesem Zusammenhang. Beim Modellieren mit P-Splines entfällt die Notwendigkeit der "exakten" optimalen Bestimmung von Knoten, da die zahlreich gewählten Spline-Funktionen in ihrer zu hohen Flexibilität durch den Penalisierungsterm eingeschränkt werden. Bei der Wahl der B-Splines als Basis schlagen Eilers & Marx (1996) äquidistante Knoten vor und unterstützen diese Wahl in einer weiteren komparativen Studie (siehe Eilers & Marx, 2005). Über die "optimale" Anzahl der Basisfunktionen (bzw. Knoten) geben die Autoren keine explizite Angabe, deuten dennoch auf deren moderat zu wählende Anzahl hin, etwa 10 bis 20, aber auch die Wahl von 50 bis 100 und mehr Funktionen sei in einigen Situationen gerechtfertigt. Von einer vergleichbaren Größe sind auch die auf truncated polynomials basierenden P-Splines. In Ruppert (2002) sowie in Ruppert, Wand & Carroll (2003) und Wand (2003) werden alternative Algorithmen, basierend auf Erfahrungswerten, zur Bestimmung der Knotenanzahl vorgeschlagen. Als Knotenstellen werden hier die Quantile des Wertebereichs des Prädiktors genommen (und sind damit im Allgemeinen nicht äquidistant). Auch in diesem Fall kann die Knotenanzahl plausibel erhöht werden, wie Ruppert, Wand & Carroll (2003) an einem Beispiel in Abschnitt 8.4 zeigen.

Von erheblicher Relevanz beim Modellieren mit P-Splines ist die optimale Bestimmung des Smoothing-Parameters  $\lambda$ . Dazu gibt es mehrere Möglichkeiten. Die klassischen, auf der Modellwahl beruhenden Ansätze, wie AIC (siehe Akaike, 1974, Burnham & Anderson, 2002) und (G)CV (siehe Craven & Wahba, 1979, Golub, Heath & Wahba, 1979), konkurrieren hierbei mit der (G)LMM-Strategie (siehe Wand, 2003, und Ngo & Wand, 2004). Hier soll die Darstellung des P-Spline Modells als generalisiertes lineares gemischtes Modell (GLMM) diskutiert werden. Gehen wir zunächst in einem GLM-Modell (McCullagh & Nelder, 1989) von einer Exponentialfamilie aus und betrachten dabei die (kanonische) Linkfunktion, so können wir die Response-Prädiktor Beziehung be-

schreiben in der Form

$$g(\mathbf{E}(y)) = \eta = f(t) \quad (2.13)$$

mit dem Response  $y$ , dem natürlichen Parameter  $\eta$ , der Linkfunktion  $g$  und der nichtparametrischen funktionalen Form  $f(\cdot)$  für die Kovariable  $t$ . Wir wählen nun:

$$f(t) = \sum_{m=1}^M \beta_m \phi_m(t) + \sum_{n=1}^N u_n \psi_n(t) \quad (2.14)$$

von der Form (2.6) mit der spezifizierten Splinebasis  $\{\phi_1, \dots, \phi_M, \psi_1, \dots, \psi_N\}$  (B-Splines oder truncated polynomials) und bilden die penalisierte *log*-Likelihood mit der Penalisierung auf die Koeffizienten  $u_n$  und dem Smoothing-Parameter  $\lambda$ . Die Parameter  $\beta_m$  bleiben unpenalisiert. (Die Darstellung (2.14) stellt einen allgemeinen Fall dar; der erste Summationsterm kann gänzlich fehlen, falls alle Parameter penalisiert werden.) Die Koeffizienten  $u_n$  werden dabei als normalverteilt vorausgesetzt:

$$\ell_p = \log \mathcal{L}(\boldsymbol{\beta}, \mathbf{u} \mid \mathbf{y}) - \lambda \mathbf{u}^T \mathbf{D} \mathbf{u}, \quad (2.15)$$

mit  $\boldsymbol{\beta} = [\beta_1, \dots, \beta_M]^T$ ,  $\mathbf{u} = [u_1, \dots, u_N]^T \sim \mathcal{N}(\mathbf{0}, \frac{1}{\lambda} \mathbf{D}^-)$ . So formuliert entpuppt sich (2.15) als generalisiertes lineares gemischtes Modell (GLMM) mit a-priori Verteilung für die Parameter  $u_n$  und Penalisierung. Dabei bezeichnet  $\mathbf{D}^-$  generalisierte Inverse von der Penalisierungsmatrix  $\mathbf{D}$ , und  $\lambda$  spielt exakt die Rolle des zu schätzenden Kovarianz-Parameters. Für die Schätzung der Parameter gibt es mehrere alternative Ansätze. In unseren Modellen benutzen wir die numerisch unaufwendige PQL-Methode (Penalized Quasi Likelihood, siehe z. B. McCulloch & Searle, 2001, und Ruppert, Wand & Carroll, 2003), deren Darstellung in Kapitel 3 erfolgt. Wie wir sehen können, liefert die Einbettung in das GLMM-Modell die automatische Schätzung von  $\lambda$  als Kovarianz-Parameter ("variance component") mithilfe von Maximum Likelihood (vgl. auch Kauermann, 2005a). Leicht erweiterbar ist das Modell nun für den Fall mehrerer Prädiktoren, die durch die spezifizierte Linkfunktion additiv (GAMM) auf den Response wirken (siehe Ruppert, Wand & Carroll, 2003, Kapitel 11).

Die Modellierung mit GAMM's und penalisierten Splines bringt klare Vorteile bezüglich des rechnerischen Aufwands. Für die Schätzung der glatten Komponenten sowie des Smoothing-Parameters existieren mehrere Algorithmen. Zu erwähnen sind an dieser Stelle die Beiträge von Marx & Eilers (1998) und Marx & Eilers (2002), in denen die Schätzung der Parameter "direkt" durchgeführt wird und somit im Vergleich zu der herkömmlichen Methode des "Backfitting" (siehe Hastie & Tibshirani, 1990) einleuchtend ist. Darauf aufbauend untersuchen Aerts, Claeskens & Wand (2002) einige theoretische Eigenschaften von GAMMs mit P-Splines (siehe auch up-to-date Darstellung der Theorie in Hall

& Opsomer, 2005). Insbesondere können die Freiheitsgrade der funktionalen Komponenten definiert werden (siehe Ruppert, Wand & Carroll, 2003, Abschnitte 8.3 und 11.4). Auch die Bayessche Analyse kann für die P-Spline Modelle erfolgreich durchgeführt werden (vgl. Fahrmeir, Kneib & Lang, 2004, Lang & Brezger, 2004, sowie zusammenfassende Dissertation von Kneib, 2006, und Crainiceanu, Ruppert & Wand, 2005). In Crainiceanu, Ruppert, Claeskens & Wand (2005) sind die Tests der Art  $H_0 : 1/\lambda = 0$  vorgestellt, die die Abweichung der funktionalen Form der glatten Komponente vom Polynom spezifizierten Grades überprüfen. Dabei ist  $\lambda$  gerade der Varianzparameter in der GLMM-Darstellung des P-Spline Modells in (2.15). In Kauermann, Ormerod & Wand (2006) wird im Rahmen einer sich auf die GAMM-Methode stützenden Diskriminanzanalyse (mit P-Splines) auf den nichtlinearen Kovariableneinfluss getestet. Für die anschließende Modellwahl wird nach dem AIC-spezifizierten Kriterium (mAIC-Kriterium, genauer in Wager, Vaida & Kauermann, 2005, für die Wahl des optimalen Parameters  $\lambda_{\text{opt}}$  dargestellt) vorgegangen. Wir bemerken hier allerdings, dass es sich dabei um die "nicht-standard" Tests handelt, da der Varianzparameter in der Nullhypothese am Rande seines Wertebereichs liegt (siehe umfassende Diskussion hierzu in Searle, Casella, & McCulloch, 1992, Lin, 1997, und Zhang & Lin, 2003).

Falls kein additives Modell zugrunde gelegt werden kann, wird zur Modellierung der (zumeist) bivariaten Struktur häufig die Kreuzprodukte-Splinebasis gewählt ("tensor product basis"; allgemeine Darstellung siehe Dierckx, 1993, und de Boor, 2001). Im Zusammenhang mit P-Splines werden die bivariaten Strukturen in Ruppert, Wand & Carroll (2003) behandelt, vgl. außerdem die Anwendungen in Eilers & Marx (2003), Currie, Durban & Eilers (2004), Marx & Eilers (2005) sowie in Eilers, Currie & Durban (2006).

## 2.6 Splinebasierte Konzepte in der Analyse von Verweildauern: Literaturüberblick

Die auf B-Splines basierende penalisierte Schätzung des relativen Risikos im Cox-Modell geht mindestens zurück auf O'Sullivan (1988b). Er benutzt die schon diskutierte Idee der penalisierten Likelihood von Good & Gaskins (1971) und penalisiert die partielle *log*-Likelihood mit dem Integral der quadrierten 2-ten Ableitung der zu schätzenden glatten Komponenten (anstatt  $\beta^T \mathbf{x}$  in (2.1) wählt er nichtparametrisch  $\theta(\mathbf{x}) = \phi_1(x_1) + \dots + \phi_p(x_p)$  im Stil von Hastie & Tibshirani, 1990). Er gibt den auf der Newton-Raphson Iteration basierenden Schätzalgorithmus an und bestimmt den optimalen Smoothing-Parameter nach der CV-Methode. Die Baseline-Hazard wird nicht simultan mitgeschätzt. Dies führen aber Etezadi-Amoli & Ciampi (1987) durch, indem die Baseline  $h_0(t)$  durch die truncated polynomials 2-ten Grades approximiert wird, auch wenn in der nichtpenalisierten Form der Likelihood. Einen weiteren Schritt nach vorn machen Zucker & Karr (1990), indem sie die Kovariableneffekte (und die Baseline Hazard) flexibel modellieren und schätzen

durch die Maximierung der penalisierten partiellen  $\log$ -Likelihood (MPPL) von der Form (2.3). Sie zeigen die asymptotische Normalität des konstruierten MPPL-Schätzers. Die Idee von Gamerman (1991) beim Modellieren der flexiblen Form von Kovariableneffekten (sowie der von Baseline-Hazard) ist deren diskrete Darstellung auf dem gesamten Überlebenszeitenintervall mit stochastischer Abhängigkeit der Koeffizienten der benachbarten Intervalle des gewählten Gitters, vgl. Formel (5) darin. Die deutlichen Vorteile beim Modellieren mit penalisierten Splines zeigt Gray (1992). Das relative Risiko  $\mathbf{x}^T \boldsymbol{\beta}$  flexibilisiert er bezüglich der einzelnen Kovariablen und wählt die B-Splines dazu. Die Parameter der penalisierten partiellen Likelihood schätzt er im Stil von Newton-Raphson mit und zeigt (unter spezifizierten Regularitätsbedingungen) die asymptotische Normalität der Schätzer. Außerdem entwickelt er einen allgemeinen Test auf Modellparameter (lineare Hypothese der Art  $\mathbf{C}\boldsymbol{\eta} = 0$ ) und definiert in Anlehnung daran die Freiheitsgrade (siehe auch Gray, 1994). Eine Reihe von Modellspezifikationen (unter anderem die Berücksichtigung von Wechselwirkungen sowie variierende Koeffizienten der Kovariablen) wird von ihm anschließend kurz erläutert. Kooperberg, Stone & Troung (1995) gehen vom flexiblen Ansatz (2.2) der logarithmierten Hazard aus und begründen diese Betrachtungsweise mit der konkaven Form von Likelihood und der Befreiung von der zusätzlichen Restriktion der Nichtnegativität der Hazardfunktion. Dabei wird die Form der  $\log$ -Hazard nicht näher spezifiziert (siehe Formel (4) darin) und kann im Prinzip eine glatte  $p$ -variate Funktion von Kovariablen  $x_1, \dots, x_p$  sein. Approximiert wird diese dann durch die Produktspline-Basis der truncated polynomials in der Form der Linearkombination von Basisfunktionen. Die ML-Schätzer der nichtpenalisierten  $\log$ -Likelihood bekommen die Autoren mit der Newton-Raphson Methode und schlagen außerdem den Algorithmus zur Bestimmung der optimalen Anzahl der Spline-Funktionen vor. Joly, Commenges & Letenneur (1998) betrachten allgemein die Intervallzensurierung (die als Spezialfall die Rechtszensurierung enthält) und modellieren (im Unterschied zu Kooperberg, Stone & Troung, 1995) direkt den Spline-basierten Schätzer für die integrierte Hazardfunktion (und nicht die  $\log$ -Hazard) im Kontext der penalisierten  $\log$ -Likelihood. Als Splinebasis wählen sie die so genannten M-Splines (als Alternative zu B- und I-Splines; die beiden M- und I-Splines werden in der Publikation ebenfalls dargestellt) und setzen die Knoten an weniger Stellen als es Daten gibt. Dies kommt dem Konzept der P-Splines nahe. Hier wird ausschließlich die (integrierte) Baseline-Hazard flexibel modelliert. Dagegen werden die Kovariablen bzw. ihre Effekte klassisch mitmodelliert.

Die PQL-Schätzung der Baseline-Hazard (im Modell ohne Kovariablen) mit der Bestimmung von Smoothing-Parametern via GLMM-Darstellung des Modells diskutieren Cai, Hyndman & Wand (2002); diese Vorgehensweise bei der datengestützten Wahl des Regulierungsparameters wurde im vorigen Abschnitt erwähnt und wird im Anwendungsteil dieser Arbeit benutzt. In der Arbeit von Tutz & Binder (2004) für den Fall der diskreten Verweilzeiten (siehe Hamerle & Tutz, 1989, zur Darstellung des diskreten Falls) wird der flexible additive Ansatz gewählt mit einer "varying-coefficients"-Komponente (zum Modellieren der zeitabhängigen Effekte) und einer additiven Kompo-

nente (zum Modellieren der funktionalen Form einiger der Kovariablen selber). Die Schätzung erfolgt im Rahmen der penalisierten *log*-Likelihood. Die Penalisierungsparameter werden nach AIC-Kriterium bestimmt. Eine weitere anwendungsorientierte Untersuchung von variierenden Faktoreinflüssen im Modell der diskreten Überlebenszeiten wird in Kauermann, Brüderl & Tutz (2005) dargestellt.

Das P-Spline Modell der Hazardfunktion in der penalisierten, aber ansonsten völlig unspezifizierten Form der Likelihood mit flexibel modellierten Kovariableneffekten ist Gegenstand der Publikation von Kauermann (2005b). Der Smoothing-Parameter wird, wie in Cai, Hyndman & Wand (2002), via GLMM-Repräsentation datengesteuert bestimmt. Diese Entwicklung wird in dieser Dissertation kontinuierlich fortgesetzt. Ein weiteres Konzept in der Analyse von Überlebenszeiten ist das der "Frailties". Wir werden das Frailty-Prinzip im Anwendungsmodell der Competing Risks ausführlich erläutern. An dieser Stelle sei lediglich auf die Publikationen von Vaida & Xu (2000) und Therneau, Grambsch & Pankratz (2003) im Kontext der penalisierten Likelihood verwiesen.

# Kapitel 3

## Modellierungsstrategie

Dieses Kapitel stellt ein Konzept für das Modellieren mit P-Splines im Bereich der Verweildauern vor und stützt sich maßgeblich auf Kauermann (2005b). Die in folgenden Kapiteln dargestellten Modelle bauen entscheidend auf dessen Ausführungen auf. Die wesentlichen Bausteine sind dabei die Transformation des Ausgangsmodells in das GLM-Modell durch einen künstlich erzeugten Poisson-Prozeß sowie die Repräsentation des P-Spline Modells als ein GLMM Modell mit normalverteilten Koeffizienten bei Splines.

### 3.1 Gestalt der Likelihoodfunktion

Wir werden den Fall rechtszensierter Daten betrachten. Im Modell der stetigen Überlebenszeit  $t$  (mit nichtinformativem Zensoringmechanismus) hat die Likelihood Funktion die Gestalt:

$$\mathcal{L} = \prod_{i=1}^n \{f(t_i)\}^{\delta_i} \{S(t_i)\}^{1-\delta_i} = \prod_{i=1}^n \{h(t_i)\}^{\delta_i} \exp \left\{ - \int_0^{t_i} h(t) dt \right\}, \quad (3.1)$$

mit Hazardrate  $h(t)$ , Überlebensfunktion  $S(t)$ , beobachteter Überlebensdauer  $t_i$ ,  $i = 1, \dots, n$ , für das  $i$ -te Individuum sowie Zensierungsindikator  $\delta_i$  mit der üblichen Kodierung: 1 für tatsächliche Überlebensdauer (d. h. interessierendes Ereignis wird beobachtet) und 0 für den zensierten Fall. Auf der  $\log$ -Skala ergibt sich dann:

$$\ell = \log(\mathcal{L}) = \sum_{i=1}^n \delta_i \log \{h(t_i)\} - \int_0^{t_i} h(t) dt,$$

oder, mit Bezeichnungen wie in (2.5), im flexiblen Fall:

$$\begin{aligned} \ell &= \sum_{i=1}^n \delta_i \log\{h(t_i)\} - \int_0^{t_i} \exp\{\log\{h(t)\}\} dt \\ &= \sum_{i=1}^n \delta_i \mathbf{z}_i^T \boldsymbol{\alpha}(t_i) - \int_0^{t_i} \exp\{\mathbf{z}_i^T \boldsymbol{\alpha}(t)\} dt. \end{aligned} \quad (3.2)$$

Bei Berücksichtigung der Ausführungen im vorigen Kapitel wird für die variierenden Effekte  $\alpha_0(t), \dots, \alpha_p(t)$  eine glatte hochdimensionale Struktur der Art wie in (2.14) gewählt, d. h. wir modellieren die Effekte in der Form

$$\alpha_k(t) = \sum_{r=1}^{R_k} \beta_r^{(k)} Z_r^{(k)}(t) + \sum_{s=1}^{S_k} b_s^{(k)} B_s^{(k)}(t) = \mathbf{Z}_k^T(t) \boldsymbol{\beta}_k + \mathbf{B}_k^T(t) \mathbf{b}_k, \quad (3.3)$$

mit Parametern  $\beta_r^{(k)}$ ,  $r = 1, \dots, R_k$ , und  $b_s^{(k)}$ ,  $s = 1, \dots, S_k$ ,  $k = 0, 1, \dots, p$ . Diese Darstellung entspricht z. B. der Wahl der truncated polynomials als Splinebasis für die Approximation der Komponenten des Modells, wobei die erste Summe auf den polynomialen Term in  $\beta$ 's in (2.7) bezogen wird. (Diese Darstellungsweise ist auch für den Fall von Vorteil, wenn ein Trend in der Überlebenszeit "exakt" modelliert und die Flexibilität des Modells durch den zweiten Teil in der Zerlegung ermöglicht wird.) Für das Modellieren mit B-Splines ist die Darstellung

$$\alpha_k(t) = \sum_{r=1}^{R_k} u_r^{(k)} B_r^{(k)}(t) = \mathbf{B}_k^T(t) \mathbf{u}_k, \quad k = 0, 1, \dots, p, \quad (3.4)$$

relevant. Dabei soll, wie im Kontext der P-Splines, die Anzahl der Funktionen (bzw. der Knoten der Splinebasis) groß genug, aber deutlich kleiner als die Anzahl der Beobachtungen festgelegt werden (siehe Diskussion in Abschnitt 2.5).

Durch die Zunahme der zu schätzenden Parameter ( $\boldsymbol{\beta}, \mathbf{b}$ ) in (3.3) bzw.  $\mathbf{u}$  in (3.4) wird die Likelihoodfunktion instabil. Die übliche Methode der Penalisierung (diskutiert in Abschnitt 2.4) schafft dabei Abhilfe, d. h. man geht über zu der penalisierten Form der *log*-Likelihood:

$$\ell_{\text{pen}} = l - \frac{1}{2} \sum_{k=0}^p \lambda_k \mathbf{b}_k^T \mathbf{D}_k \mathbf{b}_k, \quad (3.5)$$

bzw., wenn die Penalisierung auf alle zu schätzenden Parameter auferlegt wird, zu:

$$\ell_{\text{pen}} = l - \frac{1}{2} \sum_{k=0}^p \lambda_k \mathbf{u}_k^T \mathbf{D}_k \mathbf{u}_k, \quad (3.6)$$

wie im Falle der B-Splines. Dabei ist die Wahl von Penalisierungsmatrizen  $\mathbf{D}_k$  von großer Bedeutung und wird zumeist in Abhängigkeit von der gewählten Splinebasis bestimmt (die konkrete Wahl erfolgt in Abschnitt 3.3). Die Parameter  $\lambda_k$ ,  $k = 0, 1, \dots, p$ , steuern quantitativ die Glattheit der jeweiligen Modellkomponenten.

## 3.2 Approximation der Likelihood; Poisson - Modell

Die in (3.2) angegebene *log*-Likelihood enthält das Integral-Teil, das nur in wenigen speziellen Fällen exakt evaluiert werden kann. (Aber auch dann kann es von Vorteil sein, keine exakte Berechnung durchzuführen, die nur zu unnötigen komplexeren Kalkulationen und Notationen führen würde, ohne erheblichen Effizienz- und Genauigkeitsverlust.) In der Regel wird es approximiert. Für unsere Modellierungszwecke ist dazu eine gewöhnliche rechteckige bzw. trapezoide Approximation völlig ausreichend. Hierauf basieren die Modelle bei Kauermann (2005b) und Kauermann & Khomski (2006). Ein weiterer Vorteil dieser unaufwendigen Approximation ist die gewünschte Einbettung des Ausgangsmodells in das Poisson-GLMM Modell mit anschließender PQL-Schätzmethode, wie wir dies darlegen werden. Die Ausführung erfolgt in Anlehnung an die Darstellung der glatten Komponenten wie in (3.3).

Betrachten wir beispielsweise eine trapezoide Approximation des Integrals. Der Einfachheit halber wird zunächst kein Kovariablenvektor miteinbezogen, wir haben also nur die Baseline-Hazardrate im Modell. Es soll zunächst eine Gitterpunktmenge spezifiziert werden. Völlig ausreichend ist dabei die Wahl der Gitterpunkte als die Zeitpunkte  $\{0 = \kappa_0, \kappa_1, \dots, \kappa_K\}$  der beobachteten Events (einschließlich des Gitterpunktes 0). Begründet wird diese Wahl dadurch, dass das Integral eine kumulative Hazardrate darstellt, und die saturierte Information über deren Größe daher in den Zeitpunkten der Events konzentriert ist. Manchmal (insbesondere in Modellen mit so genannten "long-term survivors") bringt aber diese Wahl einen zusätzlichen rechnerischen Aufwand mit sich, der mit der Konstruktion der noch zu diskutierenden künstlichen Beobachtungen zusammenhängt, so dass die Anzahl der Stützstellen für die Integralberechnung (ohne erheblichen Genauigkeitsverlust) reduziert werden soll. Das Integral-Teil ergibt sich also wie folgt (siehe Anhang A.10):

$$\int_0^{t_i} \exp\{\alpha_0(u)du\} \approx \sum_{j=0}^{J_i} \exp\{\alpha_0(\min\{t_i, \kappa_j\}) + o_j^{(i)}\},$$

mit  $o_0^{(i)} = \log\{(\kappa_1 - \kappa_0)/2\}$ ,  $o_j^{(i)} = \log\{(\kappa_{j+1} - \kappa_{j-1})/2\}$  für  $1 \leq j \leq J_i - 1$  und  $o_{J_i}^{(i)} = \log\{(t_i - \kappa_{J_i-1})/2\}$ . Dabei sind  $\kappa$ 's die Gitterpunkte. Die mit  $o_j^{(i)}$  bezeichneten Parameter sind somit bekannt und spielen die Rolle der Offset-Parameter.  $J_i$  ist der Index, so dass  $\kappa_{J_i}$  der kleinste Gitterpunkt ist mit  $\kappa_{J_i-1} <$



$t_i \leq \kappa_{J_i}$ . Damit ergibt sich die Approximation der  $\log$ -Likelihood als

$$\ell \propto \sum_{i=1}^n \left[ \delta_i \alpha_0(t_i) - \sum_{j=0}^{J_i} \exp \left\{ \alpha_0(\min\{t_i, \kappa_j\}) + o_j^{(i)} \right\} \right]. \quad (3.7)$$

Definieren wir hier die Zufallsvariablen  $\delta_{ij}$  und Zeitpunkte  $t_{ij}$  jeweils durch

$$\delta_{ij} = \begin{cases} 0 & \text{für } j = 0, 1, \dots, J_i - 1 \\ \delta_i & \text{für } j = J_i \end{cases} \quad \text{und} \quad t_{ij} = \begin{cases} \kappa_j & \text{für } j = 0, 1, \dots, J_i - 1 \\ t_i & \text{für } j = J_i \end{cases},$$

bekommen wir für die  $\log$ -Likelihood den Ausdruck

$$\ell = \sum_{i=1}^n \sum_{j=0}^{J_i} \left[ \delta_{ij} \alpha_0(t_{ij}) - \exp \left\{ \alpha_0(t_{ij}) + o_j^{(i)} \right\} \right]. \quad (3.8)$$

Die approximative Form der  $\log$ -Likelihood in (3.8) entspricht gerade der  $\log$ -Likelihood der Poisson-verteilten Zufallsvariablen  $\delta_{ij}$  mit den Intensitätsparametern  $\xi_{ij} = \exp \left\{ \alpha_0(t_{ij}) + o_j^{(i)} \right\}$ . Wir beachten allerdings, dass die auf diese Weise künstlich erzeugten Poisson-Daten  $\delta_{ij}$  von niedriger Intensität sind, da sie nur Werte 0 oder 1 annehmen können. Wir notieren für jede Beobachtung die Anzahl der Intervalle zwischen den Integrationspunkten solange, bis das interessierende Ereignis auftritt, füllen den künstlichen Vektor mit Nullen bis zur gegebenen (evtl. zensierten) Beobachtungszeit auf und komplettieren diesen Vektor mit der Ausprägung für  $\delta_i$ , also mit 1, falls die Beobachtung nicht zensiert ist, ansonsten aber mit 0. Bild 3.1 visualisiert den Mechanismus zur Erzeugung der künstlichen Poisson-Daten.

Um die weitere Darstellung kompakter und einfacher zu gestalten, führen wir die Matrixnotationen ein. Wegen (3.3) ergibt sich zunächst für  $\alpha_0(t)$ :

$$\begin{aligned} \alpha_0(t_{ij}) &= \sum_{r=1}^{R_0} \beta_r^{(0)} Z_r^{(0)}(t_{ij}) + \sum_{s=1}^{S_0} b_s^{(0)} B_s^{(0)}(t_{ij}) \\ &= \mathbf{Z}_0(t_{ij}) \boldsymbol{\beta}_0 + \mathbf{B}_0(t_{ij}) \mathbf{b}_0 = \mathbf{W}_0(t_{ij}) \boldsymbol{\theta}_0 \end{aligned}$$

mit

$$\begin{cases} \mathbf{Z}_0(\cdot) = [Z_1^{(0)}(\cdot), \dots, Z_{R_0}^{(0)}(\cdot)] \\ \mathbf{B}_0(\cdot) = [B_1^{(0)}(\cdot), \dots, B_{S_0}^{(0)}(\cdot)] \\ \mathbf{W}_0(\cdot) = [\mathbf{Z}_0^{(0)}(\cdot), \mathbf{B}_0^{(0)}(\cdot)] \end{cases} \quad \text{und} \quad \begin{cases} \boldsymbol{\beta}_0 = [\beta_1^{(0)}, \dots, \beta_{R_0}^{(0)}]^T \\ \mathbf{b}_0 = [b_1^{(0)}, \dots, b_{S_0}^{(0)}]^T \\ \boldsymbol{\theta}_0 = [\boldsymbol{\beta}_0^T, \mathbf{b}_0^T]^T \end{cases}.$$

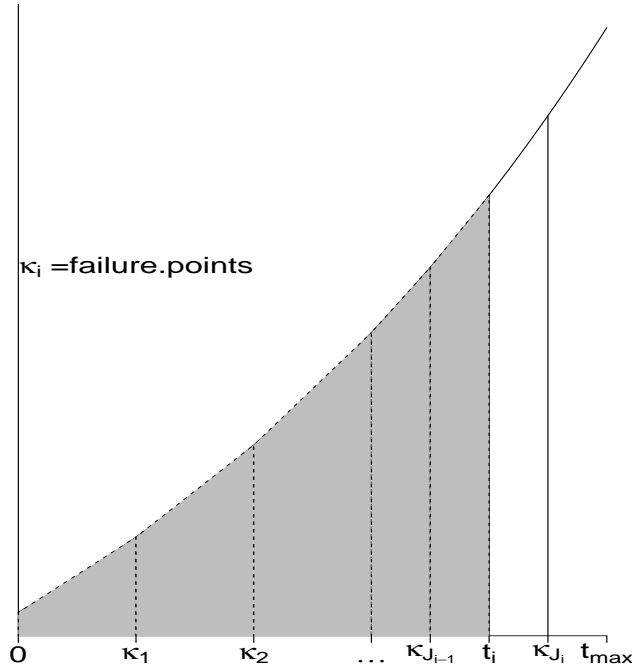


Abbildung 3.1: Approximation der integrierten Hazard durch Trapezoide.

Mit den weiteren Notationen  $\boldsymbol{\alpha}_0 = \{\alpha_0(t_{ij})\}_{\substack{i=1,\dots,n \\ j=0,1,\dots,J_i}}$  und  $\boldsymbol{o} = \{o_j^{(i)}\}_{\substack{i=1,\dots,n \\ j=0,1,\dots,J_i}}$  erhalten wir für den Poisson-Datenvektor  $\boldsymbol{\delta} = \{\delta_{ij}\}_{\substack{i=1,\dots,n \\ j=0,1,\dots,J_i}}$  die approximative *log*-Likelihood in der Form

$$\ell = \boldsymbol{\delta}^T \mathbf{W}_0 \boldsymbol{\theta}_0 - \mathbf{1}^T \exp \{ \mathbf{W}_0 \boldsymbol{\theta}_0 + \boldsymbol{o} \},$$

mit der Matrix  $\mathbf{W}_0$  zusammengesetzt aus den Zeilenvektoren  $\mathbf{W}_0(t_{ij})$ , untereinander geschrieben. Die Erweiterung auf den Fall mit Kovariablen ist jetzt einfach. Es sei dazu  $\mathbf{X} = [\mathbf{1}, \mathbf{x}_1, \dots, \mathbf{x}_p]$  die der Generierung von Poisson-Daten entsprechend aufgeblähte Kovariablenmatrix, d. h. der zur  $i$ -ten Überlebenszeit gehörige Zeilenvektor der Kovariablen wird in der resultierenden aufgeblähten Kovariablenmatrix so oft geschrieben, wie es dem oben beschriebenen Mechanismus zur Erzeugung der Poisson-Daten entspricht. Mit den Notationen

$$\mathbf{W}_{ij} = \left[ \mathbf{1} [\mathbf{Z}_{ij}^{(0)}, \mathbf{B}_{ij}^{(0)}], x_{i1} [\mathbf{Z}_{ij}^{(1)}, \mathbf{B}_{ij}^{(1)}], \dots, x_{ip} [\mathbf{Z}_{ij}^{(p)}, \mathbf{B}_{ij}^{(p)}] \right]$$

und

$$\mathbf{W} = \{ \mathbf{1} \odot [\mathbf{Z}_0, \mathbf{B}_0], \mathbf{x}_1 \odot [\mathbf{Z}_1, \mathbf{B}_1] \dots, \mathbf{x}_p \odot [\mathbf{Z}_p, \mathbf{B}_p] \} = \{ \mathbf{W}_{ij} \}_{ij},$$

wobei  $\odot$  komponentenweises Multiplizieren des Vektors  $\mathbf{x}_k$  mit der zugehörigen Matrix  $[\mathbf{Z}_k, \mathbf{B}_k]$  bezeichnet (siehe Anhang A.2), resultiert mit  $\boldsymbol{\theta} = [\boldsymbol{\theta}_0^T, \boldsymbol{\theta}_1^T, \dots, \boldsymbol{\theta}_p^T]^T$

die  $\log$ -Likelihood zu

$$\ell = \boldsymbol{\delta}^T \mathbf{W} \boldsymbol{\theta} - \mathbf{1}^T \exp \{ \mathbf{W} \boldsymbol{\theta} + \mathbf{o} \}. \quad (3.9)$$

Die penalisierte  $\log$ -Likelihood ist dann

$$\ell_{\text{pen}} = \ell - \frac{1}{2} \sum_{k=0}^p \lambda_k \mathbf{b}_k^T \mathbf{D}_k \mathbf{b}_k. \quad (3.10)$$

Die Schätzung des Parameters  $\boldsymbol{\theta}$  erfolgt hier iterativ nach dem Newton-Raphson Schema. Die Miteinbeziehung von Smoothing-Parametern  $\lambda_k$  in die Schätzroutine, wie sie hier dargestellt werden soll, erfordert die Darstellung des Modells als Poisson-GLMM Modell. Das GLMM-Modell wird in Abschnitt 3.4 beschrieben, und die Modellparameter darin werden nach der PQL-Methode ("penalized quasi-likelihood") geschätzt. Diese Methode ist numerisch unaufwendig und leicht zu implementieren.

### 3.3 Wahl der Penalisierungsmatrix

Hier wollen wir kurz auf die spezielle Form der Penalisierungsmatrizen eingehen. Wie bereits in Abschnitt 2.5 diskutiert, wird die Gestalt der Penalisierung in Abhängigkeit von der spezifizierten Splinebasis gewählt. Für die truncated polynomials wird die Ridge-Penalisierung auf die Koeffizienten  $b_s^{(k)}$  in (3.3) gewählt, so dass die resultierende Matrix  $\mathbf{D}_k$  die diagonale Gestalt hat:

$$\mathbf{D}_k = \begin{bmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & & \ddots & \\ & \cdots & & 1 \end{bmatrix} \in \mathbb{R}^{S_k \times S_k}.$$

Für die späteren Optimierungszwecke ist es bequemer, die Penalisierung durch die Parameter  $\boldsymbol{\theta}_k$  auszudrücken. Das bedeutet, dass die Parameter  $\boldsymbol{\beta}_k$  in (3.3) unpenalisiert bleiben, so dass der Penalisierungsterm in (3.10) geschrieben wird als  $\sum_{k=0}^p \lambda_k \boldsymbol{\theta}_k^T \tilde{\mathbf{D}}_k \boldsymbol{\theta}_k$  mit

$$\tilde{\mathbf{D}}_k = \begin{bmatrix} \mathbb{O}_{R_k, R_k} & \mathbb{O}_{R_k, S_k} \\ \mathbb{O}_{S_k, R_k} & \mathbf{D}_k \end{bmatrix} \in \mathbb{R}^{(R_k+S_k) \times (R_k+S_k)}.$$

Die Penalisierungsmatrix für den ganzen Parameter  $\boldsymbol{\theta}$  zu bilden ist nun einfach. Wenn mit B-Splines modelliert wird, werden in (3.4) die endlichen Differenzen der benachbarten Koeffizienten im Vektor  $\mathbf{u}_k$  gebildet, wie dies Eilers & Marx

(1996) vorschlagen. So sieht z. B. für die Differenzen 2-ter Ordnung  $u_{j+1}^{(k)} - 2u_j^{(k)} + u_{j-1}^{(k)}$  die entsprechende Matrix wie folgt aus:

$$\mathbf{L}_k = \begin{bmatrix} 1 & -2 & 1 & \cdots & \cdots \\ & 1 & -2 & 1 & \cdots \\ & & \ddots & \ddots & \ddots \\ & & & 1 & -2 & 1 \end{bmatrix} \in \mathbb{R}^{R_{k-2} \times R_k}. \quad (3.11)$$

Damit ist  $\mathbf{D}_k = \mathbf{L}_k^T \mathbf{L}_k$  in (3.6), da die Differenzen quadriert werden (hier ist  $\boldsymbol{\theta}_k = \mathbf{u}_k$ , so dass  $\mathbf{D}_k = \mathbf{D}_k$ .) Wie wir sehen, besitzen die Differenzenmatrizen keinen vollen Rang, so dass deren Kreuzprodukt  $\mathbf{D}_k$  nicht invertierbar ist. Dieser Sachverhalt wird einige Schwierigkeiten bei der datengesteuerten Bestimmung des Penalisierungsparameters bereiten, so dass eine approximative Lösung gefunden und hergeleitet werden muss.

Einige Anwendungen diktieren auch spezielle Penalisierungen, wie z. B. bei der Berücksichtigung von Periodizitäten in den Daten (siehe Eilers & Marx, 2005, und Ramsay & Silverman, 2005). Weitere Vorschläge finden sich in Aldrin (2006).

### 3.4 GLMM-Darstellung des Modells; Laplace-Approximation

In der Formel (3.10) für die penalisierte *log*-Likelihood wird der Penalisierungseffekt durch die "optimale" Wahl von  $\lambda_k$ ,  $k = 0, 1, \dots, p$ , bestimmt. Allgemein gebräuchlich sind dabei die Techniken, die auf der Modellwahl beruhen, z. B. Bestimmung nach AIC ("Akaike Information Criterion", Akaike, 1974) oder (G)CV ("Generalized) Cross Validation", Craven & Wahba, 1979), und auf den diesen beiden verwandten Kriterien. Eine ganz andere Strategie für die Bestimmung des optimalen Wertes  $\lambda_{\text{opt}}$  folgt aus der Darstellung des Modells (3.10) als GLMM-Modell, in dem die Koeffizienten  $\mathbf{b}_k$  als Realisierungen normalverteilter Zufallsvektoren betrachtet werden, mit der Kovarianzmatrix abhängig vom Parameter  $\lambda_k$ . Die GLMM-Darstellung wurde schon in Abschnitt 2.5 kurz besprochen. Einen intensiven Gebrauch davon im Kontext der P-Splines machen Wand (2003) und Ruppert, Wand & Carroll (2003). (Zur älteren Diskussion im nichtpenalisierten Fall siehe Brumback, Ruppert & Wand, 1999.) Die GLMM-Strategie wurde für den Fall der rechtszensierten Daten von Cai, Hyndman & Wand (2002) im Kontext der Überlebenszeit angewandt. Wir nehmen hier an:

$$\mathbf{b}_k \sim \mathcal{N}(\mathbf{0}, \lambda_k^{-1} \mathbf{D}_k^-), \quad (3.12)$$

mit  $\mathbf{D}_k^-$  als generalisierte Inverse von  $\mathbf{D}_k$ . Da die Wahl von  $\mathbf{D}_k$  von der Wahl der P-Spline Basis abhängt, resultiert für penalisierte truncated polynomi-

als die Identitätsmatrix als eindeutige Inverse, wohingegen für penalisierte B-Splines unendlich viele generalisierte Inversen (oder einfach g-Inversen) mit der Eigenschaft  $\mathbf{D}_k \mathbf{D}_k^- \mathbf{D}_k = \mathbf{D}_k$  existieren. Daher hat in diesem Fall die ”-“ Bezeichnung in der Kovarianzmatrix von  $\mathbf{b}_k$  eher generische Bedeutung. Wie wir aber weiter unten sehen werden, ist die Spezifizierung einer bestimmten g-Inversen von keiner Bedeutung. (Ausführliche Behandlung von g-Inversen siehe bei Rao & Mitra, 1974.)

Generell ergibt sich im GLMM-Rahmen die Likelihoodfunktion als marginale Likelihood nach dem Ausintegrieren von zufälligen Komponenten des Modells (”marginal likelihood approach”). Wir bekommen somit für den Datenvektor  $\{\boldsymbol{\delta}\}_{\substack{i=1,\dots,n \\ j=0,1,\dots,J_i}}$ , wie in Abschnitt 3.2 definiert, in Abhängigkeit von den fixen Parametern  $\boldsymbol{\beta}$  und Penalisierung  $\boldsymbol{\lambda} = [\lambda_0, \lambda_1, \dots, \lambda_p]^T$ :

$$\mathcal{L}(\boldsymbol{\beta}, \boldsymbol{\lambda} | \boldsymbol{\delta}) = \int f_{\delta|b}(\boldsymbol{\delta} | \mathbf{b}) f_b(\mathbf{b}) d\mathbf{b}, \quad (3.13)$$

mit der bedingten Poisson-Dichte  $f_{\delta|b}$  und Normalverteilungsdichte  $f_b$ . Allgemein ist die numerische Integration nur schwer zu handhaben. Eine der gebräuchlichsten Methoden ist dabei die Laplace-Approximation (siehe z. B. Tierney & Kadane, 1986). Kurz skizziert wird hier die Approximation des Integrals in der von McCulloch & Searle (2001) angegebenen Form benutzt:

$$\log \int_{\mathcal{R}^q} \exp\{f(\mathbf{b})\} d\mathbf{b} \doteq f(\hat{\mathbf{b}}) + \frac{q}{2} \log 2\pi - \frac{1}{2} \log \left| -\frac{\partial^2 f(\hat{\mathbf{b}})}{\partial \mathbf{b} \partial \mathbf{b}^T} \right|,$$

mit  $\hat{\mathbf{b}}$  als Maximumstelle der Funktion  $f(\cdot) : \mathbb{R}^q \rightarrow \mathbb{R}$ . Mit  $f(\cdot) = \log f_{\delta|b}(\cdot) + \log f_b(\cdot)$  ergibt sich als Approximation für die marginale *log*-Likelihood (vorausgesetzt die Invertierbarkeit der Matrizen):

$$\begin{aligned} \ell(\boldsymbol{\beta}, \boldsymbol{\lambda} | \boldsymbol{\delta}) &= \log \int \exp f(\mathbf{b}) d\mathbf{b} \\ &\doteq \log f_{\delta|b}(\boldsymbol{\delta} | \hat{\mathbf{b}}) - \frac{1}{2} \hat{\mathbf{b}}^T \mathbf{D} \hat{\mathbf{b}} + \frac{1}{2} \log |\mathbf{D}| - \frac{1}{2} \log \left| -\frac{\partial^2 h(\hat{\mathbf{b}})}{\partial \mathbf{b} \partial \mathbf{b}^T} \right| \\ &= \ell_{\text{pen}}(\boldsymbol{\beta}, \hat{\mathbf{b}}, \boldsymbol{\lambda} | \boldsymbol{\delta}) + \frac{1}{2} \log |\mathbf{D}(\boldsymbol{\lambda})| - \frac{1}{2} \log \left| \tilde{\mathbf{I}}(\boldsymbol{\beta}, \hat{\mathbf{b}}) + \mathbf{D}(\boldsymbol{\lambda}) \right|, \end{aligned} \quad (3.14)$$

wobei  $\mathbf{D}(\boldsymbol{\lambda})$  die blockdiagonale Matrix ist mit  $\lambda_k \mathbf{D}_k$ ,  $k = 0, 1, \dots, p$  als Einzelblöcke, sowie  $\tilde{\mathbf{I}}(\boldsymbol{\beta}, \hat{\mathbf{b}})$  die Untermatrix der Fisher-Information  $\mathbf{I}(\boldsymbol{\beta}, \mathbf{b})$  zugehörig zum Vektor  $\mathbf{b}$  der zufälligen Komponenten und  $|\cdot|$  die Matrixdeterminante. Der Ausdruck rechts in der letzten Gleichung in (3.14) enthält die penalisierte *log*-Likelihood, wie sie in (3.10) angegeben wurde. Unterstellt man weiter,

dass die Abhängigkeit des Terms  $\tilde{\mathbf{I}}(\boldsymbol{\beta}, \hat{\mathbf{b}})$  von  $\boldsymbol{\beta}$  vernachlässigbar ist, und maximiert nun hier bezüglich  $\boldsymbol{\beta}$ , kommt man auf dieselben Schätzgleichungen wie beim direkten Maximieren in (3.10) bezüglich  $\boldsymbol{\beta}$  und  $\mathbf{b}$ . Diese Vorgehensweise rechtfertigt auch die Laplace-Approximation, die (unter den getroffenen Annahmen) zu denselben Schätzern führt wie die PQL-Methode. Die in GLMM PQL-gestützte Schätzmethode wurde unter anderen von Breslow & Clayton (1993) und Wolfinger & O'Connell (1993) propagiert.

### 3.5 Regulierung des Smoothing-Effektes; Freiheitsgrade der Modellkomponenten

Die Bestimmung von  $\boldsymbol{\lambda}$ , ausgehend von (3.14), ist nun einfach. Maximiert an der Stelle  $\boldsymbol{\beta} = \hat{\boldsymbol{\beta}}$  (und  $\hat{\mathbf{b}}$  wie gehabt) ergibt sich zunächst die maximierte approximative marginale *log*-Likelihood als

$$\ell_{mm}(\boldsymbol{\lambda}|\boldsymbol{\delta}) \doteq \ell_{\text{pen}}(\hat{\boldsymbol{\theta}}, \boldsymbol{\lambda}|\boldsymbol{\delta}) - \frac{1}{2} \log \left| \tilde{\mathbf{I}}_{\text{pen}}(\hat{\boldsymbol{\theta}}, \boldsymbol{\lambda}) \mathbf{D}^-(\boldsymbol{\lambda}) \right|, \quad (3.15)$$

mit  $\hat{\boldsymbol{\theta}} = (\hat{\boldsymbol{\beta}}, \hat{\mathbf{b}})$  und  $\tilde{\mathbf{I}}_{\text{pen}}(\hat{\boldsymbol{\theta}}, \boldsymbol{\lambda}) = \tilde{\mathbf{I}}(\hat{\boldsymbol{\theta}}) + \mathbf{D}(\boldsymbol{\lambda})$ . Um nun den Parameter  $\boldsymbol{\lambda}$  zu bestimmen, sind die Ableitungen von  $\ell_{mm}(\boldsymbol{\lambda})$  zu kalkulieren. Dies ist recht einfach unter der Berücksichtigung der Struktur der Penalisierungsmatrix  $\mathbf{D}(\boldsymbol{\lambda})$  im Fall der truncated polynomials (als Referenz für das Kalkül der Matrixableitungen sei z. B. auf Harville, 1997, und Rao & Rao, 1998, verwiesen, siehe auch Rogers, 1980, sowie Anhang A.3 und A.7):

$$\hat{\lambda}_k = \frac{S_k}{\hat{\mathbf{b}}_k^T \mathbf{D}_k \hat{\mathbf{b}}_k + \text{tr} \left\{ \mathbf{D}_k \left[ \tilde{\mathbf{I}}_{\text{pen}}^{-1}(\hat{\boldsymbol{\theta}}, \boldsymbol{\lambda}) \right]_{kk} \right\}}, \quad k = 0, 1, \dots, p, \quad (3.16)$$

mit  $S_k$  als Anzahl der penalisierten Komponenten  $b_s^{(k)}$  in (3.3) und dem Sub-Index  $kk$  als  $k$ -ter Block-Komponente der Inversen der penalisierten Fisher-Information, zugehörig zum Vektor  $\mathbf{b}_k$ . Hier ist  $\mathbf{D}_k$  die Einheitsmatrix. Die Formel (3.16) erlaubt die Fixpunkt-iterative Bestimmung von  $\lambda_k$ . Die zweiten Ableitungen sind für die Schätzroutine im Stil von Newton-Raphson ebenfalls leicht zu berechnen (siehe Anhang A.7). Für den Fall der penalisierten B-Spline Basis erhält man (siehe Herleitung in Anhang A.6):

$$\hat{\lambda}_k = \frac{\dim(\mathbf{B}_k) - \text{degree}_k}{\hat{\mathbf{b}}_k^T \mathbf{D}_k \hat{\mathbf{b}}_k + \text{tr} \left\{ \mathbf{D}_k \left[ \tilde{\mathbf{I}}_{\text{pen}}^{-1}(\hat{\boldsymbol{\theta}}, \boldsymbol{\lambda}) \right]_{kk} \right\}}, \quad k = 0, 1, \dots, p, \quad (3.17)$$

wobei hier  $\dim(\mathbf{B}_k)$  die Dimension der gewählten B-Spline Basis bezeichnet,  $\mathbf{D}_k = \mathbf{L}_k^T \mathbf{L}_k$  gilt und  $\text{degree}_k$  die Differenzenordnung (gleich Anzahl der Zeilen in der Matrix  $\mathbf{L}_k$ ) notiert. Die Berechnung erfolgt hier ebenfalls nach der Fixpunkt-Methode.

In welcher Relation stehen nun die Freiheitsgrade der Modellkomponenten zu den Parametern  $\lambda_k$ ? Den Ausführungen von Ruppert, Wand & Carroll (2003), Abschnitte 8.3 und 11.4, folgend definieren wir im approximativen GLMM-Modell zunächst die totalen Freiheitsgrade als

$$\text{df}(\boldsymbol{\lambda}) = \text{tr} [\mathbf{I}_{\text{pen}}^{-1} \mathbf{I}], \quad (3.18)$$

mit Fisher-Information  $\mathbf{I}$  und penalisierter Fisher-Information  $\mathbf{I}_{\text{pen}} = \mathbf{I} + \text{diag}\{\tilde{\mathbf{D}}_k, k = 0, 1, \dots, p\}$ . (Hier sind die Matrizen  $\tilde{\mathbf{D}}_k$  von der blockdiagonalen Gestalt  $\tilde{\mathbf{D}}_k = \text{diag}[\mathbf{0}_k, \mathbf{D}_k]$  im Fall der truncated polynomials.) Die Freiheitsgrade beschreiben in quantitativer Art den benötigten gesamten Glättungsgrad. Die Freiheitsgrade der einzelnen glatten Komponenten lassen sich daraus wie folgt berechnen:

$$\text{df}_k(\lambda_k) = \text{tr} [\mathbf{E}_k \mathbf{I}_{\text{pen}}^{-1} \mathbf{I}], \quad k = 0, 1, \dots, p, \quad (3.19)$$

wobei  $\mathbf{E}_k$  die Diagonalmatrizen sind mit 1-Einträgen an den Stellen, die den korrespondierenden Parametern der glatten Komponenten des Modells entsprechen, und ansonsten 0. Natürlich stimmt in diesem Fall  $\text{df} = \sum_k \text{df}_k$ .

### 3.6 Varianz der Modellparameter; "Sandwich-Estimator"

Die Schätzer für die Modellparameter  $\boldsymbol{\beta}, \mathbf{b}$ , berechnet aus der GLMM-Darstellung (Abschnitt 3.4), sind bei fixem  $\boldsymbol{\lambda}$  ähnlich denen, die nach der PQL-Methode (Abschnitt 3.2) berechnet wurden, wie aus den angegebenen Schätzgleichungen ersichtlich. Wenn der Parameter  $\boldsymbol{\lambda} = \mathbf{0}$  gesetzt ist, greift die Standardargumentation der asymptotischen Normalität des Schätzers. Im penalisierten Fall schlagen Marx & Eilers (1998) den so genannten "Sandwich"-Schätzer für die Kovarianz der glatten Komponenten vor. Diese Form des Kovarianzschätzers wird von Ruppert, Wand & Carroll (2003) (siehe dort Abschnitte 6.4 und 11.4) begründet und analytisch hergeleitet. Tutz & Binder (2004) zeigen weitere Begründungen für diese Art der Approximation auf. In Anlehnung an Ruppert, Wand & Carroll (2003) erfolgt also die Bestimmung der Kovarianz von  $\hat{\boldsymbol{\theta}} = [\boldsymbol{\beta}_0^T, \mathbf{b}_0^T, \dots, \boldsymbol{\beta}_p^T, \mathbf{b}_p^T]^T$  durch den Sandwich-Schätzer:

$$\widehat{\text{Var}}\{\hat{\boldsymbol{\theta}}\} = \mathbf{I}_{\text{pen}}^{-1}(\hat{\boldsymbol{\theta}}, \hat{\boldsymbol{\lambda}}) \mathbf{I}(\hat{\boldsymbol{\theta}}) \mathbf{I}_{\text{pen}}^{-1}(\hat{\boldsymbol{\theta}}, \hat{\boldsymbol{\lambda}}). \quad (3.20)$$

Daraus können die (punktweisen) Konfidenzbänder für die jeweiligen glatten Komponenten des Modells leicht berechnet werden: Für den variierenden Koeffizienten  $\hat{\alpha}_k(\cdot) = \mathbf{W}_k(\cdot) \hat{\boldsymbol{\theta}}_k$  gilt:

$$\text{Var}\{\hat{\alpha}_k(t)\} = \mathbf{W}_k(t) \text{Var}\{\hat{\boldsymbol{\theta}}_k\} \mathbf{W}_k(t)^T, \quad k = 0, 1, \dots, p. \quad (3.21)$$

# Kapitel 4

## Anwendungsmodelle: Überblick

In Kapitel 3 wurden die wesentlichen Komponenten der Modellierung beschrieben. Die Methode der Schätzung durch die Überführung in ein Poisson-GLMM Modell ist ein wesentlicher Bestandteil in der folgenden Anwendung. Die ersten beiden Modelle der "Two-Way"-Art konzentrieren sich auf die glatte Modellierung der Effekte im Fall von zwei verschiedenen Zeitmessungen mit jeweils Kalenderzeit bzw. saisonaler Komponente als zweite Zeitskala. Beide Zeitachsen werden hier zunächst als additiv modelliert, also ohne Interaktion miteinander. Im Fall der Kalenderzeit (siehe Kapitel 5), interessiert die dynamische Entwicklung der Kovariableneffekte nicht nur in Bezug auf die Überlebensdauer bis zum Eintreten eines bestimmten Events, sondern auch der Unterschied in bestimmten Zeitperioden. Dies ist der Fall bei der Analyse von Arbeitslosigkeitszeiträumen. Von Interesse ist der glatte Verlauf eines Risikofaktoreffektes über Jahre hinweg. Hier wollen wir das Analyseergebnis von Kauermann & Khomski (2006) ausführlich darstellen sowie Anregung und Diskussion zur weiteren Entwicklung geben.

Im folgenden Modell, Kapitel 6, nehmen wir die Saison, also den Jahresabschnitt, als zweite Zeitkomponente und modellieren glatt die Kovariableneinflüsse, additiv in beiden Zeitachsen, in Bezug auf die (logarithmierte) Hazardrate. Als Datenmaterial dienen Mortalitätsdaten von Kindern in einem afrikanischen Land. Die Sterberaten schätzt man hier immer noch als sehr hoch ein und berücksichtigt dabei die Saison, in der ein Kind geboren wurde, als einer der Risikofaktoren (siehe dazu z. B. Moore, Cole & Poskit, 1997, und Jaffar, Leach, Greenwood & Greenwood, 2000). Bisher erfolgten dazu die meisten Studien im Rahmen des klassischen Cox-Modells, so etwa bei Becher et al. (2004). Wir wollen auf dieser Studie ansetzen und den Einfluss der saisonalen Komponente glatt in der nichtparametrischen Art in das Modell miteinbeziehen.

Eine Weiterentwicklung der beiden Modelle ist die Berücksichtigung der Wechselwirkung zwischen den beiden Zeitskalen. Die Unterschiede in Abläufen von Überlebenszeiten etwa in Bezug auf die Kalenderzeit werden erst recht dann offensichtlich, wenn die gesamte Studie über eine große Zeitspanne hindurch er-



folgt. Das Modell könnte für die Arbeitslosigkeitszeiträume Aufschluss darüber geben, wie unterschiedlich die Risikofaktoren in verschiedenen Zeitintervallen wirken, z. B. in den 1980er vs. 1990er Jahren. Da der Arbeitsmarkt und die damit verbundene Arbeitsdynamik großen Veränderungen über die Jahre hinweg unterliegen, rechnet man unter anderem auch mit ziemlichen Unterschieden in Arbeitslosigkeitszeiträumen in verschiedenen kalendarischen Zeitperioden. Derselben Fragestellung unterliegt auch das Modell der kindlichen Mortalitäten, insbesondere, wenn man untersuchen will, ob die Jahreszeit der Geburt auf die Überlebensdauer nachhaltig wirkt. (In vielen afrikanischen Ländern unterscheidet man in Bezug auf die Mortalitätsrate zwischen der trockenen Jahreszeit und der Regenzeit, als dem das Sterberisiko erhöhenden Jahresabschnitt.) Das Modell in Kapitel 7 untersucht diese Fragestellung.

In Kapitel 8 behandeln wir das Modell der Competing Risks mit Berücksichtigung von Individuum- oder clusterspezifischen latenten Faktoren, so genannten Frailties, Begriff, der sich für die Bezeichnung und Modellierung von unbeobachteten Heterogenitäten in der Literatur zu Überlebenszeiten eingebürgert hat. Da es sich hier um zwei noch nicht diskutierte Konzepte handelt, die "Competing Risks" und die "Frailties", werden wir zunächst die beiden Ideen darstellen und erläutern. Anschließend behandeln wir das erarbeitete Modell, die Methodik und das Schätzprinzip. Als Anwendungsbeispiel dienen wiederum die SOEP-Daten von Arbeitslosen. (Mehrere frühere Untersuchungen sind in diesem Kontext der Wiederbeschäftigung durchgeführt worden, siehe z. B. Narendranathan & Stewart, 1993, Jacobson & Carlin, 1995, Gianelli, 1996, Rosholm, 1997, und Tysse & Vaage, 1999).

In Kapitel 9, das den Anwendungsteil dieser Arbeit abschließt, betrachten wir das Multi-Zustandsmodell mit Berücksichtigung der Verweildauer. Dieses baut in seiner Vorgehensweise auf dem Modell der Competing Risks auf. Behandelt wird die vereinfachte Situation mit nur zwei Zuständen und Zensierung.

# Kapitel 5

## Two-Way Additives Hazard-Modell mit Kalendereffekten

Im Vordergrund der Modellierung steht die Miteinbeziehung der Kalenderzeit als zweite zeitliche Komponente, neben der der Überlebensdauer. Die Berücksichtigung von kalenderspezifischen Effekten ist insbesondere bei langfristig angelegten Studien von Bedeutung, bei denen es, im Gegensatz zum klassischen Fall, nicht möglich ist, die unter Beobachtung stehenden Einheiten homogen in Bezug auf ihre Aufnahme in die Studie zu behandeln. Deskriptiv werden solche Überlebenszeiten mithilfe des so genannten Lexis-Diagramms dargestellt, einer beliebten grafischen Repräsentation in demografischen Studien. Umfassend behandelt wird das Lexis-Diagramm in Keiding (1990) und Francis & Pritchard (1998). Die Arbeitslosigkeitszeiten aus dem Anwendungsbeispiel sind im Lexis-Diagramm, Abbildung 5.1, dargestellt.

In diesem Diagramm ist horizontal der Zeitpunkt  $b$  des Eintritts in den Arbeitslosigkeitszustand abgebildet und vertikal die bekannte bzw. zensierte Überlebensdauer, woraus sich der kalendarische Zeitpunkt des Ereignisses  $c$  (in unserem Analysefall die Wiederbeschäftigung in Vollzeit) leicht bestimmen lässt durch die Beziehung  $c = b + t$ . Grafisch verlaufen alle Verweilzeiten parallel zueinander, indem die Punkte  $(b, 0)$  und  $(c, t)$  für die jeweilige Beobachtung durch eine Gerade verbunden werden.

Zwei Zugangsarten beim Modellieren mit mehreren Zeitskalen wurden in der Literatur vorgeschlagen und diskutiert. Der Vorschlag von Cox (1972), Abschnitt 11, betrachtet eine der Skalen als primär und modelliert die anderen als zusätzliche erklärende Kovariablen. Die zweite Vorgehensweise (Farewell & Cox, 1979) behandelt die Zeitskalen "symmetrisch", indem sie durch eine geeignete Rotation des Zeitraums zu einer einzigen Zeitmessung transformiert werden, und vollzieht so die Dimensionsreduktion mit sich. Dieses Konzept ist von Oakes (1995) weiter untersucht worden. Der weiteren Frage, welche Zeitskalen überhaupt "gut" für die Analyse sein können, wid-

men sich Duchesne & Lawless (2000) (siehe dort relevante Referenzen). Im Kontext der stochastischen Zählprozesse werden die multivariaten Zeitskalen umfassend von Andersen, Borgan, Gill & Keiding (1993) behandelt. Von Efron (2002) wurde die symmetrische Behandlung von Kalenderzeit neben der Überlebensdauer vorgeschlagen, basierend auf dem Poisson-GLMM-Modell, welches er auch als "Two-Way"-Modell bezeichnet. Zum Modellieren wählt er den (semi-)parametrischen Ansatz der "proportional hazards".

In unserem Modell gilt die Überlebensdauer als primäre Zeitmessung. Der kalenderspezifische Faktor wird in der nichtparametrischen Form dargestellt. Die additive Form der Kovariableneffekte in  $t$  und  $c$  ergibt die Interpretation des Terms  $\alpha_k(t, c)x_k = \{\alpha_k^{(t)}(t) + \alpha_k^{(c)}(c)\}x_k$  als Wechselwirkung zwischen der jeweiligen Zeitkomponente und der Kovariablen  $x_k$  (Hastie & Tibshirani, 1993). Allerdings gehen wir in der Implementierung des Modells nicht direkt von der Kalenderzeit  $c$  als kalendarischem Zeitpunkt des Ereignisses aus, sondern vom kalendarischen Beginn  $b$  der Beobachtung der Überlebenszeit in der Studie, wie im Lexis-Diagramm dargestellt. Diese Betrachtungsweise steht im Einklang mit der physikalischen Interpretation von Cox (1972), der den Zeitpunkt  $b$  auch als "Stress" bezeichnet und für die  $(t, b)$ -Betrachtung die folgende Unabhängigkeitsannahme trifft:

*...Mechanism of failure is identical at the various levels of  $[b]$  but takes place on a time-scale that depends on  $[b]$ .*

## 5.1 Modellansatz

In Anlehnung an die Formeln (2.4) und (2.5) gelte für die Hazardrate auf der  $\log$ -Skala

$$\log\{h(t, b; \mathbf{x})\} = \log\{\mathbf{z}^T[\boldsymbol{\alpha}_t(t) + \boldsymbol{\alpha}_b(b)]\} \quad (5.1)$$

mit  $\mathbf{z} = [1, \mathbf{x}^T]^T$ ,  $\boldsymbol{\alpha}_t(t) = [\alpha_0^{(t)}(t), \alpha_1^{(t)}(t), \dots, \alpha_p^{(t)}(t)]^T$  und analog für  $\boldsymbol{\alpha}_b(b)$ . Die Likelihoodfunktion (für rechtszensierte Daten) hat die zu (3.1) ähnliche Form:

$$\mathcal{L} = \prod_{i=1}^n [h(t_i, b_i; \mathbf{x}_i)]^{\delta_i} \exp\left\{-\int_0^{t_i} h(t, b_i; \mathbf{x}_i) dt\right\}. \quad (5.2)$$

Für die variierenden Effekte  $\boldsymbol{\alpha}_t(t)$  und  $\boldsymbol{\alpha}_b(b)$  wird nun eine flexible hochdimensionale Spline-Form unterstellt. Benutzen wir die Darstellung

$$\boldsymbol{\alpha}_t(t) = \mathbf{W}_t(t) \boldsymbol{\theta}_t, \quad \boldsymbol{\alpha}_b(b) = \mathbf{W}_b(b) \boldsymbol{\theta}_b,$$

ähnlich der aus Abschnitt 3.2, bekommen wir die *log*-Likelihood als

$$\ell = \boldsymbol{\delta}^T \mathbf{W}_{t,b} \boldsymbol{\theta}_{t,b} - \mathbf{1}^T \exp \{ \mathbf{W}_{t,b} \boldsymbol{\theta}_{t,b} + \mathbf{o} \}, \quad (5.3)$$

mit dem Parametervektor  $\boldsymbol{\theta}_{t,b} = [\boldsymbol{\theta}_t^T, \boldsymbol{\theta}_b^T]^T$ , der Designmatrix  $\mathbf{W}_{t,b} = [\mathbf{W}_t, \mathbf{W}_b]$  und dem Offset-Term  $\mathbf{o}$ . An der Stelle  $(i, j)$ ,  $i = 1, \dots, n$ ,  $j = 0, 1, \dots, J_i$ , gilt wegen der Überführung in das Poisson-Modell  $\mathbf{W}_{t,b}(t_{ij}, b_i) = [\mathbf{W}_t(t_{ij}), \mathbf{W}_b(b_i)]$ . Da wir die variierenden Koeffizienten glatt und additiv in beiden Zeitrichtungen modellieren wollen, hat die penalisierte *log*-Likelihood die Gestalt

$$\ell_p(\boldsymbol{\theta}_t, \boldsymbol{\theta}_b, \boldsymbol{\lambda}_t, \boldsymbol{\lambda}_b) = \ell(\boldsymbol{\theta}_t, \boldsymbol{\theta}_b) - \frac{1}{2} \mathbf{b}_t^T \mathbf{D}_t(\boldsymbol{\lambda}_t) \mathbf{b}_t - \frac{1}{2} \mathbf{b}_b^T \mathbf{D}_b(\boldsymbol{\lambda}_b) \mathbf{b}_b, \quad (5.4)$$

mit den blockdiagonalen Matrizen  $\mathbf{D}_t(\boldsymbol{\lambda}_t)$ ,  $\mathbf{D}_b(\boldsymbol{\lambda}_b)$  und  $\lambda_{tk} \mathbf{D}_{tk}$  bzw.  $\lambda_{bk} \mathbf{D}_{bk}$ ,  $k = 0, 1, \dots, p$ , als Einzelblöcke, konstruiert wie in Abschnitt 3.3 angegeben. Hier sind  $\boldsymbol{\lambda}_t = [\lambda_{t0}, \dots, \lambda_{tp}]$  und  $\boldsymbol{\lambda}_b = [\lambda_{b0}, \dots, \lambda_{bp}]$ . (Zur Darstellung wurden die truncated polynomials als Splinebasis gewählt.)

Das auf diese Weise aufgestellte Modell weist ein Identifizierungsproblem auf, welches ganz allgemein für additive Modelle gilt: Der Intercept-Term ist mit den beiden Zeitskalen vermischt und soll daher einer der additiven Komponenten eindeutig zugewiesen werden. Plausibel erscheint die Zuschreibung des Intercepts der *t*-Komponente, wohingegen für die *b*-Richtung die folgende Bedingung gesetzt wird:

$$\alpha_k^{(b)}(b_{\text{start}}) = 0, \quad k = 0, 1, \dots, p,$$

wobei  $b_{\text{start}}$  als der Anfang der kalendarischen Zeitmessung (d. h. kalendarischer Studienanfang) gilt.

## 5.2 Schätzalgorithmus

Im Prinzip könnte die Schätzung der Parameter  $\boldsymbol{\theta}_t, \boldsymbol{\theta}_b$  sowie  $\boldsymbol{\lambda}_t, \boldsymbol{\lambda}_b$  direkt erfolgen, so wie im Fall mit nur einer Zeitachse. Dies kann allerdings numerisch und computer-rechnerisch aufwendig sein, insbesondere dann, wenn die Anzahl der Kovariablen relativ groß ist. Die Schätzung der Parameter soll hier deshalb in der Art der "penalized Backfitting"-Strategie durchgeführt werden. Unter "Backfitting" verstehen wir die Idee des iterativen Wechsels zwischen  $\boldsymbol{\theta}_t$  und  $\boldsymbol{\theta}_b$  beim Schätzen. (In der Literatur ist der Begriff des Backfitting vor allem im Zusammenhang mit dem Schätzverfahren für generalisierte additive Modelle bekannt, wie dies in Hastie & Tibshirani, 1990, beschrieben ist, siehe auch Nielsen & Sperlich, 2005.) In der Darstellung des von uns entwickelten Verfahrens stützen wir uns auf die Beschreibung bei Kauermann & Khomski (2006). Zunächst wird angenommen, dass die Schätzkomponenten, verbunden

mit  $t$ -Richtung, gegeben sind, d. h.,  $\boldsymbol{\alpha}_t(\cdot) = \mathbf{W}_t(\cdot)\boldsymbol{\theta}_t$  sei fix. Wir können in diesem Fall die Formel (5.4) umschreiben in der Form

$$\begin{aligned} \ell_{\text{pen},b}(\boldsymbol{\theta}_b, \boldsymbol{\lambda}_b) &= \ell(\boldsymbol{\theta}_b) - \frac{1}{2} \mathbf{b}_b^T \mathbf{D}_b(\boldsymbol{\lambda}_b) \mathbf{b}_b \\ &= \sum_{i=1}^n [\delta_i \mathbf{W}_b(b_i) \boldsymbol{\theta}_b - \exp\{\mathbf{W}_b(b_i) \boldsymbol{\theta}_b + o_{bi}\}] - \frac{1}{2} \boldsymbol{\theta}_b^T \mathbf{D}_b \boldsymbol{\theta}_b, \end{aligned} \quad (5.5)$$

mit  $o_{bi} = \log\left[\sum_{j=1}^{J_i} \exp\{\mathbf{W}_t(t_{ij}) \boldsymbol{\theta}_t + o_j^{(i)}\}\right]$  als Offset-Parameter und  $o_j^{(i)}$  wie in (3.7) definiert. Daraus wird ersichtlich, dass (5.5) gleich der penalisierten Likelihood ist für  $n$  Poisson-verteilte Daten  $\delta_i$ ,  $i = 1, \dots, n$ , d. h. der Vektor der Zensierungsindikatoren wird aufgefasst als Poisson-verteilt (mit niedriger Intensität). Die ersten und zweiten Ableitungen bezüglich  $\boldsymbol{\theta}_b$  liefern jeweils den Score-Vektor und die (beobachtete, mit  $-1$  multiplizierte) Fisher-Matrix:

$$\mathbf{s}_{\text{pen},b}(\boldsymbol{\theta}_b, \boldsymbol{\lambda}_b) = \sum_{i=1}^n \mathbf{W}_b^T(b_i) [\delta_i - \exp\{\mathbf{W}_b(b_i) \boldsymbol{\theta}_b + o_{bi}\}] - \mathbf{D}_b \boldsymbol{\theta}_b, \quad (5.6)$$

$$\mathbf{I}_{\text{pen},b}(\boldsymbol{\theta}_b, \boldsymbol{\lambda}_b) = \sum_{i=1}^n \mathbf{W}_b^T(b_i) \mathbf{W}_b(b_i) \exp\{\mathbf{W}_b(b_i) \boldsymbol{\theta}_b + o_{bi}\} + \mathbf{D}_b. \quad (5.7)$$

Wir können den update-Wert für  $\boldsymbol{\theta}_b$  bestimmen, indem wir z. B. den Newton-Raphson Iterationsschritt bilden. Nun fixieren wir die  $b$ -Richtung, d. h.  $\boldsymbol{\alpha}_b(\cdot) = \mathbf{W}_b(\cdot)\boldsymbol{\theta}_b$  wird als gegeben vorausgesetzt. Die  $\log$ -Likelihood (5.4) wird dann zu

$$\ell_{\text{pen},t}(\boldsymbol{\theta}_t, \boldsymbol{\lambda}_t) = \sum_{i=1}^n \sum_{j=1}^{J_i} [\delta_{ij} \mathbf{W}_t(t_{ij}) \boldsymbol{\theta}_t - \exp\{\mathbf{W}_t(t_{ij}) \boldsymbol{\theta}_t + o_{tij}\}] - \frac{1}{2} \boldsymbol{\theta}_t^T \mathbf{D}_t \boldsymbol{\theta}_t, \quad (5.8)$$

mit  $o_{tij} = \mathbf{W}_b(b_i) \boldsymbol{\theta}_b + o_j^{(i)}$  als Offset-Parameter. Analog werden Score-Vektor sowie (mit  $-1$  multiplizierte) Fisher-Matrix bezüglich  $\boldsymbol{\theta}_t$  bestimmt durch:

$$\mathbf{s}_{\text{pen},t}(\boldsymbol{\theta}_t, \boldsymbol{\lambda}_t) = \sum_{i=1}^n \sum_{j=1}^{J_i} \mathbf{W}_t^T(t_{ij}) [\delta_{ij} - \exp\{\mathbf{W}_t(t_{ij}) \boldsymbol{\theta}_t + o_{tij}\}] - \mathbf{D}_t \boldsymbol{\theta}_t, \quad (5.9)$$

$$\mathbf{I}_{\text{pen},t}(\boldsymbol{\theta}_t, \boldsymbol{\lambda}_t) = \sum_{i=1}^n \sum_{j=1}^{J_i} \mathbf{W}_t^T(t_{ij}) \mathbf{W}_t(t_{ij}) \exp\{\mathbf{W}_t(t_{ij}) \boldsymbol{\theta}_t + o_{tij}\} + \mathbf{D}_t. \quad (5.10)$$

In analoger Weise bestimmt sich der update-Wert für  $\boldsymbol{\theta}_t$ . Der Backfitting-Algorithmus (mit Newton-Raphson Schritt) resultiert wie folgt:

(0) Initialisiere die Anfangswerte für  $\theta_t$  und  $\theta_b$ . Diese kann man z. B. erhalten, indem zunächst die Koeffizienten  $\mathbf{b}_t \equiv 0$  sowie  $\mathbf{b}_b \equiv 0$  gesetzt werden (bei hochdimensionalem Teil der glatten Komponenten) und indem ein generalisiertes lineares Modell an die Poisson-Daten  $\delta$  angewandt wird, um die fixen Parameter  $\beta_t$  und  $\beta_b$  zu schätzen. Wir bezeichnen die resultierenden Schätzer als  $\hat{\theta}_t^{(0)}$  und  $\hat{\theta}_b^{(0)}$ .

(r.b) Der  $r$ -te Iterationsschritt ergibt sich aus dem Update-Wert für  $\theta_b$ :

$$\hat{\theta}_b^{(r)} = \hat{\theta}_b^{(r-1)} + \mathbf{I}_{\text{pen},b}^{-1} \left( \hat{\theta}_b^{(r-1)}, \lambda_b \right) \mathbf{s}_{\text{pen},b} \left( \hat{\theta}_b^{(r-1)}, \lambda_b \right),$$

mit dem Offset  $o_{bi}$  in (5.6), berechnet aus  $\theta_t^{(r-1)}$ , sowie

(r.t) dem Update des Wertes  $\hat{\theta}_t^{(r-1)}$  durch

$$\hat{\theta}_t^{(r)} = \hat{\theta}_t^{(r-1)} + \mathbf{I}_{\text{pen},t}^{-1} \left( \hat{\theta}_t^{(r-1)}, \lambda_t \right) \mathbf{s}_{\text{pen},t} \left( \hat{\theta}_t^{(r-1)}, \lambda_t \right),$$

wobei der Offset  $o_{tij}$  in (5.9) jetzt aus  $\theta_b^{(r)}$  berechnet wird.

(r.b $\leftrightarrow$ r.t) Das Iterieren zwischen (r.b) und (r.t) komplettiert die Backfitting-Strategie.

Einige Kommentare sind hier nötig:

1. Es ist (optimierungstechnisch) selbstverständlich irrelevant, welche der beiden Zeitachsen zunächst fixiert wird.
2. Man kann mehrere Newton-Raphson Iterationen für  $\theta_t$  oder  $\theta_b$  durchführen, bevor man die jeweilige Zeitachse neu fixiert.
3. Es sind gerade die neu berechneten Offsets, die den Link zwischen den beiden Updates der Schätzwerte ermöglichen.
4. Bei "gut konditionierten" Modellen funktioniert das Verfahren sehr gut. In schwierigeren Situationen behilft man sich (wie in vielen ähnlichen Fällen) durch die Reduktion des Newton-Schrittes (indem z. B. bekannte Line-Search oder Trust-Region Verfahren angewandt werden, um die optimale Schrittgröße zu bestimmen) oder setzt Levenberg-Marquardt Strategie ein, vgl. hierzu etwa Fletcher (1987) und Nocedal & Wright (1999).
5. Des Newton-Schrittes wegen kann es durchaus vorkommen, dass man in einem lokalen Minimum landet (sofern die Likelihoods kein eindeutiges Extremum aufweisen). In diesem Fall wird empfohlen, es mit mehreren Anfangswerten zu versuchen, um ein lokales Maximum zu identifizieren.

## 5.3 Berechnung der Penaliserungsparameter

Die Berechnung der Parameter  $\boldsymbol{\lambda}_t$ ,  $\boldsymbol{\lambda}_b$ , die den Glattheitsgrad der variierenden Koeffizienten steuern, erhalten wir, in Übereinstimmung mit den Ausführungen der Abschnitte 3.4 und 3.5, indem das obige Modell als GLMM-Modell dargestellt wird, in welchem die Penaliserungsmatrizen die Rolle der a priori Kovarianzmatrix der Spline-Koeffizienten spielen. Wir unterstellen die folgende Verteilung der Modellkomponenten:

$$\begin{aligned} \delta_i | \mathbf{b}_b, \boldsymbol{\alpha}_t(\cdot) &\sim \text{Poisson}(\exp\{\mathbf{W}(b_i)\boldsymbol{\theta}_b + o_{bi}\}), \\ \mathbf{b}_{bl} &\sim \mathcal{N}(0, \lambda_{bl}^{-1} \mathbf{D}_{bl}^-), \quad l = 0, 1, \dots, p, \end{aligned} \quad (5.11)$$

mit  $\mathbf{D}_{bl}^-$  als (generalisierte) Inverse von  $\mathbf{D}_{bl}$ ,  $l = 0, 1, \dots, p$ . Die Marginalisierung des Modells, d. h. das Ausintegrieren der zufälligen Komponenten  $\mathbf{b}_b$ , führt zur marginalen Likelihood. Das Integral wird durch die Laplace-Approximation ersetzt (siehe Anhang A.5), welche bei der auf die PQL-gestützten Methode ähnliche Schätzwerte liefert. Somit wird der Parameter  $\boldsymbol{\lambda}_b$  als zu optimierender Parameter in der bezüglich  $\boldsymbol{\beta}_b$  und  $\mathbf{b}_b$  maximierten marginalen  $\log$ -Likelihood (5.5) aufgefasst. Die Laplace-Approximation liefert also:

$$\begin{aligned} \ell_{m,b|t}(\boldsymbol{\lambda}_b) &\approx \ell_{\text{pen},b}(\hat{\boldsymbol{\theta}}_b, \boldsymbol{\lambda}_b) - \frac{1}{2} \log |\mathbf{D}_b^-| - \frac{1}{2} \log \left| -\frac{\partial^2 \ell_{\text{pen},b}(\hat{\boldsymbol{\theta}}_b, \boldsymbol{\lambda}_b)}{\partial \mathbf{b}_b \partial \mathbf{b}_b^T} \right| \\ &= \ell_{\text{pen},b}(\hat{\boldsymbol{\theta}}_b, \boldsymbol{\lambda}_b) - \frac{1}{2} \log \left| \tilde{\mathbf{I}}_{\text{pen},b}(\hat{\boldsymbol{\theta}}_b, \boldsymbol{\lambda}_b) \mathbf{D}_b^- \right|, \end{aligned} \quad (5.12)$$

wobei  $\hat{\boldsymbol{\theta}}_b$  die  $\log$ -Likelihood  $\ell_{\text{pen},b}$  maximiert und  $\mathbf{D}_b$  die blockdiagonale Matrix ist, mit  $\lambda_{bl} \mathbf{D}_{bl}$  als Einzelblöcke (vgl. mit Formel (3.14)). Analog ist  $\tilde{\mathbf{I}}_{\text{pen},b}(\cdot)$  die Submatrix der penalisierten Fisher-Matrix (5.7) mit den Elementen, entsprechend zu  $\mathbf{b}_b$ . Die Ableitungen bezüglich  $\boldsymbol{\lambda}_b$  sind nun leicht zu bestimmen (siehe Herleitung in Anhang A.7). Für die Fixpunkt-Iteration bekommen wir den Ausdruck

$$\frac{1}{\hat{\lambda}_{bl}} = \frac{\text{tr}\left\{ \left[ \tilde{\mathbf{I}}_{\text{pen},b}^{-1}(\hat{\boldsymbol{\theta}}_b, \boldsymbol{\lambda}_b) \right]_{ll} \mathbf{D}_{bl} \right\} + \hat{\mathbf{b}}_{bl}^T \mathbf{D}_{bl} \hat{\mathbf{b}}_{bl}}{\dim(\mathbf{B}_{bl})}, \quad (5.13)$$

wobei der Subindex  $ll$  den  $l$ -ten Block an der Diagonalen von der Inversen der Fisher-Matrix bezeichnet. Mit  $\dim(\mathbf{B}_{bl})$  kennzeichnen wir die Anzahl der penalisierten Splines in der gewählten Splinebasis der truncated polynomials in  $b$ -Richtung. Wir bemerken hier, dass für die Wahl der truncated polynomials die Matrix  $\mathbf{D}_{bl}$  die Einheitsmatrix ist. Daher vereinfacht sich (5.13) erheblich. Man kann hier auch die zweiten Ableitungen bilden, um die Berechnung nach

Newton-Raphson durchzuführen (siehe Anhang A.7). Allerdings haben wir in Simulationen und konkretem Beispiel festgestellt, dass dies keine deutliche Effizienz bezüglich der Rechenzeit und Genauigkeit mit sich bringt. Daher scheint die Evaluierung nach der Fixpunkt-Iteration Methode völlig ausreichend. In kompletter Analogie lässt sich nun der Fall

$$\begin{aligned} \delta_{ij} | \mathbf{b}_t, \boldsymbol{\alpha}_b(\cdot) &\sim \text{Poisson}(\exp\{\mathbf{W}_t(t_{ij})\boldsymbol{\theta}_t + o_{tij}\}), \\ \mathbf{b}_{tl} &\sim \mathcal{N}(0, \lambda_{tl}^{-1} \mathbf{D}_{tl}^-), \quad l = 0, 1, \dots, p, \end{aligned} \quad (5.14)$$

behandeln, mit  $\boldsymbol{\theta}_b$  fix. Die Approximation nach Laplace ergibt hier die marginale  $\log$ -Likelihood:

$$\ell_{m,t|b}(\boldsymbol{\lambda}_t) \approx \ell_{\text{pen},t}(\hat{\boldsymbol{\theta}}_t, \boldsymbol{\lambda}_t) - \frac{1}{2} \log \left| \tilde{\mathbf{I}}_{\text{pen},t}(\hat{\boldsymbol{\theta}}_t, \boldsymbol{\lambda}_t) \mathbf{D}_t^- \right|,$$

wobei  $\mathbf{D}_t$  blockdiagonal ist, mit den Einzelblöcken  $\lambda_{tl} \mathbf{D}_{tl}$  (und den Einheitsmatrizen  $\mathbf{D}_{tl}$ ). Der update-Wert für den Schätzer für  $\boldsymbol{\lambda}_t$  ergibt sich aus

$$\frac{1}{\hat{\lambda}_{tl}} = \frac{\text{tr}\left\{ \left[ \tilde{\mathbf{I}}_{\text{pen},t}^{-1}(\hat{\boldsymbol{\theta}}_t, \boldsymbol{\lambda}_t) \right]_{ll} \mathbf{D}_{tl} \right\} + \hat{\mathbf{b}}_{tl}^T \mathbf{D}_{tl} \hat{\mathbf{b}}_{tl}}{\dim(\mathbf{B}_{tl})}, \quad (5.15)$$

mit  $\dim(\mathbf{B}_{tl})$  als Anzahl der penalisierten truncated polynomials in  $t$ -Richtung. Die iterative Bestimmung von  $\boldsymbol{\lambda}_t$  und  $\boldsymbol{\lambda}_b$  lässt sich natürlich in die Backfitting-Strategie einbauen. Die Anzahl der Iterationen für update-Werte kann ebenfalls festgelegt werden.

## 5.4 Bestimmung von Varianzen und Freiheitsgraden

Die Methode erlaubt die Bestimmung der Kovarianz von  $\hat{\boldsymbol{\theta}}_{t,b}$ . Die in Abschnitt 3.6 begründete Approximation durch die Normalverteilung liefert als "Sandwich"-Schätzer (vgl. Ruppert, Wand & Carroll, 2003, Abschnitte 6.4 und 11.4):

$$\widehat{\text{Var}}(\hat{\boldsymbol{\theta}}_{t,b}) = \mathbf{I}_{\text{pen},tb}^{-1}(\hat{\boldsymbol{\theta}}_{t,b}, \hat{\boldsymbol{\lambda}}_{t,b}) \mathbf{I}_{tb}(\hat{\boldsymbol{\theta}}_{t,b}) \mathbf{I}_{\text{pen},tb}^{-1}(\hat{\boldsymbol{\theta}}_{t,b}, \hat{\boldsymbol{\lambda}}_{t,b}) \quad (5.16)$$

mit  $(\mathbf{I}_{\text{pen},tb}) \mathbf{I}_{tb}$  als (penalisierte) Fisher-Information des Modells. Daraus können die (punktweisen) Konfidenzintervalle für die zugehörigen glatten Komponenten des Modells leicht berechnet werden. Es gilt hier für den variierenden Koeffizienten  $\hat{\alpha}_{tl}(\cdot) = \mathbf{W}_{tl}(\cdot) \hat{\boldsymbol{\theta}}_{tl}$ :

$$\text{Var}\{\hat{\alpha}_{tl}(t)\} = \mathbf{W}_{tl}(t) \text{Var}\{\hat{\boldsymbol{\theta}}_{tl}\} \mathbf{W}_{tl}(t)^T, \quad l = 0, 1, \dots, p, \quad (5.17)$$



und analog für  $\hat{\alpha}_{bl}(\cdot)$ .

Eine geeignete Möglichkeit, quantitativ den Glättungsgrad auszudrücken, ist durch die Bestimmung der Freiheitsgrade ("degrees of freedom") der Modellkomponenten gegeben. In Übereinstimmung mit Hastie & Tibshirani (1990) und mit der Weiterentwicklung von Gray (1992) lassen sich diese in folgender Weise berechnen (siehe Ruppert, Wand & Carroll, 2003, 11.4). Zunächst bestimmen sich die totalen Freiheitsgrade als

$$df(\boldsymbol{\lambda}) = \text{tr} [\mathbf{I}_{\text{pen},tb}^{-1} \mathbf{I}_{tb}]. \quad (5.18)$$

Die Freiheitsgrade, die den Richtungen  $t$  und  $b$  entsprechen, lassen sich daraus wie folgt berechnen:

$$df_{tl}(\boldsymbol{\lambda}_{tl}) = \text{tr} [\mathbf{E}_{tl} \mathbf{I}_{\text{pen},tb}^{-1} \mathbf{I}_{tb}], \quad l = 0, 1, \dots, p, \quad (5.19)$$

$$df_{bl}(\boldsymbol{\lambda}_{bl}) = \text{tr} [\mathbf{E}_{bl} \mathbf{I}_{\text{pen},tb}^{-1} \mathbf{I}_{tb}], \quad l = 0, 1, \dots, p, \quad (5.20)$$

wobei  $\mathbf{E}_{tl}$  bzw.  $\mathbf{E}_{bl}$  die Diagonalmatrizen sind mit 1-Einträgen an den Stellen, die den korrespondierenden Parametern der glatten Komponenten des Modells entsprechen, und andernfalls mit 0 gefüllt. (Eine umfassende Diskussion zur Berechnung geben Ruppert, Wand & Carroll, 2003, Abschnitte 8.3 und 11.4.) Natürlich stimmt in diesem Fall  $df = df_t + df_b = \sum_l df_{tl} + \sum_l df_{bl}$ .

## 5.5 Simulationsstudie

In diesem Abschnitt beschreiben wir das zur Validitätsüberprüfung des entwickelten Modells benötigte Simulationsverfahren. Dieses wird auch im Modell in Kapitel 6 benutzt sowie (mit weiteren Spezifikationen) im Modell der Competing Risks (Kapitel 8). Der wesentliche Punkt ist dabei die Generierung der künstlichen Poisson-verteilten Beobachtungen. Das Ziel der Simulation ist die möglichst genaue Wiedergabe der simulierten glatten Komponenten, also der variierenden Koeffizienten.

Es seien zunächst die Anzahl der Beobachtungen  $N$ , die maximale Überlebensdauer (oder der Schwellenwert)  $T$ , die Geburtszeiten (also die kalendarischen Eintrittspunkte in die Studie)  $b$  sowie eine binäre Kovariable  $x$  spezifiziert. Die Überlebensdauer  $t$  sowie der Zensierungsindikator  $\delta$  sind in die Simulation hineinintegriert. Für jede Beobachtung  $i$ ,  $i = 1, \dots, N$ , bestimmen wir den zugehörigen Vektor der künstlichen Poisson-Daten wie folgt. Die (diskrete) Überlebenszeit  $t_i$  wird zuerst auf 1 gesetzt. Wir sampeln aus der Poisson-Verteilung mit der spezifizierten niedrigen Intensität; diese wird definiert (siehe Kapitel 3) durch

$$\xi = \exp\{ \alpha_0^t(t) + \alpha_0^b(b) + x[ \alpha_x^t(t) + \alpha_x^b(b) ] \}, \quad (5.21)$$

mit den glatten Komponenten  $\alpha_0^t(t)$ ,  $\alpha_0^b(b)$  für die Baseline-Hazard und  $\alpha_x^t(t)$ ,  $\alpha_x^b(b)$  für den Effekt der Kovariablen  $x$ . Da für die Elemente des Zählvektors  $\delta_i$  nur die Werte 0 oder 1 zulässig sind, notieren wir den Wert 0, falls dieser gesampelt wird, und wiederholen die Samplingroutine (dabei erhöht sich die Verweildauer  $t_i$  um eine Einheit:  $t_i \rightarrow t_i + 1$ ), anderenfalls wird dem Poisson-Vektor  $\delta_i$  der Wert 1 am Ende hinzugeschrieben, der Wert von  $t_i$  wird dem zugehörigen Iterationswert gleichgesetzt und die Prozedur terminiert. Falls bis zum Erreichen des maximalen Wertes  $t_{\max} = T$  keine 1 kommt, setzen wir die  $i$ -te Überlebensdauer  $t_i = T$  und Zensierungsindikator  $\delta_i = 0$ . Hier seien die Simulationskomponenten nochmals aufgelistet.

### Simulations-Input

- Anzahl der Beobachtungen  $N$ ,
- binäre Kovariable  $x$ ,
- Überlebenszeiten  $t$ , Eintrittspunkte  $b$ ,
- binärer Zensierungsindikator  $\delta$ ,
- glatte Komponenten  $\alpha_t$ ,  $\alpha_b$ ,
- Anzahl der Simulationen.

Die Häufigkeitsverteilung der zensierten Beobachtungen im derart generierten Datensatz wird allein durch die Spezifikation der glatten Komponenten bestimmt. Die unterschiedlichen Häufigkeitsmuster sowie die Anzahl  $N$  der Beobachtungen werden variiert. Wir setzen folgendes Verteilungsschema:  $N = 100, 500, 1000$  Beobachtungen für den jeweils variierenden Anteil zensierter Beobachtungen 2%, 10%, 25%, 50%, 75%, 90% mit der folgenden Wahl der glatten Komponenten (als Funktionenbasis sind die truncated polynomials genommen):

$$\alpha_0^t(t) = \eta - \frac{t}{30}, \quad (5.22)$$

$$\alpha_0^b(b) = 3 \left[ \left( \frac{b}{50} \right)^2 - 0.833 \left( \frac{b}{50} \right) \right], \quad (5.23)$$

$$\alpha_x^t(t) = 0.5 + \left( \frac{t}{25} \right)^2, \quad (5.24)$$

$$\alpha_x^b(b) = -\frac{b}{60}, \quad (5.25)$$

wobei der Parameter  $\eta$  im Ausdruck für die Baseline-Hazard in  $t$ -Richtung gerade die relative Häufigkeit der zensierten Beobachtungen steuert. Die Intensität ist  $\xi = \exp\{\alpha_0^t(t) + \alpha_0^b(b) + x[\alpha_x^t(t) + \alpha_x^b(b)]\}$ . Die folgenden eingesetzten Werte für  $\eta$  entsprechen den Zensierungsmustern  $h_\delta$ :

$\eta = -1.4$	$\rightarrow$	$h_\delta \simeq 2\%$ ,
$\eta = -2$	$\rightarrow$	$h_\delta \simeq 10\%$ ,
$\eta = -2.5$	$\rightarrow$	$h_\delta \simeq 25\%$ ,
$\eta = -3.3$	$\rightarrow$	$h_\delta \simeq 50\%$ ,
$\eta = -4.2$	$\rightarrow$	$h_\delta \simeq 75\%$ ,
$\eta = -5.1$	$\rightarrow$	$h_\delta \simeq 90\%$ .

Anzumerken ist dabei, dass das Zensierungsmuster  $h_\delta \simeq 50\%$  etwa dem für den Analysedatensatz der Arbeitslosen nahe kommt (siehe Beschreibung in Kapitel 11), in dem die Anzahl der zensierten Beobachtungen bei ca. 55% liegt. Für die jeweilige Anzahl der Beobachtungen  $N$  und für das jeweilige Zensierungsmuster  $h_\delta$  ergeben sich insgesamt  $3 \times 6 = 18$  Konstellationen. Für jede dieser Konstellationen erzeugen wir eine Serie von insgesamt 100 Simulationen.

Für die generierte Überlebensdauer  $t$  betrachten wir die diskreten Zeitpunkte von 1 bis 30. Zensiert sind dabei diejenigen Zeiten, die den Schwellenwert 30 übertreffen. Die kalendarischen Zeitpunkte  $b$  (also jeweiliger Beginn oder Eintritt in die Studie) wählen wir als gleichverteilte diskrete Punkte auf dem Intervall 0 bis 50. Die binäre Kovariable  $x$  wird mit Wahrscheinlichkeit  $P(x = 1) = 0.7$ ,  $P(x = 0) = 0.3$  simuliert. Als Modell-Splines wählen wir die penalisierten truncated polynomials 1-ten Grades. Die Smoothing-Parameter werden datengesteuert bestimmt, basierend auf der in Abschnitt 5.3 beschriebenen Prozedur.

Als Output der Simulation sind die "coverage probabilities" sowie die geschätzten empirischen 5%-, 50%-, 95%- Quantile der simulierten glatten Komponenten wiedergegeben; diese sind in Anhang B.1 dargestellt. Die empirischen Quantile sind als die punktweise (für jedes  $t$  und  $b$ ) berechneten Quantile aus allen Simulationen bestimmt und zeigen den Überdeckungsgrad der simulierten (also wahren) glatten Komponente. Die coverage probabilities sind die Indikatoren für die Güte der Varianzschätzung. Sie werden ermittelt als prozentualer Anteil der Simulationen, in denen die berechneten (punktweisen 95%-)Konfidenzbänder für jedes fixe  $t$  bzw.  $b$  die wahre Funktion an der jeweiligen Stelle beinhalten. Wie wir den Abbildungen in Anhang B.1 entnehmen können, sind die Gesamtanzahl sowie der prozentuale Anteil der zensierten Beobachtungen für die Modellgüte entscheidend. Die empirischen Konfidenzbänder werden bei erhöhter Anzahl der Beobachtungen enger, d. h. die simulierten Komponenten werden durch die Simulation genauer wiedergegeben. Dies wirkt sich auf bessere coverage probabilities aus, da die Überdeckung der wahren Komponente sicherer (also wahrscheinlicher) wird. Auch der empirische Bias wird kleiner, so dass die Anzahl von 1000 Beobachtungen im Datensatz als ausreichend erscheint. Mit der Zunahme der zensierten Beobachtungen und bei deren großen Gesamtanzahl verzerrt sich das Bild nur geringfügig und wird erst bei einem sehr großen Prozentsatz (also ab ca. 75%)

bemerkbar. In Anlehnung an den realen Datensatz der Arbeitslosen ist das Simulationsergebnis daher überzeugend. Einer kritischen Bemerkung bedarf es für die simulierte Baseline-Hazard (5.23) mit einer niedrigeren coverage probability bei den kleineren Werten von  $b$ . Dieser Sachverhalt ist vermutlich mit einem kleinen Bias für die Anfangswerte verlinkt, der seinerseits wegen der Spezifizierung der Splinebasis zustande kommt (quadratische Form der wahren Baseline bei der Splinebasis als "truncated lines"). In weiteren (nicht publizierten) Experimenten für eine andere funktionale Form der Baseline konnten wir tatsächlich die Reduzierung der Verzerrung feststellen, wenn die gewählte Splinebasis die zu simulierende funktionale Form der glatten Komponenten widerspiegelt. Dennoch scheint die Wahl der truncated lines passend gewählt, da der Trend in der Überlebenszeit ausgeprägter ist als in der Kalenderzeit.

## 5.6 Anwendungsbeispiel

Der Datensatz der Arbeitslosen ist in Kapitel 11 ausführlich dargestellt. Das eingangs erwähnte Lexis-Diagramm, Abbildung 5.1, sowie deskriptive Bilder des Kaplan-Meier-Schätzers und der Verteilung des Eintritts in den Arbeitslosigkeitszustand bezüglich der Kalenderjahre, Abbildung 5.2, ermöglichen einen ersten Überblick zu den Daten. So scheint die Abnahme der Perioden der Arbeitslosigkeit Ende 1980er / Anfang 1990er Jahre mit dem allgemeinen Rückgang der Arbeitslosenquoten in Zusammenhang zu stehen. Auch die Arbeitslosigkeitszeiträume scheinen bezüglich der Kalenderzeit zu differenzieren, ganz deutlich vom Kaplan-Meier-Schätzer identifiziert, zumindest für zwei Zeitabschnitte: vor den 1990er Jahren und später.

Mehr Information unter Berücksichtigung der individuellen Unterschiede bekommen wir aus dem Analyseergebnis. Die flexiblen Kovariableneffekte sind in Abbildung 5.3 geplottet. Für eine bessere visuelle Darstellung und um einen Vergleich zu ermöglichen sind hier alle Effekte bezüglich der beiden Zeitkomponenten auf derselben  $\log$ -Skala von -3 bis 3 gezeichnet, mit Ausnahme der Baselines. Es reduzieren sich die Chancen (beinahe linear auf der  $\log$ -Skala) auf eine Wiederbeschäftigung mit der Dauer der Arbeitslosigkeit, dies auch über die Jahre hinweg, wenn auch nicht deutlich signifikant, da die bestimmten Konfidenzbänder bezüglich der Kalenderzeit breit sind und das 0-Niveau überdecken. Der Effekt der Kovariablen *Nation* variiert mit der Arbeitslosigkeitsdauer, mit weniger Chancen auf eine erneute Vollbeschäftigung bei Ausländern im ersten Jahr, mit keinerlei signifikanten Unterschieden bis ungefähr 24 Monaten der Arbeitslosigkeit und nach zwei Jahren mit umgekehrter Wirkung, die die Chancen erhöht, im Vergleich zu den deutschen Arbeitnehmern. Andererseits sind keine Unterschiede in Bezug auf die Kalenderzeit festzustellen. Weiter gelten die signifikanten geschlechtsbedingten Unterschiede: Die weiblichen Personen scheinen generell schlechtere Chancen auf eine Wiederbeschäftigung zu haben als die männlichen; dies unabhängig von der Arbeitslosigkeitsdauer. Somit wirkt dieser Faktor proportional auf die Hazard. Andererseits hat sich dieser negative Effekt über die Jahre hinweg reduziert

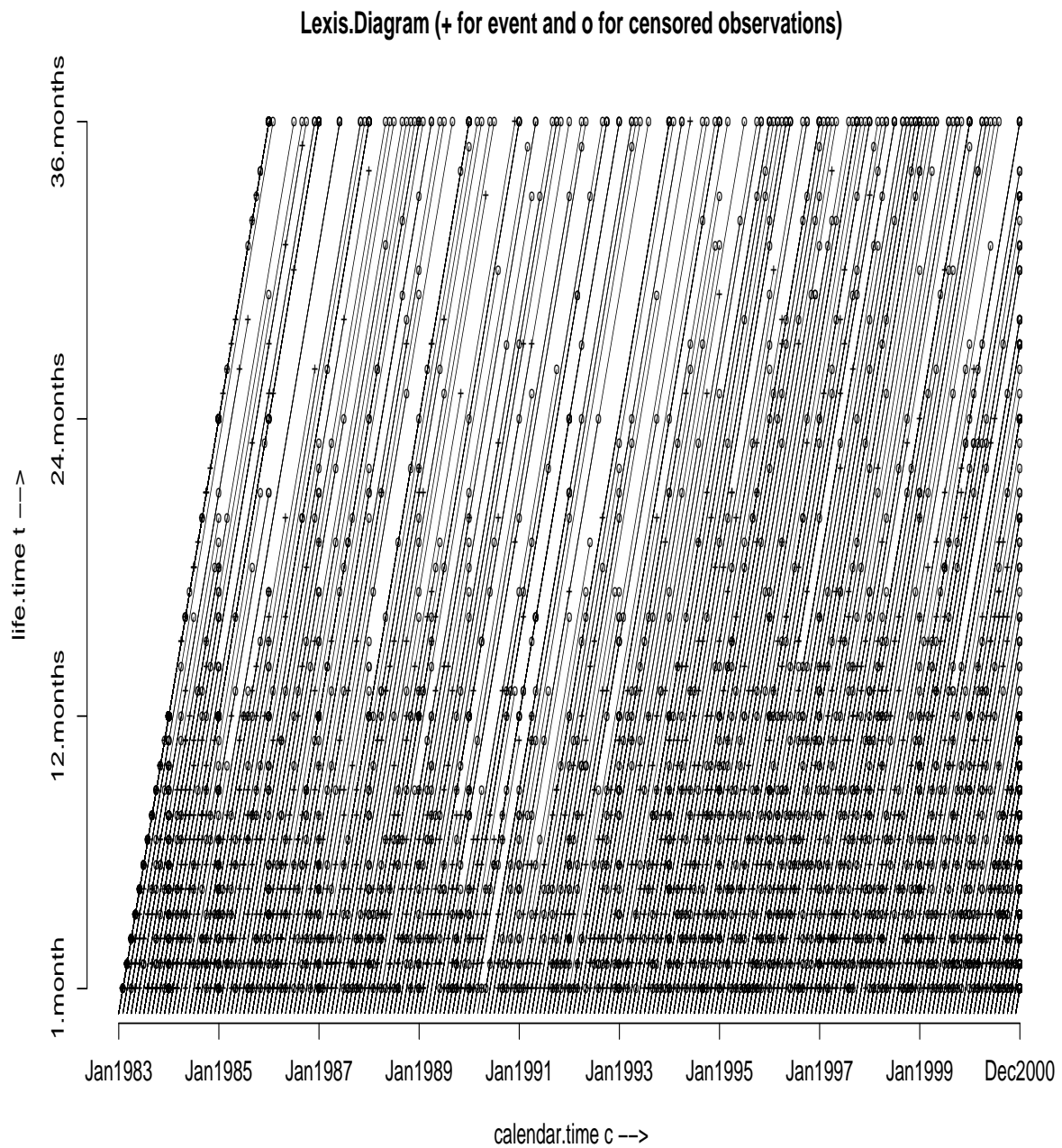


Abbildung 5.1: Lexis-Diagramm der Arbeitslosen, auf Monatsbasis. Die Arbeitslosigkeitszeiträume, die 36 Monate übersteigen, sind auf 36 gesetzt und als zensiert kodiert.

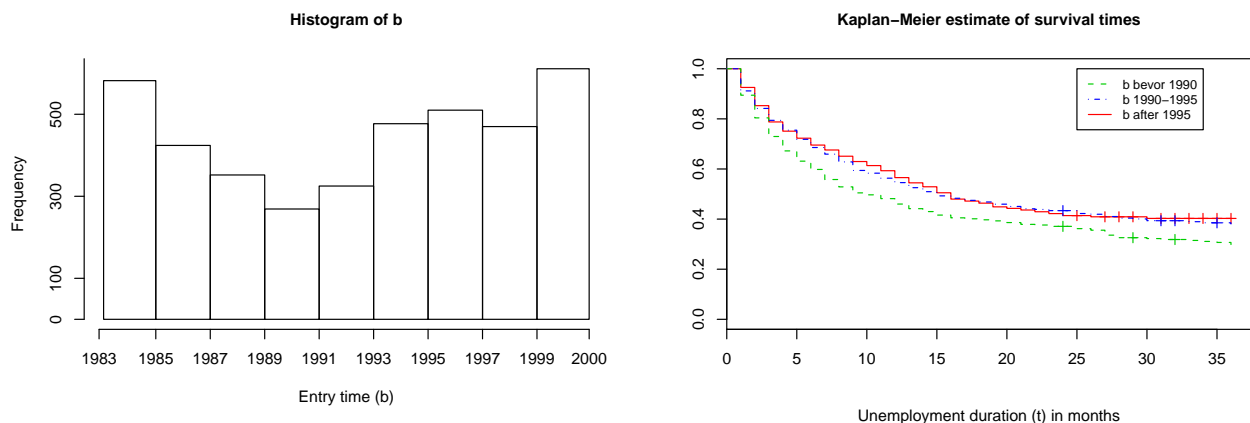


Abbildung 5.2: Histogramm für den kalendarischen Anfang der Arbeitslosigkeit sowie Kaplan-Meier-Schätzer für die Dauer der Arbeitslosigkeit, gruppiert in drei Kalenderabschnitte.

und ist seit den 1990er Jahren auf etwa gleichem Niveau geblieben. Dieser kalenderspezifische Effekt könnte bei den Frauen mit einer höheren Bereitschaft für die Aufnahme einer Beschäftigung erklärt werden. (Familiäre und personale Gründe dürften dann dieser Annahme zugrunde liegen.) Bezüglich der Kovariablen **Alter** gibt es deutliche Effektunterschiede. Die jüngeren Arbeitslosen (bis 25 Jahre) haben eine höhere Chance auf Wiederbeschäftigung im Vergleich zur Referenzkategorie der Erwerbstätigen mittleren Alters (25 bis 50). Ein umgekehrtes Bild ergibt sich für die älteren Arbeitslosen (ab 50 Jahre). In keiner der Jahresgruppen scheinen die zugehörigen Effekte eine ausgeprägte Konstanz zu besitzen. Man kann somit auch hier nicht von einem proportionalen Einfluss ausgehen. Keine klare Signifikanz lässt sich in Bezug auf die Kalenderjahre feststellen. Die berufliche Ausbildung scheint einen proportionalen Effekt zu haben und erhöht die Chancen auf Wiederbeschäftigung, wenn auch in einer nicht signifikanten Weise. Diese Situation hat sich für die nicht qualifizierten Personen mit den Jahren verschlechtert; aber auch hier gibt es keine signifikante Differenz im Vergleich zu den ausgebildeten Arbeitslosen. Personen mit Hochschulabschluss haben im ersten Jahr ihrer Arbeitslosigkeit deutlich höhere Chancen auf eine Einstellung. Der Effekt wird aber zunehmend kleiner und verschwindet gänzlich nach ca. drei Jahren. Bezüglich der Kalenderjahre hat sich die Situation bei den diplomierten Arbeitslosen nicht geändert.

## 5.7 Diskussion zur Weiterentwicklung

Die explizite Miteinbeziehung von kalenderspezifischen flexiblen Effekten ist immer dann von Bedeutung, wenn die Beobachtungen bezüglich ihres Studieneintritts (oder, allgemeiner, ihres Startpunktes auf der Kalenderskala) erheblich variieren. Die entwickelte und angewandte Technik erlaubt eine solche

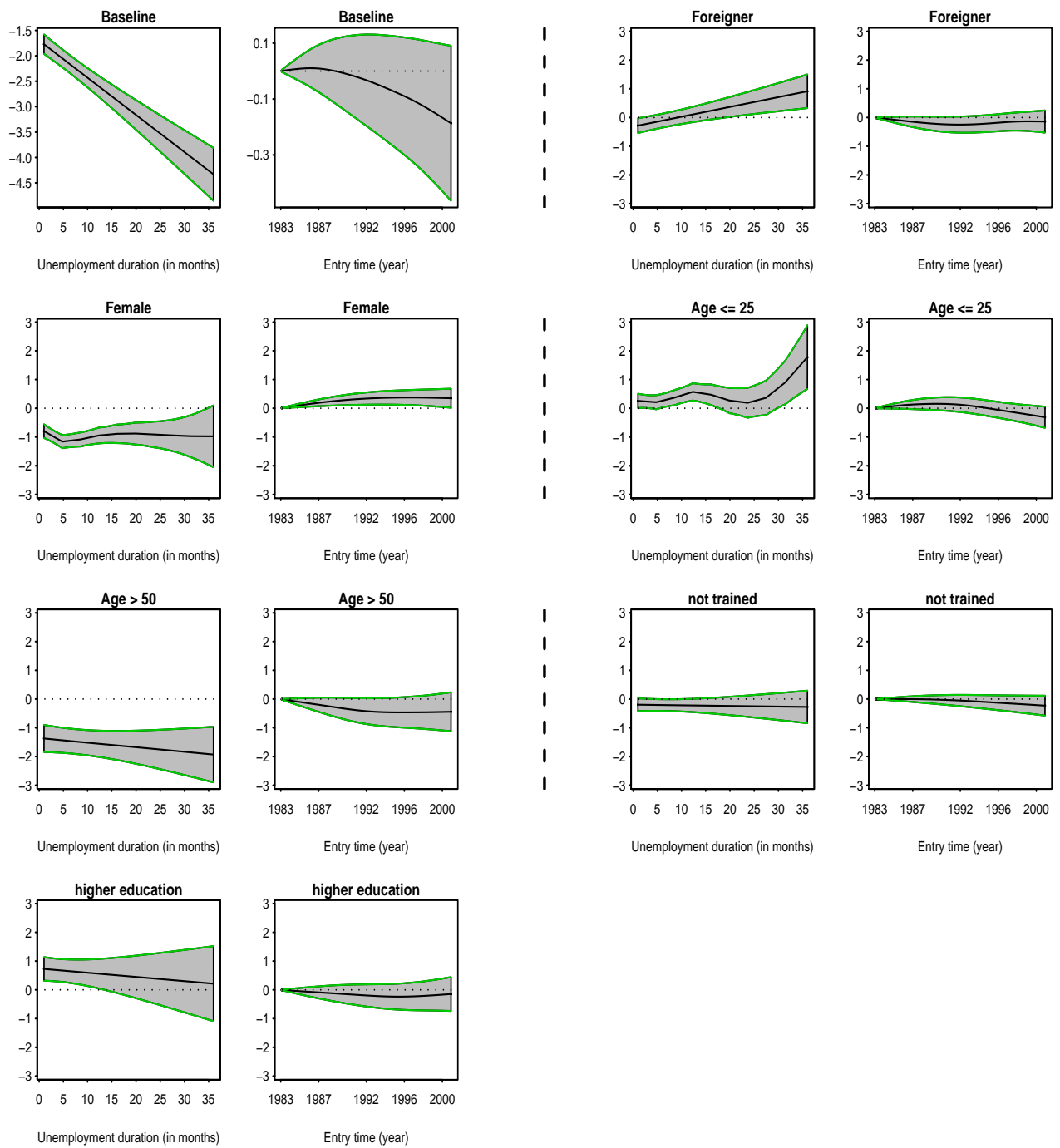


Abbildung 5.3: Funktionale Form von  $\alpha_{tl}(\cdot)$  (links) und  $\alpha_{bl}(\cdot)$  (rechts) für die Arbeitslosen.

Analyse in einer flexiblen nichtparametrischen Art und quantifiziert die Kovariableneffekte. Damit ist aber das breite Spektrum der möglichen Weiterentwicklung noch nicht ausgeschöpft. So kann z. B. die Kalenderzeit als zusätzliche Kovaribale rein parametrisch in das Modell aufgenommen werden. Man modelliert dann die Interaktion zwischen der (kategorisierten) Kalenderzeit und der Verweildauer. Weitaus allgemeiner ist in diesem Fall aber das Modell der Wechselwirkung zwischen den beiden Zeitskalen in nichtparametrischer Form. Dieses Modell wird in Kapitel 7 vorgestellt.



# Kapitel 6

## Two-Way Additives Hazard-Modell mit saisonaler Zeitkomponente

In diesem Modell (Darstellung erfolgt in Anlehnung an Becher, Kauermann & Khomski, 2007) berücksichtigen wir neben der Überlebenszeit die zweite Zeitkomponente, die den kalendarischen Zeitpunkt eines Jahres ausdrückt, also die Saison. Die Schätzroutine wiederholt im Wesentlichen die des Kapitels 5. Zur Anwendung kamen Daten, die vom Institut für Tropenmedizin zu Universität Heidelberg zur Verfügung gestellt wurden. Das Datenmaterial liefert die Sterbezeiten der Kinder im ländlichen Teil von Burkina Faso (Westafrika) und ist in Kapitel 11 beschrieben. Die immer noch viel zu großen Mortalitätsraten in diesem Land machen es erforderlich, möglichst präzise die Risikofaktoren zu identifizieren und ihre Einflüsse abzuschätzen. Dies stellt seitens der statistischen Modellbildung eine Herausforderung dar. Das entwickelte additive Verfahren scheint uns hier überlegen zu sein, um mit dessen Hilfe das Modell anzupassen. Es ist aus den zahlreich durchgeführten Studien bekannt (siehe z. B. Jaffar et al., 2000, und Moore et al., 1997), dass die Saison, in der ein Kind geboren wird, durchaus eine Rolle für seine Überlebenschancen spielt, zumindest in den ersten Lebensmonaten. Stark bedingt scheint dieser Einfluss dadurch zu sein, dass die Wetterbedingungen in den untersuchten Ländern saisonale Abhängigkeiten aufweisen. So gibt es in Burkina Faso zwei Jahresabschnitte, die Regenzeit (Juni bis Oktober) und die trockene Periode (November bis Mai), wobei die Regenzeit die Verbreitung von Krankheiten, insbesondere von Malaria, begünstigt. Um den Einfluss der Saison flexibel zu modellieren, können wir die saisonale Komponente genau in der Art ins Modell einbeziehen, wie es für die Kalenderzeit in Kapitel 5 praktiziert.

Als Benchmark dient dabei die auf dem klassischen Cox-Modell basierende Studie von Becher, Müller, Jahn, Gbangou, Kynast-Wolf & Kouyaté (2004) (siehe auch Becher & Kouyaté, 2005), in der unter anderen Kovariablen die Jahreszeit der Geburt als kategorielles Merkmal herangezogen wurde, vgl. auch Tabellen 11.1 und 11.2.

## 6.1 Modellansatz und Schätzstrategie

Zur Darstellung der glatten Komponenten wählen wir die B-Splines, sowohl für die Überlebenszeit  $t$  als auch für die Saison  $s$ . Zu beachten ist hier allerdings eine geeignete Knotenwahl der Spline-Funktionen für die saisonale Komponente. Da diese periodisch modelliert wird, soll der Funktionswert an der Stelle  $s$  mit dem an der Stelle  $s + T$  übereinstimmen, wobei  $T$  die spezifizierte Periode ist. Die B-Splines erlauben, die Periodizität in den Daten zu berücksichtigen, wenn speziell die Basis gewählt wird, wie in Abbildung 2.2, rechts, dargestellt. Die Randknoten werden also als äquidistante Punkte genommen. Um anschließend die zugehörigen Spline-Funktionen links und rechts übereinstimmend an ihren Enden glatt miteinander zu verbinden, bedient man sich der so genannten "Wrapping"-Methode. Diese simple Idee wird in Eilers & Marx (2005) beschrieben: Falls die B-Spline Basis des Grades  $d = \text{bs.degree} = \text{bs.order} - 1$  mit  $K$  Knoten `knots.s` gewählt ist, behalten wir die ersten  $K$  Spalten der Basismatrix in einer neuen und addieren die letzten  $d$  Spalten der Ausgangsmatrix zu den ersten  $d$  der neu gebildeten. Optisch ist die "gewrappte" Basis identisch mit der ursprünglichen Basis. Allerdings gehören nun die Spline-Stücke am linken und rechten Rand (in Abbildung 2.2 sind es jeweils die letzten drei links und rechts) zu einer Spalte in der Basismatrix, "gehen über glatt ineinander". Dieselbe Prozedur wird auch an die zugehörige Differenzenmatrix `Diff.s` angewandt. Der R-Code dazu ist wie folgt:

```
> library(splines)
> Basis.s<-splineDesign(knots=knots.s,ord=bs.order,x=saison)
> Basis.s[,1:bs.degree]<-Basis.s[,1:bs.degree]+
  Basis.s[, (ncol(Basis.s)-(bs.degree-1)):ncol(Basis.s)]
> Basis.s<-Basis.s[,1:(ncol(Basis.s)-bs.degree)]
> K.s<-dim(Basis.s)
> Diff.s<-matrix(0,nrow=K.s-2+bs.degree,ncol=K.s+bs.degree)
> for (i in 1:(K.s+bs.degree-2)) Diff.s[i,i:(i+2)]<-c(1,-2,1)
> Diff.s[,1:bs.degree]<-Diff.s[,1:bs.degree]+
  Diff.s[, (ncol(Diff.s)-(bs.degree-1)):ncol(Diff.s)]
> Diff.s<-Diff.s[,1:(ncol(Diff.s)-bs.degree)]
> Penalty.s<-crossprod(Diff.s)
```

mit den Knoten `knots.s`, B-Spline Ordnung `bs.order=bs.degree+1`, Dimension `K.s` der resultierenden "gewrappten" Basis `Basis.s` und der Saisonkomponente `saison` sowie der Differenzenmatrix 2-ter Ordnung `Diff.s` und Penalisierungsmatrix `Penalty.s=(Diff.s)TDiff.s`. Eilers & Marx (2005) betrachten auch Verbesserungen des Wrapping für Differenzenmatrizen. In unserer Analyse bringen sie allerdings keinen spürbaren Vorteil, so dass wir es bei der schlichten Wrapping-Methode belassen. Die Wahl der Saison-Einheiten soll ebenfalls spezifiziert werden. Dazu sind (im Kontext des Anwendungsbeispiels) die üblichen Einheiten, wie Tag, Woche oder Monat, geeignet. Die

Hazardfunktion im Modell ist

$$h(t, s; \mathbf{x}) = \exp\{\alpha_0^t(t) + \alpha_0^s(s)\} \exp\{\mathbf{x}^T \boldsymbol{\alpha}_x^t(t) + \mathbf{x}^T \boldsymbol{\alpha}_x^s(s)\}, \quad (6.1)$$

gleichet also der Form in (5.1). Hier sind  $\alpha_0^t(t)$ ,  $\alpha_0^s(s)$  die glatten Baseline-Effekte sowie  $\boldsymbol{\alpha}_x^t(t)$ ,  $\boldsymbol{\alpha}_x^s(s)$  die korrespondierenden glatten Kovariableneffekte zu den Vektoren zusammengefasst. Die penalisierte *log*-Likelihood mit den für die Wahl von B-Splines spezifizierten Differenzenmatrizen und deren Kreuzprodukten hat eine der Formel (5.4) ähnelnde Gestalt (zur Vereinfachung der Darstellung wählen wir dieselben B-Spline Basen für jede der  $p + 1$  glatten Komponenten):

$$\ell_p(\mathbf{u}_t, \mathbf{u}_s, \boldsymbol{\lambda}_t, \boldsymbol{\lambda}_s) = \ell(\mathbf{u}_t, \mathbf{u}_s) - \frac{1}{2} \mathbf{u}_t^T \mathbf{D}_t(\boldsymbol{\lambda}_t) \mathbf{u}_t - \frac{1}{2} \mathbf{u}_s^T \mathbf{D}_s(\boldsymbol{\lambda}_s) \mathbf{u}_s, \quad (6.2)$$

vgl. Formel (3.6), wobei die Matrizen  $\mathbf{D}_t(\boldsymbol{\lambda}_t)$  und  $\mathbf{D}_s(\boldsymbol{\lambda}_s)$  blockdiagonal sind mit den Einzelblöcken  $\mathbf{D}_t(\boldsymbol{\lambda}_t) = \lambda_t \mathbf{L}_t^T \mathbf{L}_t$  und  $\mathbf{D}_s(\boldsymbol{\lambda}_s) = \lambda_s \mathbf{L}_s^T \mathbf{L}_s$ . Hier sind  $\mathbf{L}_t$  und  $\mathbf{L}_s$  die jeweiligen Differenzenmatrizen, wie in Abschnitt 3.3 dargestellt. Die Modellparameter sind die zusammengesetzten Vektoren  $\mathbf{u}_t = [\mathbf{u}_0^t, \dots, \mathbf{u}_p^t]$  und  $\mathbf{u}_s = [\mathbf{u}_0^s, \dots, \mathbf{u}_p^s]$ .

Ein weiterer Schritt ist die Berücksichtigung der Restriktionen. Wie wir bereits aus dem Modell in Kapitel 5 erfahren haben, tritt beim additiven Modellieren das Identifikationsproblem auf, weil der Intercept-Term nur einer der Zeitskalen eindeutig zugewiesen werden soll. Es erscheint sinnvoll, diesen Term der  $t$ -Komponente zuzuschreiben. Außerdem soll in diesem Fall die Wahl des saisonalen Zeitpunktes in der Restriktion keine Rolle spielen (im Gegensatz zur natürlichen Wahl des kalendarischen Zeitpunktes im Modell des Kapitels 5). Um die Restriktion herzuleiten, stellen wir die additive Komponente in der Form

$$\tilde{\alpha}_l^t(t) + \tilde{\alpha}_l^s(s) = \left[ \alpha_l^t(t) + \int_{S_{\text{Anfang}}}^{S_{\text{Ende}}} \alpha_l^s(s) ds \right] + \left[ \alpha_l^s(s) - \int_{S_{\text{Anfang}}}^{S_{\text{Ende}}} \alpha_l^s(s) ds \right]$$

dar, so dass die Restriktion  $\int \tilde{\alpha}_l^s(s) = 0$ ,  $l = 0, 1, \dots, p$ , offensichtlich wird. Diese drückt die Tatsache aus, dass die saisonalen Effekte, über die ganze Periode hinweg betrachtet, verschwinden. Die Interpretation ähnelt der für die ANOVA-Modelle (Effektkodierung) und ist in gewisser Weise eine Verallgemeinerung für den stetigen Fall. (Die alternative Restriktion in der Form  $\alpha_l(s_{\text{start}}) = 0$ ,  $l = 0, 1, \dots, p$ , hätte die Interpretation der Dummy-Kodierung.)

Für das Fitten des Modells benutzen wir allerdings, wegen des numerischen Aufwands, die Darstellung der Restriktion in der Dummy-Kodierung mit dem spezifizierten Anfangszeitpunkt der Saison, also derart  $\alpha_0^s(1) = 0$ ,  $\alpha_l^s(1) = 0$ ,  $l = 1, \dots, p$ , wobei 1 den Startpunkt der Periode bezeichnet und vom

Anwendungsbeispiel abhängt. Der zugehörige Effekt der saisonalen Komponente ist entsprechend auf 0 gesetzt. Diese Restriktion bewirkt die folgende Transformation der ("gewrappten") Splinebasis: Die Bedingung  $\alpha^s(1) = 0$  (der Index  $l$  sei an dieser Stelle, der Einfachheit der Notation halber, ausgelassen) impliziert für  $\alpha^s(s) = \mathbf{B}(s)\mathbf{u}^s$  mit  $\mathbf{B}(s) = [B_1(s), \dots, B_K(s)]$  als Zeilenvektor der Dimension  $K$  die Restriktion  $\mathbf{B}(1)\mathbf{u}^s = 0$ , d. h. es gilt

$$B_1(1)u_1 + B_2(1)u_2 + \dots + B_K(1)u_K = 0,$$

so dass daraus die Abhängigkeit zwischen den Komponenten des Vektors  $\mathbf{u}^s$ ,

$$u_1 = -\frac{B_2(1)}{B_1(1)}u_2 - \dots - \frac{B_K(1)}{B_1(1)}u_K = \mathbf{c}^T \mathbf{u}_{-1}^s,$$

entsteht, mit dem Vektor  $\mathbf{c} = [-B_2(1)/B_1(1), \dots, -B_K(1)/B_1(1)]^T$  bestehend aus  $K - 1$  Komponenten. (Wir nehmen an, dass  $B_1(1) \neq 0$  erfüllt ist.) Damit gilt für den variierenden Koeffizienten:

$$\begin{aligned} \alpha(s) &= [B_1(s), B_2(s), \dots, B_K(s)] [\mathbf{c}^T \mathbf{u}_{-1}^s, u_2, \dots, u_K]^T \\ &= [B_1(s)c_2 + B_2(s), \dots, B_1(s)c_K + B_K(s)] [u_2, \dots, u_K]^T \\ &= \bar{\mathbf{B}}(s)\mathbf{u}_{-1}^s, \end{aligned}$$

wobei  $\bar{\mathbf{B}}(s)$  die  $s$ -te Zeile der transformierten Basismatrix bezeichnet. Die umgeformte Basismatrix ist also  $\bar{\mathbf{B}} = \mathbf{B}_1\mathbf{c}^T + \mathbf{B}_{-1}$  (mit natürlicher Subindex-Notation). Die Penalisierungsmatrix wird ebenfalls entsprechend transformiert. Für die geeignet partitionierte Penalisierungsmatrix  $\mathbf{D}_s$  (mit gewöhnlicher Notation) gilt

$$(\mathbf{u}^s)^T \mathbf{D}_s \mathbf{u}^s = u_1 D_{11} u_1 + u_1 \mathbf{D}_{1-1} \mathbf{u}_{-1} + \mathbf{u}_{-1}^T \mathbf{D}_{-11} u_1 + \mathbf{u}_{-1} \mathbf{D}_{-1-1} \mathbf{u}_{-1}.$$

Die Transformation des Vektors  $u_1 = \mathbf{c}^T \mathbf{u}_{-1}$  resultiert somit

$$\begin{aligned} (\mathbf{u}^s)^T \mathbf{D}_s \mathbf{u}^s &= \mathbf{u}_{-1}^T [\mathbf{c} D_{11} \mathbf{c}^T + \mathbf{c} \mathbf{D}_{1-1} + \mathbf{D}_{-11} \mathbf{c}^T + \mathbf{D}_{-1-1}] \mathbf{u}_{-1} \\ &=: \mathbf{u}_{-1}^T \bar{\mathbf{D}}_s \mathbf{u}_{-1}. \end{aligned}$$

Insgesamt ergeben sich die zur Saison gehörenden glatten Komponenten in der Form

$$\alpha_l^s(s) = \bar{\mathbf{B}}(s)\mathbf{u}_{-1}^{s,l}, \quad l = 0, 1, \dots, p,$$

und die (transformierte) Modell-*log*-Likelihood hat die Gestalt

$$\ell_p = \ell(\mathbf{u}_t, \mathbf{u}_{s,-1}) - \frac{1}{2} \mathbf{u}_t^T \mathbf{D}_t(\boldsymbol{\lambda}_t) \mathbf{u}_t - \frac{1}{2} (\mathbf{u}_{s,-1})^T \bar{\mathbf{D}}_s(\boldsymbol{\lambda}_s) \mathbf{u}_{s,-1}, \quad (6.3)$$

mit den zusammengesetzten Vektoren  $\mathbf{u}_t$ ,  $\mathbf{u}_{s,-1}$  und den blockdiagonalen Matrizen  $\mathbf{D}_t(\boldsymbol{\lambda}_t)$ ,  $\bar{\mathbf{D}}_s(\boldsymbol{\lambda}_s)$ . Die Schätzung der Modellparameter  $\mathbf{u}_t$ ,  $\mathbf{u}_s$  erfolgt nach demselben Backfitting-Algorithmus, wie in Kapitel 5 beschrieben. Die Bestimmung der Smoothing-Parameter  $\boldsymbol{\lambda}_t$ ,  $\boldsymbol{\lambda}_s$  erfolgt nach der Formel (3.17), deren Herleitung sich in Anhang A.6 befindet.

Die Umrechnung für die Effekte  $\tilde{\alpha}^t(t)$  und  $\tilde{\alpha}^s(s)$  ergibt sich nun wie folgt. Zunächst wird das Integral  $\int_{S_{\text{Anfang}}}^{S_{\text{Ende}}} \alpha_l(s) ds$ ,  $l = 0, 1, \dots, p$ , approximativ evaluiert. Dazu wählen wir ein äquidistantes Gitter der  $m$  Punkte  $\{S_{\text{Anfang}} = \tau_1, \dots, \tau_m = S_{\text{Ende}}\}$  für die Saison. Aus der Darstellung  $\alpha_l(s) = \mathbf{B}(s) \mathbf{u}_l^s$  bekommen wir mit der Rechteckapproximation den Ausdruck

$$\int_{S_{\text{Anfang}}}^{S_{\text{Ende}}} \alpha_l(s) ds = \frac{\tau_m - \tau_1}{m} \mathbf{1}_m^T \mathbf{B}_{\text{grid}} \mathbf{u}_l^s,$$

wobei  $\mathbf{B}_{\text{grid}}$  die Spline-Matrix bezeichnet, evaluiert an den Stützstellen  $\tau_j$ ,  $j = 1, \dots, m$ . Die Schätzung für  $\tilde{\alpha}_l^s(s)$  ist dann gegeben durch

$$\hat{\alpha}_l^s(s) = \left[ \mathbf{B}(s) - \frac{\tau_m - \tau_1}{m} \mathbf{1}_m^T \mathbf{B}_{\text{grid}} \right] \mathbf{u}_l^s = \tilde{\mathbf{B}}(s) \mathbf{u}_l^s, \quad (6.4)$$

und analog für  $\tilde{\alpha}_l^t(t)$ :

$$\hat{\alpha}_l^t(t) = \left[ \mathbf{B}(t), \frac{\tau_m - \tau_1}{m} \mathbf{1}_m^T \mathbf{B}_{\text{grid}} \right] \begin{bmatrix} \mathbf{u}_l^t \\ \mathbf{u}_l^s \end{bmatrix} = \tilde{\mathbf{B}}(t) \begin{bmatrix} \mathbf{u}_l^t \\ \mathbf{u}_l^s \end{bmatrix}. \quad (6.5)$$

Daraus ist die Bestimmung von Kovarianzen nicht schwierig. Es gilt

$$\text{Var} \left[ \hat{\alpha}_l^s(s) \right] = \tilde{\mathbf{B}}(s) \text{Cov} [\mathbf{u}_l^s] \tilde{\mathbf{B}}(s)^T \quad (6.6)$$

und

$$\text{Var} \left[ \hat{\alpha}_l^t(t) \right] = \tilde{\mathbf{B}}(t) \text{Cov} \begin{bmatrix} \mathbf{u}_l^t \\ \mathbf{u}_l^s \end{bmatrix} \tilde{\mathbf{B}}(t)^T. \quad (6.7)$$

Basierend auf den geschätzten Varianzen in (6.6) und (6.7) werden die punktwisen Konfidenzintervalle leicht ermittelt.

## 6.2 Simulationsstudie

Bevor wir das Datenmaterial auswerten, wollen wir das Modell auf seine Validität überprüfen. Dazu simulieren wir künstliche Beobachtungen. Das Simulationsverfahren wurde in Abschnitt 5.5 ausführlich beschrieben, so dass wir hier nur noch die zu simulierenden glatten Komponenten definieren müssen.

Wir simulieren jeweils für die Anzahl der Beobachtungen  $N = 100, 500, 1000$ . Die relative Häufigkeit der zensierten Beobachtungen lässt sich durch die Wahl des Intensitätsparameters  $\xi$  in der Poisson-Verteilung steuern. Das Simulationsschema ist also völlig identisch mit dem für das Modell in Kapitel 5. Wir spezifizieren die Geburten  $b$  (Eintritte in die Studie) als zufällige (gleichverteilte) ganze Zahlen aus dem Intervall  $[1, 50]$  und binäre Kovariable  $x$  mit den Werten  $\{0, 1\}$  und mit der Wahrscheinlichkeit  $P(x = 1) = 0.7$ . Da das Simulationsschema nicht die exakte Anzahl der zensierten Beobachtungen spezifizieren lässt, so dass von Simulation zu Simulation unterschiedliche relative Häufigkeiten produziert werden, teilen wir diese grob in sechs verschiedene Muster: bis ca. 2%, bis ca. 10%, bis ca. 20%, bis ca. 50%, bis ca. 75%, bis ca. 90% als Anteil der zensierten Beobachtungen. Die Wahl der glatten Komponenten geben wir uns wie folgt vor:

$$\alpha_0^t(t) = \eta - \frac{t}{30}, \quad (6.8)$$

$$\alpha_0^b(b) = \sin\left(2\pi\frac{b-1}{50}\right), \quad (6.9)$$

$$\alpha_x^t(t) = 0.5 + \left(\frac{t}{25}\right)^2, \quad (6.10)$$

$$\alpha_x^b(b) = 1 - \cos\left(2\pi\frac{b-1}{50}\right), \quad (6.11)$$

wobei der Parameter  $\eta$  im Ausdruck für die Baseline-Hazard in  $t$ -Richtung gerade die relative Häufigkeit der zensierten Beobachtungen steuert. Die Intensität ist  $\xi = \exp\{\alpha_0^t(t) + \alpha_0^b(b) + x[\alpha_x^t(t) + \alpha_x^b(b)]\}$ . Die eingesetzten Werte für  $\eta$  entsprechen den Zensierungsmustern  $h_\delta$ :

$$\begin{aligned} \eta = -1 &\quad \rightarrow \quad h_\delta \simeq 2\%, \\ \eta = -2 &\quad \rightarrow \quad h_\delta \simeq 10\%, \\ \eta = -2.5 &\quad \rightarrow \quad h_\delta \simeq 20\%, \\ \eta = -4 &\quad \rightarrow \quad h_\delta \simeq 50\%, \\ \eta = -5 &\quad \rightarrow \quad h_\delta \simeq 75\%, \\ \eta = -6 &\quad \rightarrow \quad h_\delta \simeq 90\%. \end{aligned}$$

Anzumerken ist, dass der Anteil der zensierten Beobachtungen bis ca. 90% etwa dem realen Datensatz der Kindersterblichkeiten (Beschreibung siehe Ab-

schnitt 11.2) nahekommt, in dem diese bei ca. 85% liegt. Für die jeweilige Anzahl der Beobachtungen  $N$  und für das jeweilige Zensierungsmuster  $h_\delta$  ergeben sich insgesamt 18 Konstellationen. Für jede dieser Konstellationen erzeugen wir eine Serie von insgesamt 100 Simulationen. Als Ergebnis bekommen wir die geschätzten empirischen 5%-, 50%- und 95%- Quantile sowie die coverage probabilities, wie in Anhang B.2 dargestellt. (Es wurden B-Splines 3-ten Grades, also der Ordnung 4, gewählt.)

Insgesamt überzeugt die Anpassungsgüte in allen simulierten Komponenten. Auch hier verkleinert sich der Bias mit Zunahme der Beobachtungen. Die Konfidenzbänder werden enger. Auf den ersten Blick überraschend wirkt die prozentuale Erhöhung der zensierten Beobachtungen (bis ca. 20% bei 100 Beobachtungen und bis ca. 10% bei 1000 Beobachtungen) der Reduzierung von Bias und der Verbesserung von coverage probabilities entgegen, ganz deutlich bei den Kovariablen-Komponenten in (6.10) und (6.11). Dieser Effekt soll aber nicht überinterpretiert werden und hat unserer Meinung nach mit der Wahl der Integrationspunkte beim Approximieren der kumulativen Hazardrate in der *log*-Likelihood zu tun. Zu dieser Wahl gibt es mehrere Möglichkeiten (z. B. als Zeitpunkte der beobachteten Events oder schlicht die Quantile aus dem spezifizierten Bereich in  $[0, t = t_{\max}]$ ). Überzeugend wirkt aber das Ergebnis bei dem sehr hoch zensierten Beobachtungsanteil (die Situation der so genannten "long-term-survivors") und ist im Fall des Anwendungsbeispiels hervorragend im wörtlichen Sinne.

## 6.3 Auswertung des Datenmaterials

Der Datensatz der Kindersterblichkeiten ist in Kapitel 11 ausführlich beschrieben (siehe dort auch die deskriptiven Statistiken sowie die Ergebnisse des Cox-Modells in Tabelle 11.2). Die geschätzten variierenden Koeffizienten sind in Abbildung 6.1 dargestellt. Die ersten beiden Grafiken, oben links, zeigen die geschätzte logarithmierte *Baseline-Hazard*  $\log\{h_0(t, s)\} = \alpha_0^t(t) + \alpha_0^s(s)$ , additiv in beiden Zeitkomponenten. Die restlichen Plots zeigen die logarithmierten Kovariableneffekte, jeweils für Überlebenszeit  $t$  (in Jahren) und Saison der Geburt  $s$  (in Monaten). Das Sterberisiko ist bei Geburt sehr hoch, also ca.  $\exp(-3) = 0.05$ , das die demografische Interpretation als 50 pro 1000 Personenjahre hat und das danach bis Ende des fünften Lebensjahres auf den Wert von ca.  $\exp(-4.7) = 9$  pro 1000 Personenjahre kontinuierlich absinkt. Die Geburtsmonate Mai und Juni (Ende der trockenen Periode und Beginn der Regenzeit) haben eine höhere saisonale Wirkung als der Durchschnitt.

Der Effekt des *Geschlechts* scheint über die Jahre hinweg eine konstante Wirkung auf das Sterberisiko zu haben und ist bei Jungen leicht höher als bei Mädchen. Insbesondere ist die Hazard für die Geburten in den Monaten Juni-September höher als für den Rest des Jahres. Ein stark ausgeprägter Effekt des *Zwillingstatus* unterstützt die auf dem Cox-Modell basierende Analyse (siehe Becher et al., 2004). Die ungewöhnliche Steigerung des Risikos mit fortschrei-

tendem Alter soll aber nicht überinterpretiert werden und geht offensichtlich auf die relativ niedrige Zahl der Zwillinge in der Stichprobe zurück. Der saisonale Effekt ist im Monat Juni am stärksten. Kinder der *jüngeren Mütter* (bis 18 Jahre) haben nur in den ersten zwei Lebensjahren ein erhöhtes Risiko. Es zeigt sich keine Variation bezüglich der Saison. Dagegen haben die Kinder der *älteren Mütter* kein erhöhtes Risiko zu Beginn ihres Lebens. Der Saison-effekt ist in den Monaten Februar-Mai am kleinsten. Die *ethnischen Gruppen* weisen eine gewisse Heterogenität in ihren Hazards, im Vergleich zu der Referenzgruppe der *Markas*, auf, dies in beiden Zeitskalen. *Bwaba* (BWA) haben ein leicht niedrigeres Risiko für alle Lebensjahre mit einem stark differenzierten saisonalen Muster zwischen den Jahresabschnitten Februar-April (klein) und August-Oktober (hoch). Bei *Mossi* (MOS) wird das kleinere Risiko noch deutlicher ab dem dritten Lebensjahr. Die saisonale Schwankung hat ähnliches aber weniger ausgeprägtes Muster als bei den *Bwaba*. Die *Peulh* (PLH) zeigen eine erhöhte Sterberate nur im ersten Lebensjahr, aber mit klaren Differenzen bezüglich des Geburtsmonats; höher in November-Januar. Die Gruppe der *Samo* (SAM) einschließlich einige andere ethnische Völker haben gleiche Überlebenschancen wie die *Marka*, der saisonale Effekt variiert aber und ist für die in Februar bis April geborenen Kinder am niedrigsten. Die Muslime als Referenzkategorie für *Religion* haben eine ähnliche Hazard wie die Christen. Dennoch besteht zwischen den beiden ein Unterschied bezüglich der Saison der Geburt, die Christen haben ein kleineres Risiko in den Geburtsmonaten April-Juli und ansonsten ein höheres Sterberisiko. Bei Animisten sind die Überlebenschancen geringer, besonders in den ersten drei Lebensjahren. Schließlich spielt die Entfernung vom nächstgelegenen Krankenhaus eine erhebliche Rolle bei den Überlebenschancen, dies unabhängig vom Geburtsmonat.



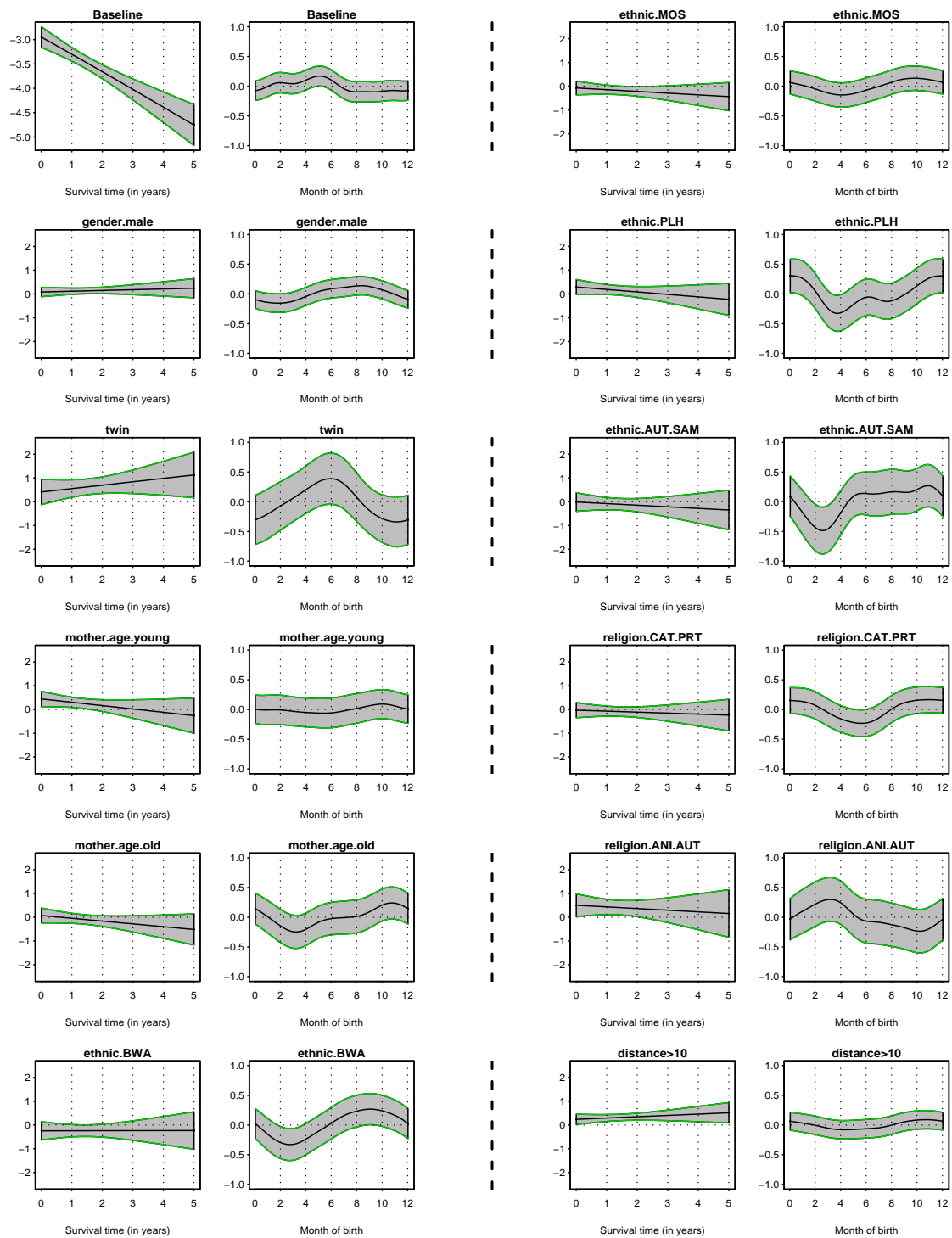


Abbildung 6.1: Geschätzte glatte dynamische Effekte mit Konfidenzbändern für die Überlebenszeit in Jahren, links, sowie für die Saison der Geburt in Monaten, rechts. Als Referenzniveau gilt die gestrichelte horizontale 0-Linie.

# Kapitel 7

## Two-Way-Interaction Hazard-Modell

Das Modell dieses Kapitels untersucht den Wechselwirkungseffekt zweier zeitlicher Komponenten auf die Hazardrate, die somit auf der *log*-Skala nicht mehr als additive Funktion modelliert wird, sondern rein bivariat. Zum Modellieren benutzen wir das Tensor-Produkt von Splines, wie es bei Dierckx (1993) oder bei de Boor (2001) definiert wird. Es gibt inzwischen zahlreiche Anwendungen der P-Splines für den Fall der bivariaten Modellierung; zu erwähnen sind z. B. die Arbeiten von Eilers & Marx (2003) und Marx & Eilers (2005) im Bereich der Signal-Prozesse sowie die Arbeiten von Currie, Durban & Eilers (2004), Richards, Kirkby & Currie (2005), Eilers, Currie & Durban (2006) und Currie, Durban & Eilers (2006) im Bereich der Sterbetafeln (mit Erweiterungen auf den allgemeinen multivariaten Fall in den letzten beiden Artikeln). Im Anwendungsteil dieses Kapitels gehen wir wieder auf die Daten zu den Kindersterblichkeiten ein und versuchen, den Wechselwirkungseffekt der Saison bei Geburt mit der Überlebenszeit zu schätzen. Anzumerken ist bei einer solchen Fragestellung, dass es inzwischen einige Untersuchungen darüber gibt, ob auch in westlichen Ländern der nachträgliche Effekt des Geburtsmonats auf das erreichte Lebensalter der Person statistisch signifikant feststellbar ist, vgl. hierzu Gavrilov & Gavrilova (1999) und Doblhammer & Vaupel (2001). (Dieser Sachverhalt sollte dann aber primär den aktuariellen Bereich interessieren.)

### 7.1 Modell-Likelihood

Für die Überlebenszeiten hat die Hazardfunktion folgende Gestalt:

$$\log h(t, s; \mathbf{x}) = \log\{h_0(t; s)\} + \mathbf{x}^T \boldsymbol{\beta}(t; s) = \mathbf{z}^T \boldsymbol{\alpha}(t; s), \quad (7.1)$$

mit  $\mathbf{z}^T = [1, \mathbf{x}^T]$  und  $\boldsymbol{\alpha}(t; s) = [\log\{h_0(t; s)\}, \boldsymbol{\beta}^T(t; s)]^T$  (konsistent mit Notationen der Abschnitte 2.1 und 3.1). Der nächste Schritt ist die bivariate

Darstellung der Komponenten von  $\boldsymbol{\alpha}(t; s)$  durch die Tensor-Produkte der Splinbasen  $\mathbf{B}^{(k),t}$  der Dimension  $M_k$  und  $\mathbf{B}^{(k),s}$  der Dimension  $N_k$ ,  $k = 0, 1, \dots, p$ , d. h. wir modellieren die glatten Komponenten des Modells in der bivariaten Form:

$$\begin{aligned} \alpha_k(t, s) &= \sum_{m=1}^{M_k} \sum_{n=1}^{N_k} u_{mn}^{(k)} B_m^{(k),t}(t) B_n^{(k),s}(s) \\ &= \left[ B_1^{(k),t \otimes s}(t, s), \dots, B_{M_k N_k}^{(k),t \otimes s}(t, s) \right] \begin{bmatrix} u_1^{(k)} \\ \vdots \\ u_{M_k N_k}^{(k)} \end{bmatrix} \\ &=: \mathbf{B}_k^{t \otimes s}(t, s) \mathbf{u}_k, \quad k = 0, 1, \dots, p. \end{aligned} \quad (7.2)$$

Mit  $\boldsymbol{\alpha}(t; s) = [\mathbf{B}_0^{t \otimes s}(t, s) \mathbf{u}_0, \dots, \mathbf{B}_p^{t \otimes s}(t, s) \mathbf{u}_p]^T =: \mathbf{B}_{t \otimes s}(t, s) \mathbf{u}_{t \otimes s}$  und somit  $\log h(t, s; \mathbf{x}) = \mathbf{z}^T \mathbf{B}_{t \otimes s}(t, s) \mathbf{u}_{t \otimes s}$  resultiert die Likelihoodfunktion (für rechtszensierte Daten) als

$$\begin{aligned} \ell &= \sum_i \left[ \delta_i \log h(t_i, s_i) - \int_0^{t_i} h(t, s_i) dt \right] \\ &= \sum_i \left[ \delta_i \mathbf{z}^T \mathbf{B}_{t \otimes s}(t, s) \mathbf{u}_{t \otimes s} - \int_0^{t_i} \exp \{ \mathbf{z}^T \mathbf{B}_{t \otimes s}(t, s) \mathbf{u}_{t \otimes s} \} dt \right]. \end{aligned} \quad (7.3)$$

Das Integral in (7.3) wird nun approximiert. Mit den in Abschnitt 3.2 eingeführten Notationen bekommen wir die Likelihoodfunktion der Poisson-Daten  $\boldsymbol{\delta}$ :

$$\ell = \boldsymbol{\delta}^T \mathbf{W}_{t \otimes s} \mathbf{u}_{t \otimes s} - \mathbf{1}^T \exp \{ \mathbf{W}_{t \otimes s} \mathbf{u}_{t \otimes s} + \mathbf{o} \},$$

(Darstellung mit B-Splines) bzw. wenn die Trends in den zeitlichen Komponenten nicht penalisiert modelliert werden:

$$\ell = \boldsymbol{\delta}^T \mathbf{W}_{t \otimes s} \boldsymbol{\theta}_{t \otimes s} - \mathbf{1}^T \exp \{ \mathbf{W}_{t \otimes s} \boldsymbol{\theta}_{t \otimes s} + \mathbf{o} \},$$

(Darstellung mit truncated polynomials) mit Parametern  $\boldsymbol{\beta}_{t \otimes s}$ ,  $\mathbf{b}_{t \otimes s}$  zusammengefasst zum Vektor  $\boldsymbol{\theta}_{t \otimes s}$ . Nach wie vor ist  $\mathbf{o}$  der Offset-Parameter des Poisson-Modells und  $\mathbf{W}_{t \otimes s}$  die Design-Matrix.

Für die weiteren Ausführungen soll die Struktur der Penalisierungsmatrix für den bivariaten Fall ebenfalls berücksichtigt werden. Dafür können wir zunächst (7.2) darstellen in der Form

$$\alpha_k(t, s) = \{ [\mathbf{B}^{(k),t}(t) \otimes \mathbf{1}_{N_k}^T] \odot [\mathbf{1}_{M_k}^T \otimes \mathbf{B}^{(k),s}(s)] \} \text{vec}\{\mathbf{U}_k\},$$

wobei  $\otimes$  das bekannte Kronecker-Produkt kennzeichnet und  $\odot$  das komponentenweise Multiplizieren bedeutet (siehe Anhang, Formel A.4). Hier sind außerdem die Koeffizienten  $u_{mn}^{(k)}$  zu einer Matrix  $\mathbf{U}_k$  zusammengesetzt, und der  $\text{vec}$ -Operator schreibt deren Spalten untereinander in einen Vektor. Es liegt nun nahe, die benachbarten Elemente  $u_{mn}^{(k)}$  in der Matrix  $\mathbf{U}_k$  sowohl entlang der Spalten als auch entlang der Zeilen zu penalisieren, d. h. als Penalisierung (für den ganzen Vektor  $\mathbf{u}_{t \otimes s}$ ) gilt nun:

$$\nabla_{\mathbf{u}_{t \otimes s}} = \text{Block.diag} \left[ \nabla_{\mathbf{u}_{t \otimes s}^0}^0, \dots, \nabla_{\mathbf{u}_{t \otimes s}^p}^p \right], \quad (7.4)$$

mit jeweils  $\nabla_{\mathbf{u}_{t \otimes s}^k}^k = \lambda_k^t \nabla_k^t + \lambda_k^s \nabla_k^s$  und  $\nabla_k^t = \mathbf{I}_{\dim(\mathbf{B}_k^s)}^t \otimes \mathbf{D}_k^t \mathbf{D}_k^t$ ,  $\nabla_k^s = \mathbf{D}_k^s \mathbf{D}_k^s \otimes \mathbf{I}_{\dim(\mathbf{B}_k^t)}^s$ ,  $k = 0, 1, \dots, p$ . Hier sind  $\mathbf{I}_{\dim(\mathbf{B}_k^s)}^t$  bzw.  $\mathbf{I}_{\dim(\mathbf{B}_k^t)}^s$  Identitätsmatrizen angegebener Dimension sowie  $\mathbf{D}_k^t$  bzw.  $\mathbf{D}_k^s$  Differenzenmatrizen spezifizierter Ordnung (im Fall der B-Splines) oder Einheitsmatrizen (im Fall der truncated polynomials) jeweils für  $t$ - bzw.  $s$ - Zeitskala (vgl. ausführliche Darstellung bei Eilers, Currie & Durban, 2006). Die penalisierte  $\log$ -Likelihood hat die Form:

$$\ell_{\text{pen}} = l - \frac{1}{2} \mathbf{u}_{t \otimes s}^T \nabla_{\mathbf{u}_{t \otimes s}} \mathbf{u}_{t \otimes s}. \quad (7.5)$$

Es gibt nun alternative Vorgehensweisen, die Schätzer der Parameter  $\boldsymbol{\theta}_{t \otimes s}$  bzw.  $\mathbf{u}_{t \otimes s}$  zu bestimmen. Die Schätzprozedur hängt maßgeblich mit der Bestimmung der Smoothing-Parameter zusammen. Bei den bekannten fixen Werten für Penalisierungen lassen sich die Parameter nach der (Fisher-)Scoring-Methode berechnen; bei der beobachteten Fisher-Matrix ist dies das Newton-Raphson-Schema.

## 7.2 Berechnung der Smoothing-Parameter

Mehrere Möglichkeiten für die Wahl der Smoothing-Parameter  $\boldsymbol{\lambda}^t = [\lambda_0^t, \dots, \lambda_p^t]$  und  $\boldsymbol{\lambda}^s = [\lambda_0^s, \dots, \lambda_p^s]$  stehen zur Verfügung. Die nach AIC- bzw. GCV-Kriterium zu bestimmenden Werte brauchen zumeist ein Wertegitter. (Denn dann ist keine implizite komplexe Analytik erforderlich). Angesichts der Parameterdimensionen, hier also  $= (p+1) \times (p+1)$ , ist dies eine sehr aufwendige Aufgabe. Einen Ausweg aus der Situation bietet die Festsetzung gleicher Werte für  $\lambda$ 's (zumindest  $\lambda_t^{(k)} = \lambda_s^{(k)}$  bei der  $k$ -ten Komponente des Modells, so genannter isotropischer Fall). Beim AIC-Kriterium:

$$\text{AIC}(\boldsymbol{\lambda}^t, \boldsymbol{\lambda}^s) = -\ell_p(\boldsymbol{\lambda}^t, \boldsymbol{\lambda}^s) + \text{df}(\boldsymbol{\lambda}^t, \boldsymbol{\lambda}^s), \quad (7.6)$$

mit den Freiheitsgraden

$$df(\boldsymbol{\lambda}^t, \boldsymbol{\lambda}^s) = \text{tr}\{\mathbf{I}_{\text{pen}}^{-1}(\boldsymbol{\lambda}^t, \boldsymbol{\lambda}^s)\mathbf{I}(\boldsymbol{\lambda}^t, \boldsymbol{\lambda}^s)\}, \quad (7.7)$$

wobei  $\mathbf{I}(\boldsymbol{\lambda}^t, \boldsymbol{\lambda}^s)$  die Fisher-Matrix ist und  $\mathbf{I}_{\text{pen}}(\boldsymbol{\lambda}^t, \boldsymbol{\lambda}^s)$  die penalisierte Fisher-Matrix bezeichnet (vgl. Gray, 1992, und Ghidry, Lesaffre & Eilers, 2004), fällt die Entscheidung  $\boldsymbol{\lambda} = \boldsymbol{\lambda}_{\text{opt}}$  für den Parameter mit minimalem AIC-Wert.

Alternativ gibt es die Möglichkeit einer GLMM-Darstellung mit zufälliger Komponente  $\mathbf{u}$  und Penalisierung. Die anschließende Marginalisierung (also Ausintegrieren von  $\mathbf{u}$ ) erlaubt die automatische Schätzung der Penalisierungsparameter als Varianzkomponenten. Diese Darstellung bedarf der Verteilung von  $\mathbf{u}$ . Im Fall der B-Splines besteht eine Schwierigkeit im Zusammenhang mit der Penalisierungsmatrix, da diese nicht invertierbar ist. Eine Lösung besteht in der Darstellung der auf B-Splines basierenden Designmatrix äquivalent in einer neuen Basis mit fixem  $\boldsymbol{\beta}$ -Teil und zufälligem  $\mathbf{b}$ -Teil mit Normalverteilung der zufälligen Komponenten. Dieses Vorgehen ist ausführlich in Currie, Durban & Eilers (2006) beschrieben. Eine Vereinfachung dieser Prozedur besteht im Folgenden: Betrachten wir zunächst den Fall mit nur Baseline-Hazard im Modell, also ohne Kovariablen. Die Penalisierung  $\nabla_{\mathbf{u}_{t \otimes s}}$  besteht dann nur aus dem Block  $\mathbf{P} = \lambda_0^t \nabla_0^t + \lambda_0^s \nabla_0^s$ . Es ist nun möglich, die Penalisierung  $\mathbf{u}^T \mathbf{P} \mathbf{u}$  zu der Form  $\boldsymbol{\alpha}^T \mathbf{F} \boldsymbol{\alpha}$  mit der invertierbaren Matrix  $\mathbf{F}$  zu bringen. Für die zufällige Komponente  $\boldsymbol{\alpha}$  spezifizieren wir:

$$\boldsymbol{\alpha} \sim \mathcal{N}(\mathbf{0}, \mathbf{F}^{-1}),$$

mit der Kovarianzmatrix  $\mathbf{F}$ . Diese lässt sich komplett aus den Komponenten von  $\mathbf{P}$  bestimmen, ist invertierbar und hat die diagonale Gestalt der Art:

$$\mathbf{F} = \lambda_0^t \boldsymbol{\Psi}_t + \lambda_0^s \boldsymbol{\Psi}_s$$

mit bekannten diagonalen Matrizen  $\boldsymbol{\Psi}_t$  und  $\boldsymbol{\Psi}_s$ . Wir komplettieren nun den Parametervektor  $\boldsymbol{\alpha}$  mit dem fixen Parameter  $\boldsymbol{\beta}$  so, dass eine nichtausgeartete Transformation  $\tilde{\mathbf{L}}$  mit

$$\tilde{\mathbf{u}} = \begin{bmatrix} \boldsymbol{\beta} \\ \boldsymbol{\alpha} \end{bmatrix} = \begin{bmatrix} \mathbf{S}_{\boldsymbol{\beta}} \mathbf{u} \\ \mathbf{S}_{\boldsymbol{\alpha}} \mathbf{u} \end{bmatrix} = \tilde{\mathbf{L}} \mathbf{u}$$

entsteht. Die Matrix  $\tilde{\mathbf{L}}$  lässt sich auch komplett bestimmen. Den Herleitungen und Notationen in Anhang A.6 folgend, resultiert mit der Bezeichnung

$$\mathbf{R}(\mathbf{u}; \lambda_0^t, \lambda_0^s) = (\mathbf{0}, \mathbf{I})(\tilde{\mathbf{L}}^{-1})^T \mathcal{I}_p(\mathbf{u}; \lambda_0^t, \lambda_0^s) \tilde{\mathbf{L}}^{-1}(\mathbf{0}^T, \mathbf{I}^T)^T$$

die maximierte marginale *log*-Likelihood:

$$\ell_{m,m}(\lambda_0^t, \lambda_0^s) = \ell_{\text{pen}}(\mathbf{u}) + \frac{1}{2} \log |\mathbf{F}| - \frac{1}{2} \log |\mathbf{R}(\mathbf{u}; \lambda_0^t, \lambda_0^s)|. \quad (7.8)$$

Die Ableitung bezüglich  $\lambda_0^t$  resultiert in

$$\begin{aligned} \frac{\partial \ell_{m,m}}{\partial \lambda_0^t} &= -\frac{1}{2} \hat{\mathbf{u}}^T \nabla_0^t \hat{\mathbf{u}} + \frac{1}{2} \text{tr}\{\mathbf{F}^{-1} \Psi_t\} - \frac{1}{2} \text{tr} \left\{ \mathbf{R}^{-1}(\hat{\mathbf{u}}; \lambda_0^t, \lambda_0^s) \frac{\partial \mathbf{R}(\hat{\mathbf{u}})}{\partial \lambda_0^t} \right\} \\ &\stackrel{!}{=} 0, \end{aligned} \quad (7.9)$$

und analog bezüglich  $\lambda_0^s$ . Eine wesentliche Vereinfachung der Formel ergibt sich im isotropen Fall, also wenn  $\lambda_0^t = \lambda_0^s = \lambda$ . Dann ist  $\mathbf{F} = \lambda \mathbf{I}$  (siehe Currie, Durban & Eilers, 2006) und es gilt mit (7.9) die folgende Fix-Punkt Iterationsgleichung:

$$\frac{1}{\lambda} = \frac{\hat{\mathbf{u}}^T [\nabla_0^t + \nabla_0^s] \hat{\mathbf{u}} + \text{tr}\{\mathbf{R}^{-1} \frac{\partial \mathbf{R}}{\partial \lambda}\}}{\dim(\mathbf{F})}.$$

Analoge Berechnungsformeln ergeben sich für den Fall der Kovariablen im Modell.

### 7.3 Anwendung an die Mortalitätsdaten

In Abbildung 7.1 sind auf der *log*-Skala die Baseline-Hazard sowie die bivariaten Effekte der Zwillinge, der jüngeren und älteren Mütter, der Jungen sowie der großen Entfernung zum Krankenhaus dargestellt. Dabei wurde der Penalisierungsparameter in die *t*-Richtung auf 100 und in die *s*-Richtung auf 1 gesetzt. Man kann dem Plot der Baseline-Hazard einen leicht nachhaltigen Effekt des Geburtsmonats entnehmen: die Kinder, die zu Beginn der Regenzeit geboren wurden (Monate Mai-Juni), haben im Vergleich zu Kindern, die in den übrigen Monaten geboren wurden, ein höheres Sterberisiko, auch nach den ersten fünf Lebensjahren. Der Verlauf der Lebensdauern variiert geringfügig bezüglich jeweiliger Geburtsmonate. Die Interpretation der Kovariableneffekte ist weitaus schwieriger. Zwar sinken generell die Risiken, jeweils verbunden mit dem Effekt des Alters der Mutter, in den folgenden Jahren, aber in Bezug auf den Geburtsmonat fallen sie unterschiedlich stark ab. Der Effekt der Zwillinge soll (wie schon in der Analyse des additiven Modells erwähnt) nicht überinterpretiert werden: das ansteigende Risiko ist offensichtlich durch das Modellartefakt hervorgerufen und geht auf die kleine Anzahl der Zwillinge im Datensatz zurück. Die komplexe Struktur der Effekte der ‘‘Jungen’’ sowie der ‘‘Entfernung’’ ist anscheinend durch unzureichende Penalisierung hervorgerufen.

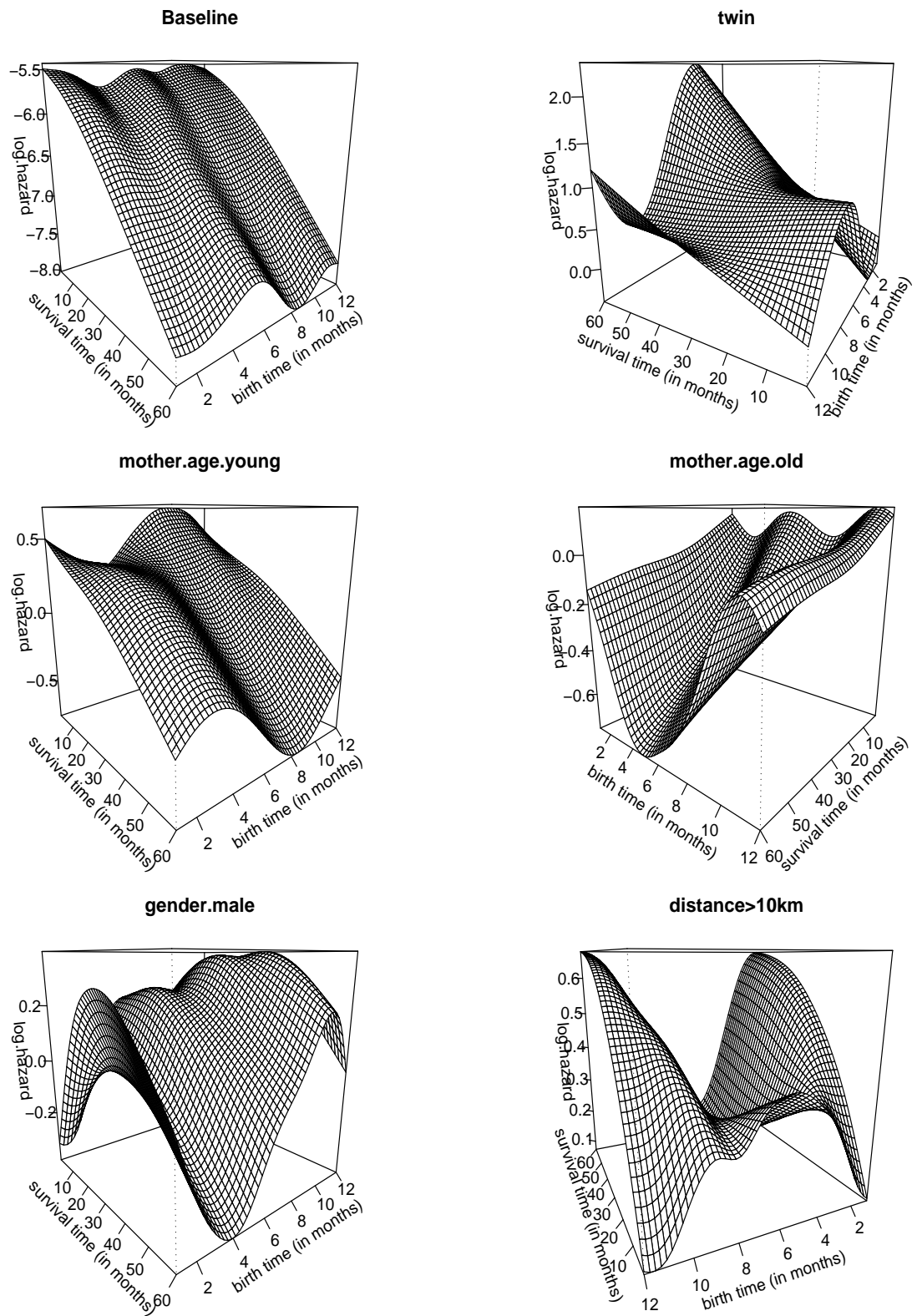


Abbildung 7.1: Plots der bivariaten Wechselwirkungseffekte zwischen Überlebensdauer und Saison der Geburt mit der Wahl von Penalisierung  $\lambda_t^{(k)} = 10^2$  und  $\lambda_s^{(k)} = 1$ .

# Kapitel 8

## Competing-Risks-with-Frailties Modell

### 8.1 Allgemeiner Überblick

”Competing-Risks”-Modelle sind bereits etabliert auf dem Gebiet der Analyse von Überlebenszeiten. Einen allgemeinen Überblick hierzu geben Hougaard (2000), Crowder (2001) und Lawless (2003). (Eine ältere systematische Behandlung liegt vor in David & Moeschberger, 1978, vgl. auch Cox & Oakes, 1984.) Gemeinsam ist diesen Modellen die Situation, in der jede Beobachtungseinheit in der Studie unter dem Risiko steht, eines der interessierenden Events unterschiedlicher Art zu erfahren. Ein gleichzeitiger Auftritt mehrerer Events ist dabei ausgeschlossen. Damit ist die Situation beschrieben, in der die Ereignisarten, also die Risiken, im ”Wettbewerb” miteinander stehen, die Ursache für den Ausfall (oder Zustandsänderung) der Beobachtung zu sein. Das Konzept ist nicht ohne Interpretationsschwierigkeiten und wird ausführlich in Hougaard (2000) diskutiert. In unserer Modellierung gehen wir von einem additiven Ansatz für die Hazardfunktion aus, d. h. es wird für das  $i$ -te Individuum die folgende Beziehung unterstellt:

$$h^{(i)}(t_i, \mathbf{d}_i; \mathbf{x}_i) = h_1(t_i; \mathbf{x}_i) + \cdots + h_L(t_i; \mathbf{x}_i), \quad (8.1)$$

mit  $L$  Competing Risks, der Hazardrate  $h_l(t_i; \mathbf{x}_i)$  für das  $l$ -te Risiko und dem Indikatorvektor (für die Ausfallart)  $\mathbf{d}_i$  mit  $\{0, 1\}$ -Elementen. Eine generelle, die Modellierung erleichternde Annahme ist dabei die der Unabhängigkeit von Competing Risks. Das bedeutet, dass die latenten (oder theoretischen) Überlebensdauern  $Y_l$ ,  $l = 1, \dots, L$  im (hypothetischen) Fall der Präsenz nur eines einzigen Risikos  $l$  stochastisch unabhängig sind. Beobachtet wird nur die Größe  $T_i = \min\{Y_l, l = 1, \dots, L\}$ . (Für den Fall von nicht unabhängigen Risiken sei z. B. auf David & Moeschberger, 1978, oder Crowder, 2001, verwiesen.) In Bezug auf die (im Weiteren zu konstruierende)  $\log$ -Likelihood Funktion, welche in diesem Fall in  $L$  Komponenten zerfällt, bringt dies einen wesentli-



chen Vorteil beim Schätzen: Sollten diese Komponenten keine gemeinsamen Parameter besitzen, kann die Schätzung der Modellparameter der jeweiligen Komponenten unabhängig voneinander durchgeführt werden. Hier, wie in den übrigen Anwendungen auch, wird für  $h_l(t_i; \mathbf{x}_i)$ ,  $l = 1, \dots, L$ , eine glatte Struktur unterstellt.

Eine weitere Modellspezifizierung betrifft den Fall von so genannten "Frailty"-Parametern. Das Konzept von Frailty spielt in der Überlebenszeitanalyse eine ähnliche Rolle wie die zufälligen Komponenten in einem (G)LMM-Modell ("generalized linear mixed model"). (Die GLMM-Modelle sind in der statistischen Literatur umfassend dargestellt, zu zitieren sind die Standardwerke wie z. B. Pinheiro & Bates, 2000, Verbeke & Molenberghs, 2000, und McCulloch & Searle, 2001, Molenberghs & Verbeke, 2005.) Es wird demnach versucht, die unbeobachteten Heterogenitäten in den spezifizierten Gruppen (oder Clustern) durch die Einführung eines zusätzlichen stochastischen Terms im Modell zu beschreiben. Dieses Thema wird ausführlich in Therneau & Grambsch (2000) und Hougaard (2000) behandelt (vgl. hier auch zahlreiche Referenzen). Wir erweitern die additive Hazardstruktur in (8.1) wie folgt:

$$h^{(i)}(t_{ij}, \mathbf{d}_{ij}; \mathbf{x}_{ij}) = h_1(t_{ij}; \mathbf{x}_{ij})w_{i1} + \dots + h_L(t_{ij}; \mathbf{x}_{ij})w_{iL}. \quad (8.2)$$

Hier bezeichnet der Index  $i$  das  $i$ -te Cluster mit  $j = 1, \dots, J_i$  Beobachtungseinheiten und  $\mathbf{w}_i = [w_{i1}, \dots, w_{iL}]^T$  den Frailty-Vektor, spezifisch für dieses Cluster  $i$ . Die Zeiten der Events für unterschiedliche Cluster werden als unabhängig voneinander angesehen; innerhalb des  $i$ -ten Clusters sind diese unabhängig bedingt auf das  $i$ -te Frailty. Für  $\mathbf{w}_i$  selbst wird eine multivariate Verteilung unterstellt. Um das Modellieren einerseits flexibel genug und andererseits numerisch relativ unaufwendig zu gestalten, wählen wir eine "Misch"-Struktur ("finite mixture distribution") der Art:

$$\mathbf{w}_i \sim \sum_{\mathbf{m}=1}^M c_{\mathbf{m}} \prod_{l=1}^L f_{\Gamma}(w_{il}; \mu_{m_l}, \sigma_{m_l}^2), \quad (8.3)$$

mit dem Multiindex  $\mathbf{m} = [m_1, \dots, m_L]_{\substack{m_1=1, \dots, M_1 \\ m_L=1, \dots, M_L}}$ ,  $\mathbf{M} = [M_1, \dots, M_L]$ , und den Gamma-Dichten  $f_{\Gamma}(w_{il}; \mu_{m_l}, \sigma_{m_l}^2)$  mit Erwartungswert  $\mu_{m_l}$  und Varianz  $\sigma_{m_l}^2$  (vgl. die äquivalente Darstellungsform der Dichte in Anhang A.4). Die Wahl der Frailty-Dichte legt außerdem einige Restriktionen auf die unbekanntenen Gewichte  $c_{\mathbf{m}}$ . Diese Gewichte sowie die Frailties werden bestimmt durch die Schätzzroutine im Rahmen des EM-Algorithmus, vgl. Dempster, Laird & Rubin (1977) und McLachlan & Krishnan (1997).

Bevor wir fortfahren, bedarf es an dieser Stelle einer kurzen Anmerkung. Die Art der Daten, denen unser Interesse gilt, stellt Angaben über die Zeiträume von Arbeitslosigkeit dar und die diesen Zustand terminierenden Ereignisse. Demnach besteht jedes Cluster aus mehreren Dauern oder Spells, die das  $i$ -te Individuum (also  $i$ -tes Cluster) aufweist, weil man auch mehr als nur einmal

arbeitslos werden kann. Die Beobachtungen (oder Spells) im jeweiligen Cluster weisen alle denselben Frailty-Wert auf ("shared-Frailty"-Modell). Das Konzept solcher Modelle ist z. B. in Hougaard (2000) ausführlich diskutiert. Wir wollen hier ansetzen und Abhängigkeitsstrukturen in den Risiken durch die flexible Modellierung von Frailties mit in das Schätzprinzip hineinintegrieren.

## 8.2 Identifizierbarkeit; Spezifizierung der Likelihood

Es existieren grundsätzlich zwei Ansätze beim Modellieren von Competing Risks. Der erste, auch als "klassischer Ansatz" bezeichnet, konzentriert sich auf die Bestimmung der gemeinsamen multivariaten Survivorfunktion  $S(\mathbf{t}) = S(t_1, \dots, t_L)$  mit  $L$  Competing Risks und  $t_l$ , dem Zeitpunkt des Events  $l$ , wäre dieses Event aufgetreten. Die Zeiten  $t_l$  sind die latenten (also unbeobachtbaren) Eventzeitpunkte, es wird lediglich die Zeit  $t^* = \min\{t_1, \dots, t_L\}$  beobachtet und der zugehörige Eventtyp (falls die Beobachtung nicht zensiert). Dieses multivariate Vorgehen hat einen wesentlichen Nachteil: es führt unvermeidlich zum so genannten Identifizierbarkeitsproblem, das weitreichende Konsequenzen haben kann; es gibt auf der Beobachtungsbasis von  $\{t_i, C_i\}$  mit Überlebensdauer  $t_i$  und Ausfalltyp  $C_i$  zu jedem Modell mit abhängigen Competing Risks und multivariater Survivorfunktion  $S(\mathbf{t})$  ein weiteres eindeutig bestimmtes so genanntes "proxy"-Modell der unabhängigen Events und Überlebensfunktion  $S^*(\mathbf{t})$ , das dieselben Sub-Survivorfunktionen  $\bar{F}(j, t) = P(C = j, T \geq t)$  produziert. Daher kann man auf der Basis des beobachteten Datenmaterials nicht sagen, welches Modell korrekt ist. Eine der ersten systematischen Überlegungen hierzu findet sich in Tsiatis (1975), weitere Ausführungen und Beispiele bei Crowder (2001). Es gibt allerdings Szenarien, bei denen die Identifizierbarkeit möglich ist, d. h. die Kenntnis der Sub-Survivorfunktionen erlaubt eine eindeutige Rekonstruktion von  $S(\mathbf{t})$ , vgl. Heckman & Honore (1989) und (bezüglich Erweiterungen für den Fall von mehreren Spells ?). Auch wenn die Identifizierbarkeit im Fall der unabhängigen Competing Risks nicht auftritt, empfiehlt es sich, von dieser Modellierungsart Abstand zu nehmen (siehe z. B. Crowder, 2001, Kapitel 6, darin weitere Referenzen).

Der zweite, "moderne" Ansatz ist hazardbasiert. Dieser erscheint auch natürlich, verursacht keine Interpretationsschwierigkeiten und wird hier herangezogen. Die Likelihoodfunktion wird daher in den risikospezifischen Hazardraten ausgedrückt. Dabei gehen wir von unabhängigen Risiken aus. Um die Herleitung (in Anlehnung an Lawless, 2003; vgl. auch David & Moeschberger, 1978, für die äquivalente Darstellung) einfach zu gestalten betrachten wir zunächst die Situation, in der für jedes Individuum (oder Cluster) nur eine Beobachtung (oder Spell) gegeben ist. Die Daten sind in der Form  $\{t_i, \mathbf{d}_i; \mathbf{x}_i\}$  gegeben, mit der Überlebenszeit  $t_i$ , dem Kovariablenvektor  $\mathbf{x}_i$  und dem vektorwertigen Ereignisindikator  $\mathbf{d}_i$  für die Markierung des Endes der Verweildauer. Es

gilt  $\mathbf{d}_i = (d_{i1}, \dots, d_{iL})$ , wobei die Elemente  $d_{il} \in \{0, 1\}$  sind und im Falle des Events nur eines von  $d_{il}$  gleich 1 ist. Im zensierten Fall gilt  $d_{il} = 0$ ,  $l = 1, \dots, L$ . Für die weitere Herleitung brauchen wir die Definition von risikospezifischen Hazardraten. Notationsmäßig ist es an dieser Stelle von Vorteil, mit  $C$  den beobachteten Eventtyp zu bezeichnen, d.h. die diskrete Zufallsvariable  $C$  habe die Menge  $\{1, \dots, L\}$  als Wertebereich. Es gilt dann:

$$h_l(t) = \lim_{\Delta t \rightarrow 0} \frac{P(T < t + \Delta t, C = l | T \geq t)}{\Delta t}.$$

Wegen  $P(T < t + \Delta t | T \geq t) = \sum_{l=1}^L P(T < t + \Delta t, C = l | T \geq t)$  ergibt sich die marginale, d. h. von der Ereignisart unabhängige Hazardrate als

$$h(t; \mathbf{x}) = \sum_{l=1}^L h_l(t; \mathbf{x}).$$

Analog zum univariaten Fall schreiben wir die Likelihoodfunktion zunächst in der Form

$$\mathcal{L} = \prod_{i=1}^n \{f(C_i, t_i)\}^{\delta_i} \{S(t_i)\}^{1-\delta_i},$$

wobei  $f(C_i, t_i)$  die gemeinsame Dichte von  $(t_i, C_i)$  ist und  $\delta_i$  den Zensierungsindikator bezeichnet. Somit ist der Beitrag zur Likelihood des  $i$ -ten Individuums gleich  $f(t_i, C_i)$  im unzensierten Fall und  $S(t_i)$  im zensierten. (Wir unterstellen hier den Standardfall des unabhängigen uninformativen Zensoringmechanismus; Nach wie vor werden die rechtszensierten Daten betrachtet.) Die Dichte  $f(C_i, t_i)$  können wir auch schreiben als

$$f(C_i, t_i) = \prod_{l=1}^L \{f(C_i = l, t_i)\}^{d_{il}},$$

mit dem ereignisspezifischen Indikator  $\mathbf{d}_i = [d_{i1}, \dots, d_{iL}]$ ,  $d_{il} = 1_{\{C_i=l\}}$ . Aus der Definition für die ursachenspezifische Hazardrate können wir herleiten:

$$h_l(t) = \lim_{\Delta t \rightarrow 0} \frac{P(t \leq T < t + \Delta t, C = l)}{\Delta t} \frac{1}{P(T \geq t)} = \frac{f(t, C = l)}{S(t)},$$

so dass  $f(t, C = l) = h_l(t)S(t)$  gilt. Damit ist

$$\begin{aligned}\mathcal{L} &= \prod_{i=1}^n \prod_{l=1}^L \{h_l(t_i) S(t_i)\}^{d_{il}} \{S(t_i)\}^{1-\delta_i} \\ &= \prod_{i=1}^n \prod_{l=1}^L \{h_l(t_i)\}^{d_{il}} S(t_i).\end{aligned}$$

Mit  $S(t) = \exp\{-\int_0^t h(u) du\}$  ergibt sich letztendlich:

$$\mathcal{L} = \prod_{i=1}^n \prod_{l=1}^L \{h_l(t_i)\}^{d_{il}} \exp\{-\int_0^{t_i} h_l(t) dt\}.$$

Nun ist es einfach, die Likelihood auf den Fall mit mehreren Spells (bei deren bedingten Unabhängigkeit) pro Individuum zu erweitern. Es resultiert (für  $i = 1, \dots, n, j = 1, \dots, n_i$ )

$$\mathcal{L} = \prod_{i=1}^n \prod_{j=1}^{n_i} \prod_{l=1}^L \left[ \{h_l(t_{ij}; \mathbf{x}_{ij})\}^{d_{ijl}} \exp\{-\int_0^{t_{ij}} h_l(t; \mathbf{x}_{ij}) dt\} \right].$$

Die Likelihood kann nun äquivalent umgeschrieben werden als

$$\begin{aligned}\mathcal{L} &= \prod_{l=1}^L \prod_{i=1}^n \prod_{j=1}^{n_i} \left[ \{h_l(t_{ij}; \mathbf{x}_{ij})\}^{d_{ijl}} \exp\{-\int_0^{t_{ij}} h_l(t; \mathbf{x}_{ij}) dt\} \right] \quad (8.4) \\ &= \prod_{l=1}^L \mathcal{L}_l.\end{aligned}$$

Diese Form ist für die Parameterschätzung in einer Situation geeignet, in der jede Likelihood durch einen unabhängigen Set von Parametern spezifiziert ist. Die Schätzroutine zerfällt also in die Anpassung von  $L$  voneinander unabhängigen Sub-Problemen.

### 8.3 Frailty als unbeobachtete Heterogenität

Das Konzept des "Frailty" spielt in der modernen Analyse von Verweildauern eine außerordentlich wichtige Rolle. Der Begriff wurde in die statistische Literatur von Vaupel, Manton & Stallard (1979) eingeführt und wird umfassend in Hougaard (2000) und Therneau & Grambsch (2000) diskutiert (siehe auch Übersichtsartikel in Aalen, 1988 und Aalen, 1994). Allgemein wird damit versucht, die spezifizierten Gruppen (oder Cluster) in Bezug auf die

Überlebensdauern zu charakterisieren bzw. zu unterscheiden. Die Klassenheterogenität wird dabei durch den stochastischen Term, den Frailty-Parameter, beschrieben. In Bezug auf die Modellierung profitiert man von der Darstellung des Modells, indem der Frailty-Term als unbeobachtete (latente) Größe oder der fehlende Kovariablenvektor modelliert wird. Am häufigsten geschieht dies in der multiplikativen Weise:

$$h^{(ij)}(t) = h_0(t) \exp\{\boldsymbol{\beta}^T(t)\mathbf{x}_{ij} + \omega_i\} = h_0(t) \exp\{\boldsymbol{\beta}^T(t)\mathbf{x}_{ij}\}w_i,$$

mit der (nicht weiter spezifizierten) Baseline-Hazard  $h_0(t)$ , dem Frailty-Term  $\omega_i = \log(w_i)$ ,  $i = 1, \dots, n$ ,  $j = 1, \dots, n_i$ , dem Kovariablenvektor  $\mathbf{x}_{ij}$  sowie den (zeitabhängigen) Effekten  $\boldsymbol{\beta}$ . In dieser Form kann man eine Analogie zum gemischten Modell mit zufälligen Effekten ("random effects model") durchführen. Man spricht auch vom so genannten "Shared-Frailty"-Modell, da die Individuen bzw. Einheiten in einem Cluster den gleichen Frailty-Wert besitzen. Die großen Werte deuten dabei auf eine höhere Chance hin, das Event des Interesses zu erfahren, im Vergleich zu den Individuen aus den Clustern mit niedrigem Frailty-Wert. Ein weiterer unmittelbar einleuchtender Nutzen von Frailties liegt in der Vermeidung einer Fehlspezifizierung des Kovariablenvektors bzw. in der Nichtberücksichtigung der weiteren nichteinbezogenen Faktoren bezüglich ihres Gesamteffektes auf die Hazard (siehe aber auch Keiding, Andersen & Klein, 1997).

Zur Spezifikation der Frailty-Verteilung gibt es zahlreiche Alternativen, einen breiten Überblick gibt Hougaard (2000). Einer der populärsten Ansätze ist dabei die Gamma-Verteilung (für den Frailty-Parameter  $w$ ). Die Schätzstrategie erfolgt dann numerisch relativ unaufwendig via EM-Algorithmus, siehe z. B. Klein (1992) und Nielsen, Gill, Andersen & Sørensen (1992). Als weitere mögliche Verteilung gilt die Normalverteilung (für  $\omega$ ), vgl. McGilchrist & Aisbett (1991), McGilchrist (1993) und Erweiterungen in Vaida & Xu (2000) für den multivariaten Fall von Frailty mit beliebiger Designmatrix. Penalisierte Likelihood-Konzepte für die Proportional Hazards mit Penalisierung des Frailty-Parameters sowie Implementierungen in R werden in Therneau & Grambsch (2000) diskutiert (siehe auch Ripatti & Palmgren, 2000).

Wie bereits in Abschnitt 8.1 erwähnt, sind, bedingt auf den Frailty-Effekt, die Überlebenszeiten im selben Cluster stochastisch unabhängig. Damit dient die Einführung des Frailty-Parameters nicht nur der Berücksichtigung der Heterogenitäten, sondern erlaubt auch die Modellierung von Abhängigkeiten für geclusterte oder multiple Überlebenszeiten im marginalen (oder unbedingten) Modell. Im Fall der Competing Risks wird der Frailty-Parameter zum Frailty-Vektor  $\mathbf{w}_i = [w_{i1}, \dots, w_{iL}]$  durch die Einbeziehung des risikospezifischen Frailty-Parameters  $w_{il}$ ,  $l = 1, \dots, L$ . Dabei setzen wir die Unabhängigkeit von Risiken, bedingt auf den Frailty-Parameter, voraus. Den zunächst gewählten additiven Ansatz erweitern wir durch den folgenden:

$$h^{(i)}(t_{ij}; \mathbf{x}_{ij}) = h_1(t_{ij}; \mathbf{x}_{ij})w_{i1} + \dots + h_L(t_{ij}; \mathbf{x}_{ij})w_{iL}. \quad (8.5)$$

Hier bezeichnet der Index  $i$  das  $i$ -te Cluster mit  $j = 1, \dots, J_i$  Beobachtungseinheiten sowie  $\mathbf{w}_i = [w_{i1}, \dots, w_{iL}]$  Frailty-Vektor, spezifisch für  $i$ . Die Zeiten der Events werden als unabhängig zwischen den Clustern angesehen; innerhalb des  $i$ -ten Clusters sind diese unabhängig bedingt auf das  $i$ -te Frailty. Für  $\mathbf{w}_i$  wird eine multivariate Verteilung gewählt, die in Abschnitt 8.5 spezifiziert wird. Die Likelihoodfunktion lässt sich nun schreiben als

$$\mathcal{L} = \prod_{l=1}^L \prod_{i=1}^n \prod_{j=1}^{n_i} \left[ \{h_l(t_{ij}; \mathbf{x}_{ij}) w_{il}\}^{d_{ijl}} \exp\left\{-w_{il} \int_0^{t_{ij}} h_l(t; \mathbf{x}_{ij}) dt\right\} \right] = \prod_{l=1}^L \mathcal{L}_l(\mathbf{w}_l). \quad (8.6)$$

Dies ist die Likelihood für den gegebenen Frailty-Wert  $\mathbf{w} = [\mathbf{w}_1, \dots, \mathbf{w}_L]^T$ . Dabei zerfällt sie in das Produkt der  $L$  Komponenten, die dann (wie schon oben diskutiert) unabhängig voneinander maximiert werden können.

## 8.4 Wahl der glatten Komponenten; Approximation der Likelihood

Bevor die Schätzroutine ausführlich dargestellt wird, wollen wir an dieser Stelle, wie bereits in vorigen Anwendungen geschehen, die glatten Modellkomponenten sowie die Approximation der Likelihood beschreiben.

Die ursachenspezifischen Hazardraten  $h_l(t; \mathbf{x})$  werden mithilfe der P-Spline Ansätze modelliert. Als Basis werden die truncated polynomials gewählt. Wir setzen:

$$h_l(t; \mathbf{x}) = \exp\{\alpha_{l0}(t) + x_1 \alpha_{l1}(t) + \dots + x_p \alpha_{lp}(t)\},$$

mit  $p$  zeitabhängigen Kovariableneffekten  $\alpha_{lr}(t)$ ,  $r = 1, \dots, p$ , und der logarithmierten Baseline-Hazard  $\alpha_{l0}(t)$  für das  $l$ -te Risiko. Es gilt also

$$\alpha_{lr}(t) = \mathbf{Z}_{lr}(t) \boldsymbol{\beta}_{lr} + \mathbf{B}_{lr}(t) \mathbf{b}_{lr}, \quad r = 0, 1, \dots, p,$$

mit der üblichen Wahl  $\mathbf{Z}_{lr}(t) = [1, t]$ ,  $\mathbf{B}_{lr}(t) = [(t - \kappa_1)_+, \dots, (t - \kappa_{q_{lr}})_+]$  und den Spline-Knoten  $\kappa_1^{(l)}, \dots, \kappa_{q_{lr}}^{(l)}$ ,  $r = 0, \dots, p$ ,  $l = 1, \dots, L$ , für das jeweilige Risiko. Dabei sind  $\boldsymbol{\beta}_l = [\boldsymbol{\beta}_{l0}^T, \dots, \boldsymbol{\beta}_{lp}^T]^T$  und  $\mathbf{b}_l = [\mathbf{b}_{l0}^T, \dots, \mathbf{b}_{lp}^T]^T$  für  $l = 1, \dots, L$ . Benutzen wir nun z. B. die Rechteckapproximation des Integrals in (8.6), so bekommen wir die approximierte  $\log$ -Likelihood:

$$\ell \doteq \sum_{i=1}^n \sum_{l=1}^L \sum_{j=1}^{n_i} \left\{ d_{ijl} \log \{h_l(t_{ij}, \mathbf{x}_{ij}) w_{il}\} - \sum_{k=1}^{K_{ij}} h_l(t_{ijk}; \mathbf{x}_{ij}) [t_{ijk} - \tau_{k-1}] w_{il} \right\}. \quad (8.7)$$

Dabei sind  $0 = \tau_0 < \tau_1 < \dots < \tau_K$  die Knoten für die Rechteckapproximation, die die Menge der beobachteten Zeitpunkte überdecken, der Index  $K_{ij} = \operatorname{argmin}\{\tau_k \geq t_{ij}\}$  und  $t_{ijk} = \min\{t_{ij}, \tau_k\}$  (zur Vereinfachung der Notation). Für die Konstruktion sei auf Abschnitt 3.2 verwiesen. Die Parameter  $\boldsymbol{\beta}_l, \mathbf{b}_l$  setzen wir zum Vektor  $\boldsymbol{\theta}_l = [\boldsymbol{\beta}_l^T, \mathbf{b}_l^T]^T$  zusammen. Weiterhin notieren wir mit  $\mathbf{W}_l(t)$  die Matrix

$$\mathbf{W}_l(t) = \begin{pmatrix} Z_{l0}(t) & \dots & 0 & B_{l0}(t) & \dots & 0 \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ 0 & \dots & Z_{lp}(t) & 0 & \dots & B_{lp}(t) \end{pmatrix} = [\mathbf{Z}_l(t), \mathbf{B}_l(t)].$$

Damit schreibt sich die Approximation der *log*-Likelihood in der Form

$$\begin{aligned} \ell(\boldsymbol{\theta}|\mathbf{w}) &= \sum_{l=1}^L \ell_l(\boldsymbol{\theta}_l|\mathbf{w}_l) \\ &= \sum_{l=1}^L \sum_{i=1}^n \sum_{j=1}^{n_i} \left\{ d_{ijl} \mathbf{x}_{ij} \mathbf{W}_l(t_{ij}) \boldsymbol{\theta}_l - \sum_{k=1}^{K_{ij}} \exp\{\mathbf{x}_{ij} \mathbf{W}_l(t_{ijk}) \boldsymbol{\theta}_l + o_{ijkl}\} \right\} \end{aligned} \quad (8.8)$$

mit dem Offset-Term  $o_{ijkl} = \log\{[t_{ijk} - \tau_{k-1}] w_{il}\}$ . Mit der Einführung der Zufallsvariablen  $\delta_{ijkl} = 0$  für  $k < K_{ij}$  und  $\delta_{ijkl} = d_{ijl}$  für  $k = K_{ij}$  erreichen wir, analog zu den Überlegungen aus Abschnitt 3.2, dass diese Form der (approximativen) Likelihood genau der der (Pseudo)-Poissonverteilten Zufallsvariablen  $\delta_{ijkl}$  entspricht mit den (für den gegebenen Frailty-Parameter  $\mathbf{w}_i$ ) Intensitäten  $\exp\{\mathbf{x}_{ij} \mathbf{W}_l(t_{ijk}) \boldsymbol{\theta}_l + o_{ijkl}\}$ ,  $i = 1, \dots, n$ ;  $j = 1, \dots, n_i$ ;  $k = 1, \dots, K_{ij}$ ;  $l = 1, \dots, L$ .

Um die Likelihood kompakter darzustellen, multiplizieren wir den Zeilenvektor der Kovariablen  $\mathbf{x}_{ij}$  mit der Matrix  $\{\mathbf{W}_l(t_{ijk})\}_{k=1, \dots, K_{ij}}$  den Dimensionen entsprechend und speichern das Ergebnis (über alle  $i, j, k$ ) in der Matrix  $\tilde{\mathbf{W}}_l$ . Mit den Notationen aus Abschnitt 3.2 bezeichnen wir mit  $\boldsymbol{\delta}_l = \{\delta_{ijkl}\}_{ijk}$  den  $l$ -ten risikospezifischen Poisson-Vektor. Die  $l$ -te Komponente der Likelihood resultiert dann als

$$\ell_l(\boldsymbol{\theta}_l|\mathbf{w}_l) = \boldsymbol{\delta}_l^T \tilde{\mathbf{W}}_l \boldsymbol{\theta}_l - \mathbf{1}^T \exp\{\tilde{\mathbf{W}}_l \boldsymbol{\theta}_l + \mathbf{o}_l\}. \quad (8.9)$$

Auf die Parameter  $\mathbf{b}_l$  wird die Penalisierung aufgelegt. Dafür wählen wir als Penalisierungsterm die Komponente

$$\frac{1}{2} \sum_{r=0}^p \lambda_{lr} \mathbf{b}_{lr}^T \mathbf{D}_{lr} \mathbf{b}_{lr} = \frac{1}{2} \mathbf{b}_l^T \mathbf{D}_l(\boldsymbol{\lambda}_l) \mathbf{b}_l$$

und schreiben die penalisierte Likelihood als

$$\begin{aligned} \ell_{\text{pen}} &= \sum_{l=1}^L \ell_{l,\text{pen}}(\boldsymbol{\theta}_l, \boldsymbol{\lambda}_l | \mathbf{w}_l) \\ &= \sum_{l=1}^L \left[ \ell_l(\boldsymbol{\theta}_l, \boldsymbol{\lambda}_l | \mathbf{w}_l) - \frac{1}{2} \mathbf{b}_l^T \mathbf{D}_l(\boldsymbol{\lambda}_l) \mathbf{b}_l \right]. \end{aligned} \quad (8.10)$$

Es wird die Ridge-Penalisierung gewählt, so dass die Matrizen  $\mathbf{D}_{lr}$  die Einheitsmatrizen sind. Die Penalisierungsparameter  $\lambda_{lr}$  sind zum Vektor  $\boldsymbol{\lambda}_l$  zusammengesetzt, die Matrix  $\mathbf{D}_l(\boldsymbol{\lambda}_l)$  weist eine entsprechende blockdiagonale Struktur auf.

## 8.5 Spezifizierung der Frailty-Verteilung

Ein besonderes Augenmerk richtet sich auf die Wahl der Frailty-Verteilung. Wir verfolgen hier den flexiblen Ansatz zum Modellieren der unbeobachteten Heterogenität. Dabei folgen wir der Idee von Ghidry, Lesaffre & Eilers (2004) und Komarek, Lesaffre & Hilton (2005). In diesen Artikeln wurde ein ähnlicher Ansatz gewählt, im Rahmen der AFT-Modelle ("Accelerated Failure Time" Modelle), zum flexiblen Modellieren der Fehlergröße. Wir wählen also eine multivariate Mischstruktur der Form:

$$\mathbf{w}_i \sim \sum_{\mathbf{m}=\mathbf{0}}^{\mathbf{M}} c_{\mathbf{m}} \prod_{l=1}^L f_{\Gamma}(w_{il}; \mu_{lm_l}, \sigma_{lm_l}^2), \quad (8.11)$$

mit  $\mathbf{M} = [M_1, \dots, M_L]$  und  $\mathbf{0} = [0, \dots, 0]$ . (Beachte an dieser Stelle die Notationsänderung im Vergleich zu der Formel (8.3): der Multiindex  $\mathbf{m}$  durchläuft nun den Punkt  $\mathbf{0}$ . Der Grund hierfür wird in Abschnitt 8.6 erläutert.) Hier durchläuft der Multiindex  $\mathbf{m} = \{m_1, \dots, m_L\}$  das  $L$ -dimensionale Gitter mit Dimensionen  $(M_1 + 1) \otimes \dots \otimes (M_L + 1)$ , so dass für die fixe Kombination  $[m_1, \dots, m_L]$  das entsprechende Produkt von Gamma-Dichten gebildet wird. Die Parametersets  $\{\mu_{lm_l}\}$  und  $\{\sigma_{lm_l}^2\}$  sind Erwartungswerte bzw. Varianzen der jeweiligen Dichten. Diese Werte sind fix gesetzt und werden in der Optimierungsroutine nicht geändert. Dabei wird die erste Komponente  $\sigma_{l0}^2 \approx 0$ ,  $l = 1, \dots, L$ , gesetzt. Wir begründen diese Wahl in Unterabschnitt 8.6.1. (Die konkrete Wahl des Vektors für Varianzen erfolgt in Abschnitt 8.9.)



Eine allgemeine Restriktion, welche die Schätzung des Frailty-Parameters ermöglicht, ist dabei  $\mu_{lm_i} = 1$ . Eine umfassende Diskussion dieser Identifizierbarkeit findet sich in Hougaard (2000). Somit gilt für die Shape- und Scale-Parameter der Gamma-Dichte die Beziehung  $\alpha_{lm_i} = \beta_{lm_i} = \frac{1}{\sigma_{lm_i}^2}$  (siehe Definition der Dichte in Anhang A.4). Das Gitter für die Werte von  $\alpha_{lm_i}$  bzw.  $\sigma_{lm_i}^2$  legen wir eindeutig fest. Dabei ist die Forderung  $\alpha_{lm_i} > 1$  bzw.  $\sigma_{lm_i}^2 < 1$  zu berücksichtigen. In diesem Fall ist dann die Annahme  $E\{e^{w_i}\} < \infty$  nicht verletzt, da diese auch für die Identifizierbarkeit wichtig ist, vgl. dazu ?).

Um die Kriterien für die Definition einer Dichte zu erfüllen, benötigen wir außerdem die Restriktionen

$$\sum_{m=0}^M c_m = 1, \quad c_m \geq 0. \quad (8.12)$$

Im Fall  $L = 1$ , also nur ein Risiko, ergibt sich die gewöhnliche eindimensionale gemischte Dichte aus  $M_1$  Komponenten. Die Anpassung der Mischstruktur erfolgt allein durch die Bestimmung von Mischkomponenten  $c_m$ .

Die weiter oben eingeführte Frailty-Verteilung hat wegen der bekanntlich flexiblen Form sehr weitreichende Vorteile. Zunächst hat die unterstellte Flexibilität eine Wirkung auf das Problem der Identifizierbarkeit, wie in Abschnitt 8.2 erläutert. In vager Formulierung bedeutet dies: Je variabler die Frailty-Verteilung ist, desto höher ist die Chance, dass eine Identifizierbarkeit vorliegt (siehe dazu ?). Zum anderen wird damit die Problematik der Überparametrisierung vermieden. Mischverteilungen eignen sich hervorragend zur Beschreibung unbekannter Strukturen. Wie in Heckman & Singer (1984) diskutiert und mit ökonomischen Beispielen belegt, führt die falsche Spezifikation von Frailties zu verzerrten Ergebnissen. Die Schätzung der Parameter der Überlebensdauer ist sehr sensitiv bezüglich der Wahl der Frailty-Verteilung. Heckman & Singer (1984) diskutieren diese Problematik und studieren im Zusammenhang damit die diskrete multinomiale Frailty-Verteilungen als (nicht-parametrischen) ML-Schätzer im speziellen gemischten Modell. Die multinomiale Verteilung wird auch von van den Berg & Lindeboom (1998) unterstellt. Weitere Überlegungen bezüglich der Wahl von Frailty-Verteilung gibt van den Berg (2001), begründet dies aus der Sicht der ökonomischen Theorie ("job search" Modelle) und diskutiert Vor- und Nachteile sowohl von Gamma-Verteilung als auch von diskreter Multinomialverteilung. Unsere Wahl der Frailty-Verteilung vervollständigt dies auf den Fall von Competing Risks mit mehreren Spells pro Cluster (bzw. Individuum) und benötigt weniger Strukturparameter (also nur  $c_m$ ) zum Anpassen der Verteilung. Außerdem ermöglicht sie eine numerisch unaufwendige Schätzzroutine via EM-Algorithmus, welche im nächsten Abschnitt ausführlich beschrieben wird.

## 8.6 Schätzroutine; EM-Algorithmus

Wie schon erwähnt, wird für die Schätzung der unbekannt Parameter der EM-Algorithmus benutzt (siehe hierzu Dempster, Laird & Rubin, 1977, und McLachlan & Krishnan, 1997; kurz in Anhang A.9 skizziert). Die Vorgehensweise erscheint hier wegen der Form der unterstellten Frailty-Verteilung überlegen und verursacht insofern keinerlei numerischen Aufwand, als die Berechnungsformeln für Update-Werte und Verteilungsdichten in analytischer Form darstellbar sind. Außerdem sprechen für den Einsatz gute (unter milden Regularitätsannahmen) Konvergenzeigenschaften und eine relativ einfache Implementierung.

Beschreiben wir zunächst zusammenfassend die einzelnen Schritte. Um den M-Schritt ("Maximization-Step") zu realisieren, brauchen wir die Schätzwerte für Frailty-Parameter. Es wird dann, unter der Kenntnis von Frailty  $\mathbf{w}$ , die (komplette)  $\log$ -Likelihood bezüglich der Modellparameter  $\boldsymbol{\theta}$  maximiert. Die Schätzwerte für Frailties bekommen wir aus dem zugehörigen E-Schritt ("Expectation-Step"). In diesem Schritt werden die Erwartungswerte für die unbeobachteten Größen bestimmt, ausgehend von der vorherigen EM-Iteration. Dazu brauchen wir die durch die Kenntnis von Daten bedingte Dichte von Frailties. Wegen der Verteilungsannahme sind die Schätzwerte allein durch die Gewichte  $\mathbf{c}_m$  bestimmbar. Da diese nicht bekannt sind, werden sie aus den Daten geschätzt. Die zweifache Ausnutzung der beobachteten Daten ist unter dem Begriff "Empirical Bayes" bekannt (siehe dazu Carlin & Louis, 2000; ein kurzer Abriss der Bayes-Prozedur befindet sich in Anhang A.8).

Die einzelnen Gamma-Dichten haben die Gestalt:

$$f(\alpha_{lm_l}, \beta_{lm_l}; w_{il}) = \frac{\beta_{lm_l}^{\alpha_{lm_l}}}{\Gamma(\alpha_{lm_l})} w_{il}^{\alpha_{lm_l}-1} \exp\{-\beta_{lm_l} w_{il}\},$$

mit Shape-Parameter  $\alpha_{lm_l}$  und Scale-Parameter  $\beta_{lm_l}$ . Daraus berechnen sich Erwartungswert bzw. Varianz (vgl. Anhang A.4) durch  $\mu_{lm_l} = \frac{\alpha_{lm_l}}{\beta_{lm_l}} = 1$  bzw.  $\sigma_{lm_l}^2 = \frac{\alpha_{lm_l}}{\beta_{lm_l}^2} = \frac{1}{\alpha_{lm_l}}$  (wegen der in Abschnitt 8.5 eingeführten Restriktion). Die bedingte Dichte von Frailty  $\mathbf{w}_i$ , gegeben die Pseudo-Poisson Daten  $\boldsymbol{\delta}_i$  für das  $i$ -te Cluster (oder Individuum), resultiert aus

$$f(\mathbf{w}_i | \boldsymbol{\delta}_i) = \frac{f(\mathbf{w}_i, \boldsymbol{\delta}_i)}{f(\boldsymbol{\delta}_i)} = \frac{f(\boldsymbol{\delta}_i | \mathbf{w}_i) f(\mathbf{w}_i)}{f(\boldsymbol{\delta}_i)}, \quad i = 1, \dots, n. \quad (8.13)$$

Die interessierenden Dichten in (8.13) können wir wie folgt bestimmen. Wir fixieren zunächst den Index  $i$ , d. h. für das  $i$ -te Individuum und das  $l$ -te Risiko gilt:  $\delta_{ijkl} | w_{il} \sim \text{Poisson}(\xi_{ijkl} w_{il})$  mit  $\xi_{ijkl} = \exp\{x_{ij} W_l(t_{ijk}) \theta_l + \log[t_{ijk} - \tau_{k-1}]\}$ . Es gilt dann für den Vektor  $\boldsymbol{\delta}_{il}$ , bestehend aus den stochastisch unabhängigen Komponenten  $\{\delta_{ijkl}\}_{k=1, \dots, K_{ij}}^{j=1, \dots, n_i}$ :

$$\begin{aligned}
 f(\boldsymbol{\delta}_{il} | w_{il}) &= \prod_{j=1}^{n_i} \prod_{k=1}^{K_{ij}} \frac{(\xi_{ijkl} w_{il})^{\delta_{ijkl}}}{\delta_{ijkl}!} \exp\{-\xi_{ijkl} w_{il}\} \\
 &= \prod_{j=1}^{n_i} \prod_{k=1}^{K_{ij}} (\xi_{ijkl} w_{il})^{\delta_{ijkl}} \exp\{-\xi_{ijkl} w_{il}\},
 \end{aligned}$$

da  $\delta_{ijkl}! = 1$  ist, wegen  $\delta_{ijkl} \in \{0, 1\}$ , und damit für  $\boldsymbol{\delta}_i | \mathbf{w}_i$  (unter der bedingten Unabhängigkeit von Risiken bzw. von korrespondierenden Verweilzeiten):

$$f(\boldsymbol{\delta}_i | \mathbf{w}_i) = \prod_{l=1}^L f(\boldsymbol{\delta}_{il} | w_{il}) = \prod_{l=1}^L \prod_{j=1}^{n_i} \prod_{k=1}^{K_{ij}} (\xi_{ijkl} w_{il})^{\delta_{ijkl}} \exp\{-\xi_{ijkl} w_{il}\}. \quad (8.14)$$

Die marginale Dichte von  $\boldsymbol{\delta}_i$  ist dann

$$\begin{aligned}
 f(\boldsymbol{\delta}_i) &= \int f(\boldsymbol{\delta}_i | \mathbf{w}_i) f(\mathbf{w}_i) d\mathbf{w}_i = \int \prod_{l=1}^L f(\boldsymbol{\delta}_{il} | w_{il}) f(\mathbf{w}_i) d\mathbf{w}_i \\
 &= \int \left\{ \prod_{l=1}^L f(\boldsymbol{\delta}_{il} | w_{il}) \right\} \left\{ \sum_{m=0}^M c_m \prod_{l=1}^L f_{\Gamma}(w_{il}; \alpha_{lm_l}, \beta_{lm_l}) \right\} d\mathbf{w}_i \\
 &= \int \sum_{m=0}^M c_m \prod_{l=1}^L f(\boldsymbol{\delta}_{il} | w_{il}) f_{\Gamma}(w_{il}; \alpha_{lm_l}, \beta_{lm_l}) d\mathbf{w}_i \\
 &= \int \sum_{m=0}^M c_m \prod_{l=1}^L R_l(w_{il}) d\mathbf{w}_i = \sum_{m=0}^M c_m \prod_{l=1}^L \int R_l(w_{il}) dw_{il}, \quad (8.15)
 \end{aligned}$$

mit  $R_l(w_{il}) := f(\boldsymbol{\delta}_{il} | w_{il}) f_{\Gamma}(w_{il}; \alpha_{lm_l}, \beta_{lm_l})$ . Die Zeichen des Integrals und des Produktes sind umtauschbar. Es ist nun:

$$\begin{aligned}
 R_l(w_{il}) &= \prod_{j=1}^{n_i} \prod_{k=1}^{K_{ij}} (\xi_{ijkl} w_{il})^{\delta_{ijkl}} \exp\{-\xi_{ijkl} w_{il}\} \frac{\beta_{lm_l}^{\alpha_{lm_l}}}{\Gamma(\alpha_{lm_l})} w_{il}^{\alpha_{lm_l}-1} \exp\{-\beta_{lm_l} w_{il}\} \\
 &= \frac{\beta_{lm_l}^{\alpha_{lm_l}}}{\Gamma(\alpha_{lm_l})} \cdot \prod_{j=1}^{n_i} \xi_{ijl}^{d_{ijl}} w_{il}^{\sum_{j=1}^{n_i} d_{ijl} + \alpha_{lm_l} - 1} \cdot \exp\left\{-\left[\sum_{j=1}^{n_i} \sum_{k=1}^{K_{ij}} \xi_{ijkl} + \beta_{lm_l}\right] w_{il}\right\}.
 \end{aligned}$$

Mit den Bezeichnungen  $\alpha_{lm_l | \boldsymbol{\delta}_i} = \sum_{j=1}^{n_i} d_{ijl} + \alpha_{lm_l}$  und  $\beta_{lm_l | \boldsymbol{\delta}_i} = \sum_{j=1}^{n_i} \sum_{k=1}^{K_{ij}} \xi_{ijkl} + \beta_{lm_l}$  bekommen wir:

$$R_l(w_{il}) = \frac{\beta_{lm_l}^{\alpha_{lm_l}}}{\Gamma(\alpha_{lm_l})} \prod_{j=1}^{n_i} \xi_{ijlK_{ij}}^{d_{ijl}} w_{il}^{\alpha_{lm_l}|\delta_i - 1} \exp\{-\beta_{lm_l}|\delta_i w_{il}\},$$

und nach dem Integrieren über  $w_{il}$ :

$$\begin{aligned} \int R_l(w_{il}) dw_{il} &= \frac{\beta_{lm_l}^{\alpha_{lm_l}}}{\Gamma(\alpha_{lm_l})} \prod_{j=1}^{n_i} \xi_{ijlK_{ij}}^{d_{ijl}} \frac{\Gamma(\alpha_{lm_l}|\delta_i)}{\beta_{lm_l}^{\alpha_{lm_l}|\delta_i}} \\ &= \frac{\exp\{\alpha_{lm_l} \log(\beta_{lm_l})\}}{\exp\{\alpha_{lm_l}|\delta_i \log(\beta_{lm_l}|\delta_i)\}} \frac{\Gamma(\alpha_{lm_l}|\delta_i)}{\Gamma(\alpha_{lm_l})} \prod_{j=1}^{n_i} \xi_{ijlK_{ij}}^{d_{ijl}}. \end{aligned}$$

Damit ergibt sich insgesamt in (8.15)

$$\begin{aligned} f(\boldsymbol{\delta}_i) &= \sum_{m=0}^M c_m \left[ \prod_{l=1}^L \left\{ \frac{\exp\{\alpha_{lm_l} \log(\beta_{lm_l})\}}{\exp\{\alpha_{lm_l}|\delta_i \log(\beta_{lm_l}|\delta_i)\}} \frac{\Gamma(\alpha_{lm_l}|\delta_i)}{\Gamma(\alpha_{lm_l})} \prod_{j=1}^{n_i} \xi_{ijlK_{ij}}^{d_{ijl}} \right\} \right] \\ &= \sum_{m=0}^M c_m \Delta_{im}, \end{aligned} \quad (8.16)$$

wobei wir mit  $\Delta_{im}$  den Ausdruck in den eckigen Klammern in (8.16) bezeichnen. Die marginale (d. h. mit dem ausintegrierten  $\mathbf{w}_i$ ) Likelihood der Pseudo-Poisson Daten  $\boldsymbol{\delta}$  erhalten wir nun als

$$\mathcal{L}(\boldsymbol{\delta}) = \prod_{i=1}^n \sum_{m=0}^M c_m \Delta_{im}. \quad (8.17)$$

Die gemeinsame Dichte  $f(\boldsymbol{\delta}_i | \mathbf{w}_i) f(\mathbf{w}_i)$  resultierte bereits als Nebenergebnis in der Herleitung von (8.16). Es gilt nämlich:

$$\begin{aligned} f(\boldsymbol{\delta}_i | \mathbf{w}_i) f(\mathbf{w}_i) &= \sum_{m=0}^M c_m \prod_{l=1}^L f(\delta_{il} | w_{il}) f_{\Gamma}(w_{il}; \alpha_{lm_l}, \beta_{lm_l}) \\ &= \sum_{m=0}^M c_m \left[ \Delta_{im} \prod_{l=1}^L f_{\Gamma}(w_{il}; \alpha_{lm_l}|\delta_i, \beta_{lm_l}|\delta_i) \right]. \end{aligned} \quad (8.18)$$

Mit (8.16) und (8.18) bekommen wir schließlich den Ausdruck für die posteriori Dichte in (8.13):

$$\begin{aligned}
 f(\mathbf{w}_i | \boldsymbol{\delta}_i) &= \frac{\sum_{\mathbf{m}=0}^M c_{\mathbf{m}} \Delta_{i\mathbf{m}} \prod_{l=1}^L f_{\Gamma}(w_{il}; \alpha_{lm_l | \boldsymbol{\delta}_i}, \beta_{lm_l | \boldsymbol{\delta}_i})}{\sum_{\mathbf{m}=0}^M c_{\mathbf{m}} \Delta_{i\mathbf{m}}} \\
 &= \sum_{\mathbf{m}=0}^M \tilde{c}_{\mathbf{m}}^{(i)} \prod_{l=1}^L f_{\Gamma}(w_{il}; \alpha_{lm_l | \boldsymbol{\delta}_i}, \beta_{lm_l | \boldsymbol{\delta}_i}), \tag{8.19}
 \end{aligned}$$

mit  $\tilde{c}_{\mathbf{m}}^{(i)} = c_{\mathbf{m}} \Delta_{i\mathbf{m}} / \sum_{\mathbf{m}=0}^M c_{\mathbf{m}} \Delta_{i\mathbf{m}} = c_{\mathbf{m}} \bar{\Delta}_{i\mathbf{m}}$ ;  $\bar{\Delta}_{i\mathbf{m}} = \Delta_{i\mathbf{m}} / \sum_{\mathbf{m}=0}^M c_{\mathbf{m}} \Delta_{i\mathbf{m}}$ .

Aus (8.19) lassen sich nun leicht die posteriori Randdichten von  $w_{il}$ ,  $l = 1, \dots, L$ , bestimmen durch

$$f(w_{il} | \boldsymbol{\delta}_i) = \sum_{m_l=0}^{M_l} \tilde{c}_{(l)m_l}^{(i)} f_{\Gamma}(w_{il}; \alpha_{lm_l | \boldsymbol{\delta}_i}, \beta_{lm_l | \boldsymbol{\delta}_i}), \tag{8.20}$$

mit  $\tilde{c}_{(l)m_l}^{(i)} = \sum_{l'=1, l' \neq l}^L \sum_{\mathbf{m}_{-l}=0}^{M_{-l}} \tilde{c}_{\mathbf{m}}^{(i)}$ ,  $\mathbf{m}_{-l} = [m_1, \dots, m_{l-1}, m_{l+1}, \dots, m_L]$ . Dies bedeutet, der Multiindex  $\mathbf{m}$  durchläuft die Multi-Indexmenge, mit Ausnahme vom  $l$ -ten Element. Visualisiert am Beispiel  $L = 2$  sind die Koeffizienten  $\tilde{c}_{(l)m_l}^{(i)}$  einfach die Randsummen über der Matrix der Elemente  $\{\tilde{c}_{m_1 m_2}^{(i)}\}_{\substack{m_1=1, \dots, M_1 \\ m_2=1, \dots, M_2}}$ , so dass z. B. die Randsumme  $\tilde{c}_{1\bullet}^{(i)} = \tilde{c}_{11}^{(i)} + \dots + \tilde{c}_{1M_2}^{(i)} = \tilde{c}_{(1)m_1=1}^{(i)}$  in der obigen Notation und der reduzierte Index  $\mathbf{m}_{-1} = m_2 \in \{1, \dots, M_2\}$ .

Den Schätzer für den posteriori Frailty-Effekt  $w_{il}$  bekommen wir nun als den posteriori Erwartungswert:

$$\hat{w}_{il} = E(w_{il} | \boldsymbol{\delta}_i) = \sum_{m_l=0}^{M_l} \tilde{c}_{(l)m_l}^{(i)} \frac{\alpha_{lm_l | \boldsymbol{\delta}_i}}{\beta_{lm_l | \boldsymbol{\delta}_i}}. \tag{8.21}$$

Im hergeleiteten Ausdruck (8.21) hängt der Erwartungswert von den unbekanntem Mixture-Koeffizienten  $\mathbf{c}_{\mathbf{m}}$  ab. Deren Schätzung bekommt man grundsätzlich aus der Maximierung der marginalen  $\log$ -Likelihood (8.17), unter Restriktionen (8.12) (empirische Bayes-Prozedur, vgl. Anhang A.8). Die flexible Gestalt der unterstellten priori Verteilung für Frailties kann andererseits zu instabilen Schätzern für die Gewichte führen. Deshalb sollten diese, in Übereinstimmung mit dem allgemeinen Vorgehen, penalisiert werden, da wir an einem glatten Schätzer für die Frailty-Dichte interessiert sind. Wir führen daher die Penalisierungskomponente in der Form  $\frac{1}{2} \lambda_{\mathbf{c}} \mathbf{c}^T \boldsymbol{\Sigma}_{\mathbf{c}} \mathbf{c}$  ein, mit der noch weiter zu spezifizierenden Penalisierungsmatrix  $\boldsymbol{\Sigma}_{\mathbf{c}}$ . Damit wird der Komponentenvektor  $\mathbf{c}$  bestimmt aus der Maximierung der penalisierten (marginalen)

Likelihood mit Restriktionen:

$$\left\{ \begin{array}{l} \ell_{\text{pen}}(\boldsymbol{\delta}) = \log \left[ \prod_{i=1}^n \sum_{\mathbf{m}=0}^M c_{\mathbf{m}} \Delta_{i\mathbf{m}} \right] - \frac{1}{2} \lambda_{\mathbf{c}} \mathbf{c}^T \boldsymbol{\Sigma}_{\mathbf{c}} \mathbf{c}, \\ \sum_{\mathbf{m}=0}^M c_{\mathbf{m}} = 1, \quad c_{\mathbf{m}} \geq 0. \end{array} \right. \quad (8.22)$$

(8.22) ist ein typisches nichtlineares (konvexes) Optimierungsproblem mit Restriktionen, das zu lösen ist. Die Optimierungsstrategie ist in Anhang (A.11) skizziert. Bevor wir uns mit dem M-Schritt befassen, soll hier über die Wahl der Penalisierungsmatrix diskutiert werden.

### 8.6.1 Wahl der Penalisierungsmatrix

An dieser Stelle wollen wir die Wahl der Penalisierungsmatrix  $\boldsymbol{\Sigma}_{\mathbf{c}}$  sowie des Penalisierungsparameters  $\lambda_{\mathbf{c}}$  diskutieren. Die Wahl des Parameters  $\lambda_{\mathbf{c}}$  erfolgt diskret auf dem Gitter der Werte (konkret gesetzt werden sie in Abschnitt 8.9). Die großen Werte  $\lambda_{\mathbf{c}} \rightarrow \infty$  weisen zwingend auf das Modell ohne Frailties, also ohne (clusterspezifische) Heterogenitäten. Dies wird aus der Form des Penalisierungsterms unmittelbar deutlich, da in diesem Fall der Ausdruck  $\lambda_{\mathbf{c}} \mathbf{c}^T \boldsymbol{\Sigma}_{\mathbf{c}} \mathbf{c}$  das stärkste Gewicht hat und dessen Minimierung zu  $\boldsymbol{\Sigma}_{\mathbf{c}}^{1/2} \mathbf{c} = 0$  führt, also zur Lösung  $c_{\mathbf{m}} = 0$ , und zwar für alle  $c_{\mathbf{m}}$  außer dem einem, das der Varianz  $\sigma_{i_0}^2 \approx 0$  entspricht. Die Bestimmung des optimalen Wertes  $\lambda_{\mathbf{c}}$  erfolgt nach dem AIC-Kriterium, vgl. Formel (8.27).

Die geeignete Wahl der Penalisierungsmatrix hat einen entscheidenden Einfluss auf die "Güte" des Modells. Wenn keine Heterogenität bezüglich der Verweildauern in den Daten vorliegt, sollte zugunsten eines Modells ohne Frailties entschieden werden. In diesem Fall soll die (geschätzte) priori Frailty-Verteilung zu einer 1-Punkt Verteilung ausarten können. Dies kann eigentlich nur dann passieren, wenn alle, bis auf eins,  $c_{\mathbf{m}}$  stark penalisiert werden, so dass wegen der Restriktionen in (8.22) lediglich für ein einziges  $\mathbf{m}^*$  gilt:  $c_{\mathbf{m}^*} \approx 1$ . Dies ist, wiederum wegen der Wahl von  $\sigma_{l_{m_l}}^2$ , nur für  $\mathbf{m} = \mathbf{0}$  möglich, da in diesem Fall die korrespondierenden Gamma-Verteilungen die Varianzen  $\sigma_{i_0}^2 \approx 0$  haben. Das Wahrscheinlichkeitsmaß konzentriert sich also um den Wert 1. Umgekehrt sollen die mixture-Komponenten im Heterogenitätsfall weniger penalisiert werden, um die Variabilität aufdecken zu können. Diese Überlegung führt zu einer Ridge-Penalisierung auf die Gewichte  $c_{\mathbf{m}}$ . Um diese möglichst optimal zu wählen, betrachten wir zunächst die bedingte Survivor-Funktion (für ein Cluster  $i$ ; der Index  $i$  sei an dieser Stelle ausgelassen):

$$S(t | w) = \exp \left\{ - \sum_{l=1}^L w_l H_l(t; \mathbf{x}) \right\},$$

mit den integrierten Hazardraten  $H_l(t; \mathbf{x}) = \int_0^t h_l(u; \mathbf{x}) du$ . Die marginale Survivor-Funktion ergibt sich dann als

$$\begin{aligned} S(t) &= \int S(t | \mathbf{w}) f(\mathbf{w}) d\mathbf{w} \\ &= \sum_{\mathbf{m}=0}^M c_{\mathbf{m}} \int \exp \left\{ - \sum_{l=1}^L w_l H_l(t; \mathbf{x}) \right\} \prod_{l=1}^L f_{\Gamma}(w_l; \alpha_{lm_l}, \beta_{lm_l}) d\mathbf{w} \\ &= \sum_{\mathbf{m}=0}^M c_{\mathbf{m}} \exp \left\{ \sum_{l=1}^L \alpha_{lm_l} [\log(\beta_{lm_l}) - \log(\beta_{lm_l} + H_l(t; \mathbf{x}))] \right\}. \end{aligned}$$

Die Taylor-Entwicklung der Logarithme jeweils um die Stelle  $\beta_{lm_l}$  sowie die anschließende Approximation  $\exp(x) \approx 1 + x$  für kleine Argumente  $x$  resultiert in der Approximation

$$S(t) = \sum_{\mathbf{m}=0}^M c_{\mathbf{m}} \exp \left( - \sum_{l=1}^L \frac{\alpha_{lm_l}}{\beta_{lm_l}} H_l(t; \mathbf{x}) \right) \left\{ 1 + \frac{1}{2} \sum_{l=1}^L \frac{\alpha_{lm_l}}{\beta_{lm_l}^2} H_l^2(t; \mathbf{x}) + \dots \right\}.$$

Mit der Wahl des Erwartungswertes für mixture-Komponenten  $\alpha_{lm_l}/\beta_{lm_l} = 1$  (und damit für die Varianz  $\alpha_{lm_l}/\beta_{lm_l}^2 = \sigma_{lm_l}^2$ ) erhalten wir:

$$S(t) = S_0(t) \left\{ 1 + \frac{1}{2} \sum_{\mathbf{m}=0}^M c_{\mathbf{m}} \sum_{l=1}^L \sigma_{lm_l}^2 H_l^2(t; \mathbf{x}) + \dots \right\}, \quad (8.23)$$

wobei  $S_0(t) = \exp\{-\sum_{l=1}^L H_l(t; \mathbf{x})\}$  die Survivor-Funktion ohne Frailties bezeichnet. Der Ausdruck in den geschweiften Klammern in (8.23) ergibt also eine (multiplikative) Abweichung vom Modell ohne Frailties. Daraus ist ersichtlich, dass die Summen  $\sum_{l=1}^L \sigma_{lm_l}^2$  als Ridging-Faktoren oder Penalisierungen auf  $c_{\mathbf{m}} = c_{m_1 \dots m_L}$  gelten. (Der numerische Wert der integrierten Hazard  $H_l(t; \mathbf{x})$  ist dabei unerheblich.) In der Matrixform expandieren die Penalisierungen zu der Diagonalmatrix  $\text{diag}\{\sum_{l=1}^L \sigma_{lm_l}^2\}$ , so dass die Penalisierungsmatrix  $\Sigma_c$  die Gestalt hat:

$$\Sigma_c = \left[ \text{diag} \left\{ \sum_{l=1}^L \sigma_{lm_l}^2 \right\} \right]^2,$$

und die Indizes der Elemente in den Summen  $\sigma_{1m_1}^2 + \dots + \sigma_{Lm_L}^2$  mit dem entsprechenden Multiindex der Elemente  $c_{\mathbf{m}} = c_{m_1 \dots m_L}$  korrespondieren.

**Kommentar:** Die Stärke der Penalisierung steuert damit die Abweichung vom Modell ohne Frailties (also Heterogenitäten) und kann qualitativ eine Art Test für die Heterogenität darstellen, vgl. hierzu die Arbeit von Blossfeld

& Hamerle (1987). Wir wollen allerdings diese Sicht nicht weiter diskutieren und belassen es im weiteren Vorgehen bei der Ermittlung des optimalen Penalisierungswertes  $\lambda_c$  und der Schätzung der Frailties.

Damit ist der E-Schritt des EM-Algorithmus abgeschlossen. Im nächsten Abschnitt 8.7 wird die (unter der Kenntnis von Frailty  $\mathbf{w}$ ) komplette Modell-Likelihood maximiert. Die Smoothing-Parameter, die den Glättungsgrad der variierenden Koeffizienten steuern, werden aus der GLMM-Darstellung des Modells gewonnen. Dabei werden die als Random-Effekte behandelten Koeffizienten  $\mathbf{b}_l$ ,  $l = 1, \dots, L$ , in der Likelihood ausintegriert; es resultiert die (approximative marginale) Modell-Likelihood, die die Monotonie-Eigenschaft des EM-Algorithmus besitzt.

## 8.7 GLMM-Darstellung; Bestimmung der Penalisierungsparameter

Die Kenntnis der im E-Schritt geschätzten Frailty-Werte ermöglicht die Schätzung der Parameter  $\boldsymbol{\theta}_l = [\boldsymbol{\beta}_l, \mathbf{b}_l]$ ,  $l = 1, \dots, L$ , in der  $\log$ -Likelihood (8.10). Dabei werden die Parameter  $\mathbf{b}_l$  als zufällige Komponenten behandelt. Dies ermöglicht den Link zu gemischten Modellen und die Schätzung der Smoothing-Parameter  $\boldsymbol{\lambda}_l$  in derselben Art, wie dies bereits für das Modell in Kapitel 5 gezeigt wurde. Wir setzen also die Normalverteilung für die Komponenten  $\mathbf{b}_l$  voraus:

$$\mathbf{b}_{lr} \sim \mathcal{N}\left(\mathbf{0}, \frac{1}{\lambda_{lr}} \mathbf{D}_{lr}^{-1}\right), \quad l = 1, \dots, L; r = 0, \dots, p,$$

wobei für den Fall der truncated polynomials als Splinebasis die Matrizen  $\mathbf{D}_{lr}$  die Identitäten sind. Die Schätzung der Parameter  $\boldsymbol{\theta}_l$  und  $\boldsymbol{\lambda}_l$  erfolgt in zwei Schritten. Zunächst, bedingt auf  $\mathbf{b}_l$ , werden die Schätzer für  $\boldsymbol{\theta}_l = [\boldsymbol{\beta}_l, \mathbf{b}_l]$  aus den jeweiligen approximativen  $\log$ -Likelihood Funktionen  $\ell_l$  in (8.10) ermittelt (die Schätzung erfolgt dabei für jedes  $l = 1, \dots, L$  unabhängig). Um nun die Schätzer für  $\boldsymbol{\lambda}_l$  zu berechnen, werden die zufälligen Komponenten  $\mathbf{b}_l$  in der GLMM-Darstellung

$$\delta_l | \mathbf{b}_l, \mathbf{w}_l \sim \text{Poisson}(\exp\{\tilde{\mathbf{W}}_l \boldsymbol{\theta}_l + \mathbf{o}_l\})$$

aus der gemeinsamen Dichte von  $\boldsymbol{\beta}_l$ ,  $\mathbf{b}_l$  ausintegriert (hier  $\mathbf{w}_l = [w_{l1}, \dots, w_{ln}]$ ). Dabei benutzen wir die Laplace-Approximation des Integrals. Hier steckt implizit der Frailty-Wert  $\mathbf{w}_l$  im Offset  $\mathbf{o}_l$ . Es gilt also für die marginale Likelihood in diesem GLMM-Modell:



$$\begin{aligned}\mathcal{L}_{\text{marg},l}(\boldsymbol{\beta}_l, \boldsymbol{\lambda}_l | \mathbf{w}_l) &= \int \mathcal{L}_l(\boldsymbol{\theta}_l | \mathbf{w}_l) \mathcal{L}_l(\mathbf{b}_l) d\mathbf{b}_l \\ &= \int \exp \left\{ \ell_l(\boldsymbol{\theta}_l, \boldsymbol{\lambda}_l | \mathbf{w}_l) + \frac{1}{2} \log |\mathbf{D}_l(\boldsymbol{\lambda}_l)| - \frac{1}{2} \mathbf{b}_l^T \mathbf{D}_l(\boldsymbol{\lambda}_l) \mathbf{b}_l \right\} d\mathbf{b}_l,\end{aligned}$$

so dass sich für die marginale *log*-Likelihood die approximative Gestalt ergibt:

$$\begin{aligned}\ell_{\text{marg},l}(\boldsymbol{\beta}_l, \boldsymbol{\lambda}_l | \mathbf{w}_l) &= \log \{ \mathcal{L}_{\text{marg},l}(\boldsymbol{\beta}_l, \boldsymbol{\lambda}_l | \mathbf{w}_l) \} \\ &\approx \ell_l(\boldsymbol{\beta}_l, \hat{\mathbf{b}}_l, \boldsymbol{\lambda}_l | \mathbf{w}_l) - \frac{1}{2} \hat{\mathbf{b}}_l^T \mathbf{D}_l \hat{\mathbf{b}}_l + \frac{1}{2} \log |\mathbf{D}_l(\boldsymbol{\lambda}_l)| \\ &\quad - \frac{1}{2} \log |\mathbf{I}_{\text{pen},l}(\hat{\mathbf{b}}_l, \boldsymbol{\lambda}_l | \mathbf{w}_l)|.\end{aligned}\tag{8.24}$$

Hier maximiert  $\hat{\mathbf{b}}_l$  die penalisierte *log*-Likelihood  $\ell_{\text{pen},l}(\boldsymbol{\theta}_l, \boldsymbol{\lambda}_l | \mathbf{w}_l)$  bei gegebenem fixem  $\boldsymbol{\beta}_l$ . Außerdem bezeichnet  $\mathbf{I}_{\text{pen},l}(\hat{\mathbf{b}}_l, \boldsymbol{\lambda}_l | \mathbf{w}_l)$  den zu  $\hat{\mathbf{b}}_l$  gehörigen Block in der mit  $\mathbf{D}_l(\boldsymbol{\lambda}_l)$  penalisierten Fisher-Information ("–" 2-te Ableitung von  $\ell_{\text{pen},l}$  in (8.10)). Die Maximierung bezüglich  $\boldsymbol{\beta}_l$  führt zu dem ML-Schätzer  $\hat{\boldsymbol{\theta}}_l$ , gewonnen aus der penalisierten Form der *log*-Likelihood  $\ell_{\text{pen},l}(\boldsymbol{\theta}_l, \boldsymbol{\lambda}_l | \mathbf{w}_l)$ . Die resultierende approximative, an der ML-Stelle  $\hat{\boldsymbol{\theta}}_l$  evaluierte (bei gegebenem fixem  $\mathbf{w}_l$ ), marginale *log*-Likelihood hängt somit nur noch von  $\boldsymbol{\lambda}_{lr}$ ,  $r = 0, \dots, p$ , ab. Die jeweiligen ersten Ableitungen produzieren beim Setzen auf  $\stackrel{!}{=} 0$  die Fixpunkt-Iteration Gleichungen:

$$\frac{1}{\hat{\lambda}_{lr}} = \frac{\text{tr} \left\{ \left[ \mathbf{I}_{\text{pen},l}^{-1}(\hat{\mathbf{b}}_l, \boldsymbol{\lambda}_l | \mathbf{w}_l) \right]_{rr} \mathbf{D}_{lr} \right\} + \hat{\mathbf{b}}_{lr}^T \mathbf{D}_{lr} \hat{\mathbf{b}}_{lr}}{\dim\{\mathbf{b}_l\}},$$

wobei  $\dim\{\mathbf{b}_l\}$  die Anzahl der Komponenten im Vektor  $\hat{\mathbf{b}}_l$  bezeichnet. Der Subindex  $rr$  notiert den  $r$ -ten Unterblock der angegebenen Matrix.

Die *log*-Likelihood Funktionen  $\ell_{\text{marg},l}$ ,  $l = 1, \dots, L$ , in (8.24) hängen von den jeweiligen Frailty-Komponenten ab. Daher sollen die Frailties im Rahmen des M-Schrittes aus den Likelihoodfunktionen ausintegriert werden. Dazu schreiben wir die vom Frailty-Wert  $\mathbf{w}$  abhängende Likelihood zunächst in der Form:

$$\begin{aligned}\mathcal{L}_{\text{marg}}(\hat{\boldsymbol{\theta}}, \hat{\boldsymbol{\lambda}} | \mathbf{w}) &= \prod_{l=1}^L \mathcal{L}_{\text{marg},l}(\hat{\boldsymbol{\theta}}_l, \hat{\boldsymbol{\lambda}}_l | \mathbf{w}_l) \\ &= \prod_{l=1}^L \exp \left[ \ell_l(\hat{\boldsymbol{\theta}}_l, \hat{\boldsymbol{\lambda}}_l | \mathbf{w}_l) - \frac{1}{2} \hat{\mathbf{b}}_l^T \mathbf{D}(\hat{\boldsymbol{\lambda}}_l) \hat{\mathbf{b}}_l \right] \exp \left[ \left( \frac{\log |\mathbf{D}(\hat{\boldsymbol{\lambda}}_l)|}{\log |\mathbf{I}_{\text{pen},l}(\hat{\mathbf{b}}_l, \hat{\boldsymbol{\lambda}}_l)|} \right)^{1/2} \right] \\ &= \left\{ \prod_{l=1}^L \mathcal{L}_l(\hat{\boldsymbol{\theta}}_l, \hat{\boldsymbol{\lambda}}_l | \mathbf{w}_l) \right\} \left\{ \prod_{l=1}^L \exp \left[ -\frac{1}{2} \hat{\mathbf{b}}_l^T \mathbf{D}(\hat{\boldsymbol{\lambda}}_l) \hat{\mathbf{b}}_l + \left( \frac{\log |\mathbf{D}(\hat{\boldsymbol{\lambda}}_l)|}{\log |\mathbf{I}_{\text{pen},l}(\hat{\mathbf{b}}_l, \hat{\boldsymbol{\lambda}}_l)|} \right)^{1/2} \right] \right\}.\end{aligned}\tag{8.25}$$

Die penalisierte Fisher-Information  $\mathbf{I}_{\text{pen},l}$  hängt implizit von  $\mathbf{w}_l$  ab. Wir wollen aber diese Abhängigkeit vernachlässigen, indem wir im Ausdruck für  $\mathbf{I}_{\text{pen},l}$   $\mathbf{w}_l \equiv 1$  setzen. Es sind somit die Frailties nur noch aus den Likelihoods  $\mathcal{L}_l$  auszuintegrieren. Das Produkt  $\prod_{l=1}^L \mathcal{L}_l(\hat{\boldsymbol{\theta}}_l, \hat{\boldsymbol{\lambda}}_l | \mathbf{w}_l)$  ist nichts anderes als die (bedingte) Likelihood der Pseudo-Poisson Daten  $\boldsymbol{\delta}$ , gegeben Frailty  $\mathbf{w}$ . Die entsprechende Marginalisierung ergibt somit die Dichte wie in (8.16). Die anschließende Logarithmierung resultiert im Ausdruck

$$\begin{aligned} \ell_{\text{margin}}(\hat{\mathbf{c}}, \lambda_{\mathbf{c}}) &= \sum_{i=1}^n \log \left\{ \sum_{m=0}^M \hat{c}_m \Delta_{im} \right\} \\ &+ \sum_{l=1}^L \left\{ -\frac{1}{2} \hat{\mathbf{b}}_l^T \mathbf{D}(\boldsymbol{\lambda}_l) \hat{\mathbf{b}}_l + \frac{1}{2} \log |\mathbf{D}(\boldsymbol{\lambda}_l)| - \frac{1}{2} \log |\mathbf{I}_{\text{pen},l}(\hat{\mathbf{b}}_l, \boldsymbol{\lambda}_l)| \right\}. \end{aligned} \quad (8.26)$$

In (8.26) sind die mixture-Koeffizienten  $\hat{c}_m$  die aus dem E-Schritt geschätzten Werte. Die Monotonie-Eigenschaft von  $\ell_{\text{margin}}(\cdot, \lambda_{\mathbf{c}})$  ist unmittelbar an den Simulationen und am Anwendungsbeispiel überprüft worden (siehe Abschnitte 8.9 und 8.10).

Wir sind nun mit dem Problem konfrontiert, den Smoothing-Parameter  $\lambda_{\mathbf{c}}$  geeignet zu wählen. Unsere Idee besteht in einem Modellvergleich bezüglich des AIC-Kriteriums. Dazu sei das Gitter der Werte für  $\lambda_{\mathbf{c}}$  gewählt. Als optimaler Wert bestimmt sich dann  $\hat{\lambda}_{\mathbf{c}} = \operatorname{argmin}_{\lambda} \{ \text{AIC}(\lambda) \}$ . Den AIC-Wert berechnen wir nach der Formel

$$\text{AIC}(\lambda_{\mathbf{c}}) = -\ell_{\text{margin}}(\hat{\mathbf{c}}, \lambda_{\mathbf{c}}) + \text{df}(\lambda_{\mathbf{c}}), \quad (8.27)$$

wobei  $\text{df}(\lambda_{\mathbf{c}})$  die spezifizierten Freiheitsgrade (also die Komplexität des Modells) bestimmt. Üblicherweise ist dies die Spur des Produktes der beiden Inversen, der Fisher-Matrix und der penalisierten Fisher-Matrix:

$$\text{df}(\lambda_{\mathbf{c}}) = \text{tr} \left\{ \left[ -\frac{\partial^2 \ell_{\text{margin}}(\hat{\mathbf{c}}, \lambda_{\mathbf{c}})}{\partial \mathbf{c} \partial \mathbf{c}^T} + \lambda_{\mathbf{c}} \boldsymbol{\Sigma}_{\mathbf{c}} \right]^{-1} \left[ -\frac{\partial^2 \ell_{\text{margin}}(\hat{\mathbf{c}}, \lambda_{\mathbf{c}})}{\partial \mathbf{c} \partial \mathbf{c}^T} \right] \right\}. \quad (8.28)$$

Die zweiten Ableitungen in (8.28) haben wir bereits beim Maximieren der  $\log$ -Likelihood in (8.22) bestimmt, so dass kein zusätzlicher Aufwand entsteht. Theoretisch denkbar wäre auch eine optimale Wahl von  $\lambda_{\mathbf{c}}$  im Stile der "Mixed-Model"-Darstellung, indem die Verteilung für die Gewichte  $c_m$  spezifiziert ist. Dies würde aber den Rahmen der Arbeit bei Weitem sprengen.

## 8.8 Kovarianzen der Modellparameter

Befassen wir uns nun mit der Bestimmung der Varianzen von Parametern  $\boldsymbol{\theta}_l$ . Dies ist notwendig, um die Konfidenzbänder für die variierenden Koeffi-

zienten zu bestimmen. Die Berechnung im Rahmen des EM-Algorithmus erfolgt standardisiert nach der Louis-Formel (Louis, 1982, siehe auch Anhang A.9). Wegen der Unabhängigkeit erfolgt die Berechnung separat für jedes  $l = 1, \dots, L$ . In den Notationen des Anhangs A.9 stellt der Vektor  $\boldsymbol{\delta}_l$  die beobachteten Daten dar; der vollständige Beobachtungsvektor ist  $[\boldsymbol{\delta}_l, \boldsymbol{w}_l]$ , der interessierende Parametervektor ist  $\boldsymbol{\theta}_l$  und die komplette *log*-Likelihood ist  $\log \mathcal{L}_c(\boldsymbol{\delta}_l, \boldsymbol{w}_l) = \log \mathcal{L}(\boldsymbol{\delta}_l | \boldsymbol{w}_l) + \log \mathcal{L}(\boldsymbol{w}_l)$ . Da  $\log \mathcal{L}(\boldsymbol{w}_l)$  keinen Parameter  $\boldsymbol{\theta}_l$  enthält, geht die EM-Optimierung zurück auf die Schätzung aus der bedingten *log*-Likelihood  $\log \mathcal{L}(\boldsymbol{\delta}_l | \boldsymbol{w}_l)$ , die von  $\boldsymbol{w}_l$  linear abhängt. Daher wird im E-Schritt der (geschätzte) Erwartungswert für  $\boldsymbol{w}_l | \boldsymbol{\delta}_l$  in diese *log*-Likelihood eingesetzt. In der Louis-Formel für die Fisher-Information der beobachteten Daten stehen in den beiden Summanden die bedingten *log*-Likelihoods, da die kompletten *log*-Likelihoods nach  $\boldsymbol{\theta}_l$  abgeleitet werden. (Die *log*-Likelihood von  $\boldsymbol{w}_l$  hängt nicht von  $\boldsymbol{\theta}_l$  ab.) Wir haben also an der ML-Stelle  $\boldsymbol{\theta}_l$ ,  $l = 1, \dots, L$ :

$$\mathbf{I}(\hat{\boldsymbol{\theta}}_l; \boldsymbol{\delta}_l) = \mathbb{E}_{\boldsymbol{\theta}_l} \left\{ - \frac{\partial^2 \ell_l(\hat{\boldsymbol{\theta}}_l | \boldsymbol{w}_l)}{\partial \boldsymbol{\theta}_l \partial \boldsymbol{\theta}_l^T} \right\} - \text{Cov}_{\boldsymbol{\theta}_l} \left\{ \frac{\partial \ell_l(\hat{\boldsymbol{\theta}}_l | \boldsymbol{w}_l)}{\partial \boldsymbol{\theta}_l} \right\}, \quad (8.29)$$

mit der *log*-Likelihood  $\ell_l$  aus (8.9). Leiten wir  $\ell_l$  zwei mal nach  $\boldsymbol{\theta}_l$  ab, bekommen wir den Ausdruck für die (beobachtete) Fisher-Matrix der kompletten Daten:

$$\tilde{\mathbf{W}}_l^T \text{diag}[\exp\{\tilde{\mathbf{W}}_l \boldsymbol{\theta}_l + \boldsymbol{o}_l\}] \tilde{\mathbf{W}}_l. \quad (8.30)$$

Dabei ist der Vektor der Intensitäten:

$$\boldsymbol{\xi}_l = \exp\{\tilde{\mathbf{W}}_l \boldsymbol{\theta}_l + \boldsymbol{o}_l\} = \exp\{\tilde{\mathbf{W}}_l \boldsymbol{\theta}_l + \tilde{\boldsymbol{o}}_l\} \odot \tilde{\boldsymbol{w}}_l = \tilde{\boldsymbol{\xi}}_l \odot \tilde{\boldsymbol{w}}_l, \quad (8.31)$$

mit  $\tilde{\boldsymbol{o}}_l$  konstruiert wie  $\boldsymbol{o}_l$ , ohne den Frailty-Wert  $\boldsymbol{w}_l$  (vgl. Notationen in (8.8) und Anhang A.4 für die Operation  $\odot$ ; jedes  $i$ -te Element im Ausdruck für  $\tilde{\boldsymbol{w}}_l$  in (8.31) wird der Konstruktion von  $\tilde{\boldsymbol{o}}_l$  entsprechend oft repliziert). Der Erwartungswert von (8.30) ist nun leicht zu bilden. Mit Formel (8.21) bekommen wir für den ersten Term in (8.29)

$$\begin{aligned} \mathbb{E}_{\boldsymbol{\theta}_l} \left\{ - \frac{\partial^2 \ell_l(\hat{\boldsymbol{\theta}}_l | \boldsymbol{w}_l)}{\partial \boldsymbol{\theta}_l \partial \boldsymbol{\theta}_l^T} \right\} &= \tilde{\mathbf{W}}_l^T \text{diag}[\tilde{\boldsymbol{\xi}}_l \odot \mathbb{E}\{\tilde{\boldsymbol{w}}_l\}] \tilde{\mathbf{W}}_l \\ &= \tilde{\mathbf{W}}_l^T \text{diag}[\tilde{\boldsymbol{\xi}}_l] \text{diag}[\mathbb{E}\{\tilde{\boldsymbol{w}}_l\}] \tilde{\mathbf{W}}_l, \end{aligned} \quad (8.32)$$

mit  $\mathbb{E}\{\tilde{\boldsymbol{w}}_l\} = \left\{ \sum_{m_l=0}^{M_l} \tilde{c}_{(l)m_l}^{(i)} \frac{\alpha_{lm_l} \delta_i}{\beta_{lm_l} \delta_i} \right\}_{ijk}$  entsprechend aufgebläht. Der zweite Term in (8.29) wird analog berechnet, weil mit  $\partial \ell_l(\boldsymbol{\theta}_l | \boldsymbol{w}_l) / \partial \boldsymbol{\theta}_l = \tilde{\mathbf{W}}_l^T [\boldsymbol{\delta}_l - \boldsymbol{\xi}_l]$  gilt:

$$\begin{aligned} \text{Cov}_{\boldsymbol{\theta}_l} \left\{ \frac{\partial \ell_l(\widehat{\boldsymbol{\theta}}_l | \boldsymbol{w}_l)}{\partial \boldsymbol{\theta}_l} \right\} &= \widetilde{\boldsymbol{W}}_l^T \text{Cov}\{\tilde{\boldsymbol{\xi}}_l \odot \tilde{\boldsymbol{w}}_l\} \widetilde{\boldsymbol{W}}_l \\ &= \widetilde{\boldsymbol{W}}_l^T \text{diag}[\tilde{\boldsymbol{\xi}}_l] \text{diag}[\text{Var}\{\tilde{\boldsymbol{w}}_l\}] \text{diag}[\tilde{\boldsymbol{\xi}}_l] \widetilde{\boldsymbol{W}}_l, \end{aligned} \quad (8.33)$$

wobei  $\text{Var}\{\tilde{\boldsymbol{w}}_l\} = \left\{ \sum_{m_l=0}^{M_l} \tilde{c}_{(l)m_l}^{(i)} \frac{\alpha_{lm_l|\boldsymbol{\delta}_i}}{\beta_{lm_l|\boldsymbol{\delta}_i}^2} \right\}_{ijkl}$  entsprechend expandiert. Die gewonnenen Ausdrücke nun in (8.29) eingesetzt, bekommen wir

$$\begin{aligned} \boldsymbol{I}(\widehat{\boldsymbol{\theta}}_l; \boldsymbol{\delta}_l) &= \widetilde{\boldsymbol{W}}_l^T \text{diag}[\tilde{\boldsymbol{\xi}}_l] \left\{ \text{diag}[\text{E}\{\tilde{\boldsymbol{w}}_l\}] - \text{diag}[\tilde{\boldsymbol{\xi}}_l] \text{diag}[\text{Var}\{\tilde{\boldsymbol{w}}_l\}] \right\} \widetilde{\boldsymbol{W}}_l \\ &= \widetilde{\boldsymbol{W}}_l^T \text{diag}[\tilde{\boldsymbol{\xi}}_l] \text{diag} \left[ \left\{ \sum_{m_l=0}^{M_l} \tilde{c}_{(l)m_l}^{(i)} \frac{\alpha_{lm_l|\boldsymbol{\delta}_i}}{\beta_{lm_l|\boldsymbol{\delta}_i}} \left( 1 - \frac{\xi_{ijkl}}{\beta_{lm_l|\boldsymbol{\delta}_i}} \right) \right\}_{ijkl} \right] \widetilde{\boldsymbol{W}}_l. \end{aligned} \quad (8.34)$$

Wenn wir uns erinnern, dass  $\beta_{lm_l|\boldsymbol{\delta}_i} = \sum_{j=1}^{n_i} \sum_{k=1}^{K_{ij}} \xi_{ijkl} + \beta_{lm_l}$  ist, sehen wir unmittelbar die Konsequenzen der Formel (8.34). Mit wachsenden Clustergrößen  $n_i$  und steigender Überlebensdauer, also steigendem  $K_{ij}$ , wird die Fisher-Information größer und die Parameterschätzung von  $\boldsymbol{\theta}_l$  präziser. Berücksichtigen wir nun die Penalisierung, erhalten wir die Kovarianz für  $\widehat{\boldsymbol{\theta}}_l$  in der Form des Sandwich-Schätzers:

$$\text{Cov}\{\widehat{\boldsymbol{\theta}}_l\} = \left\{ \boldsymbol{I}(\widehat{\boldsymbol{\theta}}_l) + \boldsymbol{D}_l(\boldsymbol{\lambda}_l) \right\}^{-1} \boldsymbol{I}(\widehat{\boldsymbol{\theta}}_l) \left\{ \boldsymbol{I}(\widehat{\boldsymbol{\theta}}_l) + \boldsymbol{D}_l(\boldsymbol{\lambda}_l) \right\}^{-1}. \quad (8.35)$$

## 8.9 Simulationen

Wir beschreiben hier die zur Validierung des Modells notwendigen Simulationen. Es sollen die folgenden Steuerungsparameter festgelegt werden:

### Simulations-Input

- Anzahl der Cluster (oder Individuen)  $I$  sowie jeweilige Clustergrößen  $n_i, i = 1, \dots, I$ ; dabei ist  $n = \sum_{i=1}^I n_i$  die Gesamtzahl der Beobachtungen (Verweildauern).
- Anzahl der Competing Risks  $L$ .
- Glatte Komponenten, also die variierenden Koeffizienten  $\boldsymbol{\alpha}_l, l = 1, \dots, L$ .
- Frailty-Verteilung für  $\boldsymbol{w}_l, l = 1, \dots, L$ .
- Anzahl der zensierten Beobachtungen.

Sehr wichtig erscheint, die folgenden Simulationsparameter zu variieren: die Clustergrößen und -anzahl  $\{I, n_i\}$ , die Verteilung der Frailties sowie die Zensurierung. Wir beschränken uns auf die Situation  $L = 2$ . Für Frailty spezifizieren wir eine Gamma-Verteilung (siehe außerdem Kauermann & Khomski, 2007, für die weiteren Spezifikationen). Dabei wird der Erwartungswert in allen Verteilungen auf 1 gesetzt. Die Gesamtzahl an Clustern (bzw. Individuen) setzen wir auf jeweils  $I = 100, 1000$ . Zwei Alternativen zur Wahl der Clusterumfänge geben wir uns wie folgt vor. Zum einen setzen wir die Clustergrößen in Anlehnung an die Anwendungsdaten (vgl. dazu Tabelle 8.1) und simulieren dabei höchstens 5 Spells pro Cluster. Zum anderen werden die Spells einheitlich auf 10 pro Cluster gesetzt. Eine Begründung hierfür ist durch die Formel (8.34) gegeben. Wir bekommen eine bessere Anpassung der glatten Komponenten, indem die Anzahl der Spells pro Cluster erhöht wird. Dies wollen wir anhand der Simulationen überprüfen. Die glatten Komponenten sollen eine möglichst einfache Struktur haben, etwa Polynome 1-ten bzw. 2-ten Grades. Der Anteil der zensierten Überlebenszeiten soll dabei in der jeweiligen Simulationsserie ca. 5%, 20%, 50% und 75% betragen. Die Anzahl der Simulationen für jede Konstellation der Simulationsparameter wird auf 100 gesetzt.

Für den Fall  $L = 2$  simulieren wir die Überlebenszeiten wie folgt: Für jedes Cluster (oder Individuum)  $i$  werden unabhängig voneinander die Frailty-Werte  $w_{i1}$  und  $w_{i2}$  aus den spezifizierten Verteilungen gezogen. Wir starten zunächst mit dem Zeitpunkt  $t = 1$  und ziehen die pseudo-Poisson Beobachtung  $\delta_{ijtl}$  mit der Intensität  $h_l(t; x)w_{il}$ , wobei  $\log\{h_l(t; x)\}$  die spezifizierte glatte Komponente ist und  $x$  die simulierte binäre Kovariable bezeichnet. Die  $x$ -Werte werden für jedes Cluster unabhängig simuliert mit der Wahrscheinlichkeit  $P(x = 1) = 0.3$ . Falls beide  $\delta_{ijtl} = 0$ ,  $l = 1, 2$ , wiederholen wir die Prozedur der Ziehung für den Zeitpunkt  $t + 1$ . Falls  $\delta_{ijt1} = 1$  für  $l = 1$  gesampelt wird und zugleich  $\delta_{ijt2} = 0$  für  $l = 2$ , schreiben wir die realisierte Überlebenszeit  $t_{ij} = t$ , mit der Markierung  $d_{ij} = \{1, 0\}$ , in den simulierten Datensatz aus. Ein analoges Vorgehen gilt im umgekehrten Fall, wenn das Event vom Typ  $l = 2$  eintritt. Für den Fall, dass  $\delta_{ijtl} = 1$  für beide  $l = 1, 2$ , ziehen wir zufällig eines der beiden Events mit der Wahrscheinlichkeit, proportional zu den Intensitäten  $h_l(t)w_{il}$  und zwar:

$$P(\text{Event} = l) = \frac{h_l(t)w_{il}}{\sum_{l=1}^2 h_l(t)w_{il}}.$$

(Falls dabei für  $\delta_{ijtl}$  numerisch die Werte  $> 1$  gesampelt werden, sind diese auf 1 zu setzen.) Die Funktionen  $h_l(t)$  haben die Form:

$$h_l(t) = \exp\{p_l(t) + x q_l(t)\},$$

wobei  $p_l(\cdot)$  und  $q_l(\cdot)$  die Polynome des Grades höchstens 2 und weiter unten spezifiziert sind. Außerdem setzen wir den cut-off Wert für  $t$  auf 30, so dass Beobachtungen mit der Realisierung  $t > 30$  auf 30 gesetzt werden und als

zensiert notiert. Die realisierten Lebensdauern seien dabei für die Interpretationszwecke in *Monaten*, also auf Monatsbasis, gemessen.

Die weiteren Simulationsparameter sind:

- *Splinebasis*. Als übliche Wahl gelten die truncated polynomials 1-ten Grades mit  $K = 20$  Knoten, auch wenn die B-Splines ebenfalls gut geeignet erscheinen.
- *Anzahl der mixture-Komponenten*  $M_1 \times \dots \times M_L$ . Die Standardwahl ist dabei  $M_1 = \dots = M_L = 7$ .
- *Varianzen* der einzelnen Gamma-Verteilungen in der Mixture werden auf dem Gitter  $\{0.001, 0.05, 0.15, 0.4, 0.6, 0.8, 0.99\}^2$  gewählt.
- *Penalisierung*  $\lambda_c$  wird auf dem Gitter  $\{0.001, 0.1, 1, 10, 100, 1000, 5000, 10000, 100000, 1000000\}$  vorgenommen.

Der Fall  $L = 1$  entspricht nur einem Risiko im Modell und kann auch für die Situation angewandt werden, in der alle anderen Events, die nicht von Interesse sind, zusammengefasst werden, und in der die entsprechende Verweildauer als zensiert betrachtet wird. Hier unterscheidet sich das Simulationsschema nicht von denen der vorigen Modelle. Der Wert  $h_1(t)$  wird lediglich mit dem gesampelten Frailty-Wert multipliziert.

Für  $L = 2$  Competing Risks sei die Wahl der zu simulierenden glatten Komponenten wie folgt festgelegt:

$$\begin{aligned}\alpha_{01}^t(t) &= \eta_1 - \frac{t}{30}, & \alpha_{02}^t(t) &= \eta_2 + \frac{t}{50}, \\ \alpha_{x1}^t(t) &= 0.5 + \left(\frac{t}{25}\right)^2, & \alpha_{x2}^t(t) &= -\frac{t}{60},\end{aligned}$$

wobei die Parameter  $\eta_1, \eta_2$  im Ausdruck für die logarithmierten Baseline-Hazards gerade die relative Häufigkeit der zensierten Beobachtungen steuern. Die Intensitäten der jeweilig zu simulierenden Poisson-Daten sind  $\xi_1 = \exp\{\alpha_{01}^t(t) + x \alpha_{x1}^t(t)\}w_1$  und  $\xi_2 = \exp\{\alpha_{02}^t(t) + x \alpha_{x2}^t(t)\}w_2$  mit simulierten Frailty-Werten  $w_1$  und  $w_2$ . Die folgenden eingesetzten Werte für  $\eta_1, \eta_2$  entsprechen den Zensierungsmustern  $h_\delta$ :

$$\begin{aligned}\eta_1 = -2, \eta_2 = -3 & \rightarrow h_\delta \simeq 5\%, \\ \eta_1 = -6, \eta_2 = -3 & \rightarrow h_\delta \simeq 20\%, \\ \eta_1 = -5, \eta_2 = -4 & \rightarrow h_\delta \simeq 50\%, \\ \eta_1 = -5, \eta_2 = -6 & \rightarrow h_\delta \simeq 75\%.\end{aligned}$$

Als Ergebnis der Simulationen (siehe Anhang B.3.1) bekommen wir neben den empirischen Konfidenzbändern und coverage probabilities für die simu-

lierten glatten Komponenten auch die empirischen Konfidenzbänder für die Verteilungsfunktion und die Verteilungsdichte der simulierten Frailties.

Die obigen Simulationen beruhen maßgeblich auf den diskreten Werten für die Überlebenszeit. Um die Stetigkeit der Lebensdauer zu imitieren, müssen wir in der jeweiligen Simulation das Gitter der möglichen Realisierungswerte für die Zeit  $t$  verfeinern, d. h., in der obigen Simulationsvorschrift gehen wir zu einem größeren Wertebereich für  $t$  über, z. B. von der *Monatsskala* der Werte 1 bis 30 zu einer *Tagesskala* der Werte von 1 bis 900, jeder Monat bestehe dabei aus 30 Tagen. Die zu simulierenden glatten Komponenten werden dann wie folgt umgerechnet:

$$\alpha_{01}^t(t) = \eta_1 - \frac{t/30}{30}, \quad \alpha_{02}^t(t) = \eta_2 + \frac{t/30}{50},$$

$$\alpha_{x1}^t(t) = 0.5 + \left(\frac{t/30}{25}\right)^2, \quad \alpha_{x2}^t(t) = -\frac{t/30}{60}.$$

Die Hazardraten der beiden Competing Risks, also die Intensitäten der jeweiligen zu simulierenden Poisson-Daten werden nun zu  $\xi_1 = \frac{1}{30} \exp\{\alpha_{01}^t(t) + x \alpha_{x1}^t(t)\}w_1$  und  $\xi_2 = \frac{1}{30} \exp\{\alpha_{02}^t(t) + x \alpha_{x2}^t(t)\}w_2$  mit simulierten Frailty-Werten  $w_1$  und  $w_2$ . Das Zensierungsmuster  $h_\delta$  bleibt unverändert. Die Simulationsergebnisse (mit derselben Parameterkonstellation) sind in Anhang B.3.2 abgebildet. Für die Vergleichszwecke mit den Simulationen aus B.3.1 zeichnen wir die simulierten variierenden Koeffizienten auf der Monatsskala, geschätzt auf der Tagesbasis.

## 8.10 Anwendungsbeispiel

Zur Analyse wurden die Daten mit den Arbeitslosigkeitszeiten des SOEP herangezogen (siehe Abschnitt 11.1). Betrachtet werden die Individuen ohne Beschäftigung (und arbeitslos gemeldet) in den Jahren 1990-2000. Jede in Betracht zu ziehende Periode der Arbeitslosigkeit kann aus mehreren Gründen terminiert werden. Unser Interesse gilt aber nur zwei konkurrierenden Austrittszuständen (also Competing Risks), der Voll- bzw. Teilzeitbeschäftigung. Alle übrigen Terminierungen werden als zensiert betrachtet. Die Dauer der Arbeitslosigkeit wird in Monaten gemessen, was die Kalkulation der Risiken technisch möglich macht. Tabelle 8.1 zeigt die Verteilung der Spells (Perioden) der Arbeitslosigkeit bei 2714 Individuen. Mehr als die Hälfte war nur einmal arbeitslos. Zwei Individuen wechselten sogar 10 mal zwischen Beschäftigung und Arbeitslosigkeit. Einen ersten Eindruck von den Überlebensdauern gibt der Kaplan-Meier-Schätzer in Abbildung 8.1. Daraus können wir unmittelbar auf die im Vergleich zur Beschäftigung in Teilzeit erhöhten Chancen bei voller Wiederbeschäftigung schließen. Intuitiv könnte man dies mit einer niedrigeren Bereitschaft der Menschen, einen Teilzeitjob zu übernehmen, begründen,

was bis Mitte der 1990-er Jahre noch hätte gelten können. Die geschätzten priori-Dichten sowie die Frailty-Schätzer sind in Abbildung 8.2 dargestellt.

Die Kovariablen im Modell sind dieselben wie im Modell der Two-Way Hazards und ausführlich in Abschnitt 11.1 beschrieben. In der Dummy-Kodierung mit derselben Kategorisierung sind in Abbildung 8.3 die geschätzten logarithmierten Kovariableneffekte der Ausländer im Vergleich zu den Deutschen, Frauen vs. Männer, der jüngeren bzw. der älteren Arbeitnehmer im Vergleich zu den Personen mittleren Alters sowie der nicht Ausgebildeten im Vergleich zu den ausgebildeten Personen und derjenigen mit Hochschulabschluss im Vergleich zu den Arbeitslosen ohne Hochschulbildung dargestellt. Aus dem linken oberen Plot (logarithmierte Baseline Hazard für die Vollbeschäftigung) schließen wir auf die Reduktion der Chancen mit langer Arbeitslosigkeitsdauer. Anders sieht es im Fall der Aufnahme von Teilzeitarbeit aus. Das Risiko wird über die ganze Verweildauer ohne Beschäftigung hindurch auf ungefähr demselben Niveau geschätzt mit einer Erhöhung der Chancen am Ende der Beobachtungszeit. Diese Steigerung könnte man damit begründen, dass nach langer Arbeitslosigkeit doch die Bereitschaft wächst, eine Teilzeitbeschäftigung aufzunehmen, weil sonst keine Leistungen vom Staat beansprucht werden können. Der Effekt der Nationalität in Bezug auf die Vollbeschäftigung zeigt, dass die Chancen der nicht deutschen Arbeitnehmer in den ersten zwei Jahren der Arbeitslosigkeit kleiner sind als die der deutschen. Dieses Bild ändert sich aber diametral bei den langzeitarbeitslosen Ausländern. Generell sind dagegen die Chancen der Ausländer auf Teilzeitarbeit konstant schlechter als bei deutschen Arbeitnehmern. Die weiblichen Arbeitslosen haben signifikant geringere Chancen auf Vollbeschäftigung, und diese Chancen sinken mit zunehmender Dauer der Arbeitslosigkeit. Umgekehrt sind ihre (beinahe konstanten) Chancen auf einen Teilzeitjob signifikant höher als bei männlichen Erwerbstätigen. In Bezug auf Vollbeschäftigung spielt das Alter eine signifikante Rolle und ist (erwartungsgemäß) bei jüngeren Personen höher und bei älteren niedriger als bei Arbeitslosen mittleren Alters. Kein signifikanter Unterschied bezüglich der Chancen besteht jedoch bei jüngeren Arbeitslosen im Zusammenhang mit einer Teilzeitbeschäftigung. Dieser Unterschied ist jedoch bei den älteren Personen signifikant auf gleichem Niveau. Ausgebildete Personen haben generell höhere Chancen auf eine Wiederbeschäftigung sowohl in Voll- als auch in Teilzeit; Chancen auf einen Teilzeitjob sind aber nicht signifikant ausgeprägt. Ein Studium steigert die Chancen auf Wiederbeschäftigung; für Teilzeitjobs sogar signifikant höher während der gesamten Dauer der Arbeitslosigkeit. Dies könnte z. B. durch eine hohe Bereitschaft, jeder in Frage kommenden Arbeit nachzugehen, erklärt werden.



number of spells	1	2	3	4	5	6	7	8	9	10
number of individuals	1944	473	191	56	27	11	6	4	0	2
proportion	0.72	0.17	0.07	0.02	0.01	<0.01	<0.01	<0.01	0	<0.01

Tabelle 8.1: Verteilung der Anzahl der Spells für den Datensatz der Competing Risks.

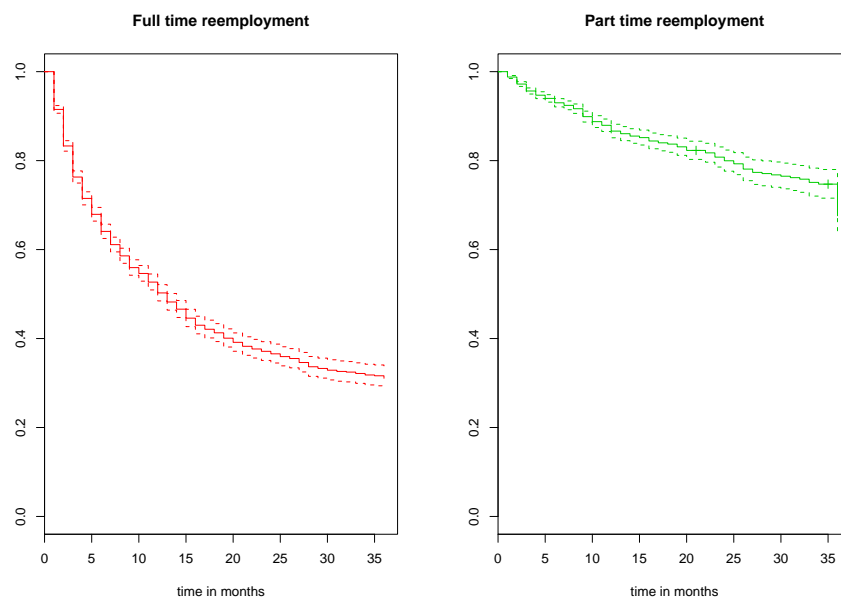


Abbildung 8.1: Kaplan-Meier-Schätzer für die zwei Competing Risks.

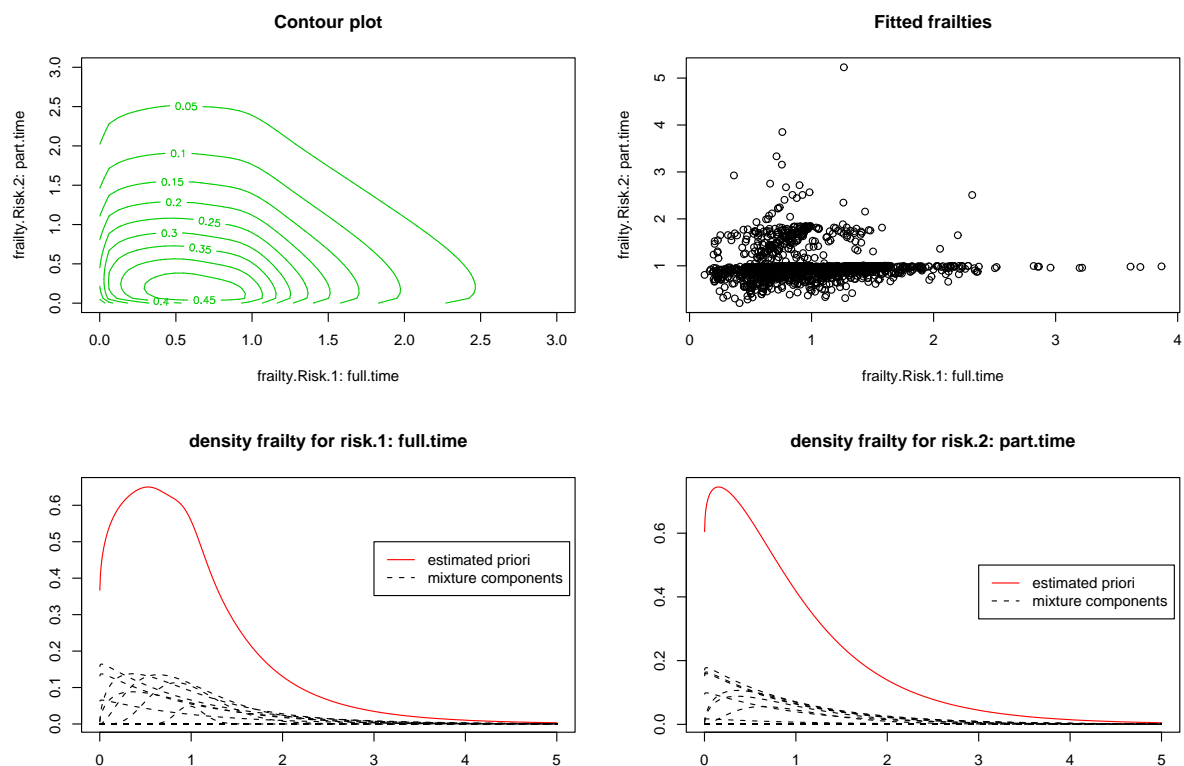


Abbildung 8.2: Geschätzte bivariate Frailty-Verteilung (oben links) sowie geschätzte posteriori Erwartungswerte der Frailties (oben rechts). Unten sind die geschätzten marginalen Frailty-Dichten (durchgezogen) sowie die einzelnen normierten (gestrichelt) mixture Dichte-Komponenten gezeichnet.

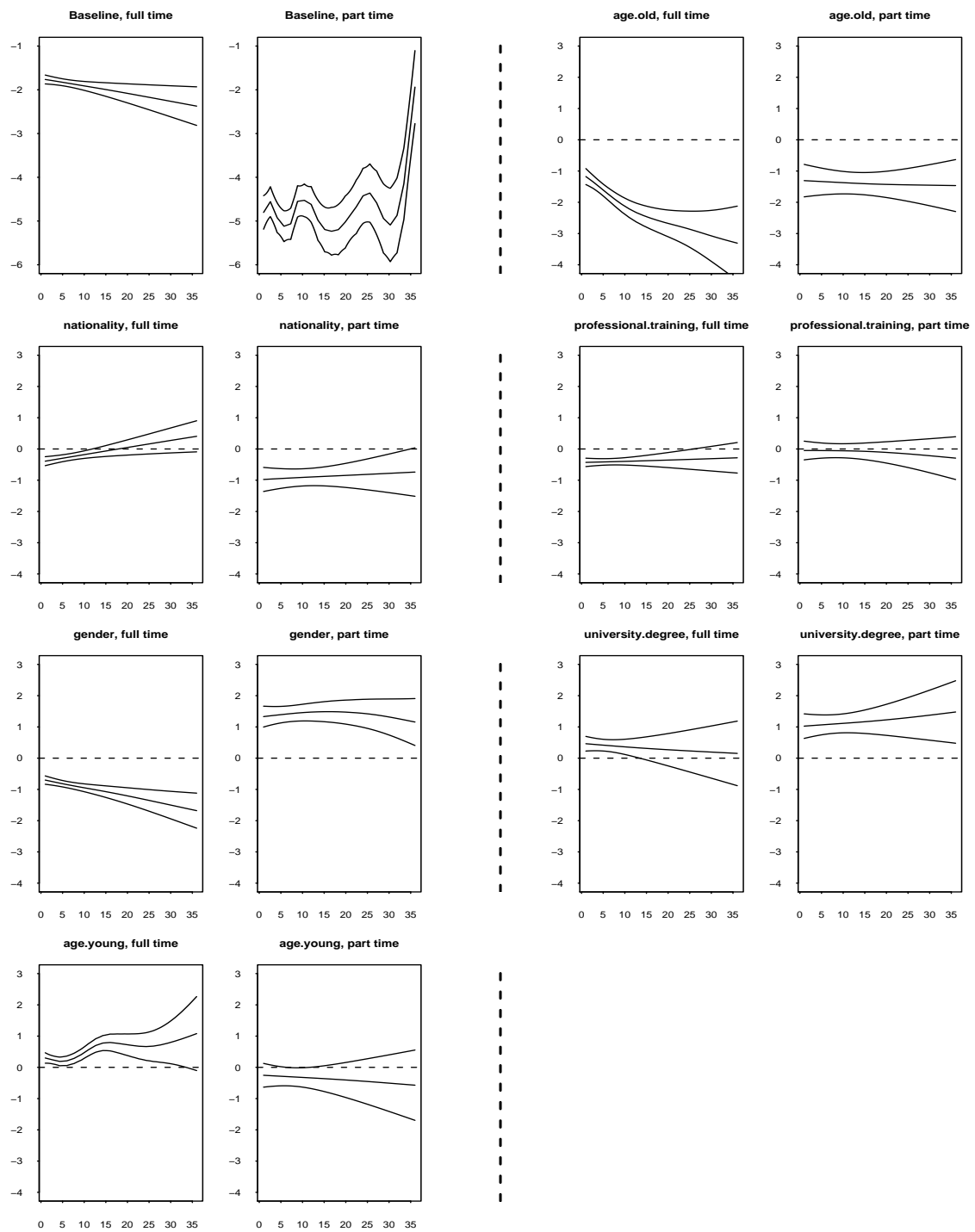


Abbildung 8.3: Geschätzte dynamische Effekte mit Konfidenzintervallen für die Vollzeitbeschäftigung, links, und für Teilzeitarbeit, rechts. Die gestrichelte 0-Linie zeigt das Referenzniveau.

# Kapitel 9

## Multi-Zustandsmodell

In diesem Kapitel betrachten wir die spezielle Situation zweier Zustände und modellieren die Hazardraten für den Übergang aus einem der Zustände in den anderen, mit möglicher Zensierung. Es ist außerdem möglich, beliebig oft zwischen den Zuständen zu wechseln; keiner der beiden ist also ein absorbierender Zustand. Dabei berücksichtigen wir die Verweildauern im betreffenden Zustand. Da mehrere Beobachtungen oder Spells (im jeweiligen Zustand) pro Individuum (oder Cluster) betrachtet werden, werden auch die individuellen Frailty-Parameter mit in das Modell integriert. In der Literatur wurde der Vorgang (ohne Berücksichtigung der Heterogenität) unter anderem in Hougaard (2000) und Kalbfleisch & Prentice (2002) diskutiert. Eine wichtige, das Modellieren erleichternde Annahme ist die Markov-Eigenschaft des Event-History-Prozesses für die Multi-Zustandsmodelle. Da die Verweildauer mitberücksichtigt wird, spricht man von einer Semi-Markov-Eigenschaft.

Die Modellierungstechnik unterscheidet sich nur geringfügig vom Modell der Competing Risks. Daher sind die im Weiteren zu benutzenden Notationen weitgehend aus dem vorigen Kapitel übernommen. Wir erläutern zunächst die zusätzlichen das weitere Vorgehen klärenden Notationen. Die Hazardrate aus dem Zustand 1 in den Zustand 2 zu geraten bezeichnen wir als  $h_{(1)} = h_{(1)2}w_{(1)2}$ . Dabei indiziert der Subindex (1)2 die (evtl. zensierte) Verweildauer im Zustand 1, welche durch den Wechsel in den Zustand 2 endet. Der Frailty-Parameter  $w_{(1)2}$  wirkt multiplikativ auf die Hazard. Mit der ähnlichen Notation für den umgekehrten Fall  $2 \rightarrow 1$  besitzt das  $i$ -te Individuum den Frailty-Parametervektor  $\mathbf{w}_i = [w_{(1)2}, w_{(2)1}]$ , welcher durch die gemischte Gamma-Verteilungsstruktur approximiert wird. Die Likelihood hat die Form

$$\mathcal{L} = \prod_{l=1}^2 \prod_{i=1}^n \prod_{j=1}^{n_i} \prod_{l' \neq l} \left[ h_{(l)l'}^{\delta_{ij}^{l'}}(t_{ij}; \mathbf{x}_{ij}) w_{(l)l'} \exp\left\{-\int_0^{t_{ij}} h_{(l)l'}(u; \mathbf{x}_{ij}) w_{(l)l'} du\right\} \right]. \quad (9.1)$$

Dabei ist  $\delta_{ij}^{l'}$  der Eventindikator mit den  $\{0, 1\}$ -Werten und hat die Ausprägung 1, falls der Übergang in den Zustand  $l'$  aus dem Zustand  $l$  stattfindet. Weitere

Bezeichnungen sind selbsterklärend. (Wir bemerken an dieser Stelle kurz, dass nicht unbedingt jedes Individuum in jedem der Zustände verweilen muss bzw. eine Beobachtung darüber im Datensatz vorliegen soll. Im jeweiligen Fall fehlt der korrespondierende Eintrag in der gebildeten Likelihood bzw. wird durch den 1-Eintrag ersetzt. Diese Konvention ist wichtig für die spätere Implementierung.) Die Dichte der künstlich generierten Poisson-Daten  $\boldsymbol{\delta}_{i.1}$  ist dann (vgl. Kapitel 3):

$$f(\boldsymbol{\delta}_{i.1} | w_{i(1)2}) = \prod_j \prod_k \frac{\{\xi_{ij1k} w_{i(1)2}\}^{\delta_{ijk1}}}{\delta_{ijk1}!} \exp\{-\xi_{ij1k} w_{i(1)2}\}, \quad (9.2)$$

mit  $\xi_{ij1k} = \exp\{\mathbf{x}_{ij} \mathbf{W}_1(t_{ijk}^{(1)}) \boldsymbol{\theta}_1 + o_{ijk}^{(1)}\}$  und  $\delta_{ijk1}! = 1$  und folgenden Komponenten: Integrationspunkte  $\tau_k^{(1)}$ , Überlebenszeiten  $t_{ij}^{(1)}$  im 1-ten Zustand,  $t_{ijk}^{(1)} = \min\{t_{ij}^{(1)}, \tau_k^{(1)}\}$ , Zeilenvektor  $\mathbf{x}_{ij}$ , Designmatrix  $\mathbf{W}_1(t_{ijk}^{(1)})$  sowie Offset-Parameter  $o_{ijk}^{(1)}$ . Analog gilt mit ähnlichen Bezeichnungen für den umgekehrten Fall:

$$f(\boldsymbol{\delta}_{i.2} | w_{i(2)1}) = \prod_j \prod_k \frac{\{\xi_{ij2k} w_{i(2)1}\}^{\delta_{ijk2}}}{\delta_{ijk2}!} \exp\{-\xi_{ij2k} w_{i(2)1}\}. \quad (9.3)$$

Aus (9.2) und (9.3) ergeben sich die marginalen Dichten von  $\boldsymbol{\delta}_{i.1}$  bzw.  $\boldsymbol{\delta}_{i.2}$  wie folgt:

$$\begin{cases} f(\boldsymbol{\delta}_{i.1}) = \int f(\boldsymbol{\delta}_{i.1} | w_i) f(w_i) dw_i = \int f(\boldsymbol{\delta}_{i.1} | w_{i(1)2}) f(w_{i(1)2}, w_{i(2)1}) dw_{i(1)2} dw_{i(2)1}, \\ f(\boldsymbol{\delta}_{i.2}) = \int f(\boldsymbol{\delta}_{i.2} | w_i) f(w_i) dw_i = \int f(\boldsymbol{\delta}_{i.2} | w_{i(2)1}) f(w_{i(1)2}, w_{i(2)1}) dw_{i(1)2} dw_{i(2)1}. \end{cases} \quad (9.4)$$

Die Randdichten von  $w_{i(1)2}$  bzw.  $w_{i(2)1}$  bestimmen sich aus der priori Dichte des Frailty-Vektors:  $f(\mathbf{w}_i) = \sum_{m=0}^M c_m \prod_{l=1}^2 f_\Gamma(w_{il}; \alpha_{lm_l}, \beta_{lm_l})$ , mit den Parametern  $(\alpha_{lm_l}, \beta_{lm_l})$  wie in Kapitel 8 definiert. Es gilt:

$$\begin{cases} f(w_{i1}) = \sum_{m_1=0}^{M_1} c_{(1)m_1} f_\Gamma(w_{i1}; \alpha_{1m_1}, \beta_{1m_1}), \\ f(w_{i2}) = \sum_{m_2=0}^{M_2} c_{(2)m_2} f_\Gamma(w_{i2}; \alpha_{2m_2}, \beta_{2m_2}). \end{cases} \quad (9.5)$$

Dabei sind (in gewohnter Notation)  $c_{(l)m_l}$  die Randsummen, im zweidimensionalen Fall einfach die Zeilen- ( $c_{m_1 \bullet}$ ) bzw. Spalten- ( $c_{\bullet m_2}$ ) summen. (Um die Notation übersichtlich zu gestalten, bezeichnen wir in (9.5)  $w_{il} := w_{i(l)l'}$  für

Frailty  $w_{i(l)l'}$  und meinen damit das Risiko, von Zustand  $l$  in Zustand  $l'$  zu geraten.) Für die marginalen Dichten ergibt sich also:

$$\begin{aligned} f(\boldsymbol{\delta}_{i.l.}) &= \int f(\boldsymbol{\delta}_{i.l.} | w_{i(l)l'}) f(w_{i(l)l'}) dw_{i(l)l'} \\ &= \sum_{m_l=0}^{M_l} c_{(l)m_l} \Delta_{i l m_l}, \end{aligned}$$

mit  $\Delta_{i l m_l}$  wie folgt:

$$\Delta_{i l m_l} = \frac{\exp\{\alpha_{lm_l} \log \beta_{lm_l}\}}{\exp\{\alpha_{lm_l|\boldsymbol{\delta}_{i.l.}} \log \beta_{lm_l|\boldsymbol{\delta}_{i.l.}}\}} \cdot \frac{\Gamma(\alpha_{lm_l|\boldsymbol{\delta}_{i.l.}})}{\Gamma(\alpha_{lm_l})} \cdot \prod_{\substack{j=1 \\ s_{ij}=l}}^{n_i} \xi_{ijl}^{\mathbf{1}_{\{d_{ijl}>0\}}}$$

Dabei bezeichnet  $d_{ijl}$  den Eventindikator und  $s_{ij} = l$  indiziert die Produktbildung nur über die Zeilen (also einzelne Spells) im Datensatz, die dem Verweilzustand  $l$  entsprechen. Die marginale (Daten-)Likelihood ergibt sich damit wie folgt:

$$\mathcal{L} = \prod_{i=1}^n f(\boldsymbol{\delta}_i) = \prod_{i=1}^n \prod_{l=1}^2 f(\boldsymbol{\delta}_{i.l.}) = \prod_{i=1}^n \prod_{l=1}^2 \sum_{m_l=0}^{M_l} c_{(l)m_l} \Delta_{i l m_l}, \quad (9.6)$$

und die marginale  $\log$ -Likelihood ist

$$\log \mathcal{L} = \sum_{l=1}^2 \sum_{i=1}^n \log \left\{ \sum_{m_l=0}^{M_l} c_{(l)m_l} \Delta_{i l m_l} \right\}. \quad (9.7)$$

(Der Vollständigkeit halber soll an dieser Stelle bemerkt werden, dass es in einem tatsächlichen Fall nicht für jedes Individuum  $i$  die Information darüber vorliegen soll, ob es im jeweiligen Zustand  $l$  jemals verweilte. Daher soll berechtigterweise zwischen den Indizes  $i_l$ ,  $l = 1, 2$ , unterschieden werden. In der Formel (9.6) sind im Produkt die jeweiligen Terme durch 1-Einträge zu ersetzen.) Die Gewichte  $c_m$  bestimmen sich hieraus als Lösung des folgenden Optimierungsproblems:

$$\left\{ \begin{array}{l} \log \mathcal{L} = \sum_{l=1}^2 \sum_{i=1}^n \log \left\{ \sum_{m_l=0}^{M_l} c_{(l)m_l} \Delta_{i l m_l} \right\} \rightarrow \max \\ \sum_{m=0}^M c_m = 1, \quad c_m \geq 0. \end{array} \right.$$

Dieses Problem ist dem Problem (8.22) ähnlich, und die Optimierungsstrategie ist in Anhang A.11 beschrieben. Im nächsten Schritt lassen sich die Schätzwerte für Frailties ermitteln. Dazu ist zunächst die bedingte Dichte zu bilden. Unter der bedingten Unabhängigkeit gilt:

$$f(\mathbf{w}_i | \boldsymbol{\delta}_i) = \frac{f(\boldsymbol{\delta}_i | \mathbf{w}_i) f(\mathbf{w}_i)}{f(\boldsymbol{\delta}_i)} = \frac{f(\boldsymbol{\delta}_{i1} | w_{i1}) f(\boldsymbol{\delta}_{i2} | w_{i2}) f(\mathbf{w}_i)}{f(\boldsymbol{\delta}_i)}, \quad (9.8)$$

mit  $f(\mathbf{w}_i) = f(w_{i1}, w_{i2}) = \sum_{m_1, m_2} c_{m_1 m_2} f_\Gamma(w_{i1}; \alpha_{1m_1}, \beta_{1m_1}) f_\Gamma(w_{i2}; \alpha_{2m_2}, \beta_{2m_2})$  und  $f(\boldsymbol{\delta}_i) = f(\boldsymbol{\delta}_{i1}) f(\boldsymbol{\delta}_{i2}) = \left[ \sum_{m_1=0}^{M_1} c_{(1)m_1} \Delta_{i1 m_1} \right] \left[ \sum_{m_2=0}^{M_2} c_{(2)m_2} \Delta_{i2 m_2} \right]$ . Damit ist die bedingte Dichte von  $w_{i1}$  gegeben durch:

$$\begin{aligned} f(w_{i1} | \boldsymbol{\delta}_i) &= \int f(\mathbf{w}_i | \boldsymbol{\delta}_i) dw_{i2} = \frac{1}{f(\boldsymbol{\delta}_i)} f(\boldsymbol{\delta}_{i1} | w_{i1}) \int f(\boldsymbol{\delta}_{i2} | w_{i2}) f(\mathbf{w}_i) dw_{i2} \\ &= \frac{1}{f(\boldsymbol{\delta}_i)} f(\boldsymbol{\delta}_{i1} | w_{i1}) \left[ \sum_{m_1, m_2} c_{m_1 m_2} f_\Gamma(w_{i1}; \alpha_{1m_1}, \beta_{1m_1}) \int f(\boldsymbol{\delta}_{i2} | w_{i2}) f_\Gamma(w_{i2}; \alpha_{2m_2}, \beta_{2m_2}) dw_{i2} \right] \\ &= \frac{1}{f(\boldsymbol{\delta}_i)} \sum_{m_1, m_2} c_{m_1 m_2} \Delta_{i2 m_2} f(\boldsymbol{\delta}_{i1} | w_{i1}) f_\Gamma(w_{i1}; \alpha_{1m_1}, \beta_{1m_1}). \end{aligned}$$

Der Übergang zum Erwartungswert liefert:

$$\begin{aligned} E\{w_{i1} | \boldsymbol{\delta}_i\} &= \frac{1}{f(\boldsymbol{\delta}_i)} \sum_{m_1, m_2} c_{m_1 m_2} \Delta_{i1 m_1} \Delta_{i2 m_2} \frac{\alpha_{1m_1} \boldsymbol{\delta}_{i1}}{\beta_{1m_1} \boldsymbol{\delta}_{i1}} \\ &= \sum_{m_1, m_2} c_{m_1 m_2} \bar{\Delta}_{i1 m_1} \bar{\Delta}_{i2 m_2} \frac{\alpha_{1m_1} \boldsymbol{\delta}_{i1}}{\beta_{1m_1} \boldsymbol{\delta}_{i1}}, \quad (9.9) \end{aligned}$$

mit  $\bar{\Delta}_{i l m_l} = \Delta_{i l m_l} / \sum_{m_l=0}^{M_l} c_{(l)m_l} \Delta_{i l m_l}$ ,  $l = 1, 2$ . Analog gilt für  $l=2$ :

$$E\{w_{i2} | \boldsymbol{\delta}_i\} = \sum_{m_1, m_2} c_{m_1 m_2} \bar{\Delta}_{i1 m_1} \bar{\Delta}_{i2 m_2} \frac{\alpha_{2m_2} \boldsymbol{\delta}_{i2}}{\beta_{2m_2} \boldsymbol{\delta}_{i2}}.$$

Der EM-Algorithmus zum Schätzen der Frailties stimmt also mit dem aus Kapitel 8 überein. Die angefangenen Simulationen überzeugen von der Güte der Modellanpassung. Es ist beabsichtigt, die Ergebnisse der Modellierung in der zur Veröffentlichung reifen Form darzustellen.

# Kapitel 10

## Zusammenfassung und Ausblick

Allen diskutierten Modellen lag die grundsätzliche Idee einer Transformation von Ausgangsdaten in einen GLM-Rahmen, basierend auf dem Poisson-Modell. Die Approximation des Integrals von der Hazardrate, also der kumulativen Hazardrate, via Rechteck- oder trapezoide Annäherung ist dabei wesentlich. Diese Überführung ermöglichte eine numerisch unaufwendige und transparente Parameterschätzung via PQL-Methode (im Gegensatz z. B. der auf dem MCMC-Konzept basierten Berechnung). Die Smoothing-Parameter, die die Glattheit der variierenden Koeffizienten steuerten, konnten ebenso leicht datengesteuert ermittelt werden, ausgehend von der GLMM-Darstellung, mittels der Marginalisierung via Laplace-Approximation. Die durchgeführten Simulationen überzeugten von der Anpassungsgüte. Die Wahl einer bestimmten P-Spline Basis konnte nicht als grundsätzlich angesehen werden; beide Basen scheinen gute approximative und numerische Eigenschaften zu besitzen.

Die additiven Modelle der Two-Way-Art in Kapiteln 5 und 6 lassen sich auf ein Modell mit Berücksichtigung der Interaktion erweitern, das in Kapitel 7 ansatzweise für die Lebensdauer und Saison (eines Jahres) dargestellt wurde. Ein wichtiger Aspekt dabei ist eine mögliche anschließende Kovariablenselektion. Zum einen kann das "beste" Kovariablenset mit herkömmlichen Modellvergleichsverfahren, basierend z. B. auf AIC-Kriterium, durchgeführt werden. Zum anderen kann die Kovariablenwahl, ausgehend von dem vollständigen Modell, auf der Basis der optimalen Wahl der Penalisierungsparameter erfolgen. Hierzu schließen sich mögliche Tests auf die funktionale Form der Effekte, parametrisch vs. nichtparametrisch, an.

Im Modell der Competing Risks erlaubte der gewählte Hazardansatz eine einfache Modellspezifikation ohne Identifizierbarkeitsprobleme, vgl. Diskussion dazu in Abschnitt 8.2. Für Frailties wurde eine gemischte flexible Struktur gewählt, die eine erhebliche Relaxierung des parametrischen Falls ermöglichte. Dabei hat sich die korrekte Wahl der Penalisierung der Gewichte in der Mixture-Struktur als wesentlich erwiesen, siehe die Herleitung in Abschnitt 8.6.1. Mit den gewählten Spezifikationen ließen sich der Schätzalgorithmus einfach implementieren und die Schätzer für Frailties leicht berechnen. Die Anpassungsgüte



wurde an den zahlreich durchgeführten Simulationen überprüft. Die Genauigkeit der Schätzer für die glatten Komponenten hängt dabei mit der Clustergröße zusammen. Dies folgt unmittelbar aus der Formel (8.34).

Auch im Fall der Competing Risks sind Erweiterungen bzw. Verallgemeinerungen des Grundmodells möglich. So kann z. B. eine zweite Zeitskala eingeführt werden und das Modellieren in der schon bekannten Two-Way-Art bzw. in der Form einer Interaktion zwischen den beiden Zeitachsen erfolgen. Außerdem könnten zeitabhängige Frailties, insbesondere im Falle langer Lebensdauern, modelliert werden, da sich die unbeobachteten (also latenten) Einflüsse, aber auch die Clusterunterschiede, mit der Zeit ändern können. In der SchätZRoutine könnte die Wahl des Penalisierungsparemeters  $\lambda_c$  auch datengesteuert erfolgen, indem für die Gewichte  $c_m$  auch eine priori-Verteilung unterstellt würde. Da diese dabei nichtnegativ sind, kann im stetigen Fall grundsätzlich keine Normalverteilung zugrunde gelegt werden. Eine Alternative wäre Gamma- oder eine andere (nicht zu schiefe) Verteilung mit positivem Träger und  $\lambda_c$  als zusätzlichem Verteilungsparemeter. Der Aufwand beim Schätzen ist dabei vermutlich wesentlich.

Die Multi-Zustandsmodelle sind in der Arbeit nur kurz angerissen worden. Die geplante Publikation der Ergebnisse steht noch aus. Weitere Forschung könnte sich stärker auf einen allgemeinen Fall mehrerer Zustände konzentrieren (absorbierende Zustände miteingeschlossen). Der Modellierungsvorgang stellt allerdings eine wesentliche Herausforderung dar, da das damit verbundene Problem - wie sich die additive Hazardstruktur einfach in das Modell integrieren lässt - gelöst werden soll.

Das Modellieren mit P-Splines erfordert insbesondere im Bereich der Überlebenszeiten eine noch umfassendere Forschung, um ein starkes theoretisches Fundament zu bilden. Darauf aufbauend kann die P-Spline gestützte Anwendung nur profitieren. Unsere Zusammenarbeit konnte hier entsprechenden Beitrag leisten.

# Kapitel 11

## Datenmaterial

Hier diskutieren wir ausführlich die für die Modellanalysen eingesetzten Daten.

### 11.1 Arbeitslosigkeitsdaten des SOEP

Die Modelle in Kapiteln 5 und 8 wurden angewandt an die Daten über die Arbeitslosigkeitsdauer aus dem deutschen Sozio-ökonomischen Panel (abgekürzt SOEP). SOEP ist eine Service-Einrichtung innerhalb des Deutschen Instituts für Wirtschaftsforschung (DIW Berlin, siehe <http://www.diw.de>), die im Jahre 1983 gegründet wurde und von der Abteilung "Längsschnittstudie Sozio-ökonomisches Panel" betrieben wird. Das Kernstück des SOEP beruht auf einer repräsentativen Wiederholungsbefragung von über 12000 Privathaushalten in Deutschland. Die Auswahl der interessierenden Merkmale ist dabei sehr groß, so dass damit sehr unterschiedliche Lebensbereiche abgedeckt werden. Der SOEP-Datensatz gibt Auskunft über objektive Lebensbedingungen, Wertvorstellungen, Risikoeinstellungen, den Wandel in verschiedenen Lebensbereichen und über die Abhängigkeiten, die zwischen verschiedenen Lebensbereichen und deren Veränderungen existieren. Das Erhebungsschema entspricht im Wesentlichen dem Längsschnittdesign (hat also ausgeprägten Panelcharakter). Das SOEP zeichnet sich durch eine hohe Stabilität aus. Kontinuierliche Informationen sind unter anderem vorhanden über Erwerbs- und Familienbiografien, Erwerbsbeteiligung und berufliche Mobilität, Einkommensverläufe, Gesundheit, Lebenszufriedenheit, gesellschaftliche Partizipation und Zeitverwendung, Persönlichkeitsmerkmale, Haushaltszusammensetzung, Wohnsituation. (Weitere Information siehe [www.diw.de/deutsch/sop](http://www.diw.de/deutsch/sop).)

Die Analyse des Modells aus Kapitel 5 besteht in der Untersuchung der Verweildauer im Zustand der Arbeitslosigkeit, terminiert durch das interessierende Ereignis: Wiederaufnahme der Vollbeschäftigung als Austritt aus diesem Zustand. Die zur Untersuchung herangezogenen Kovariablen haben einen makroökonomischen Charakter. Beschränkt wurde die Auswahl auf die Kovariablen *Alter*, *Geschlecht*, *Nationalität*, berufliche *Ausbildung* und *Hochschul-*

*abschluss.* Hiervon sind Beobachtungen aus den neuen Bundesländern ausgeschlossen. Für das uns interessierende Ereignis, Austritt aus der Arbeitslosigkeit mit Wiederaufnahme einer Vollbeschäftigung, enthält der resultierende Datensatz die Personenangaben in mehreren gleichartigen Zuständen. So tritt z. B. der Zustand der Arbeitslosigkeit bei einer Person im Durchschnitt knapp über 1.5-mal mit insgesamt von 1 bis 15 Arbeitslosigkeitsperioden (oder Spells) bei Einzelpersonen auf. Im Modell des Kapitels 5 analysieren wir die Situation mit jeweils nur einem Arbeitslosigkeitsspell, wobei wir die Spells zufällig ziehen. Außerdem werden alle diejenigen Beobachtungen aus der Analyse eliminiert, deren Eintrittsdatum in die SOEP-Studie in den Monat Januar des Jahres 1983 fällt, da die ältere Biografie einer Person nicht bekannt ist. Es resultiert ein endgültiger Datensatz mit 4020 Beobachtungen und folgenden Spalten:

- *time*: Dauer der Arbeitslosigkeit in vollen Monaten.
- *birth*: Eintrittsmonat (vom Februar 1983 bis Dezember 2000) in den Zustand der Arbeitslosigkeit.
- *status*: Zensierungsindikator mit der üblichen Kodierung: 1 als vollbeschäftigt, 0 anders (> 36 Monate arbeitslos auch als zensiert gesetzt).
- *Nationalität*: deutsch vs. nicht deutsch (Referenzkategorie: deutsch).
- *Geschlecht*: männlich vs. weiblich (Referenzkategorie: männlich).
- *Alter*: wurde kategorisiert in < 25, bis 50 und > 50 Jahre (Referenzkategorie: mittleres Alter).
- *Beruf*: erlernter Beruf (oder Ausbildung) vorhanden: ja/nein (Referenzkategorie: ja).
- *Hochschule*: Hochschulausbildung: ja/nein (Referenzkategorie: nein).

Im Modell der Competing Risks erweitert sich der Datensatz um die Spalten, die den jeweiligen Austrittszuständen (aus dem Arbeitslosigkeitszustand der beobachteten Verweildauer) entsprechen, in der  $\{0, 1\}$ -Kodierung, 1 für das Event des Interesses, 0 für zensierten Fall. Von Interesse sind zwei Zustände: Wiederaufnahme der Voll- oder Teilzeitbeschäftigung als zwei Competing Risks für die Beendigung der Arbeitslosigkeit. Zur Analyse werden alle Spells der Arbeitslosigkeit herangezogen. Wir betrachten hier die Spells ab dem Jahr 1990, so dass der resultierende Datensatz insgesamt 3982 Beobachtungen enthält. Die Häufigkeiten der Spells sind Tabelle 8.1 zu entnehmen. Da jedes Individuum mehrfach arbeitslos gewesen sein kann und somit die Information darüber in geclusteter Form vorliegt und analysiert wird, sind auch die Clusteridentifikationen (z. B. die Personennummern) mit anzugeben. In beiden Modellen werden die Dauern der Arbeitslosigkeit in vollen Monaten gemessen.

## 11.2 Kindersterblichkeiten

Zwar hatte die Feldstudie zur Erfassung aller im spezifizierten ländlichen Teil von Burkina Faso geborenen Kinder im Jahr 1993 begonnen, doch wurden später kritische Anmerkungen über die anfangs erfassten Daten geäußert. Die Datenqualität wurde dann aber erheblich verbessert, so dass sehr verlässliche Beobachtungen ab dem Jahr 1998 vorliegen. Der 1. Januar 1998 gilt somit als der Zeitpunkt des ersten Eintritts in die Studie, also die Geburt eines Kindes. Da diese Studie langfristig angelegt ist, so dass kontinuierlich neue Information in die Datenbank hineinfließt, terminieren wir für unsere Modellierungszwecke bis zum 31. Oktober 2005. In seiner endgültigen Form enthält der zu analysierende Datensatz 8986 Beobachtungen. Hiervon haben 1003 Kinder ihr vollendetes 5. Lebensjahr nicht erreicht. (Diese 5-Jahresgrenze dient auch als das Ende der "follow-up"-Periode für ein Kind.) Somit liegt die relative Häufigkeit der zensierten Beobachtungen bei 89%. Die Ergebnisse der Simulationsstudie unterstützen uns in der Annahme der Zuverlässigkeit für die resultierenden Schätzer. Als Kovariablen enthält der Datensatz folgende Merkmale:

- *Stammeszugehörigkeit*,
- *Religion*,
- *Alter der Mutter zum Zeitpunkt der Geburt*,
- *Zwillingsstatus*,
- *Entfernung zum nächstgelegenen medizinischen Behandlungsort*,
- *Geschlecht*,
- *Saison der Geburt*.

In einer früheren Studie (siehe Becher et al., 2004) wurden außerdem die zusätzlichen Kovariablen "*Jahr der Geburt*" und "*Reihenfolge der Geburt*" in die Auswertung einbezogen. Diese werden hier aber nicht berücksichtigt. Die Saison der Geburt  $s$  wird, neben der eigentlichen Überlebenszeit  $t$ , als zweite zeitliche Komponente in das Modell aufgenommen und in der nichtparametrischen (Spline-)Form dargestellt. Die beiden Zeitskalen werden in Tagen gemessen. Allerdings verwenden wir der Übersichtlichkeit halber in der grafischen Repräsentation die Jahres- (für  $t$ ) bzw. Monatsskala (für  $b$ ). Alle metrischen Kovariablen werden kategorisiert. Die *Stammeszugehörigkeit* hat die Ausprägungen {MAR:=Marka, BWA:=Bwaba, MOS:=Mossi, PLH:=Peulh, SAM.AUT:=Samo und andere zusammen} mit MAR als Referenzkategorie. Die *Religion* beinhaltet {MUS:=Moslems, CAT.PRT:=Katholiken und Protestanten zusammen, ANI.AUT=alle weiteren religiösen Minderheiten} mit MUS als Referenzkategorie. Das *Alter der Mutter* wurde kategorisiert in {jung:= < 18, mittel:= < 35 und alt:=  $\geq$  35} mit dem mittleren Alter als Referenzkategorie. Die *Entfernung zum nächsten Krankenhaus* wurde geteilt in { $\leq$  10 km, > 10 km} mit  $\leq$  10 km als Referenzkategorie. In Tabelle 11.1 sind die

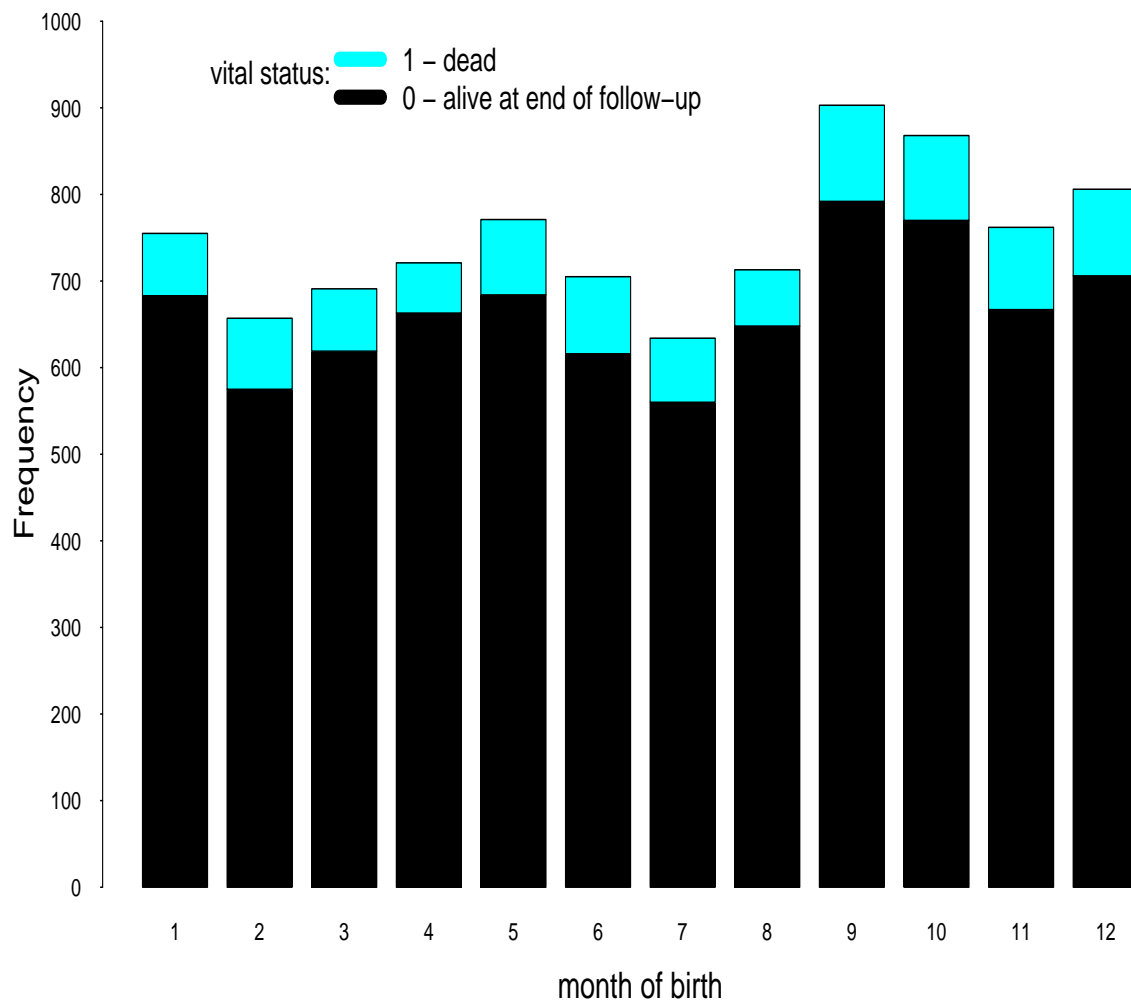


Abbildung 11.1: Histogramm der Geburtsmonate mit Zensierung.

Häufigkeitsverteilungen zusammengetragen. (Die unteren Zahlen sind die Angaben in %.) Die Verteilung der Geburten ist in Abbildung 11.1 dargestellt. Die Ergebnisse des Cox-Modells sind in Tabelle 11.2 zu entnehmen.

		Vital status		
		alive	dead	
sex	female	4078 89.55	476 10.45	4554
	male	3905 88.11	527 11.89	4432
ethnic group	Dafing	2990 88.10	404 11.90	3394
	Bwaba	1928 90.01	214 9.99	2142
	Mossi	1544 89.82	175 10.18	1719
	Peulh	810 86.17	130 13.83	940
	Samo	526 89.30	63 10.70	589
	other	185 91.58	17 8.42	202
religion	muslim	5071 88.24	676 11.76	5747
	catholic	2122 90.80	215 9.20	2337
	natural	467 85.53	79 14.47	546
	protestant	315 90.52	33 9.48	348
	other	8 100.00	0 0.00	8
age of mother at birth	<18	607 85.61	102 14.39	709
	18-34	6472 89.13	789 10.87	7261
	≥35	904 88.98	112 11.02	1016
twin status	singleton	7819 89.01	965 10.99	8784
	twin	164 81.19	38 18.81	202
distance to next health center	< 10 km	5916 89.95	661 10.05	6577
	> 10 km	2067 85.80	342 14.20	2409
Total		7983	1003	8986

Tabelle 11.1: Übersicht über die Häufigkeitsverteilungen der Kovariablenwerte. Die unteren Zahlen sind die jeweiligen Angaben in %.

Variable	Category	Coefficient	Hazard ratio	95%-CI	p-value
gender	female <sup>1</sup>	0			
	male	0.13	1.14	1.01, 1.29	0.04
ethnic group	Marka <sup>1</sup>	0			
	Bwaba	-0.19	0.83	0.65, 1.05	0.12
	Mossi	-0.18	0.84	0.67, 1.00	0.05
	Peulh	0.14	1.15	0.94, 1.40	0.18
	Samo/other	-0.08	0.93	0.73, 1.18	0.88
religion	muslim <sup>1</sup>	0			
	catholic/protestant	-0.08	0.92	0.76, 1.13	0.45
	natural/other	0.36	1.43	1.06, 1.94	0.02
twin status	singleton <sup>1</sup>	0			
	twin	0.62	1.85	1.34, 2.56	<0.01
age of mother at birth	middle <sup>1</sup>	0			
	young	0.26	1.29	1.05, 1.59	<0.01
	old	-0.08	0.92	0.76, 1.13	0.94
distance to next health center	< 10 <sup>1</sup>	0			
	> 10	0.33	1.39	1.21, 1.59	<0.01

Tabelle 11.2: Ergebnisse des Cox-Modells für die präliminarische Analyse. Dabei bezeichnet <sup>1</sup> die Referenzkategorie.

# Kapitel 12

## R-package zur Schätzung der glatten Komponenten

Der praktische Teil der Arbeit setzt unter anderem die Erstellung einer anwendungsorientierten Software voraus. Die Entwicklung und Implementierung erfolgt mithilfe der nichtkommerziellen Programmierumgebung R (vgl. R Development Core Team, 2006). Die Notwendigkeit der Implementierung ist damit begründet, dass es eine unzureichende Anzahl von entwickelten R-Paketen gibt, welche eine flexible nichtparametrische Modellierung im Bereich der Überlebenszeiten erlauben.

Im Paket `TwoWaySurvival` ist neben dem Modell der Two-Way-Hazards auch das "One-Way-Hazard"-Modell implementiert, wie es bei Kauermann (2005b) beschrieben wird. Die Befehlssyntax erfolgt in Anlehnung an Paket `survival`. Zunächst wird das Objekt der beobachteten evtl. zensierten Überlebenszeiten erzeugt durch:

```
OneWaySurv.object<-OneWaySurv(surv.time,status)
# im Modell mit nur einer Zeitskala

TwoWaySurv.object<-TwoWaySurv(surv.time, birth.time, status)
# im Modell mit zwei Zeitmessungen.
```

Anschließend wird zum Schätzen die jeweilige Hauptroutine aufgerufen durch:

```
OneWaySurvfit.object<-OneWaySurvfitCreate(formula=formula,
data=data, na.action=na.fail, control=control, basis=basis)
```

bzw.

```
TwoWaySurvfit.object<-TwoWaySurvfitCreate(formula=formula,
data=data, na.action=na.fail, control=control,
component=component)
```

Dabei ist das `formula`-Argument zu ersetzen durch  $y \sim x$  in Standardspezifikation, wobei der linke Teil  $y$  dem Objekt `OneWaySurv.object` bzw. dem Objekt `TwoWaySurv.object` entspricht. Der rechte Teil  $x$  ist durch die Kovariablen



im Modell zu spezifizieren. Das `control`-Argument steuert die vom Anwender zusätzlich zu spezifizierenden Optimierungsparameter, die in diesem Fall die Standardparameter ersetzen. So kann man die Anzahl der Integrationspunkte (für die Approximation der integrierten Hazard) durch `number.int` ändern, was den Umfang der damit erzeugten Poisson-Daten erheblich beeinflusst und eventuell verlängernd auf die Rechenzeit wirkt. Außerdem können die Anzahl der jeweiligen Iterationen sowie die Genauigkeiten der Schätzer gesteuert werden. Ein Monitoring der Optimierungsschritte ist ebenfalls möglich.

Das `basis`-Argument im Aufruf von `OneWaySurvfitCreate` bestimmt die Wahl der penalisierten Splinebasis, entweder B-Splines 3-ten Grades, durch das Setzen `basis='bspline'`, oder truncated polynomials 1-ten Grades, durch Spezifikation `basis='trunc'`. Im Aufruf von `TwoWaySurvfitCreate` hängt die Wahl der penalisierten Splines mit der Bestimmung der zweiten zeitlichen Komponente zusammen; Im Falle `component='non.periodic'`, wenn z. B. die Kalenderzeit mit in das Modell genommen wird, werden die truncated polynomials gewählt, im Falle `component='periodic'`, z. B. beim Modellieren der saisonalen Komponente, werden die B-Splines gesetzt. (Diese Auswahl der Splines kann vom Anwender nicht geändert werden.)

Als Ergebnis der Schätzung resultieren die Listen-Objekte der jeweiligen Klassen `OneWaySurvfit` bzw. `TwoWaySurvfit`, die die Information über die Schätzung sowie den Optimierungsstatus enthalten. (Näheres kann man den Hilfeseiten zu `OneWaySurvfitObject` bzw. `TwoWaySurvfitObject` entnehmen.) Jedes Objekt der jeweiligen Klasse hat die generischen Methoden `summary`, `print` und `plot` zur übersichtlichen Ausgabe sowie zum Plotten der geschätzten glatten Komponenten.

Das Modell der Competing Risks ist im Paket `CompetingRiskFrailty` implementiert. In analoger Weise wird zunächst das Objekt der Verweildauern erzeugt:

```
CompetingRiskFrailtySurv(ID=ID,surv.time=surv.time,  
status=status)
```

mit durch ID identifizierten Clustern. Das `status`-Argument ist eine Matrix oder *data frame* mit der jeweiligen Spalte als Event-Indikator mit {0,1}-Kodierung. Der Aufruf von

```
CompetingRiskFrailtySurvfitCreate(formula=formula, data=data,  
na.action=na.fail, control=control, risk.names=risk.names)
```

bewirkt die Schätzung der Parameter via EM-Algorithmus, wie in Kapitel 8 beschrieben. Zusätzlich zu dem zwingenden Argument `formula` sollen auch die Risiken, also Austrittszustände durch `risk.names=Namen` spezifiziert werden. Die Schätzung erfolgt mit truncated polynomials 1-Grades. Das Ergebnis des Fitting ist im Objekt der Klasse `CompetingRiskFrailtySurvfit` gespeichert und hat die generischen Methoden für das Printen und Plotten.

Die beiden Pakete befinden sich in ständiger Weiterentwicklung. Für alle kritischen Anmerkungen und Verbesserungsvorschläge sind deren Entwickler ausgesprochen dankbar. Es wird auch beabsichtigt, die beiden Pakete zu einem Programm zusammenzuführen, in dem sowohl ein Modell mit Frailties als auch eins ohne Heterogenitäten geschätzt werden kann. Es wird auch eine Möglichkeit zur Implementierung einer Wechselwirkung zwischen zwei Zeitachsen eingeräumt, indem z. B. die Interaktion zwischen Überlebenszeit und Kalenderzeit bzw. Überlebenszeit und saisonaler Zeitkomponente berücksichtigt wird (siehe Kapitel 7).

# Anhang A

## Mathematische und statistische Basics

### A.1 Grundlegende Begriffe der Analyse von Verweildauern

Als **Survivor-Funktion** für die Zufallsvariable  $T$  mit nichtnegativem Wertebereich definiert man die Überlebenszeitfunktion (oder Survivorfunktion)  $S(t) = 1 - F(t) = P(T \geq t)$ , wobei  $F(\cdot)$  die Verteilungsfunktion von  $T$  bezeichnet. Als **Hazardrate** definiert man

$$h(t; \mathbf{x}) = \lim_{\Delta t \rightarrow 0} \frac{P(t < T \leq t + \Delta t \mid T \geq t; \mathbf{x})}{\Delta t} \quad (\text{A.1})$$

und interpretiert als relative (oder auch Grenzwert-) Wahrscheinlichkeit bzw. Chance (oder Risiko), im Intervall  $(t, t + \Delta t]$  ein (interessierendes) Event zu beobachten, unter der Bedingung, dass bis zum Erreichen des Intervalls kein Event stattgefunden hat. Dabei ist  $\mathbf{x}$  der Kovariablenvektor. Leicht festzustellen sind die folgenden Beziehungen zwischen  $h(\cdot)$ ,  $S(\cdot)$  und der Dichte  $f(\cdot)$ :

$$h(t) = \frac{f(t)}{S(t)}, \quad S(t) = \exp\left\{-\int_0^t h(u) du\right\} = \exp\{H(t)\}, \quad (\text{A.2})$$

wobei hier  $H(\cdot) = \int_0^t h(u) du$  als **kumulative** Hazardfunktion (oder integrierte Hazardrate) bezeichnet wird.

Unter **Zensierung** versteht man allgemein die Problematik der vollständigen Beobachtungsdauer von Untersuchungseinheiten im gegebenen Untersuchungszeitrahmen. Im Fall der rechtszensierten Daten sind dies die "lost-to-follow-up"-Beobachtungen, die aus der Studie ausscheiden, bevor sie ein Event des Studieninteresses erfahren, oder auch diejenigen, die die vorgegebene Untersuchungszeit überleben (Zensierung vom Typ I). Weiter unterscheidet man zwischen **informativem** und **nichtinformativem** Zensierungsmechanismus, der der "Generierung" von zensierten Daten zugrunde liegt. Wir beschäftigen

uns nur mit nichtinformativer Zensierung und verweisen im ersten Fall auf Kalbfleisch & Prentice (2002), Abschnitt 6.2. Eine **Risikomenge** zum bestimmten Zeitpunkt  $t$  bezeichnet den Vorrat an Beobachtungen, die bis zu diesem Zeitpunkt noch kein Event des Interesses erfahren haben und in der Studie enthalten sind.

Die allgemeine Gestalt der Likelihoodfunktion beinhaltet mögliche Zensierungstypen sowie Bedingungen an Beobachtungen für die Aufnahme in die Studie, so genannte "truncated data", ein Konzept, das im Weiteren nicht verfolgt wird (siehe hierzu Klein & Moeschberger, 1997, für Definition und Beispiele). Im speziellen Fall der rechtszensierten Daten mit nichtinformativer Zensierung erhält man die Likelihood

$$\mathcal{L} = \prod_{i=1}^n h(t_i; \mathbf{x}_i)^{\delta_i} \exp\left\{-\int_0^{t_i} h(u; \mathbf{x}_i) du\right\}, \quad (\text{A.3})$$

ausgedrückt in Abhängigkeit von der Hazard  $h(\cdot)$ .

## A.2 Mathematische Funktionen

Die **Gamma-Funktion** ist definiert als

$$\Gamma(x) = \int_0^\infty t^x e^{-t} dt, \quad x > 0.$$

Es gilt die Relation  $\Gamma(x + 1) = x\Gamma(x)$ . Daraus kann man für ganzzahlige  $k = 0, 1, \dots$  berechnen:

$$\frac{\Gamma(x + k)}{\Gamma(x)} = \prod_{i=0}^{k-1} (x + i),$$

und

$$\log \left\{ \frac{\Gamma(x + k)}{\Gamma(x)} \right\} = \sum_{i=0}^{k-1} \log(x + i).$$

Eine alternative approximative Berechnung des Quotienten ergibt sich aus der Stirling-Approximation (siehe z. B. Abramovitz & Stegun, 1965)

$$\Gamma(x) = e^{-x} x^{x-1/2} \sqrt{2\pi} \left[ 1 + \frac{1}{12x} + \frac{1}{288x^2} - \frac{139}{51\,840x^3} - \frac{571}{2\,488\,320x^4} + O(x^{-5}) \right],$$

die für große Werte von  $x$  (etwa für  $x > 10$ ) geeignet ist.

Für die Vektoren  $\mathbf{a}$  und  $\mathbf{b}$  der Dimension  $n$  sei die Operation  $\odot$  wie folgt definiert:

$$\mathbf{a} \odot \mathbf{b} = [a_1 b_1, \dots, a_n b_n]^T \quad (\text{Hadamard Produkt}). \quad (\text{A.4})$$

Analog kann man diese Operation auf die Matrizen gleicher Dimensionen übertragen. Mit einer notationellen Übereinkunft kann diese Operation für eine Matrix  $\mathbf{A} = [\mathbf{A}_1, \dots, \mathbf{A}_m] \in \mathbb{R}^{n,m}$  und einen Vektor  $\mathbf{x} \in \mathbb{R}^n$  definiert werden:

$$\mathbf{A} \odot \mathbf{x} = [\mathbf{A}_1 \odot \mathbf{x}, \dots, \mathbf{A}_m \odot \mathbf{x}].$$

### A.3 Vektor- und Matrixableitungen

Hier werden einige im Weiteren angewandte Relationen formal angegeben. Zu beachten ist im vektorwertigen Fall die korrekte Anordnung der Matrix- bzw. Vektorelemente.

Es seien  $\mathbf{x}$  ein Vektor der Länge  $n$  und  $\mathbf{y}$  ein Vektor der Länge  $m$ . Wir definieren formal den Ableitungs-Operator:

$$\frac{\partial \mathbf{x}}{\partial \mathbf{y}^T} = \mathbf{x} \otimes \frac{\partial}{\partial \mathbf{y}^T} = \begin{bmatrix} \frac{\partial x_1}{\partial y_1} & \dots & \frac{\partial x_1}{\partial y_m} \\ \dots & \dots & \dots \\ \frac{\partial x_n}{\partial y_1} & \dots & \frac{\partial x_n}{\partial y_m} \end{bmatrix}.$$

Es seien außerdem  $\mathbf{A}$  eine konstante Matrix passender Dimension,  $f(\mathbf{x})$  eine skalare Funktion sowie  $\mathbf{f}(\mathbf{x})$  eine vektorwertige Funktion. Der Vektor  $\mathbf{x}$  sei funktional vom Vektor  $\boldsymbol{\xi}$  abhängig. Dann gelten formal die folgenden Beziehungen (diese sind auch leicht nachzuprüfen):

$$\begin{aligned} \frac{\partial \mathbf{A}\mathbf{x}}{\partial \mathbf{x}^T} &= \mathbf{A}, \\ \frac{\partial f(\boldsymbol{\xi})}{\partial \boldsymbol{\xi}} &= \left( \frac{\partial \mathbf{x}^T}{\partial \boldsymbol{\xi}} \right) \frac{\partial f(\mathbf{x})}{\partial \mathbf{x}} = \left( \frac{\partial \mathbf{x}}{\partial \boldsymbol{\xi}^T} \right)^T \frac{\partial f(\mathbf{x})}{\partial \mathbf{x}}, \\ \frac{\partial \mathbf{f}^T(\boldsymbol{\xi})}{\partial \boldsymbol{\xi}} &= \left( \frac{\partial \mathbf{x}^T}{\partial \boldsymbol{\xi}} \right) \left( \frac{\partial \mathbf{f}^T(\mathbf{x})}{\partial \mathbf{x}} \right), \\ \frac{\partial^2 f(\boldsymbol{\xi})}{\partial \boldsymbol{\xi} \partial \boldsymbol{\xi}^T} &= \left( \frac{\partial \mathbf{x}}{\partial \boldsymbol{\xi}^T} \right)^T \frac{\partial^2 f(\mathbf{x})}{\partial \mathbf{x} \partial \mathbf{x}^T} \left( \frac{\partial \mathbf{x}}{\partial \boldsymbol{\xi}^T} \right) + \sum_{r=1}^{\dim(\mathbf{x})} \frac{\partial^2 \mathbf{x}}{\partial \boldsymbol{\xi} \partial \boldsymbol{\xi}^T} \frac{\partial f(\mathbf{x})}{\partial x_r}. \quad (\text{A.5}) \end{aligned}$$

Im Fall eines lokalen Extremums an der Stelle  $\mathbf{x}$  verschwindet die Summe in (A.5); für die *log*-Likelihoodfunktion  $\ell$  anstelle von  $f(\cdot)$  gilt dann mit dem Übergang zum Erwartungswert in (A.5):

$$\mathbb{E} \left[ \frac{\partial^2 \ell(\boldsymbol{\xi})}{\partial \boldsymbol{\xi} \partial \boldsymbol{\xi}^T} \right] = \left( \frac{\partial \mathbf{x}}{\partial \boldsymbol{\xi}^T} \right)^T \frac{\partial^2 \ell(\mathbf{x})}{\partial \mathbf{x} \partial \mathbf{x}^T} \left( \frac{\partial \mathbf{x}}{\partial \boldsymbol{\xi}^T} \right).$$

(Hier sind dann  $\mathbf{x}$  und  $\boldsymbol{\xi}$  nichtstochastische Modellparameter mit funktionaler Abhängigkeit  $\mathbf{x} = \mathbf{x}(\boldsymbol{\xi})$ .)

Für die Ableitung des Logarithmus der Determinanten einer (nichtsingulären) Matrix  $\mathbf{A}$  gilt (vgl. auch A.7)

$$\frac{\partial \log |\mathbf{A}(\lambda)|}{\partial \lambda} = \text{tr} \left[ \mathbf{A}^{-1} \frac{\partial \mathbf{A}}{\partial \lambda} \right]. \quad (\text{A.6})$$

Als Referenz für die Kalkulation der Matrixableitungen sei auf Harville (1997) und Rao & Rao (1998) verwiesen (siehe auch Rogers, 1980).

## A.4 Verteilungen

(1) Die Poisson-verteilte Zufallsvariable  $\mathbf{x}$  hat die diskrete Verteilungsdichte:

$$f(x) = P(\mathbf{x} = x) = \frac{\xi^x}{x!} e^{-\xi}, \quad x = 0, 1, \dots$$

mit Intensitätsparameter  $\xi$ . Es gilt  $E(\mathbf{x}) = \text{Var}(\mathbf{x}) = \xi$ . Umfassend werden die Poisson-Verteilung und darauf basierende Poisson-Prozesse in Haight (1967) und Kingman (1993) dargestellt.

(2) Die Dichtefunktion eines  $p$ -variaten normalverteilten Vektors  $\mathbf{x}_p$  mit Erwartungswert  $\boldsymbol{\mu}$  und Kovarianzmatrix  $\boldsymbol{\Sigma}$  ist gegeben durch

$$f(\mathbf{x}; \boldsymbol{\mu}, \boldsymbol{\Sigma}) = \frac{1}{(2\pi)^{\frac{p}{2}} |\boldsymbol{\Sigma}|^{1/2}} \exp \left\{ -\frac{1}{2} (\mathbf{x} - \boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1} (\mathbf{x} - \boldsymbol{\mu}) \right\},$$

wobei  $|\cdot|$  die Determinante der Matrix bezeichnet. Hier wird angenommen, dass die Kovarianzmatrix positiv definit ist, so dass die Verteilung nicht ausgeartet ist. Es gibt auch die Möglichkeit, die Definition auf den Fall einer positiv semidefiniten Matrix zu erweitern (siehe z. B. Mardia, Kent & Bibby, 1979).

(3) Die univariate zweiparametrische Gamma-Verteilungsdichte hat die Form

$$f(x; \alpha, \beta) = \frac{\beta^\alpha}{\Gamma(\alpha)} x^{\alpha-1} \exp\{-\beta x\},$$

mit dem Shape-Parameter  $\alpha$  und Scale-Parameter  $\beta$ . Der Erwartungswert bestimmt sich als Quotient  $\mu = \alpha/\beta$  und die Varianz  $\sigma^2 = \alpha/\beta^2$ . Damit lässt sich die Gamma-Dichte eindeutig durch die Festlegung von  $\mu$  und  $\sigma^2$  als  $f_\Gamma(\mu, \sigma^2)$  spezifizieren. Diese Notation benutzen wir kontinuierlich im Model der Competing Risks. Falls  $\mu = 1$  ist, gilt  $\alpha = \beta$  und  $\sigma^2 = 1/\alpha$ .

(4) Allgemeine Form der Mischverteilungsdichte eines Zufallsvektors  $\mathbf{x}$  ist gegeben durch

$$f(\mathbf{x}) = \sum_{i=1}^I \pi_i f_i(\mathbf{x}; \boldsymbol{\theta}_i), \quad \sum_{i=1}^I \pi_i = 1, \quad 0 \leq \pi_i \leq 1.$$

Hier heißen die Dichten  $f_i(\cdot; \boldsymbol{\theta}_i)$  ("Mixture-") Dichtekomponenten, und die Faktoren  $\pi_i$  sind die Gewichte. In der Regel wird die Form der Dichtekomponenten bis auf die Parameter  $\boldsymbol{\theta}_i$  spezifiziert und die letzten sowie die Gewichte  $\pi_i$  aus den Daten geschätzt. Deren Anzahl  $I$  kann ebenfalls geschätzt werden. In unserer Modellierung setzen wir die Anzahl der Komponenten sowie die parametrische Form der Dichten fest und schätzen nur die Gewichte. Für die zahlreichen Anwendungen der Mischverteilungen sei auf Titterington, Smith & Makov (1985) und McLachlan & Peel (2000) verwiesen.

## A.5 Laplace-Approximation

Es sei  $f(\mathbf{x})$  eine unimodale zweimal stetig differenzierbare Funktion auf  $\mathbb{R}^p$  mit der Maximumstelle  $\hat{\mathbf{x}}$ . Wir entwickeln  $f(\mathbf{x})$  um  $\hat{\mathbf{x}}$ :

$$f(\mathbf{x}) \doteq f(\hat{\mathbf{x}}) + \frac{1}{2}(\mathbf{x} - \hat{\mathbf{x}})^T \frac{\partial^2 f(\hat{\mathbf{x}})}{\partial \mathbf{x} \partial \mathbf{x}^T} (\mathbf{x} - \hat{\mathbf{x}}).$$

Dann gilt approximativ:

$$\log \int_{\mathbb{R}^p} \exp\{f(\mathbf{x})\} d\mathbf{x} \doteq f(\hat{\mathbf{x}}) + \frac{p}{2} \log 2\pi - \frac{1}{2} \log \left| -\frac{\partial^2 f(\hat{\mathbf{x}})}{\partial \mathbf{x} \partial \mathbf{x}^T} \right|.$$

Die Approximation ist umso besser, je "konzentrierter" die Funktion  $f(\mathbf{x})$  um  $\hat{\mathbf{x}}$  ist. Im Fall der Likelihoodfunktion anstelle von  $f(\cdot)$  hängt die Approximation von der Stichprobengröße ab. Weitere Details bezüglich der Genauigkeit der Approximation geben Tierney & Kadane (1986).

## A.6 Laplace-Approximation mit nichtinvertierbarer Penalisierungsmatrix

Wir betrachten die penalisierte *log*-Likelihood

$$\ell_p(\mathbf{b}, \lambda) = \ell(\mathbf{b}) - \frac{1}{2} \lambda \mathbf{b}^T \mathbf{D} \mathbf{b},$$

mit nichtinvertierbarer Matrix  $\mathbf{D} = \mathbf{L}^T \mathbf{L}$  und  $\dim(\mathbf{L}) = (K - p, K)$ . (Im Fall der B-Splines ist  $\mathbf{L}$  die Differenzenmatrix und  $p$  die Differenzenordnung). Für den transformierten Vektor  $\mathbf{u} = \mathbf{L} \mathbf{b}$  sei die Normalverteilung unterstellt:

$$\mathbf{u} = \mathbf{L} \mathbf{b} \sim \mathcal{N} \left( 0, \frac{1}{\lambda} \mathbf{I}_{K-p} \right),$$

mit  $\mathbf{I}_{K-p}$  als Identitätsmatrix der Dimension  $K - p$ . Wir komplettieren  $\mathbf{u}$  mit  $p$  linear unabhängigen fixen Parametern  $\mathbf{u}_0$ :

$$\tilde{\mathbf{u}} = \begin{pmatrix} \mathbf{u}_0 \\ \mathbf{u} \end{pmatrix} = \begin{pmatrix} \mathbf{L}_0 \\ \mathbf{L} \end{pmatrix} \mathbf{b} = \tilde{\mathbf{L}} \mathbf{b},$$

so dass  $\tilde{\mathbf{L}} \in \mathbb{R}^{K \times K}$  invertierbar ist. Die Penalisierung  $\frac{1}{2}\lambda \mathbf{b}^T \mathbf{D} \mathbf{b}$  wird zu  $\frac{1}{2}\lambda \mathbf{u}^T \mathbf{u}$ . Damit ist (wegen der Bijektivität von  $\tilde{\mathbf{L}}$ ):

$$\ell_p(\mathbf{b}, \lambda) = \ell(\tilde{\mathbf{u}}) - \frac{1}{2}\lambda \mathbf{u}^T \mathbf{u} = \ell_p(\tilde{\mathbf{u}}, \lambda),$$

mit Penalisierung auf die letzten  $K - p$  Komponenten von  $\tilde{\mathbf{u}}$ . Um nun die marginale Likelihoodfunktion, abhängig von den Parametern  $\mathbf{u}_0$  und  $\lambda$ , zu bestimmen, integrieren wir die zufällige Komponente  $\mathbf{u}$  in der Likelihood  $\mathcal{L}(\tilde{\mathbf{u}})$  aus. Es ist

$$\mathcal{L}_{\text{marg}}(\mathbf{u}_0, \lambda) = \int \mathcal{L}(\tilde{\mathbf{u}}) d\mathbf{u} = \int \mathcal{L}(\mathbf{u}_0|\mathbf{u}) f(\mathbf{u}) d\mathbf{u}.$$

Hier ist  $\mathcal{L}(\mathbf{u}_0|\mathbf{u})$  die Likelihood der Daten, abhängig vom Parameter  $\mathbf{u}_0$  gegeben der Wert von  $\mathbf{u}$ , sowie  $f(\mathbf{u}) = (2\pi)^{-\frac{K-p}{2}} |\frac{1}{\lambda} \mathbf{I}_{K-p}^{-1}|^{-\frac{1}{2}} \exp\{-\frac{1}{2}\lambda \mathbf{u}^T \mathbf{u}\}$ . Nun wird das Integral durch seine Laplace-Approximation ersetzt. Es gilt also für die marginale *log*-Likelihood (vgl. A.5):

$$\ell_{\text{marg}}(\mathbf{u}_0, \lambda) = \ell_p(\mathbf{u}_0, \hat{\mathbf{u}}; \lambda) + \frac{K-p}{2} \log \lambda - \frac{1}{2} \log \left| -\frac{\partial^2 \ell_p(\mathbf{u}_0, \hat{\mathbf{u}}; \lambda)}{\partial \mathbf{u} \partial \mathbf{u}^T} \right|.$$

Die Maximierung bezüglich  $\mathbf{u}_0$  in diesem Ausdruck liefert:

$$\ell_{\text{marg,max}}(\lambda) = \ell_p(\hat{\mathbf{b}}; \lambda) + \frac{K-p}{2} \log \lambda - \frac{1}{2} \log \left| -\frac{\partial^2 \ell_p(\hat{\mathbf{b}}; \lambda)}{\partial \mathbf{u} \partial \mathbf{u}^T} \right|.$$

Die 2. Ableitung bezüglich  $\mathbf{u}$  wird nach der Transformationsregel (A.5) bestimmt:

$$\frac{\partial^2 \ell_p(\mathbf{b}; \lambda)}{\partial \mathbf{u} \partial \mathbf{u}^T} = \left( \frac{\partial \mathbf{b}^T}{\partial \mathbf{u}} \right) \frac{\partial^2 \ell_p(\mathbf{b})}{\partial \mathbf{b} \partial \mathbf{b}^T} \left( \frac{\partial \mathbf{b}}{\partial \mathbf{u}^T} \right) + \sum_{r=1}^{\dim(\mathbf{b})} \frac{\partial^2 \mathbf{b}}{\partial \mathbf{u} \partial \mathbf{u}^T} \frac{\partial \ell(\mathbf{b})}{\partial b_r}.$$

Der zweite Summen-Term ist identisch mit 0, da, wegen der Abhängigkeit  $\mathbf{b} = \tilde{\mathbf{L}}^{-1} \tilde{\mathbf{u}}$  und  $\tilde{\mathbf{u}} = \tilde{\mathbf{u}}(\mathbf{u})$ , mit den Formeln aus A.3 gilt:

$$\frac{\partial \mathbf{b}^T}{\partial \mathbf{u}} = \frac{\partial \tilde{\mathbf{u}}^T}{\partial \mathbf{u}} \frac{\partial \mathbf{b}^T}{\partial \tilde{\mathbf{u}}} = \left( \frac{\partial \mathbf{u}_0^T}{\partial \mathbf{u}}, \frac{\partial \mathbf{u}^T}{\partial \mathbf{u}} \right) \left( \tilde{\mathbf{L}}^{-1} \right)^T = (\mathbf{0}_{K-p,p}, \mathbf{I}_{K-p}) \left( \tilde{\mathbf{L}}^{-1} \right)^T.$$

Damit ergibt sich:

$$\begin{aligned} -\frac{\partial^2 \ell_p(\hat{\mathbf{b}}; \lambda)}{\partial \mathbf{u} \partial \mathbf{u}^T} &= (\mathbf{0}_{K-p,p}, \mathbf{I}_{K-p}) \left( \tilde{\mathbf{L}}^{-1} \right)^T \left( -\frac{\partial^2 \ell_p(\hat{\mathbf{b}}; \lambda)}{\partial \mathbf{b} \partial \mathbf{b}^T} \right) \tilde{\mathbf{L}}^{-1} \begin{pmatrix} \mathbf{0}_{p,K-p} \\ \mathbf{I}_{K-p} \end{pmatrix} \\ &= (\mathbf{0}_{K-p,p}, \mathbf{I}_{K-p}) \left( \tilde{\mathbf{L}}^{-1} \right)^T \mathcal{I}_p(\mathbf{b}, \lambda) \tilde{\mathbf{L}}^{-1} \begin{pmatrix} \mathbf{0}_{p,K-p} \\ \mathbf{I}_{K-p} \end{pmatrix} =: \mathbf{F}(\mathbf{b}, \lambda) \end{aligned}$$

mit der penalisierten Fisher-Matrix  $\mathcal{I}_p(\mathbf{b}, \lambda)$ . Die an der Stelle  $\hat{\mathbf{u}}$  maximierte marginale *log*-Likelihood  $\ell_{\text{marg,max}}(\lambda)$  wird nun nach  $\lambda$  abgeleitet. Wegen der Formel (A.6) haben wir mit der Approximation  $\frac{\partial \mathbf{F}(\hat{\mathbf{b}}, \lambda)}{\partial \lambda} \approx \mathbf{I}_{K-p}$ :

$$\begin{aligned} \frac{\partial \ell_{\text{marg,max}}(\lambda)}{\partial \lambda} &= -\frac{1}{2} \hat{\mathbf{b}}^T \mathbf{L}^T \mathbf{L} \hat{\mathbf{b}} + \frac{K-p}{2} \frac{1}{\lambda} - \frac{1}{2} \text{tr} \left\{ \mathbf{F}^{-1}(\hat{\mathbf{b}}, \lambda) \frac{\partial \mathbf{F}(\hat{\mathbf{b}}, \lambda)}{\partial \lambda} \right\} \\ &= \frac{1}{2} \hat{\mathbf{b}}^T \mathbf{L}^T \mathbf{L} \hat{\mathbf{b}} + \frac{K-p}{2} \frac{1}{\lambda} - \frac{1}{2} \text{tr} \left\{ \mathbf{F}^{-1}(\hat{\mathbf{b}}, \lambda) \right\}. \end{aligned}$$



Damit bestimmt sich  $\lambda$  iterativ durch

$$\frac{1}{\hat{\lambda}} = \frac{\hat{\mathbf{b}}^T \mathbf{D} \hat{\mathbf{b}} + \text{tr}\{\mathbf{F}^{-1}(\hat{\mathbf{b}}, \lambda)\}}{K - p}.$$

Mit der zusätzlichen Approximation:

$$\begin{aligned} \text{tr}\{\mathbf{F}^{-1}(\hat{\mathbf{b}}, \lambda)\} &\approx \text{tr}\left\{\left[\left(\tilde{\mathbf{L}}^{-1}\right)^T \mathcal{I}_p(\mathbf{b}, \lambda) \tilde{\mathbf{L}}^{-1}\right]^{-1}\right\} \\ &\approx \text{tr}\{\mathbf{L}^T \mathbf{L} \mathcal{I}_p^{-1}(\mathbf{b}, \lambda)\} \end{aligned}$$

resultiert die gesuchte Form für die iterative Berechnung von  $\lambda$  als

$$\frac{1}{\hat{\lambda}} = \frac{\hat{\mathbf{b}}^T \mathbf{D} \hat{\mathbf{b}} + \text{tr}\{\mathcal{I}_p^{-1}(\mathbf{b}, \lambda) \mathbf{D}\}}{K - p}.$$

## A.7 I. und II. Ableitungen zur Bestimmung des Smoothing-Parameters

Es gilt für eine quadratische invertierbare Matrix  $\mathbf{A}(\boldsymbol{\lambda})$ , abhängig vom Vektor  $\boldsymbol{\lambda} = [\lambda_1, \dots, \lambda_p]$ :

$$\frac{\partial \log |\mathbf{A}(\boldsymbol{\lambda})|}{\partial \lambda_l} = \text{tr}\left\{\mathbf{A}^{-1}(\boldsymbol{\lambda}) \frac{\partial \mathbf{A}(\boldsymbol{\lambda})}{\partial \lambda_l}\right\}.$$

Der detaillierte Beweis findet sich in Harville (1997).

Es sei nun die Matrix  $\mathbf{A}(\boldsymbol{\lambda}) = \mathbf{I} + \mathbf{D}(\boldsymbol{\lambda})$  mit  $\boldsymbol{\lambda} = [\lambda_1, \dots, \lambda_p]$  und  $\mathbf{D}(\boldsymbol{\lambda}) = \text{diag}[\mathbf{D}_1(\lambda_1), \dots, \mathbf{D}_p(\lambda_p)]$ , wobei  $\mathbf{D}_l(\lambda_l) = \text{diag}[\lambda_l, \dots, \lambda_l]$  der Dimension  $q_l$  ist. Es gilt dann

$$\begin{aligned} \frac{\partial \log |\mathbf{I} + \mathbf{D}(\boldsymbol{\lambda})|}{\partial \lambda_l} &= \text{tr}\left\{[\mathbf{I} + \mathbf{D}(\boldsymbol{\lambda})]^{-1} \frac{\partial \mathbf{D}(\boldsymbol{\lambda})}{\partial \lambda_l}\right\} \\ &= \text{tr}\left\{[\mathbf{I} + \mathbf{D}(\boldsymbol{\lambda})]^{-1} \text{diag}[\mathbf{0}_{q_1}, \dots, \mathbf{1}_{q_l}, \dots, \mathbf{0}_{q_p}]\right\}. \end{aligned}$$

Die entsprechende Partitionierung  $[\mathbf{I} + \mathbf{D}(\boldsymbol{\lambda})]^{-1}$  ergibt folglich

$$\frac{\partial \log |\mathbf{I} + \mathbf{D}(\boldsymbol{\lambda})|}{\partial \lambda_l} = \text{tr}\{[\mathbf{I} + \mathbf{D}(\boldsymbol{\lambda})]_{ll}^{-1}\},$$

wobei der Subindex  $ll$  den  $l$ -ten diagonalen Unterblock der Inversen  $[\mathbf{I} + \mathbf{D}(\boldsymbol{\lambda})]^{-1}$  bezeichnet. Für die multiplikative Gestalt  $\mathbf{D}_l(\lambda_l) = \lambda_l \mathbf{D}_l$  gilt entsprechend

$$\frac{\partial \log |\mathbf{I} + \mathbf{D}(\boldsymbol{\lambda})|}{\partial \lambda_l} = \text{tr}\{[\mathbf{I} + \mathbf{D}(\boldsymbol{\lambda})]_{ll}^{-1} \mathbf{D}_l\}.$$

Hier soll die explizite Skizze zum Kalkül der obigen Form dargestellt werden. Partitionieren wir die Matrizen  $\mathbf{A} = [\mathbf{I} + \mathbf{D}(\boldsymbol{\lambda})]^{-1}$  und  $\mathbf{B} = \frac{\partial \mathbf{D}(\boldsymbol{\lambda})}{\partial \lambda_l}$  jeweils als

$$\mathbf{A} = \begin{bmatrix} \mathbf{A}_1 & \mathbf{A}_2 & \mathbf{A}_3 \\ \mathbf{A}_4 & \mathbf{A}_{ll} & \mathbf{A}_5 \\ \mathbf{A}_6 & \mathbf{A}_7 & \mathbf{A}_8 \end{bmatrix}, \quad \mathbf{B} = \begin{bmatrix} \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{D}_l & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} \end{bmatrix},$$

bekommen wir nach dem Multiplizieren unter Berücksichtigung der Dimensionen

$$\mathbf{AB} = \begin{bmatrix} \mathbf{0} & \mathbf{A}_2 \mathbf{D}_l & \mathbf{0} \\ \mathbf{0} & \mathbf{A}_{ll} \mathbf{D}_l & \mathbf{0} \\ \mathbf{0} & \mathbf{A}_7 \mathbf{D}_l & \mathbf{0} \end{bmatrix},$$

so dass  $\text{tr}[\mathbf{AB}] = \text{tr}[\mathbf{A}_{ll} \mathbf{D}_l]$ .

Trivial ergibt sich weiter:

$$\frac{\partial^2 \log |\mathbf{A}(\boldsymbol{\lambda})|}{\partial \lambda_m \partial \lambda_n} = \text{tr} \left\{ \mathbf{A}^{-1}(\boldsymbol{\lambda}) \frac{\partial^2 \mathbf{A}(\boldsymbol{\lambda})}{\partial \lambda_m \partial \lambda_n} - \mathbf{A}^{-1}(\boldsymbol{\lambda}) \frac{\partial \mathbf{A}(\boldsymbol{\lambda})}{\partial \lambda_m} \mathbf{A}^{-1}(\boldsymbol{\lambda}) \frac{\partial \mathbf{A}(\boldsymbol{\lambda})}{\partial \lambda_n} \right\}.$$

Unter der expliziten Berücksichtigung der Form von  $\mathbf{A}(\boldsymbol{\lambda}) = \mathbf{I} + \mathbf{D}(\boldsymbol{\lambda})$  mit  $\mathbf{D}(\boldsymbol{\lambda})$  als blockdiagonal, wie oben definiert, bekommt man

$$\frac{\partial^2 \log |\mathbf{A}(\boldsymbol{\lambda})|}{\partial \lambda_m \partial \lambda_n} = -\text{tr} \left\{ \mathbf{A}^{-1}(\boldsymbol{\lambda}) \frac{\partial \mathbf{A}(\boldsymbol{\lambda})}{\partial \lambda_m} \mathbf{A}^{-1}(\boldsymbol{\lambda}) \frac{\partial \mathbf{A}(\boldsymbol{\lambda})}{\partial \lambda_n} \right\}.$$

Diese Form lässt sich nicht vereinfachen, da die Strukturen von  $\frac{\partial \mathbf{A}(\boldsymbol{\lambda})}{\partial \lambda_l}$ ,  $l = 1, \dots, p$ , unterschiedlich sind.

## A.8 Grundriss der Bayes-Inferenz

In Anlehnung an Carlin & Louis (2000) sei hier kurz die (empirische) Bayes-Prozedur skizziert. Zunächst ist für den beobachteten Daten-Vektor  $\mathbf{y}$  mit der Dichtefunktion  $f(\mathbf{y}|\boldsymbol{\theta})$ , abhängig vom Parameter-Vektor  $\boldsymbol{\theta}$  mit der priori Dichte  $\pi(\boldsymbol{\theta})$ , die posteriori Dichte von  $\boldsymbol{\theta}$  gegeben durch (die Anwendung des Bayes-Theorems):

$$p(\boldsymbol{\theta}|\mathbf{y}) = \frac{f(\mathbf{y}|\boldsymbol{\theta})\pi(\boldsymbol{\theta})}{\int_{\Theta} f(\mathbf{y}|\boldsymbol{\theta})\pi(\boldsymbol{\theta}) d\boldsymbol{\theta}},$$

wobei das Integral im Ausdruck die marginale Dichte von  $\mathbf{y}$  darstellt und die Rolle einer Normierungskonstante spielt. Als (posteriori) Schätzwert für  $\boldsymbol{\theta}$  kann man z. B. den Erwartungswert der posteriori Verteilung, spezifiziert durch  $p(\boldsymbol{\theta}|\mathbf{y})$ , nehmen. Unter der expliziten Berücksichtigung der Abhängigkeit der Verteilung für  $\boldsymbol{\theta}$  von weiteren so genannten Hyperparametern  $\boldsymbol{\eta}$  lässt sich die posteriori Dichte schreiben als

$$p(\boldsymbol{\theta}|\mathbf{y}, \boldsymbol{\eta}) = \frac{f(\mathbf{y}|\boldsymbol{\theta})\pi(\boldsymbol{\theta}|\boldsymbol{\eta})}{\int_{\Theta} f(\mathbf{y}|\boldsymbol{\theta})\pi(\boldsymbol{\theta}|\boldsymbol{\eta}) d\boldsymbol{\theta}}.$$

Dabei ist das Integral im Ausdruck die marginale Dichte  $m(\mathbf{y}|\boldsymbol{\eta})$  von Daten  $\mathbf{y}$  gegeben Parameter  $\boldsymbol{\eta}$ . Das besondere Augenmerk richtet sich auf die Bestimmung vom (Hyper-)parameter  $\boldsymbol{\eta}$ . Falls dieser nicht bekannt ist, wird er nach dem **empirischen Bayes-Konzept** aus den Daten geschätzt und zwar durch einen (marginalen, z. B. ML-) Schätzer  $\hat{\boldsymbol{\eta}} \equiv \hat{\boldsymbol{\eta}}(\mathbf{y})$  der Verteilung, spezifiziert durch  $m(\mathbf{y}|\boldsymbol{\eta})$ . Diese Schätzung resultiert die geschätzte posteriori Dichte  $p(\boldsymbol{\theta}|\mathbf{y}, \hat{\boldsymbol{\eta}})$ , auf der nun die Inferenz über den Parameter  $\boldsymbol{\theta}$  beruht.

## A.9 Grundriss des EM-Algorithmus

In Anlehnung an McLachlan & Krishnan (1997) und unter der Verwendung derselben Notationen skizzieren wir hier kurz die wesentlichen Komponenten des EM-Algorithmus. Es bezeichne  $\mathbf{y}$  den beobachteten Datenvektor. Der Vektor  $\mathbf{z}$  sei der Vektor der unbeobachteten oder latenten Größen (im weiteren Sinne), so dass der Vektor  $\mathbf{x} = [\mathbf{y}, \mathbf{z}]$  den kompletten Datenvektor bezeichnet. Mit den Notationen  $\mathcal{L}(\boldsymbol{\Psi} | \mathbf{y})$  und  $\mathcal{L}_c(\boldsymbol{\Psi} | \mathbf{x})$  sei die Likelihoodfunktion der beobachteten Größen bzw. der kompletten Daten bezeichnet, wobei  $\boldsymbol{\Psi}$  den zu schätzenden Parametervektor darstellt. Für die so genannte  $Q$ -Funktion, die den bedingten Erwartungswert der kompletten  $\log$ -Likelihood  $\mathcal{L}_c(\boldsymbol{\Psi})$  unter der Kenntnis der beobachteten Daten  $\mathbf{y}$  darstellt, in Notationen also

$$Q(\boldsymbol{\Psi}, \boldsymbol{\Psi}^{(0)}) = E_{\boldsymbol{\Psi}^{(0)}} \{ \log \mathcal{L}_c(\boldsymbol{\Psi}) | \mathbf{y} \}, \quad (\text{A.7})$$

wobei hier  $\boldsymbol{\Psi}^{(0)}$  den bekannten Parameterwert repräsentiert, besteht das Problem des E-Schrittes gerade in der Berechnung dieses Erwartungswertes, also der  $Q$ -Funktion. Das Problem wird zurückgeführt auf die Ermittlung der Erwartung von  $\log \mathcal{L}_c(\boldsymbol{\Psi}^{(0)})$  als stochastische (wegen  $\mathbf{z}$ ) Größe, bei bekanntem  $\mathbf{y}$ . Die vom Parameter  $\boldsymbol{\Psi}^{(0)}$  abhängende  $Q$ -Funktion wird im nächsten Schritt, als M-Schritt bezeichnet, bezüglich  $\boldsymbol{\Psi}^{(0)}$  maximiert, d. h. man wählt den Update-Wert für  $\boldsymbol{\Psi}$  als

$$\boldsymbol{\Psi}^{(1)} = \underset{\boldsymbol{\Psi}}{\operatorname{argmax}} Q(\boldsymbol{\Psi}, \boldsymbol{\Psi}^{(0)}).$$

Das Verfahren ist notwendig iterativ und besitzt die Monotonie-Eigenschaft für die Likelihood der beobachteten Daten, also  $\mathcal{L}(\boldsymbol{\Psi}^{(k+1)}) \geq \mathcal{L}(\boldsymbol{\Psi}^{(k)})$ , so dass (im Fall der Beschränktheit von oben) die Likelihood konvergiert. Die Forderung des M-Schrittes kann geschwächt werden, indem als Update-Wert für  $\boldsymbol{\Psi}$  in der  $k$ -ten Iteration jeder Wert sein kann, für den die Relation  $Q(\boldsymbol{\Psi}^{(k+1)}, \boldsymbol{\Psi}^{(k)}) \geq Q(\boldsymbol{\Psi}^{(k)}, \boldsymbol{\Psi}^{(k)})$  erfüllt ist (GEM-Algorithmus). Die Berechnung der Maximumstelle ist somit nicht notwendig und kann beispielsweise mit einigen wenigen Newton-Raphson Iterationen realisiert werden. In vielen praktischen Anwendungen ist dabei die Limes-Stelle  $\mathcal{L}^*$  auch die stationäre Stelle, d. h.,  $\mathcal{L}^* = \mathcal{L}(\boldsymbol{\Psi}^*)$  mit  $\boldsymbol{\Psi}^*$  als Lösung der Score-Gleichung  $\partial \log \mathcal{L}(\boldsymbol{\Psi}) / \partial \boldsymbol{\Psi} = 0$ , (aber nicht notwendigerweise, siehe dazu Beispiele bei McLachlan & Krishnan, 1997). Im Fall der penalisierten  $\log$ -Likelihood bleiben die Herleitungen gültig, unter der Berücksichtigung des additiven Penalisierungsterms für  $\boldsymbol{\Psi}$ .

Der weitere Schritt besteht in der Berechnung von Kovarianzen des Modellparameters  $\Psi$  an der ML-Stelle. Dafür seien zunächst die Notationen für die (beobachtete) Fisher-Matrix der unvollständigen (also der beobachteten) Daten  $\mathbf{y}$  sowie die (beobachtete) Fisher-Matrix der kompletten Daten  $\mathbf{x}$  eingeführt:

$$\begin{cases} \mathbf{I}(\Psi; \mathbf{y}) = -\frac{\partial^2 \mathcal{L}(\Psi)}{\partial \Psi \partial \Psi^T} & \text{unvollständig,} \\ \mathbf{I}_c(\Psi; \mathbf{x}) = -\frac{\partial^2 \mathcal{L}_c(\Psi)}{\partial \Psi \partial \Psi^T} & \text{komplett.} \end{cases} \quad (\text{A.8})$$

Dann gilt nach Louis (1982) die folgende Formel für die Berechnung von  $\mathbf{I}(\Psi; \mathbf{y})$ :

$$\mathbf{I}(\Psi; \mathbf{y}) = \mathbb{E}_{\Psi} \left\{ \mathbf{I}_c(\Psi; \mathbf{X}) \mid \mathbf{y} \right\} - \text{Cov}_{\Psi} \left\{ \frac{\partial \mathcal{L}_c(\mathbf{X}; \Psi)}{\partial \Psi} \mid \mathbf{y} \right\}, \quad (\text{A.9})$$

wobei an der ML-Stelle  $\hat{\Psi}$  von  $\mathcal{L}(\Psi \mid \mathbf{y})$  gilt:

$$\text{Cov}_{\Psi} \left\{ \frac{\partial \mathcal{L}_c(\mathbf{X}; \hat{\Psi})}{\partial \Psi} \mid \mathbf{y} \right\} = \mathbb{E}_{\Psi} \left\{ \frac{\partial \mathcal{L}_c(\mathbf{X}; \hat{\Psi})}{\partial \Psi} \frac{\partial \mathcal{L}_c(\mathbf{X}; \hat{\Psi})}{\partial \Psi^T} \mid \mathbf{y} \right\}.$$

## A.10 Trapezoide Approximation des Integrals

Für die trapezoide Approximation des Integrals von einer (stetigen) Funktion  $f(\cdot)$  seien zunächst die Gitterpunkte  $\{0 = \tau_0, \tau_1, \dots, \tau_K\}$  geeignet gewählt. Für die Integration von  $0 = \tau_0$  bis  $t$  gibt es folgende Möglichkeiten (vgl. Abbildung 3.1):

- $\tau_{K-1} < t < \tau_K$  für ein  $K$
- $\tau_{K-1} < t = \tau_K$ .

Im ersten Fall gilt für die Approximation des Integrals:

$$\begin{aligned} I &\doteq \frac{\tau_1 - \tau_0}{2} f_0 + \frac{\tau_1 - \tau_0}{2} f_1 + \frac{\tau_2 - \tau_1}{2} f_1 + \frac{\tau_2 - \tau_1}{2} f_2 + \dots + \\ &= \frac{\tau_{K-1} - \tau_{K-2}}{2} f_{K-2} + \frac{\tau_{K-1} - \tau_{K-2}}{2} f_{K-1} + \frac{t - \tau_{K-1}}{2} f_{K-1} + \frac{t - \tau_{K-1}}{2} f_t, \end{aligned}$$

wobei mit  $f_k$  die Stelle  $f(\tau_k)$ ,  $k = 1, \dots, K - 1$ , und mit  $f_t$  die Stelle  $f(t)$  bezeichnet seien. Die Vereinfachung der rechten Seite dieses Ausdrucks liefert:

$$\begin{aligned}
 I &\doteq \frac{\tau_1 - \tau_0}{2} f_0 + \\
 &= \left[ \frac{\tau_2 - \tau_0}{2} f_1 + \frac{\tau_3 - \tau_1}{2} f_2 + \cdots + \frac{\tau_{K-1} - \tau_{K-3}}{2} f_{K-2} + \frac{t - \tau_{K-2}}{2} f_{K-1} + \frac{t - \tau_{K-1}}{2} f_t \right] \\
 &= \frac{\tau_1 - \tau_0}{2} f_0 + \sum_{k=1}^K \frac{\min\{\tau_{k+1}, t\} - \min\{\tau_{k-1}, t\}}{2} f_{\min\{\tau_k, t\}} \\
 &= \sum_{k=0}^K \frac{\min\{\tau_{k+1}, t\} - \min\{\tau_{k-1}, t\}}{2} f_{\min\{\tau_k, t\}} \\
 &=: \sum_{k=0}^K o_k f_{\min\{\tau_k, t\}} = \sum_{k=0}^K \exp[\log\{o_k\} + \log\{f_{\min\{\tau_k, t\}}\}]
 \end{aligned}$$

mit zusätzlicher Notation  $\tau_{-1} = 0 = \tau_0$ .

Im zweiten Fall ( $t = \tau_K$ ) ergibt sich in völliger Analogie eine identische Berechnungsformel.

## A.11 Grundriss der Optimierungsstrategie für mixture-Gewichte in Competing Risks

Wir betrachten das Optimierungsproblem (8.22). Es gibt verschiedene Methoden, dieses typische konvexe Optimierungsproblem zu lösen (allgemeine Formulierung siehe in Boyd & Vandenberghe, 2004), z. B. mittels des so genannten SQP-Programms ("sequential quadratic programming"), indem in jedem Optimierungsschritt ein simpleres approximierendes quadratisches Problem (mit Restriktionen) gelöst wird, oder man kann sich auch der so genannten "barrier-functions"-Methode bedienen, indem die iterative Lösung desto mehr penalisiert wird, je weiter sie von der Restriktion abweicht (siehe dazu Fletcher, 1987, oder Nocedal & Wright, 1999). Im Fall der konvexen Zielfunktion (sowie der Restriktionen) gibt es speziell entwickelte Verfahren (siehe Boyd & Vandenberghe, 2004), da in diesem Fall ein eindeutiges (globales auf dem Definitionsbereich) Extremum existiert. Dabei kann die optimale Stelle am Rande des Definitionsbereichs liegen. Dieser Sachverhalt stellt ein Hindernis für den Fall der scheinbar vereinfachenden Reparametrisierung der Art  $\mathbf{c}_m = \exp(a_m) / \sum_{m=0}^M \exp(a_m)$  dar. Diese Reparametrisierung befreit zwar von den auferlegten Restriktionen, die Maximierung der geänderten Form der log-Likelihood kann aber durchaus ungeeignet für die konventionelle Optimierung via Newton-Raphson sein (man veranschauliche sich dies etwa für  $M = 2$  mit  $c_1 = 1 - c_2$ ). Die Optimierung wurde durchgeführt basierend auf der so genannten "primal-dual interior-point"-Methode, die in Boyd & Vandenberghe (2004) beschrieben ist (siehe dort Abschnitt 11.7 und besonders den Algorithmus in 11.7.3). Für die Berechnungen brauchen wir die 1-te und die

2-te Ableitungen der Zielfunktion und der *inequality*-Restriktion sowie die 1-te Ableitung der *equality*-Restriktion. Für die (zu minimierende) Zielfunktion  $f_0(\mathbf{c}) = -\ell_{\text{pen}}(\mathbf{c} | \boldsymbol{\delta})$  gilt:

$$\nabla_{\mathbf{c}} f_0 = \begin{bmatrix} -\sum_{i=1}^n \bar{\Delta}_{i1} \\ \vdots \\ -\sum_{i=1}^n \bar{\Delta}_{iM} \end{bmatrix} + h_{\mathbf{c}} \boldsymbol{\Sigma} \mathbf{c}, \quad \nabla_{\mathbf{cc}}^2 f_0 = \begin{bmatrix} \sum_i \bar{\Delta}_{i1}^2 & \cdots & \sum_i \bar{\Delta}_{i1} \bar{\Delta}_{iM} \\ \cdots & \cdots & \cdots \\ \sum_i \bar{\Delta}_{iM} \bar{\Delta}_{i1} & \cdots & \sum_i \bar{\Delta}_{iM}^2 \end{bmatrix} + h_{\mathbf{c}} \boldsymbol{\Sigma}.$$

(Wir schreiben hier den Multiindex  $\mathbf{m}$  in der expandierten eindimensionalen Form.) Für die *inequality*-Restriktion der Form  $\mathbf{f} = -\mathbf{c} \leq \mathbf{0}$  gilt mit  $f_m = c_m$ :

$$\mathcal{D}\mathbf{f} = \mathbf{1}\mathbf{1}^T, \quad \nabla_{\mathbf{cc}}^2 f_m = \mathbf{0}, \quad m = 0, \dots, M.$$

Die *equality*-Restriktion  $\sum_m c_m = 1$  ist durch die Form  $\mathbf{A}\mathbf{c} = \mathbf{b}$  gegeben mit der Matrix  $\mathbf{A} = [1, \dots, 1]$  vom Rang 1 und  $\mathbf{b} = 1$ . Man bildet nun die aus der Optimierungsliteratur wohl bekannten KKT-Gleichungen (siehe z. B. Fletcher, 1987):

$$\mathbf{r}_t(\mathbf{c}, \boldsymbol{\lambda}, \boldsymbol{\nu}) := \begin{bmatrix} \nabla_{\mathbf{c}} f_0 + \mathcal{D}\mathbf{f}^T \boldsymbol{\lambda} + \mathbf{A}^T \boldsymbol{\nu} \\ -\text{diag}\{\boldsymbol{\lambda}\} \mathbf{f} - (1/t)\mathbf{1} \\ \mathbf{A}\mathbf{c} - \mathbf{b} \end{bmatrix}, \quad (\text{A.10})$$

wobei  $t > 0$  und  $\boldsymbol{\lambda}, \boldsymbol{\nu}$  die dualen Lösungen der Gleichung  $\mathbf{r}_t(\mathbf{c}, \boldsymbol{\lambda}, \boldsymbol{\nu}) = \mathbf{0}$  mit dem Dualschritt proportional zu  $1/t$ . Der Newton-Schritt  $\boldsymbol{\Delta} = (\Delta\mathbf{c}, \Delta\boldsymbol{\lambda}, \Delta\boldsymbol{\nu})$  resultiert aus  $\mathbf{r}_t(\mathbf{c} + \Delta\mathbf{c}, \boldsymbol{\lambda} + \Delta\boldsymbol{\lambda}, \boldsymbol{\nu} + \Delta\boldsymbol{\nu}) \approx \mathbf{r}_t(\mathbf{c}, \boldsymbol{\lambda}, \boldsymbol{\nu}) + \mathcal{D}\mathbf{r}_t(\mathbf{c}, \boldsymbol{\lambda}, \boldsymbol{\nu})\boldsymbol{\Delta} = \mathbf{0}$  als

$$\begin{bmatrix} \nabla_{\mathbf{cc}}^2 f_0 + \sum_m \lambda_i \nabla_{\mathbf{cc}}^2 f_m & \mathcal{D}\mathbf{f}^T & \mathbf{A}^T \\ -\text{diag}\{\boldsymbol{\lambda}\} \mathcal{D}\mathbf{f} & -\text{diag}\{\mathbf{f}\} & \mathbf{0} \\ \mathbf{A} & \mathbf{0} & \mathbf{0} \end{bmatrix} \begin{bmatrix} \Delta\mathbf{c} \\ \Delta\boldsymbol{\lambda} \\ \Delta\boldsymbol{\nu} \end{bmatrix} = -\mathbf{r}_t(\mathbf{c}, \boldsymbol{\lambda}, \boldsymbol{\nu}). \quad (\text{A.11})$$

Man beginnt mit dem Initialisierungsschritt (der nicht unbedingt die Restriktionen erfüllt) und iteriert, indem der Wert für  $t$  erhöht wird (multipliziert z. B. mit 10 oder 100, siehe Boyd & Vandenberghe, 2004, Abschnitte 11.7.2-3) und der Newton-Schritt ermittelt wird, bis der Update-Wert klein genug wird. Für die numerisch effiziente Berechnung des Schrittes in (A.11) kann z. B. Harville (1997), Seite 468, konsultiert werden.

# Anhang B

## Simulationsergebnisse

### B.1 Simulation mit truncated polynomials

Die folgenden glatten Funktionen werden simuliert:

$$\begin{aligned}\alpha_0^t(t) &= \eta - \frac{t}{30}, \\ \alpha_0^b(b) &= 3 \left[ \left( \frac{b}{50} \right)^2 - 0.833 \left( \frac{b}{50} \right) \right], \\ \alpha_x^t(t) &= 0.5 + \left( \frac{t}{25} \right)^2, \\ \alpha_x^b(b) &= -\frac{b}{60},\end{aligned}$$

wobei der Parameter  $\eta$  im Ausdruck für die Baseline-Hazard in  $t$ -Richtung gerade die relative Häufigkeit der zensierten Beobachtungen steuert. Die folgenden eingesetzten Werte für  $\eta$  entsprechen den Zensierungsmustern  $h_\delta$ :

$$\begin{aligned}\eta = -1.4 &\rightarrow h_\delta \simeq 2\%, \\ \eta = -2 &\rightarrow h_\delta \simeq 10\%, \\ \eta = -2.5 &\rightarrow h_\delta \simeq 25\%, \\ \eta = -3.3 &\rightarrow h_\delta \simeq 50\%, \\ \eta = -4.2 &\rightarrow h_\delta \simeq 75\%, \\ \eta = -5.1 &\rightarrow h_\delta \simeq 90\%.\end{aligned}$$

Wir setzen folgendes Verteilungsschema:  $N = 100, 500, 1000$  Beobachtungen für den jeweiligen Anteil zensierter Beobachtungen 2%, 10%, 25%, 50%, 75%, 90%. Als Output der Simulation sind die *coverage probabilities* sowie die geschätzten empirischen 5%, 50%, 95%-Quantile der simulierten glatten Komponenten wiedergegeben. (Ausführlich ist das Simulationsverfahren in Kapitel 5 beschrieben.)

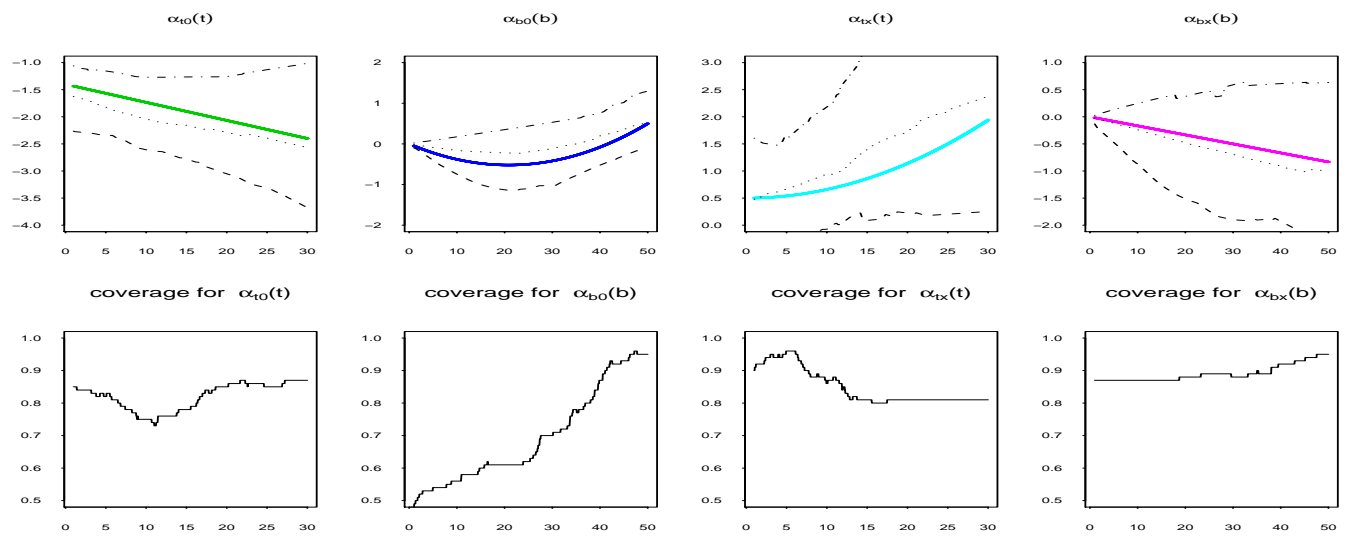


Abbildung B.1: Simulierte glatte Komponenten und coverage probabilities. 100 Beobachtungen, davon ca. 2% zensiert.

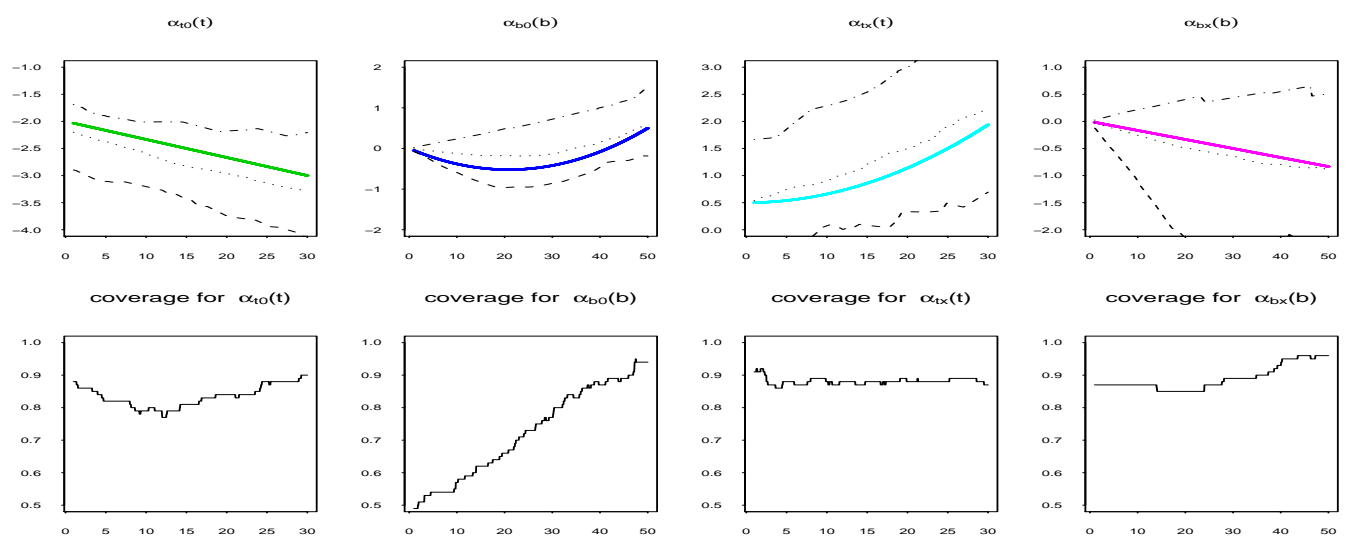


Abbildung B.2: Simulierte glatte Komponenten und coverage probabilities. 100 Beobachtungen, davon ca. 10% zensiert.



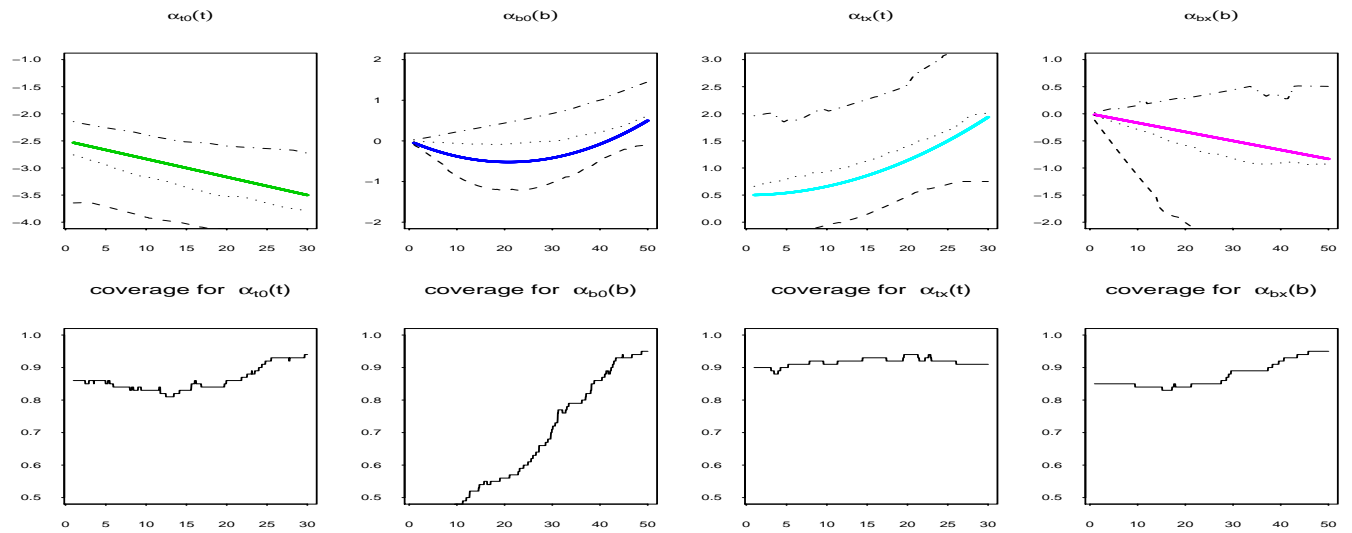


Abbildung B.3: Simulierte glatte Komponenten und coverage probabilities. 100 Beobachtungen, davon ca. 25% zensiert.

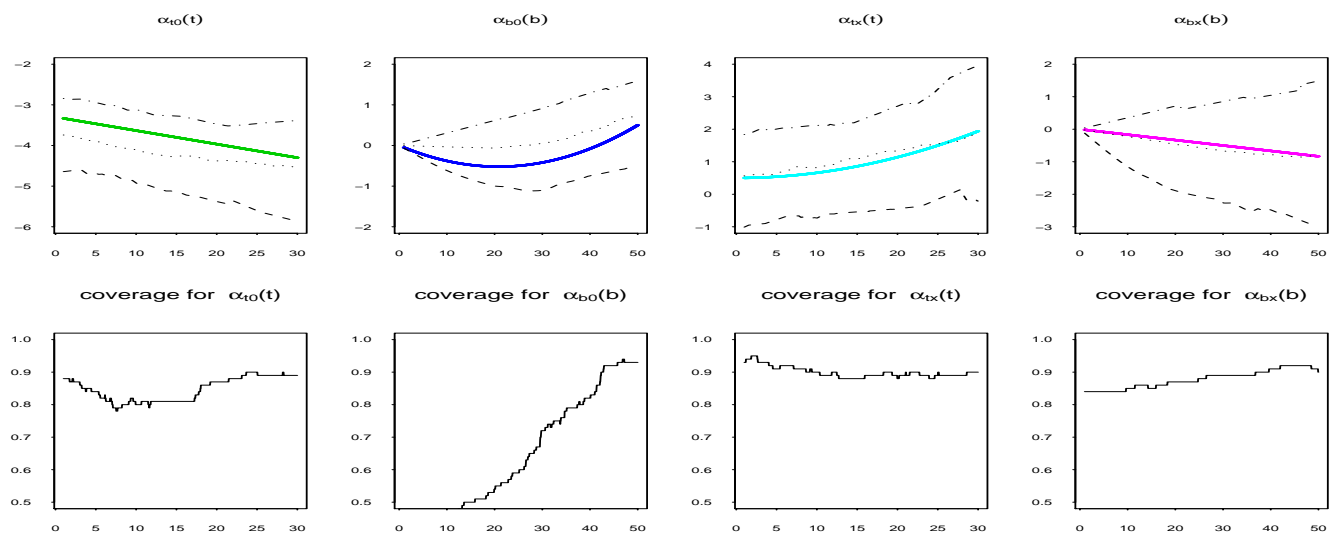


Abbildung B.4: Simulierte glatte Komponenten und coverage probabilities. 100 Beobachtungen, davon ca. 50% zensiert.

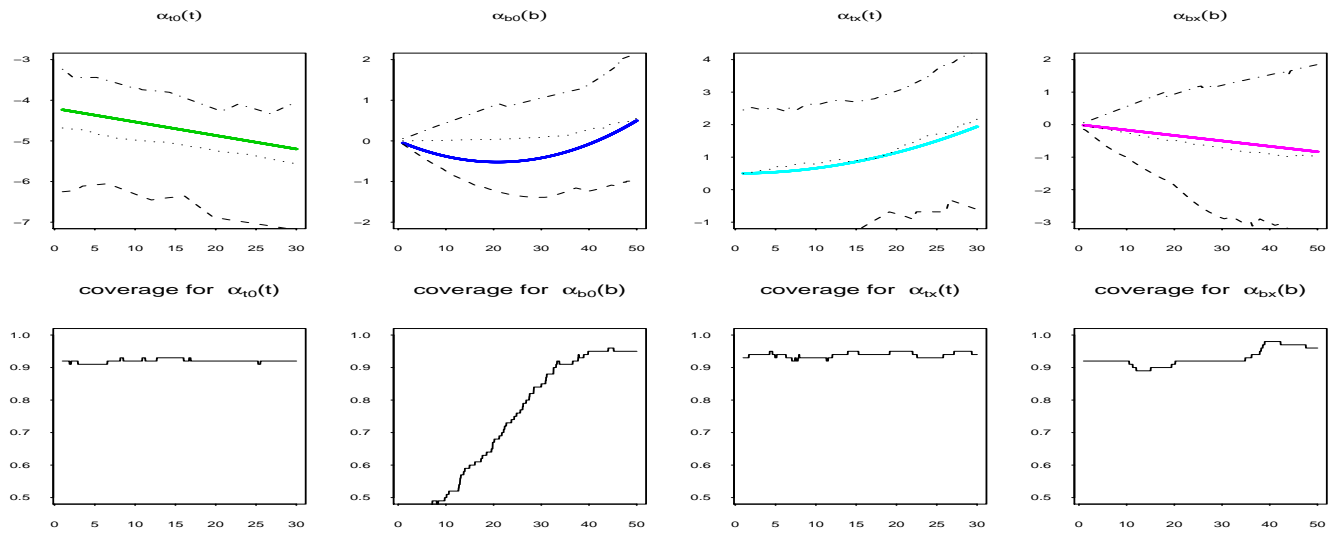


Abbildung B.5: Simulierte glatte Komponenten und coverage probabilities. 100 Beobachtungen, davon ca. 75% zensiert.

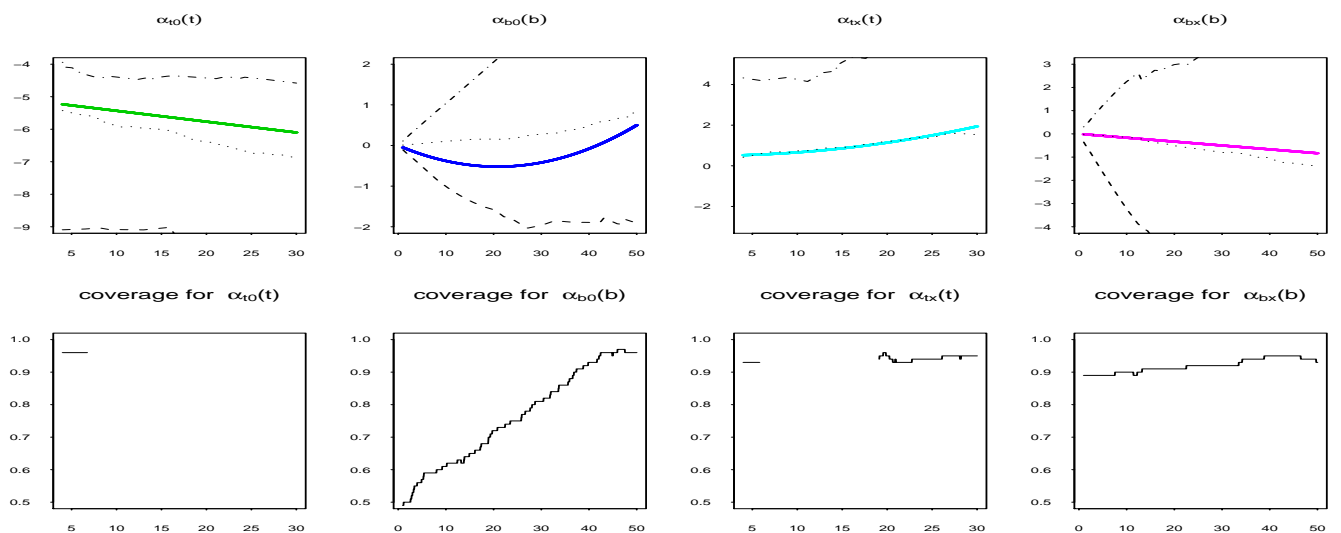


Abbildung B.6: Simulierte glatte Komponenten und coverage probabilities. 100 Beobachtungen, davon ca. 90% zensiert.

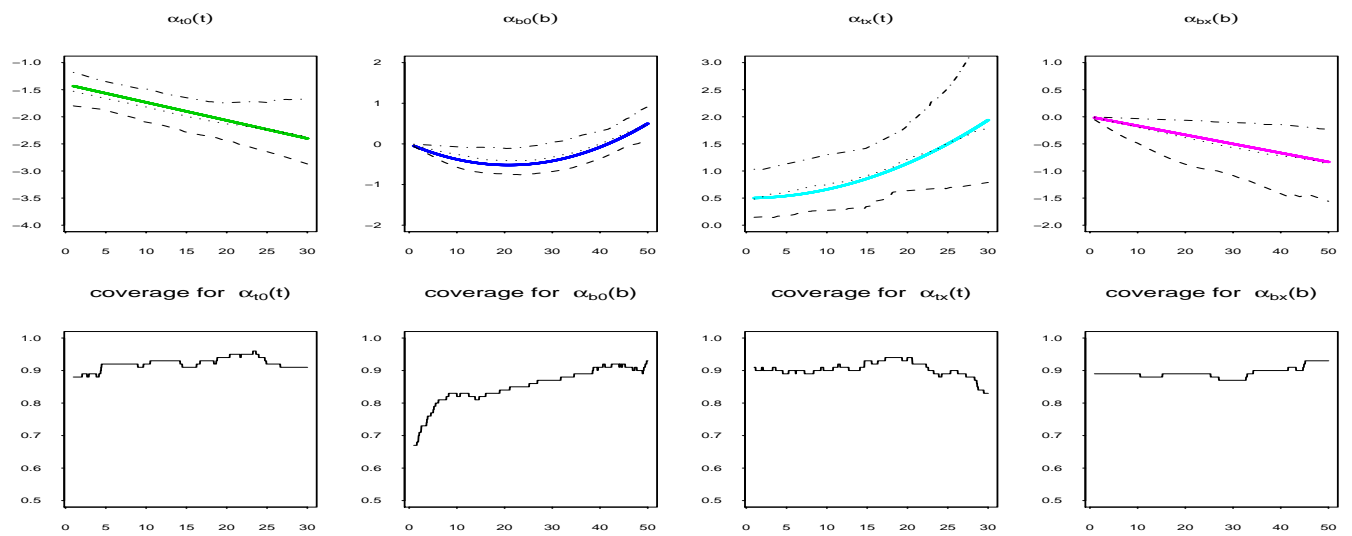


Abbildung B.7: Simulierte glatte Komponenten und coverage probabilities. 500 Beobachtungen, davon ca. 2% zensiert.

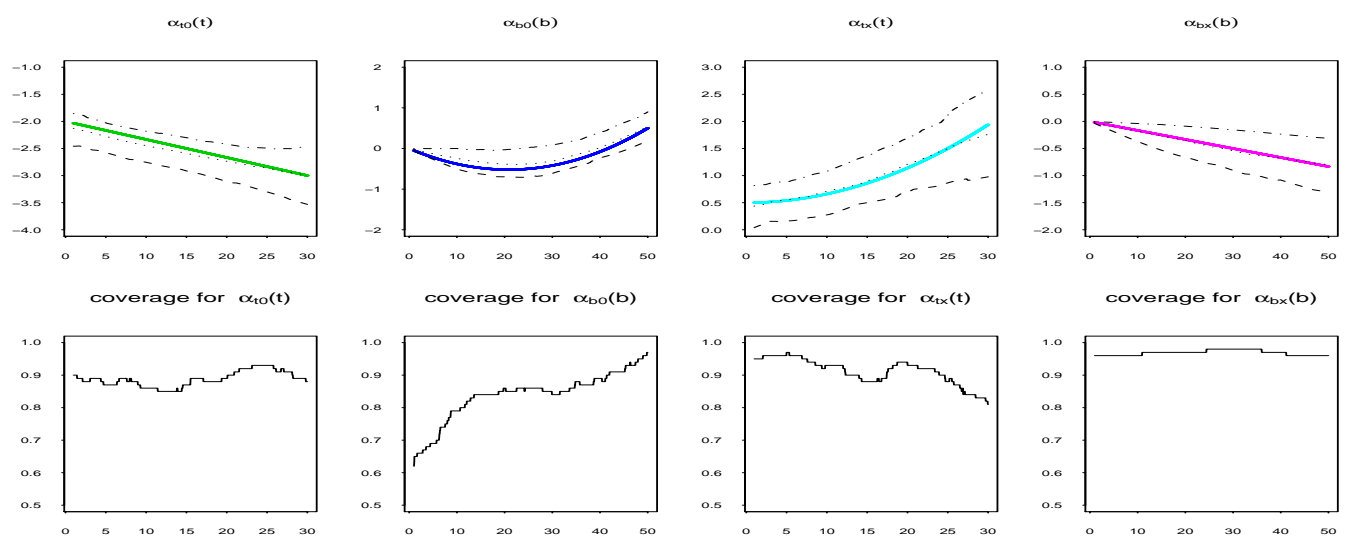


Abbildung B.8: Simulierte glatte Komponenten und coverage probabilities. 500 Beobachtungen, davon ca. 10% zensiert.

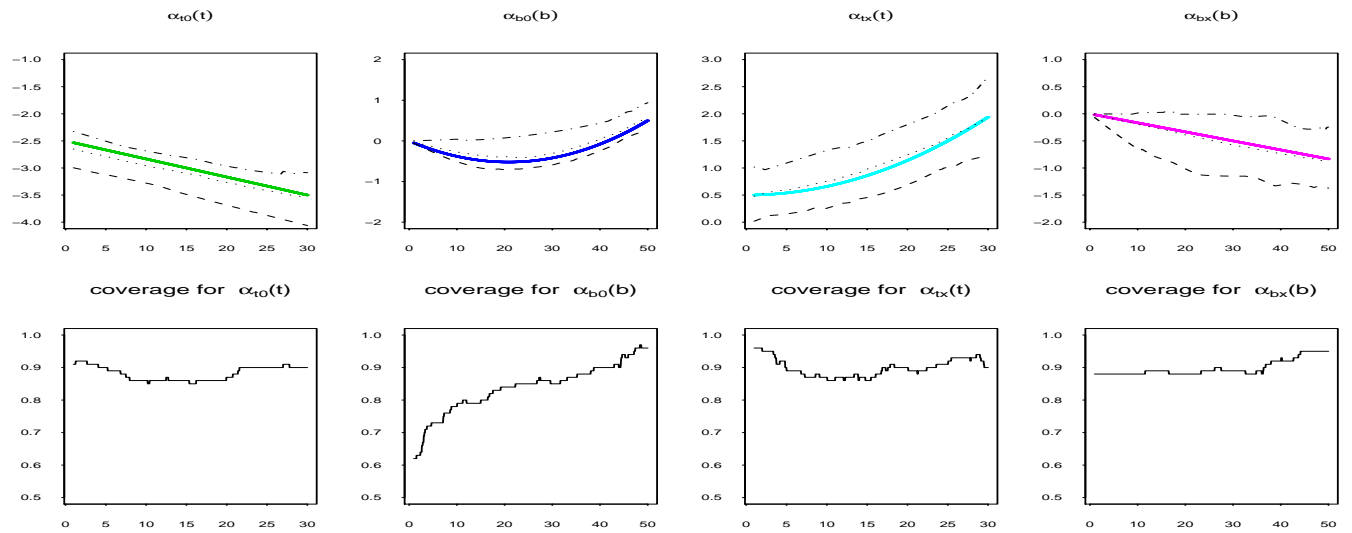


Abbildung B.9: Simulierte glatte Komponenten und coverage probabilities. 500 Beobachtungen, davon ca. 25% zensiert.

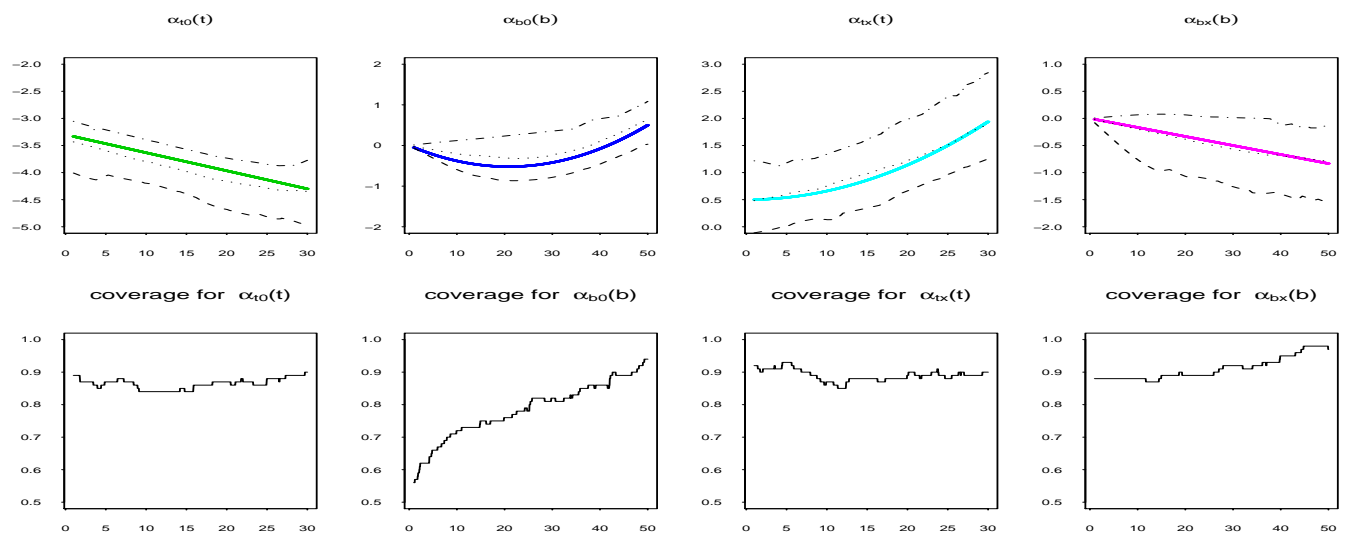


Abbildung B.10: Simulierte glatte Komponenten und coverage probabilities. 500 Beobachtungen, davon ca. 50% zensiert.

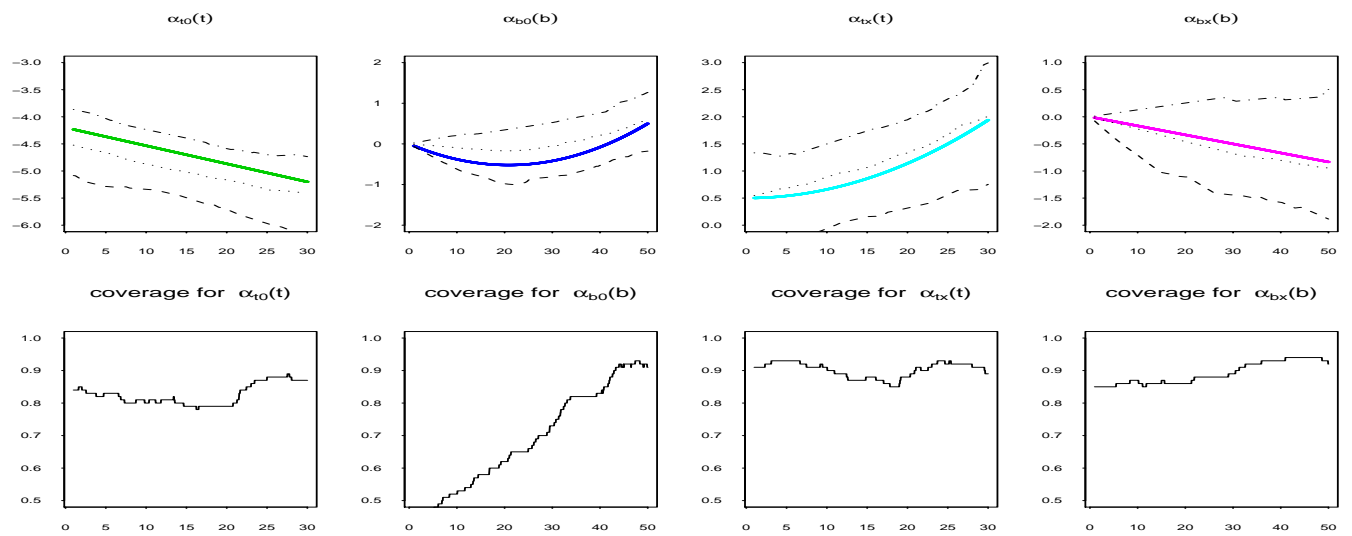


Abbildung B.11: Simulierte glatte Komponenten und coverage probabilities. 500 Beobachtungen, davon ca. 75% zensiert.

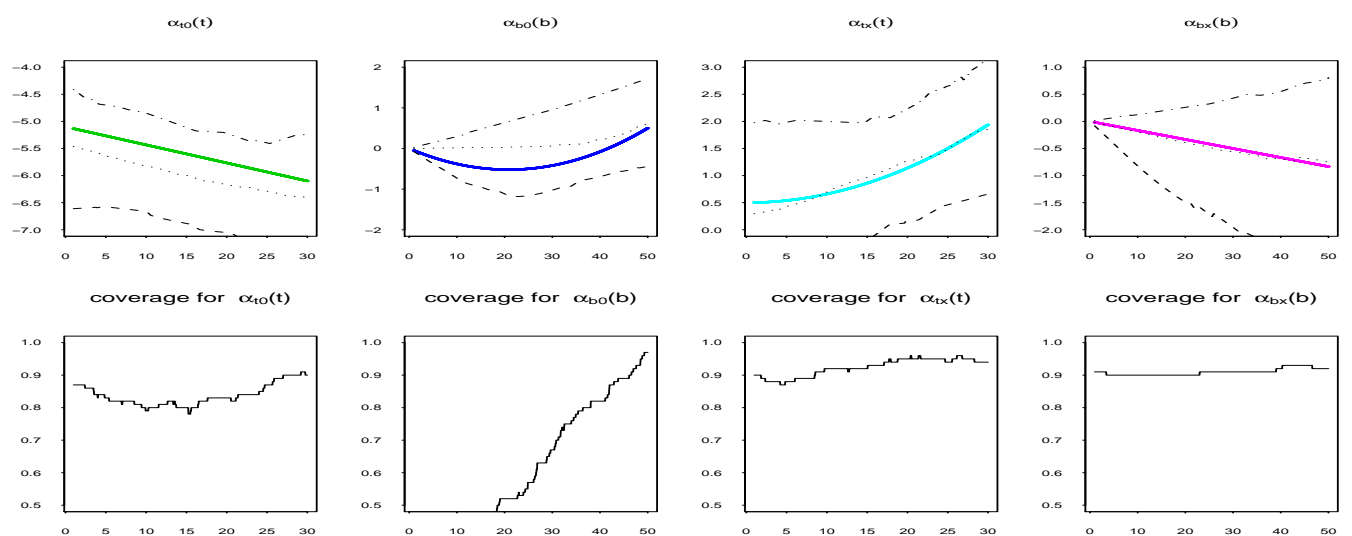


Abbildung B.12: Simulierte glatte Komponenten und coverage probabilities. 500 Beobachtungen, davon ca. 90% zensiert.

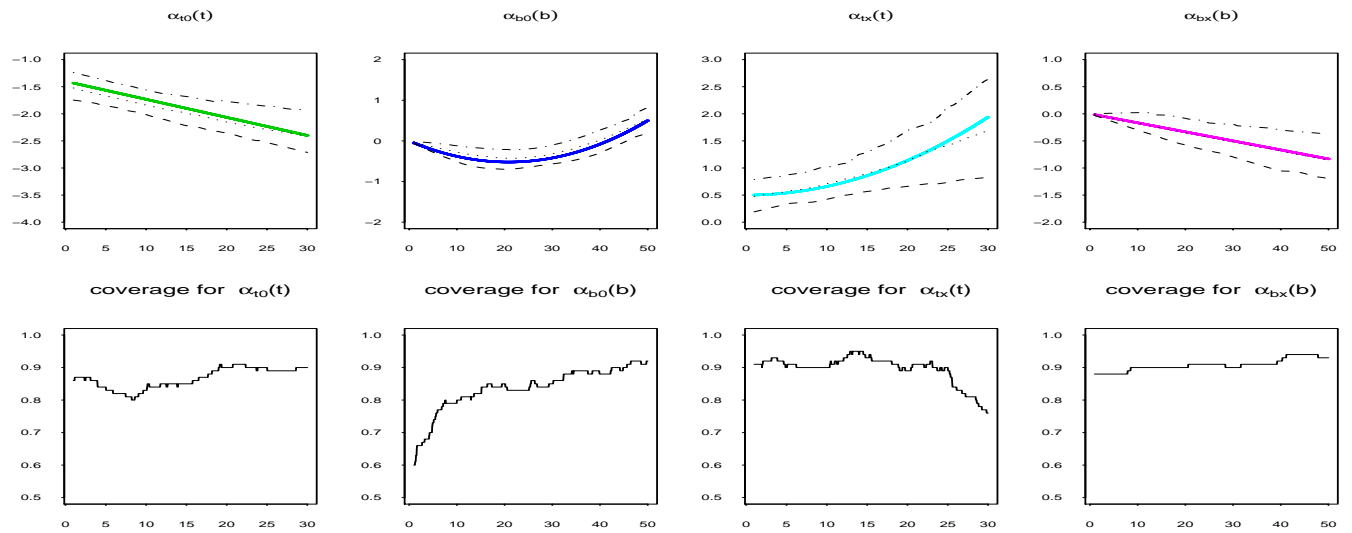


Abbildung B.13: Simulierte glatte Komponenten und coverage probabilities. 1000 Beobachtungen, davon ca. 2% zensiert.

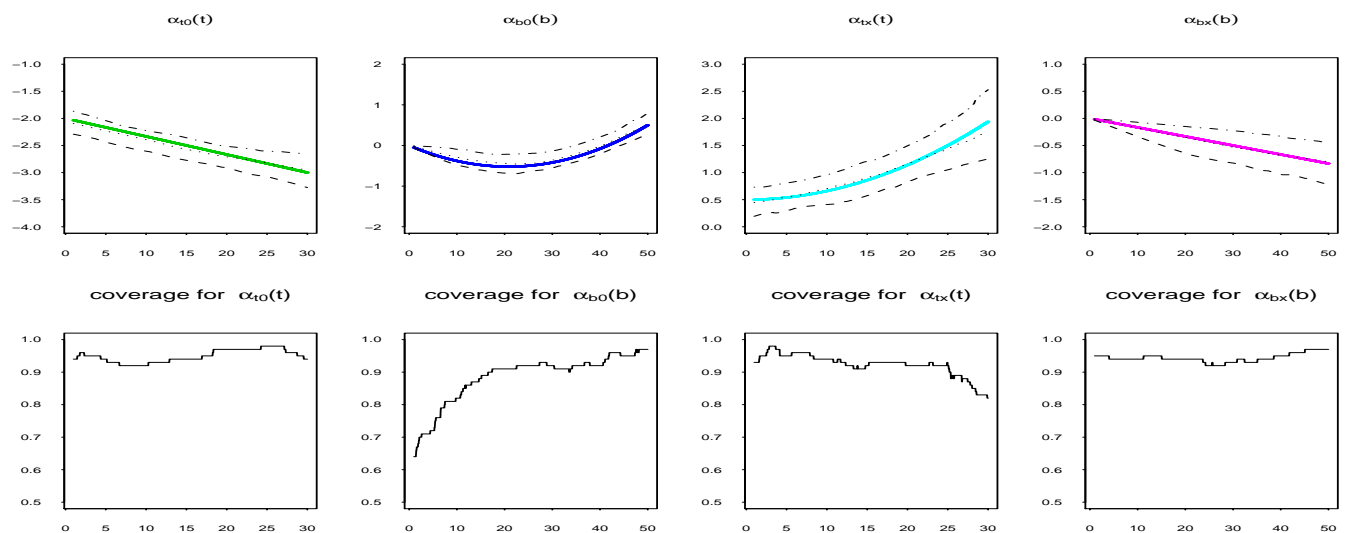


Abbildung B.14: Simulierte glatte Komponenten und coverage probabilities. 1000 Beobachtungen, davon ca. 10% zensiert.

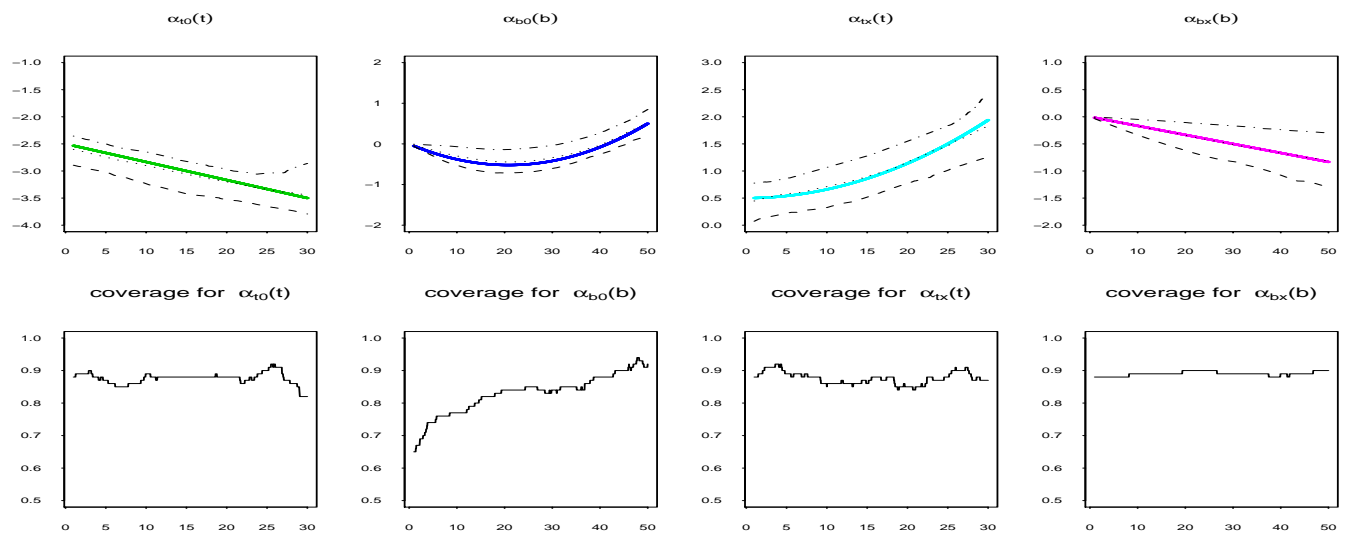


Abbildung B.15: Simulierte glatte Komponenten und coverage probabilities. 1000 Beobachtungen, davon ca. 25% zensiert.

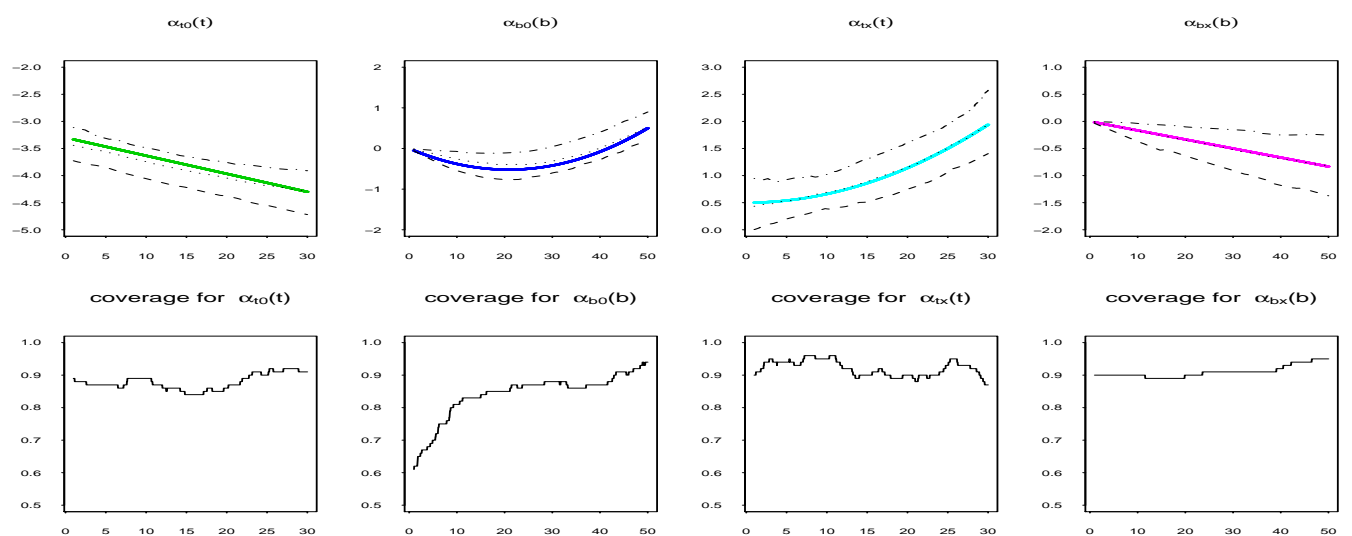


Abbildung B.16: Simulierte glatte Komponenten und coverage probabilities. 1000 Beobachtungen, davon ca. 50% zensiert.

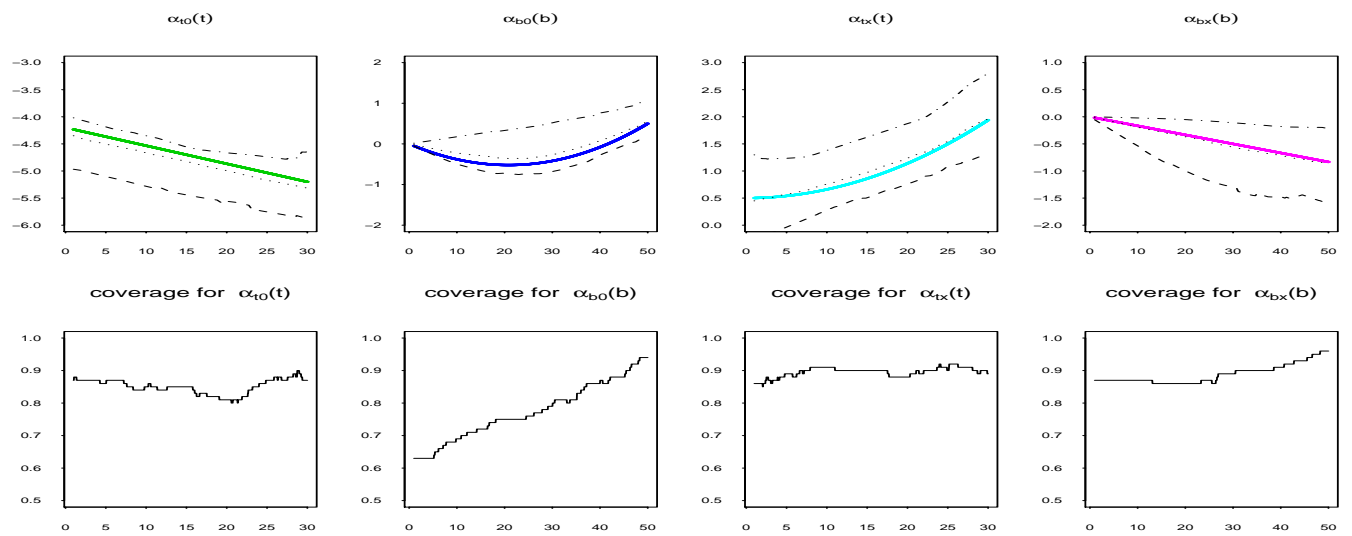


Abbildung B.17: Simulierte glatte Komponenten und coverage probabilities. 1000 Beobachtungen, davon ca. 75% zensiert.

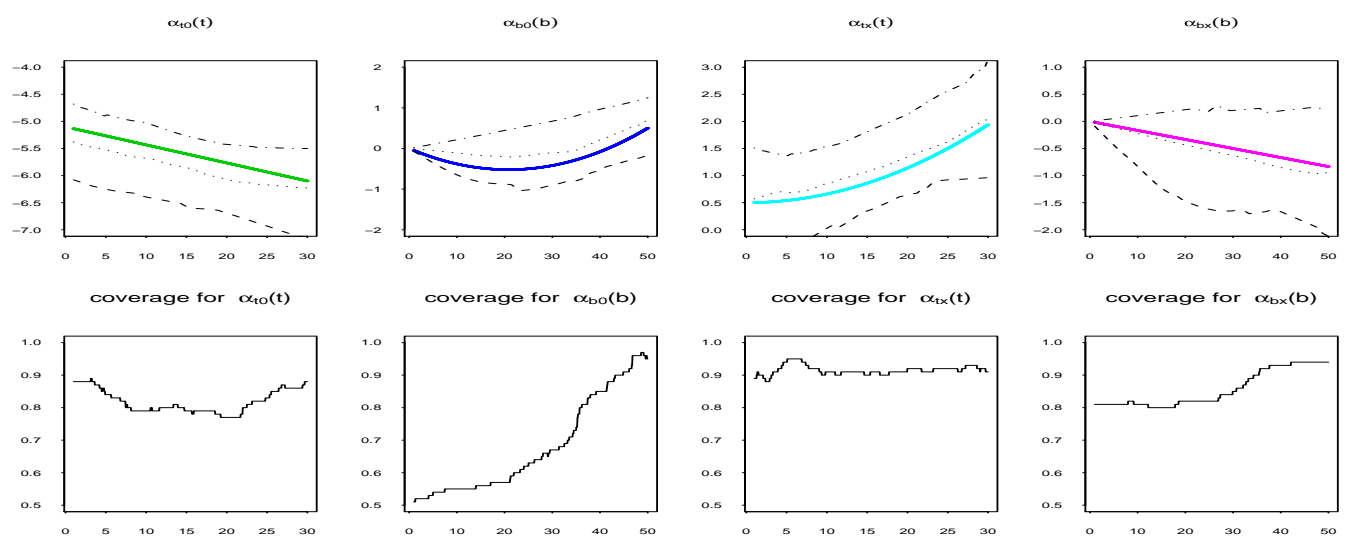


Abbildung B.18: Simulierte glatte Komponenten und coverage probabilities. 1000 Beobachtungen, davon ca. 90% zensiert.



## B.2 Simulation mit B-splines

Die Wahl der zu simulierenden glatten Komponenten ist wie folgt festgelegt:

$$\begin{aligned}\alpha_0^t(t) &= \eta - \frac{t}{30}, \\ \alpha_0^b(b) &= \sin\left(2\pi\frac{b-1}{50}\right), \\ \alpha_x^t(t) &= 0.5 + \left(\frac{t}{25}\right)^2, \\ \alpha_x^b(b) &= 1 - \cos\left(2\pi\frac{b-1}{50}\right),\end{aligned}$$

wobei der Parameter  $\eta$  im Ausdruck für die Baseline-Hazard in  $t$ -Richtung gerade die relative Häufigkeit der zensierten Beobachtungen steuert. Die folgenden eingesetzten Werte für  $\eta$  entsprechen den Zensierungsmustern  $h_\delta$  des Anteils der zensierten Beobachtungen:

$$\begin{aligned}\eta = -1 &\quad \rightarrow \quad h_\delta \simeq 2\%, \\ \eta = -2 &\quad \rightarrow \quad h_\delta \simeq 10\%, \\ \eta = -2.5 &\quad \rightarrow \quad h_\delta \simeq 20\%, \\ \eta = -4 &\quad \rightarrow \quad h_\delta \simeq 50\%, \\ \eta = -5 &\quad \rightarrow \quad h_\delta \simeq 75\%, \\ \eta = -6 &\quad \rightarrow \quad h_\delta \simeq 90\%.\end{aligned}$$

Für die jeweilige Anzahl der Beobachtungen  $N = 100, 500, 1000$  und jeweiliges Zensierungsmuster  $h_\delta$  ergeben sich insgesamt 18 Konstellationen. Für jede solche Konstellation erzeugen wir eine Serie von insgesamt 100 Simulationen. Als Ergebnis bekommen wir die geschätzten empirischen 5%, 50%, 95%-Quantile sowie die *coverage probabilities*. (Näher ist das Simulationsschema in Kapitel 6 beschrieben.)

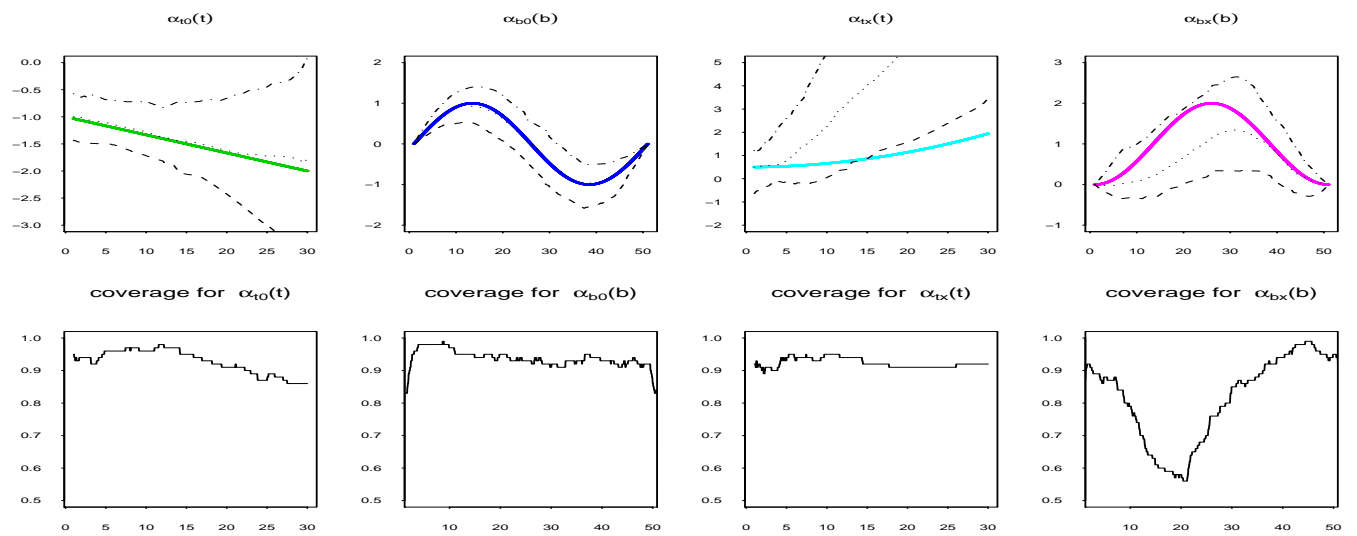


Abbildung B.19: Simulierte glatte Komponenten und coverage probabilities. 100 Beobachtungen, davon ca. 2% zensiert.

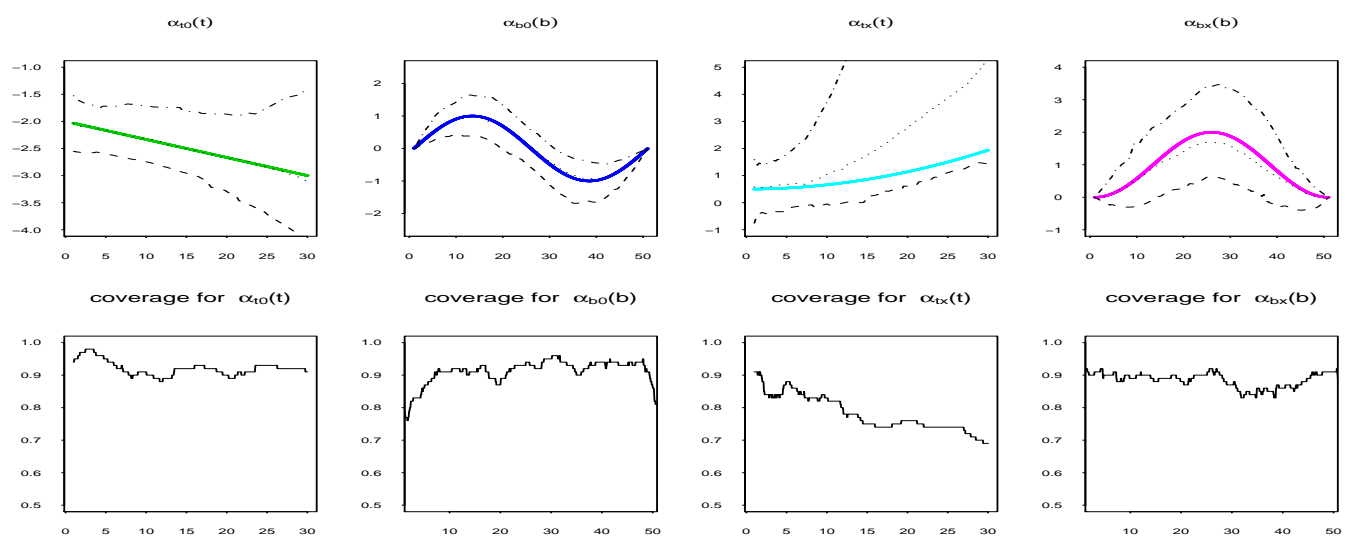


Abbildung B.20: Simulierte glatte Komponenten und coverage probabilities. 100 Beobachtungen, davon ca. 10% zensiert.

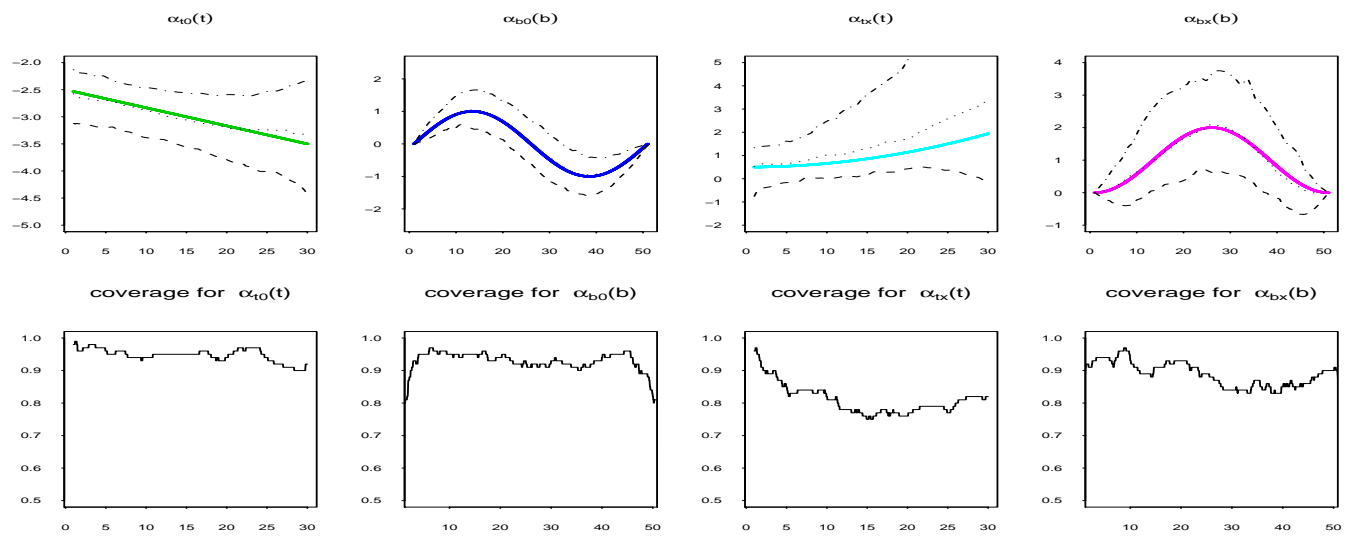


Abbildung B.21: Simulierte glatte Komponenten und coverage probabilities. 100 Beobachtungen, davon ca. 20% zensiert.

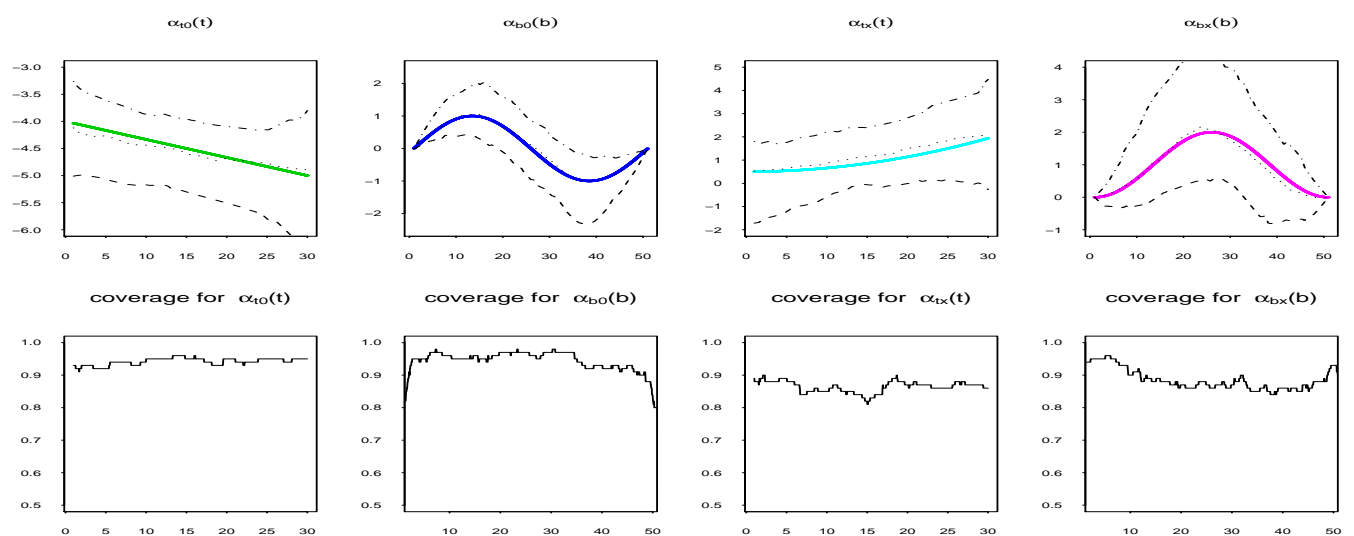


Abbildung B.22: Simulierte glatte Komponenten und coverage probabilities. 100 Beobachtungen, davon ca. 50% zensiert.

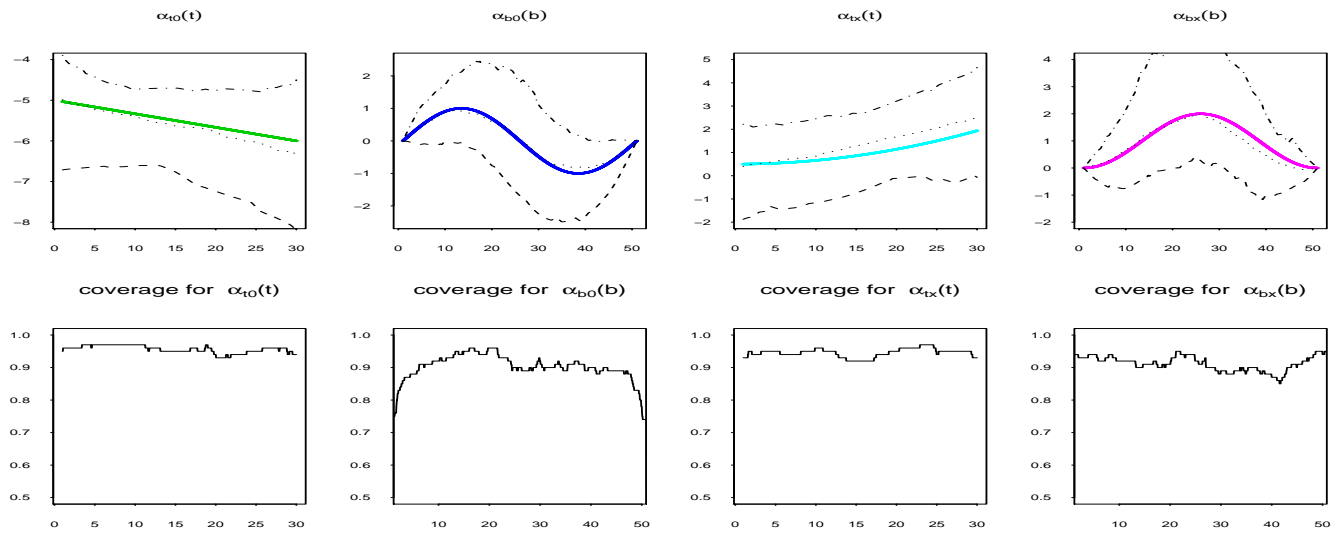


Abbildung B.23: Simulierte glatte Komponenten und coverage probabilities. 100 Beobachtungen, davon ca. 75% zensiert.

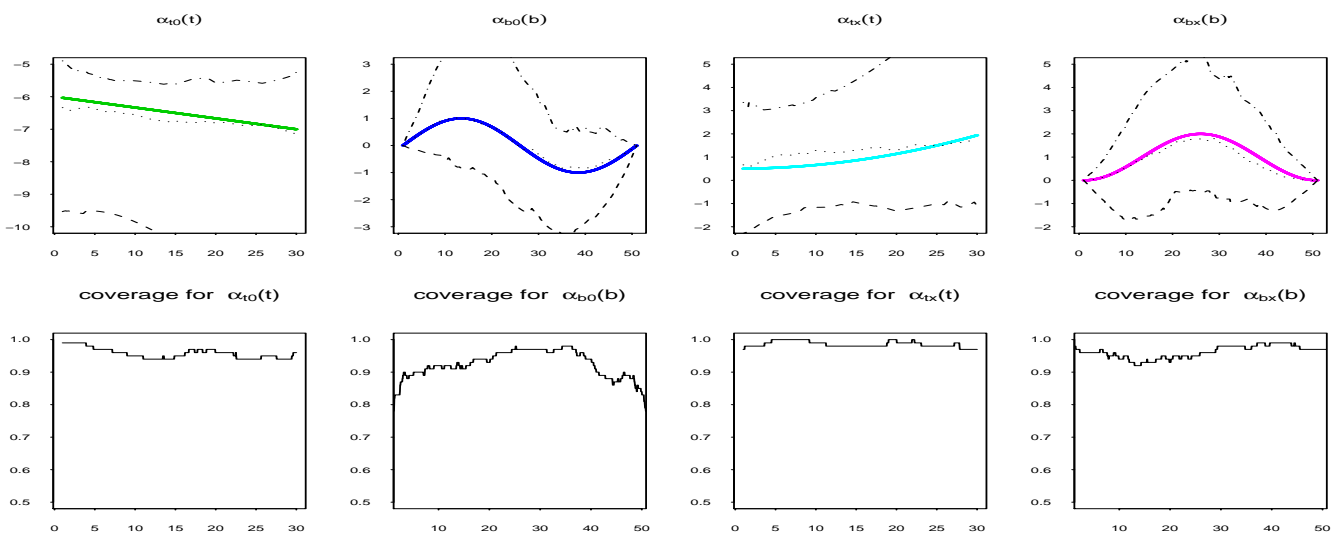


Abbildung B.24: Simulierte glatte Komponenten und coverage probabilities. 100 Beobachtungen, davon ca. 90% zensiert.

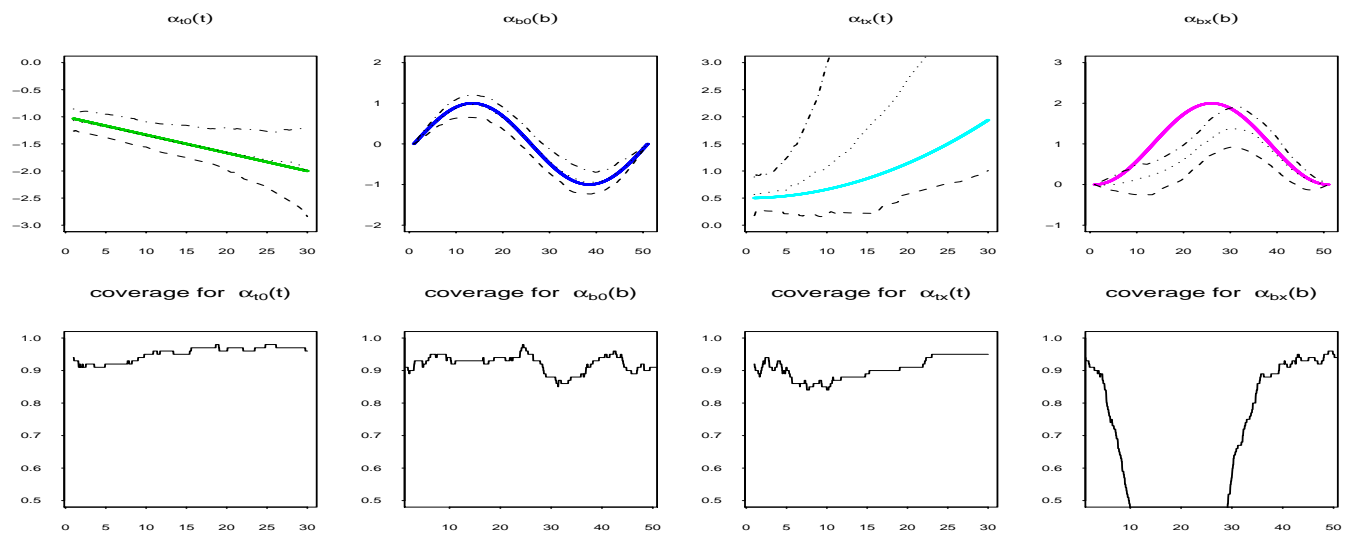


Abbildung B.25: Simulierte glatte Komponenten und coverage probabilities. 500 Beobachtungen, davon ca. 2% zensiert.

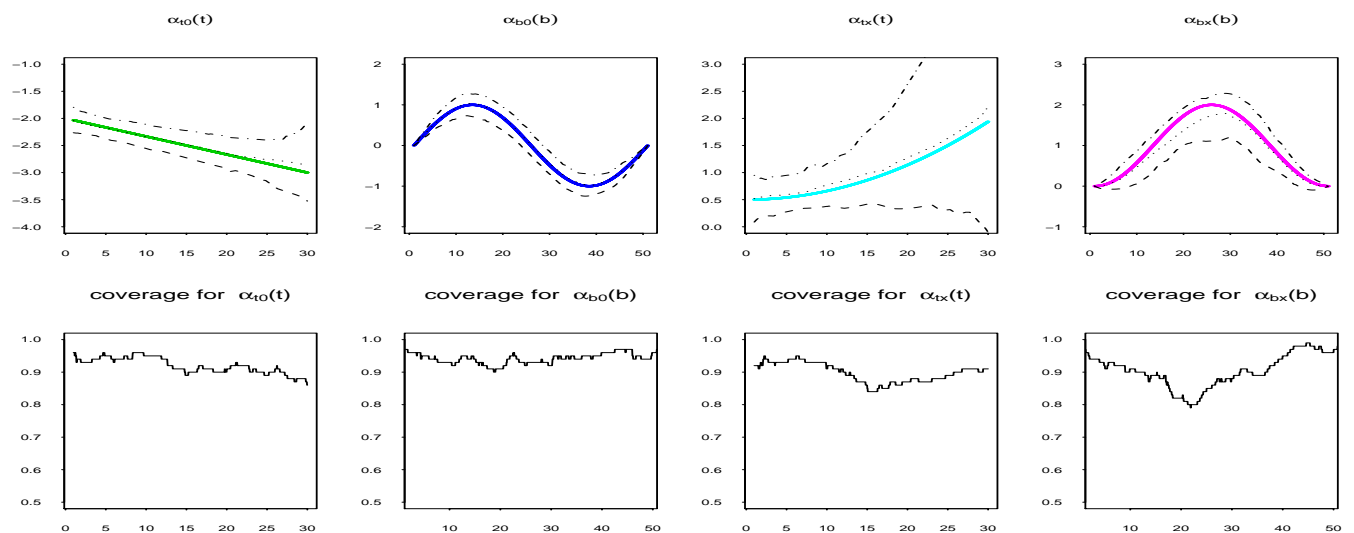


Abbildung B.26: Simulierte glatte Komponenten und coverage probabilities. 500 Beobachtungen, davon ca. 10% zensiert.

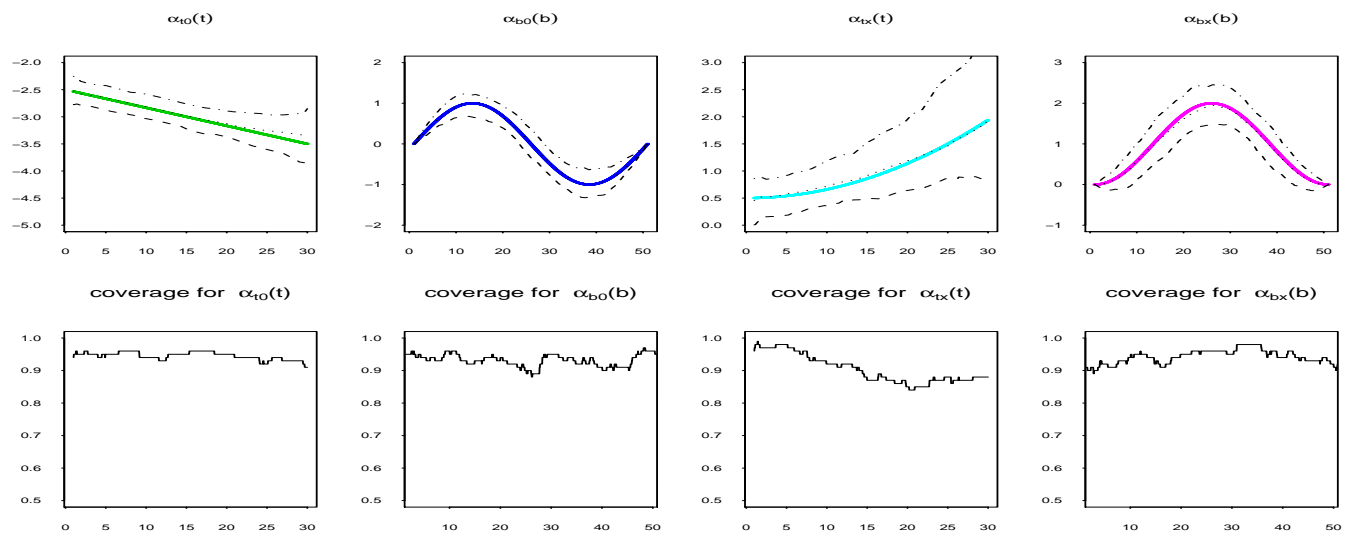


Abbildung B.27: Simulierte glatte Komponenten und coverage probabilities. 500 Beobachtungen, davon ca. 20% zensiert.

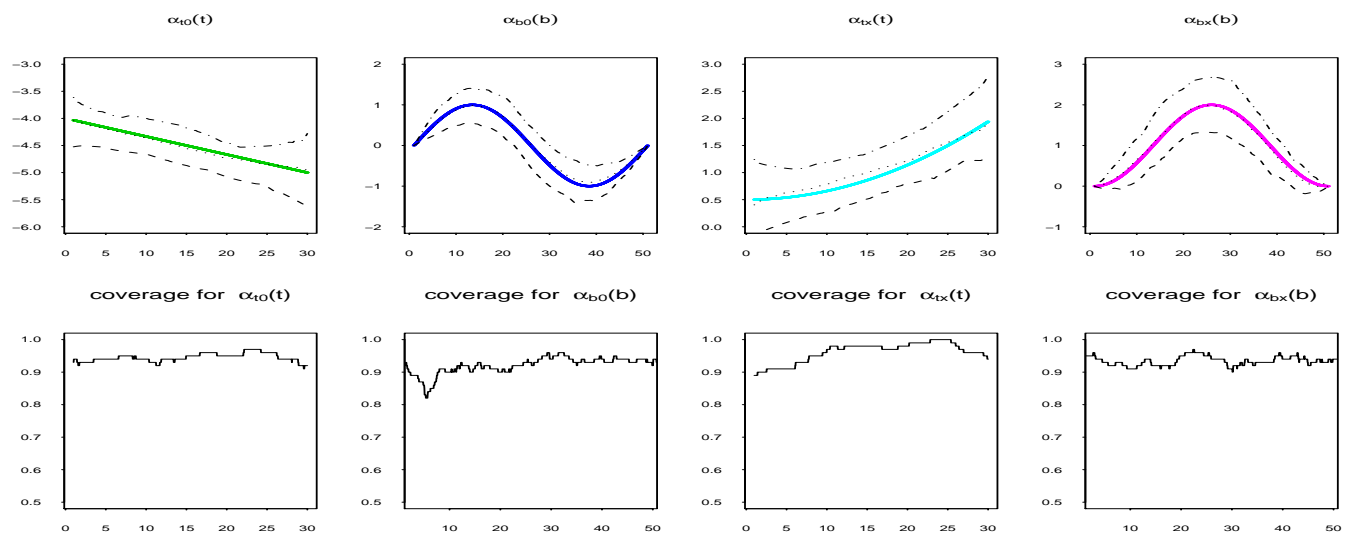


Abbildung B.28: Simulierte glatte Komponenten und coverage probabilities. 500 Beobachtungen, davon ca. 50% zensiert.

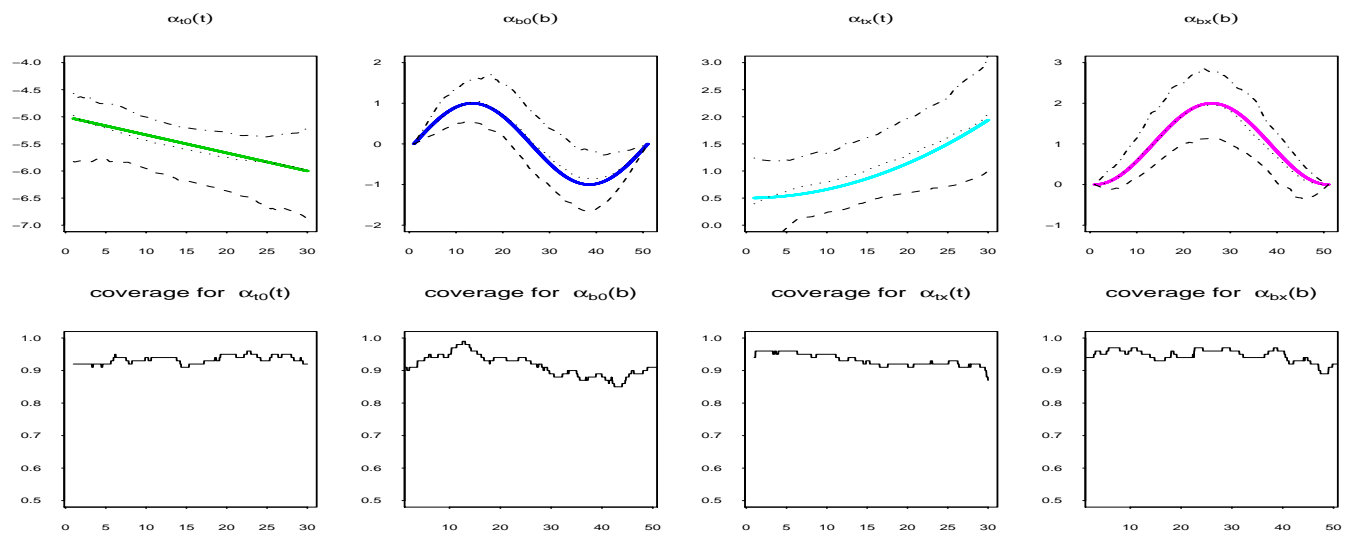


Abbildung B.29: Simulierte glatte Komponenten und coverage probabilities. 500 Beobachtungen, davon ca. 75% zensiert.

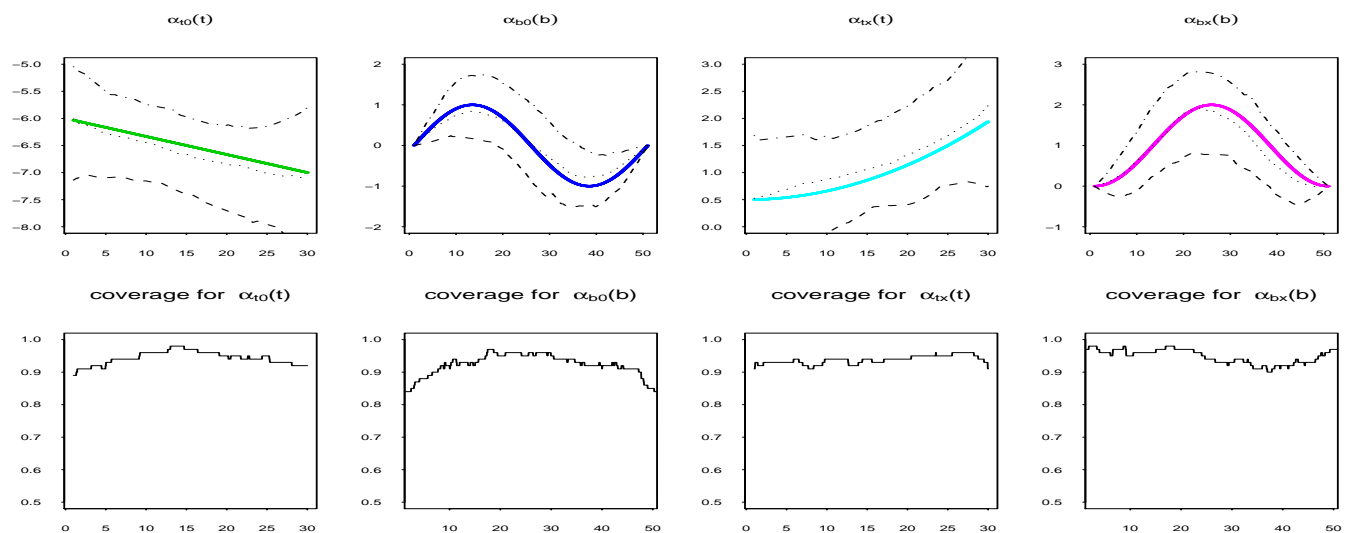


Abbildung B.30: Simulierte glatte Komponenten und coverage probabilities. 500 Beobachtungen, davon ca. 90% zensiert.

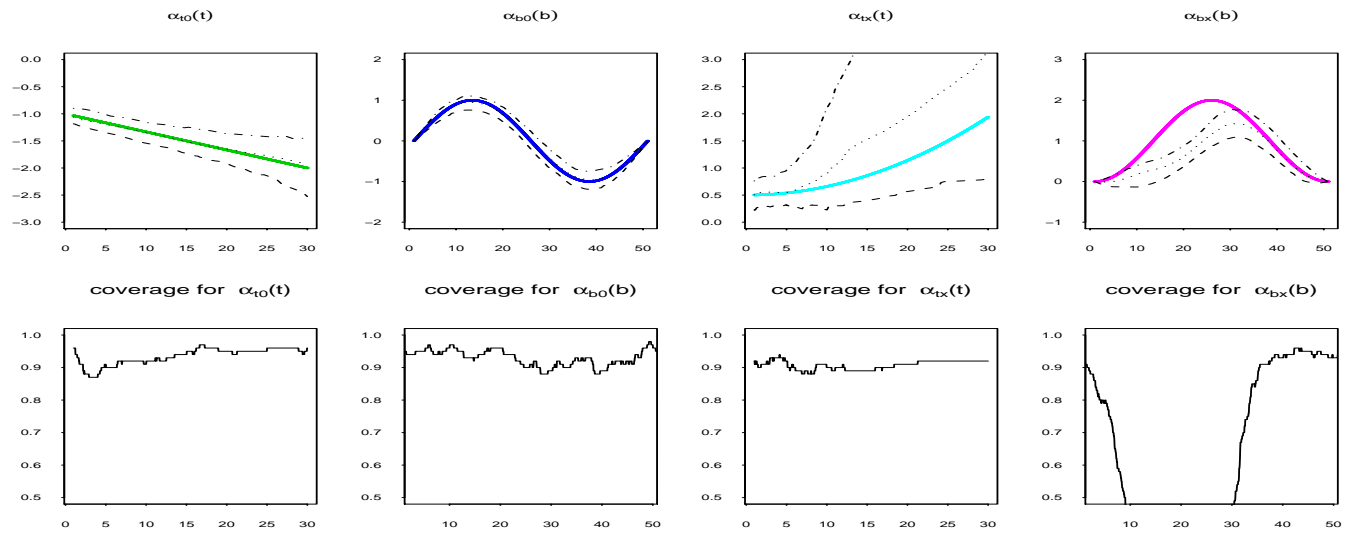


Abbildung B.31: Simulierte glatte Komponenten und coverage probabilities. 1000 Beobachtungen, davon ca. 2% zensiert.

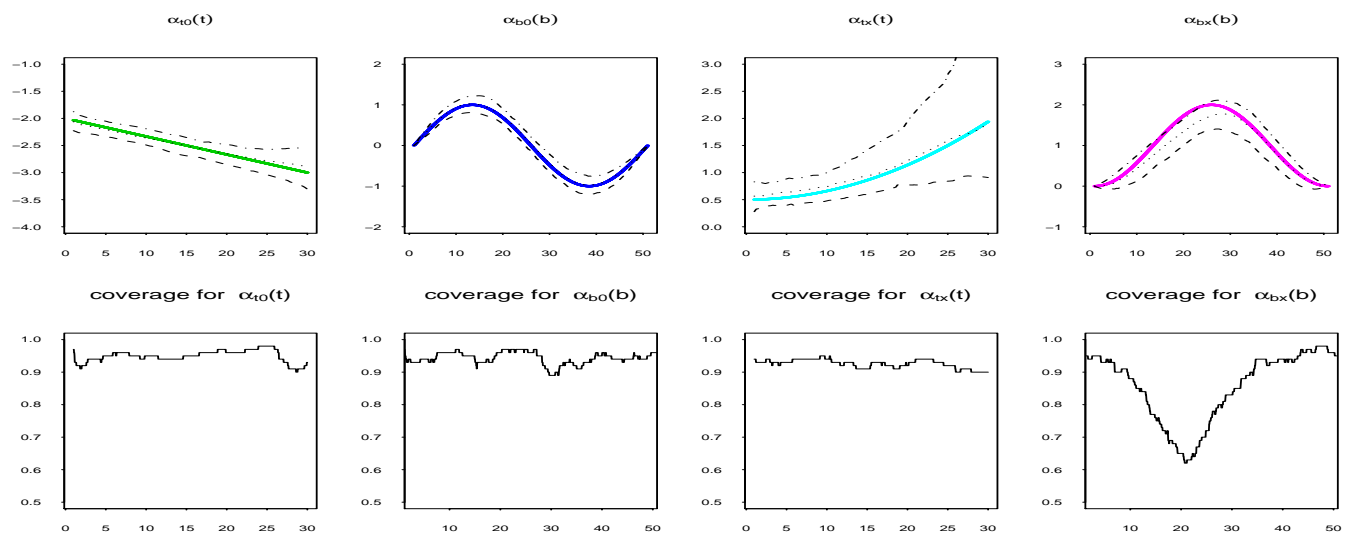


Abbildung B.32: Simulierte glatte Komponenten und coverage probabilities. 1000 Beobachtungen, davon ca. 10% zensiert.



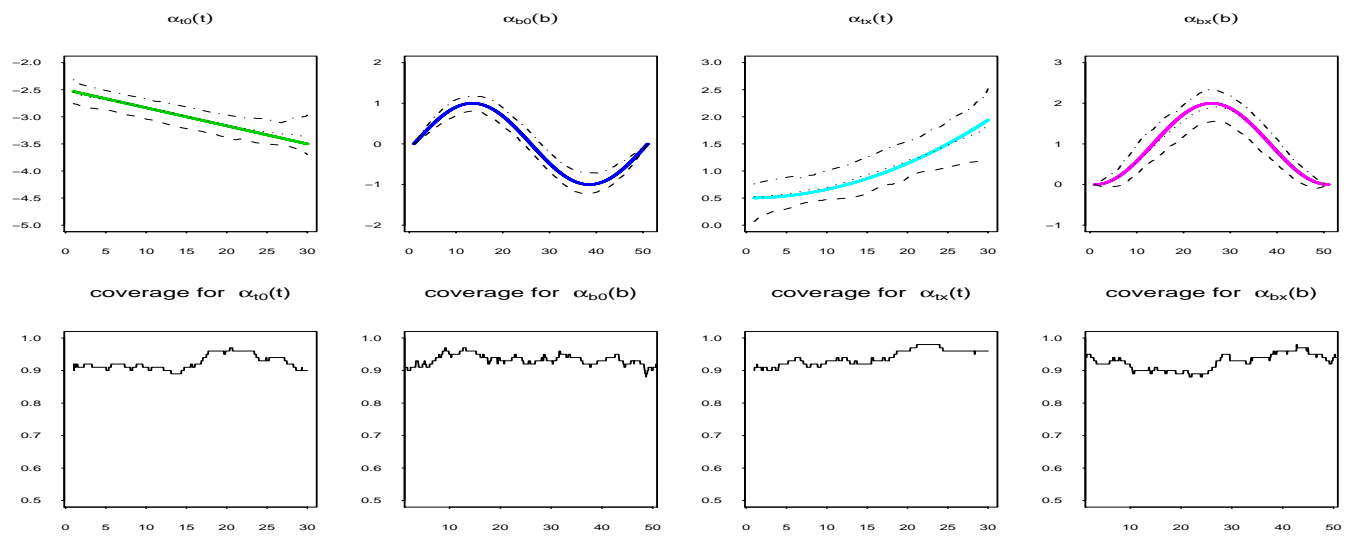


Abbildung B.33: Simulierte glatte Komponenten und coverage probabilities. 1000 Beobachtungen, davon ca. 20% zensiert.

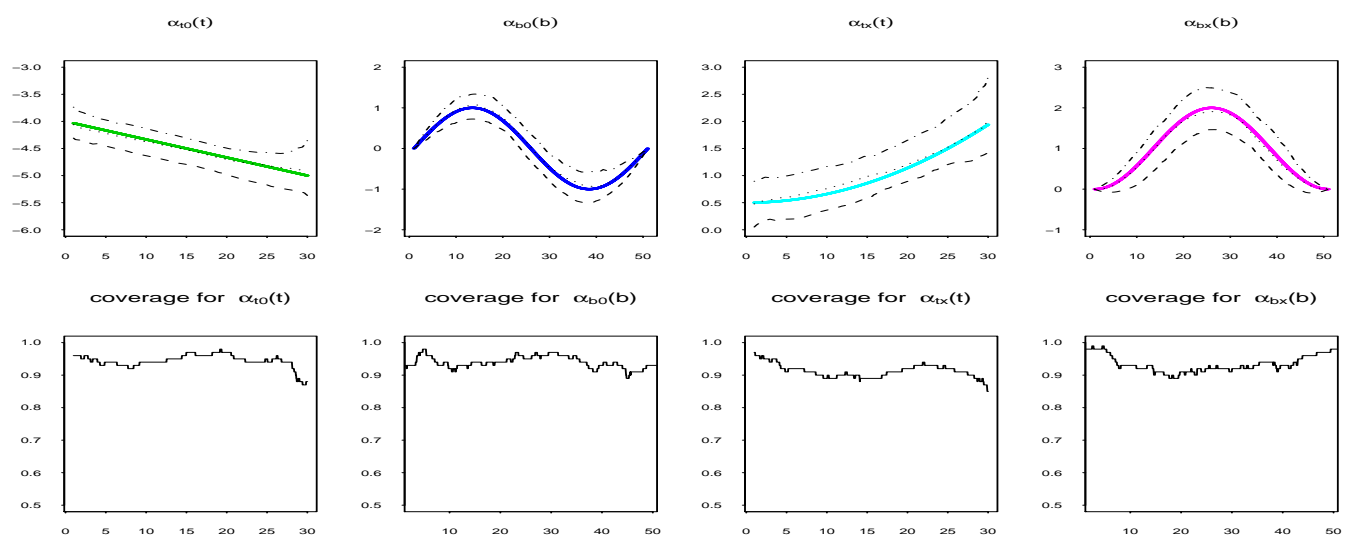


Abbildung B.34: Simulierte glatte Komponenten und coverage probabilities. 1000 Beobachtungen, davon ca. 50% zensiert.

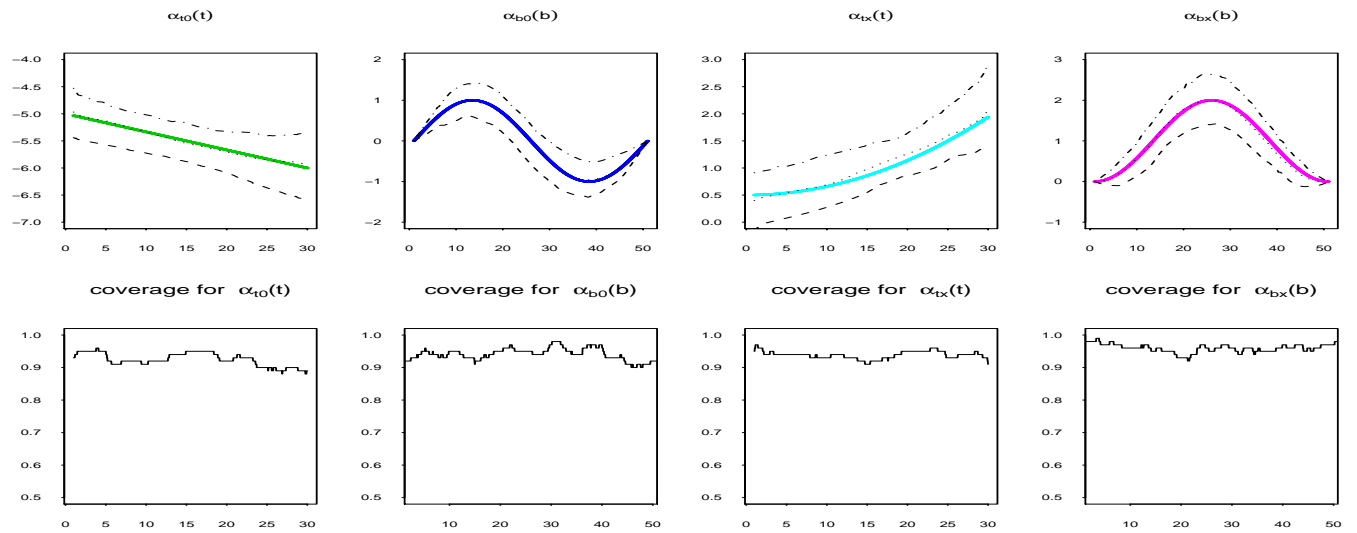


Abbildung B.35: Simulierte glatte Komponenten und coverage probabilities. 1000 Beobachtungen, davon ca. 75% zensiert.

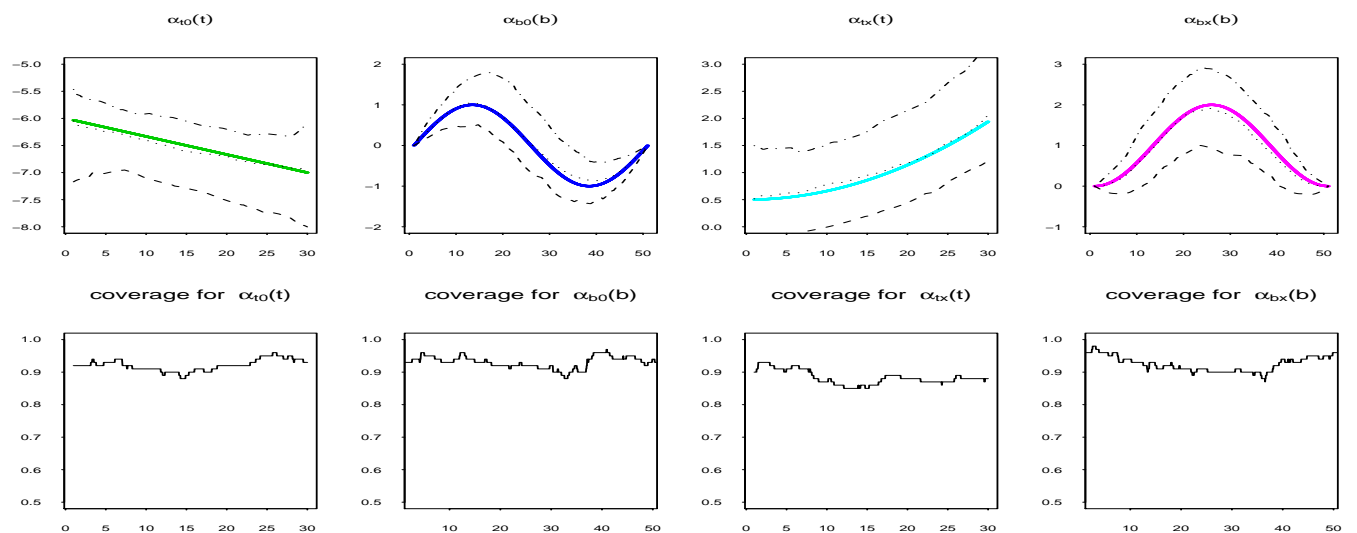


Abbildung B.36: Simulierte glatte Komponenten und coverage probabilities. 1000 Beobachtungen, davon ca. 90% zensiert.

## B.3 Simulation für das Modell der Competing Risks

### B.3.1 Simulation auf Monatsbasis

Für  $L = 2$  Competing Risks ist die Wahl der zu simulierenden glatten Komponenten wie folgt festgelegt:

$$\alpha_{01}^t(t) = \eta_1 - \frac{t}{30}, \quad \alpha_{02}^t(t) = \eta_2 + \frac{t}{50},$$

$$\alpha_{x1}^t(t) = 0.5 + \left(\frac{t}{25}\right)^2, \quad \alpha_{x2}^t(t) = -\frac{t}{60},$$

wobei die Parameter  $\eta_1, \eta_2$  im Ausdruck für die logarithmierte Baseline-Hazard gerade die relative Häufigkeit der zensierten Beobachtungen steuern und das Argument  $t$  die ganzzahligen Werte aus dem Intervall  $[1, 30]$  annimmt (interpretiert als Monate 1 bis 30). Die Intensitäten jeweiliger zu simulierender Poisson-Daten sind  $\xi_1 = \exp\{\alpha_{01}^t(t) + x \alpha_{x1}^t(t)\}w_1$  und  $\xi_2 = \exp\{\alpha_{02}^t(t) + x \alpha_{x2}^t(t)\}w_2$  mit simulierten Frailty-Werten  $w_1$  und  $w_2$ . Die folgenden eingesetzten Werte für  $\eta_1, \eta_2$  entsprechen den Zensierungsmustern  $h_\delta$ :

$$\begin{aligned} \eta_1 = -2, \eta_2 = -3 &\rightarrow h_\delta \simeq 5\%, \\ \eta_1 = -6, \eta_2 = -3 &\rightarrow h_\delta \simeq 20\%, \\ \eta_1 = -5, \eta_2 = -4 &\rightarrow h_\delta \simeq 50\%, \\ \eta_1 = -5, \eta_2 = -6 &\rightarrow h_\delta \simeq 75\%. \end{aligned}$$

Weitere Simulationsparameter sind:

- *Frailty-Verteilung.* Für  $L=2$  unabhängige Gamma-Frailties mit jeweils *shape*-Parameter =2.5= *scale*-Parameter (für Risk 1) und *shape*-Parameter =5=*scale*-Parameter (für Risk 2) werden simuliert.
- *Anzahl der Individuen  $N$ ,* jeweils 100 oder 1000.
- *Anzahl der Spells* pro Individuum. Zwei Muster stehen zur Wahl: jeweils 10 Spells pro Individuum und Spell-Verteilung ähnlich der wie im Anwendungsbeispiel, z. B. für 1000 Individuen (vgl. mit Tabelle 8.1):

number of spells	1	2	3	4	5
number of individuals	720	170	70	30	10
proportion	0.72	0.17	0.07	0.03	0.01

Näher ist das Simulationsverfahren in Abschnitt 8.9 erläutert. Als Ergebnis jeweiliger Simulationsserie (je 50 Simulationen pro Parameterkonstellation) resultieren die geschätzten empirischen 5%-, 50%-, 95%- Quantile (also die empirischen Konfidenzbänder) sowie *coverage probabilities*. Außerdem sind für Frailty-Verteilung und -Dichte die empirischen 95%- Konfidenzbänder und medialer Wert gezeichnet (vgl. außerdem Kauermann & Khomski, 2007, für weitere Simulationen).

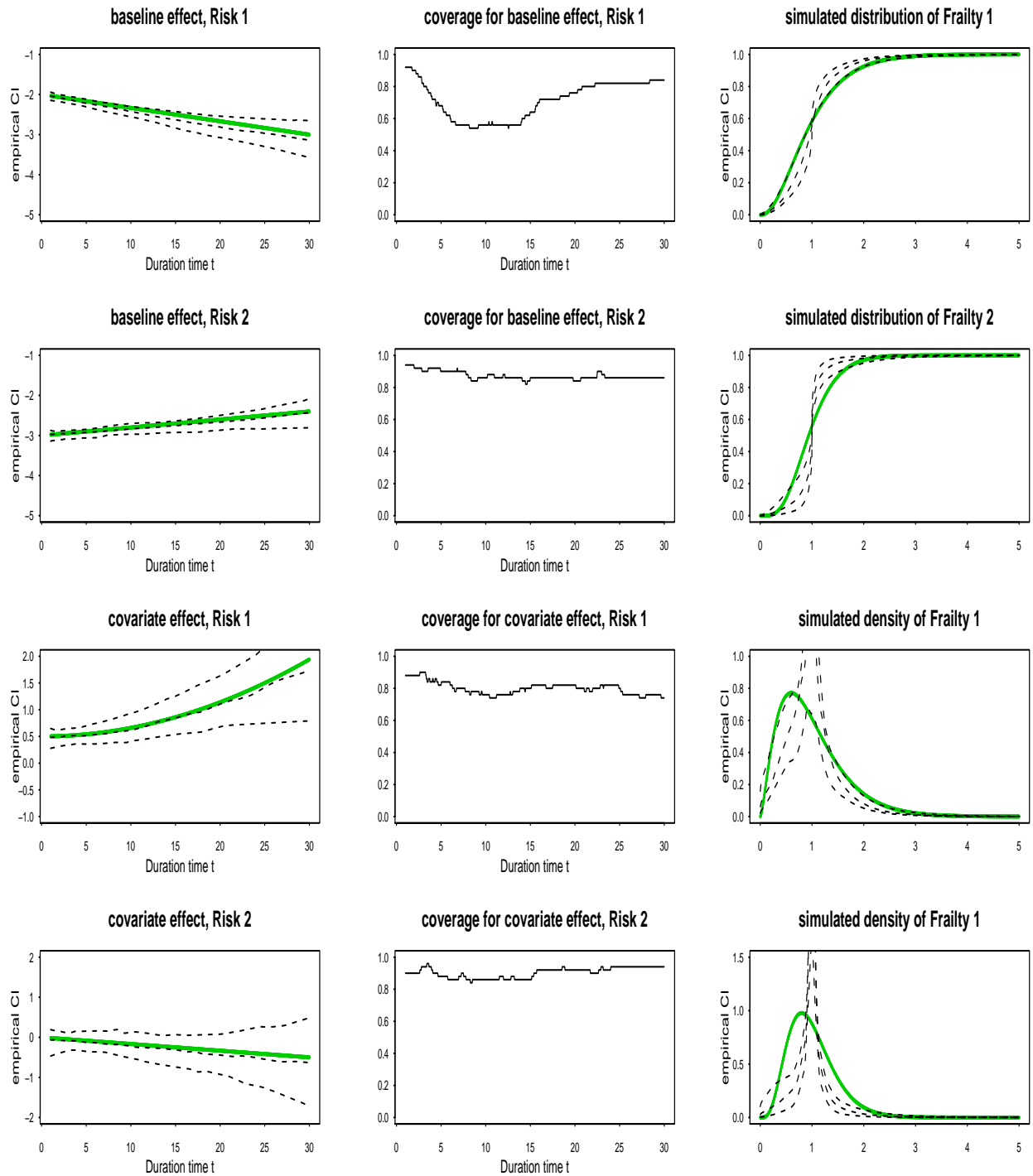


Abbildung B.37: Simulierte glatte Komponenten, coverage probabilities und Frailties. 1000 Individuen, davon ca. 5% zensiert. Häufigkeiten der Spells wie im Anwendungsbeispiel, Kapitel 8.

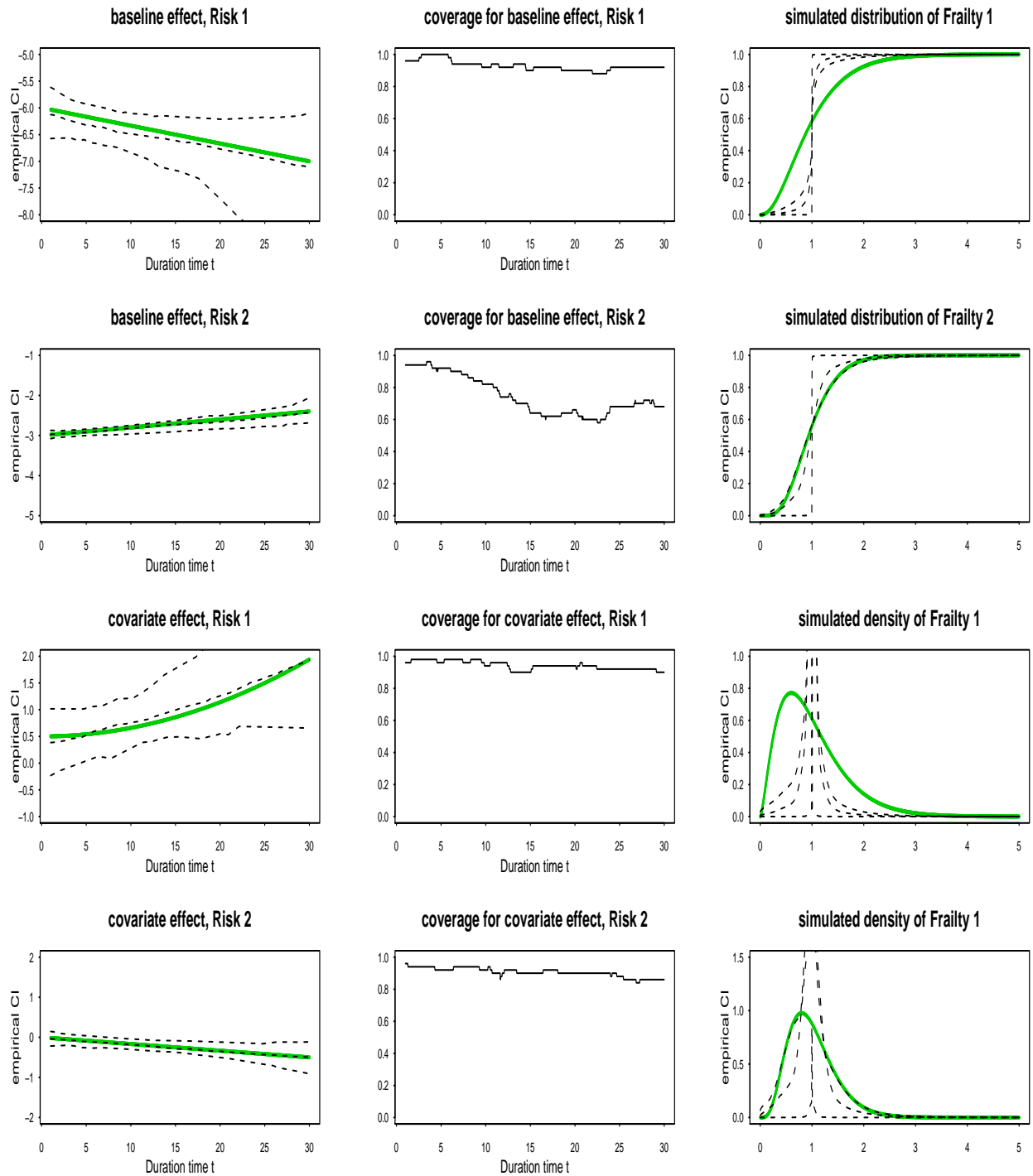


Abbildung B.38: Simulierte glatte Komponenten, coverage probabilities und Frailties. 1000 Individuen, davon ca. 20% zensiert. Häufigkeiten der Spells wie im Anwendungsbeispiel, Kapitel 8.

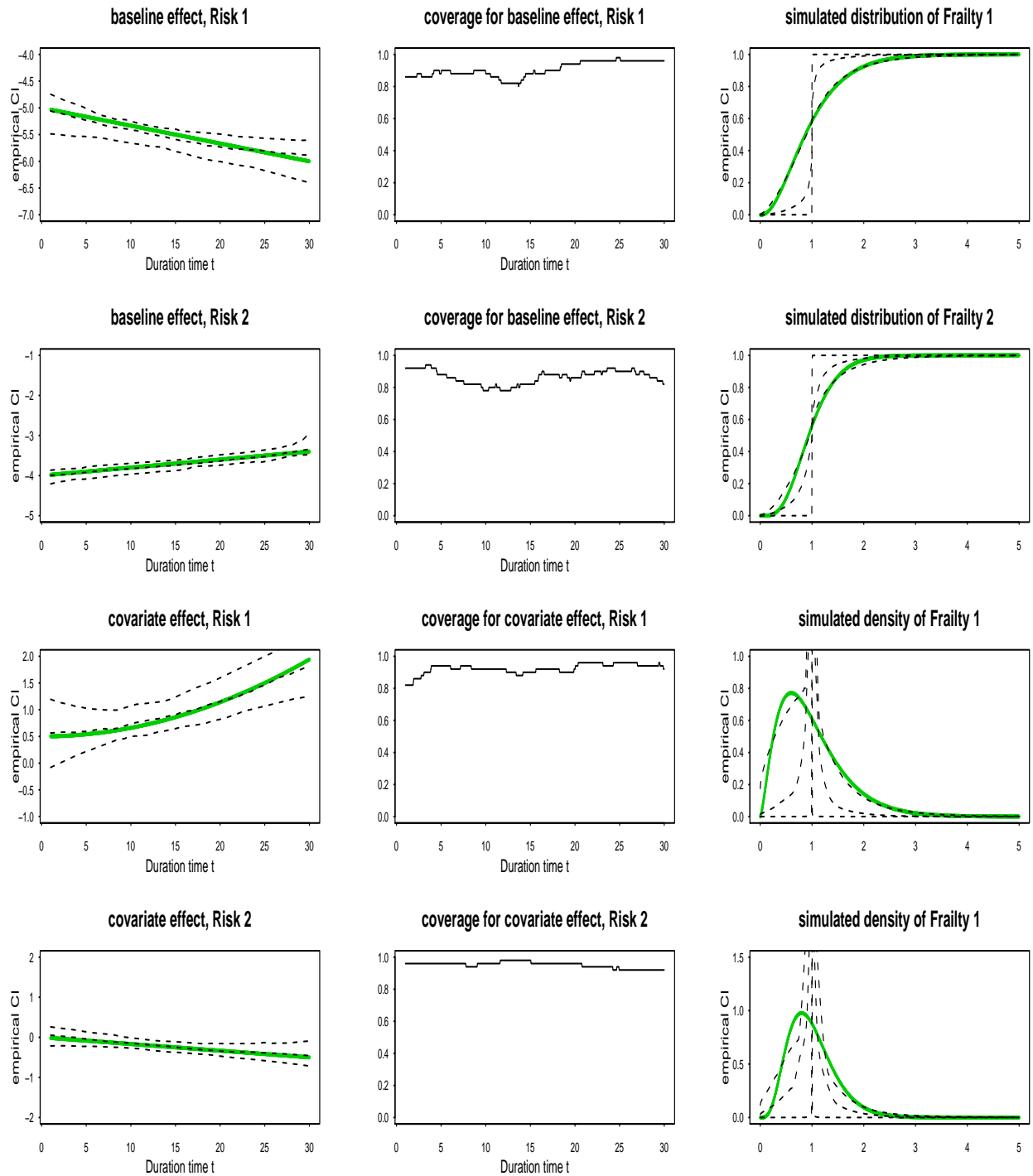


Abbildung B.39: Simulierte glatte Komponenten, coverage probabilities und Frailties. 1000 Individuen, davon ca. 50% zensiert. Häufigkeiten der Spells wie im Anwendungsbeispiel, Kapitel 8.

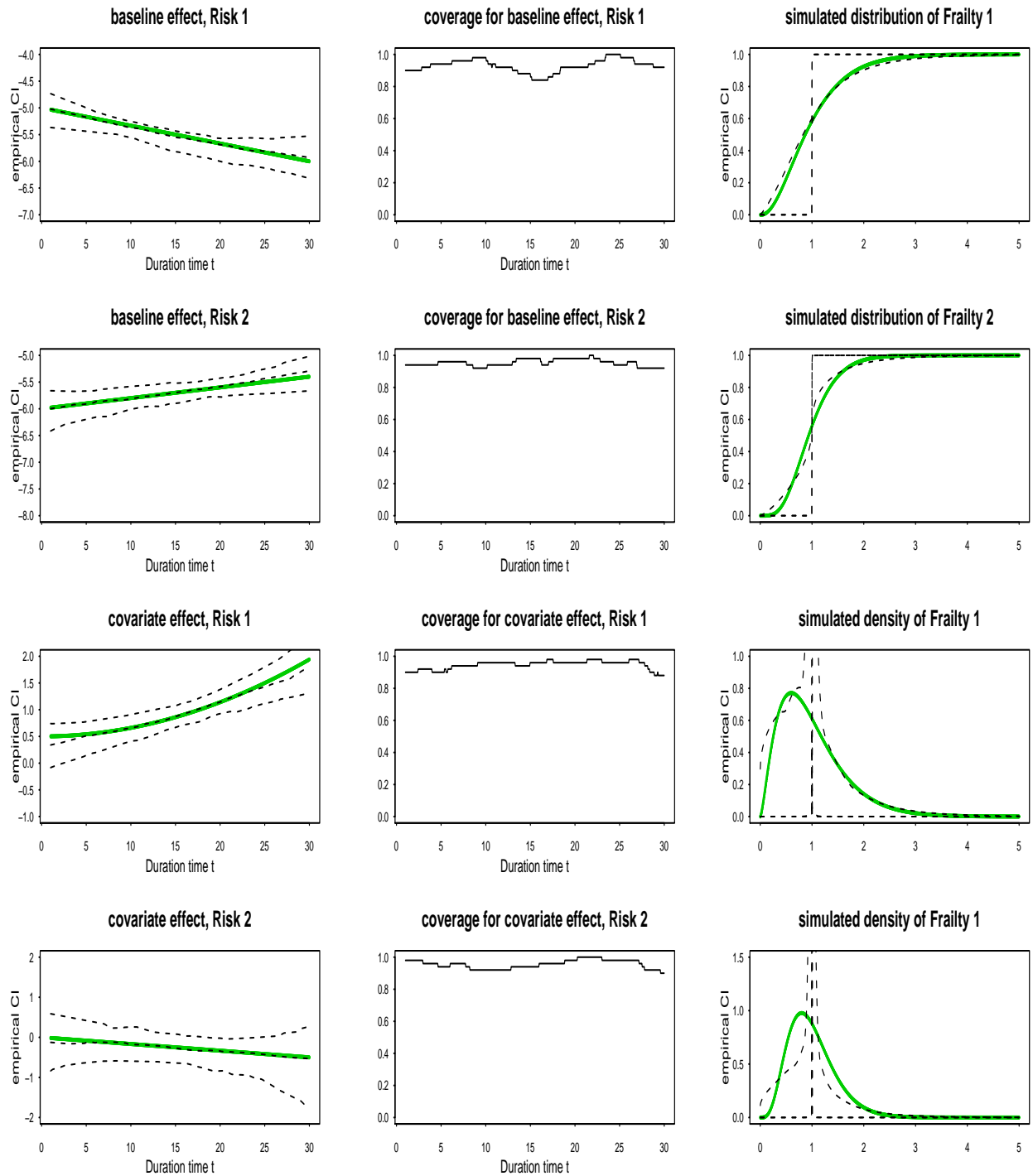


Abbildung B.40: Simulierte glatte Komponenten, coverage probabilities und Frailties. 1000 Individuen, davon ca. 75% zensiert. Häufigkeiten der Spells wie im Anwendungsbeispiel, Kapitel 8.

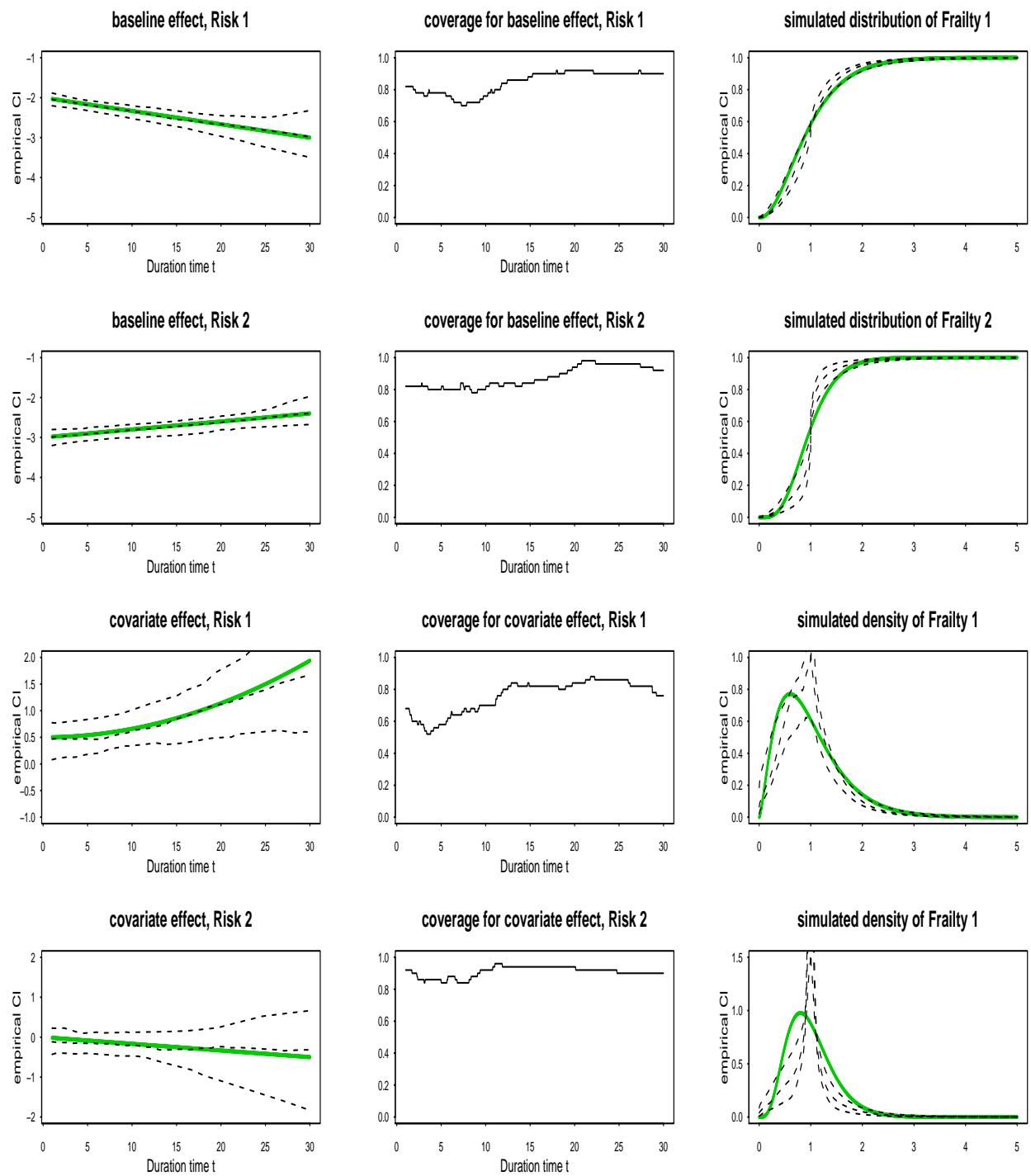


Abbildung B.41: Simulierte glatte Komponenten, coverage probabilities und Frailties. 100 Individuen, davon ca. 5% zensiert, 10 Spells pro Individuum.



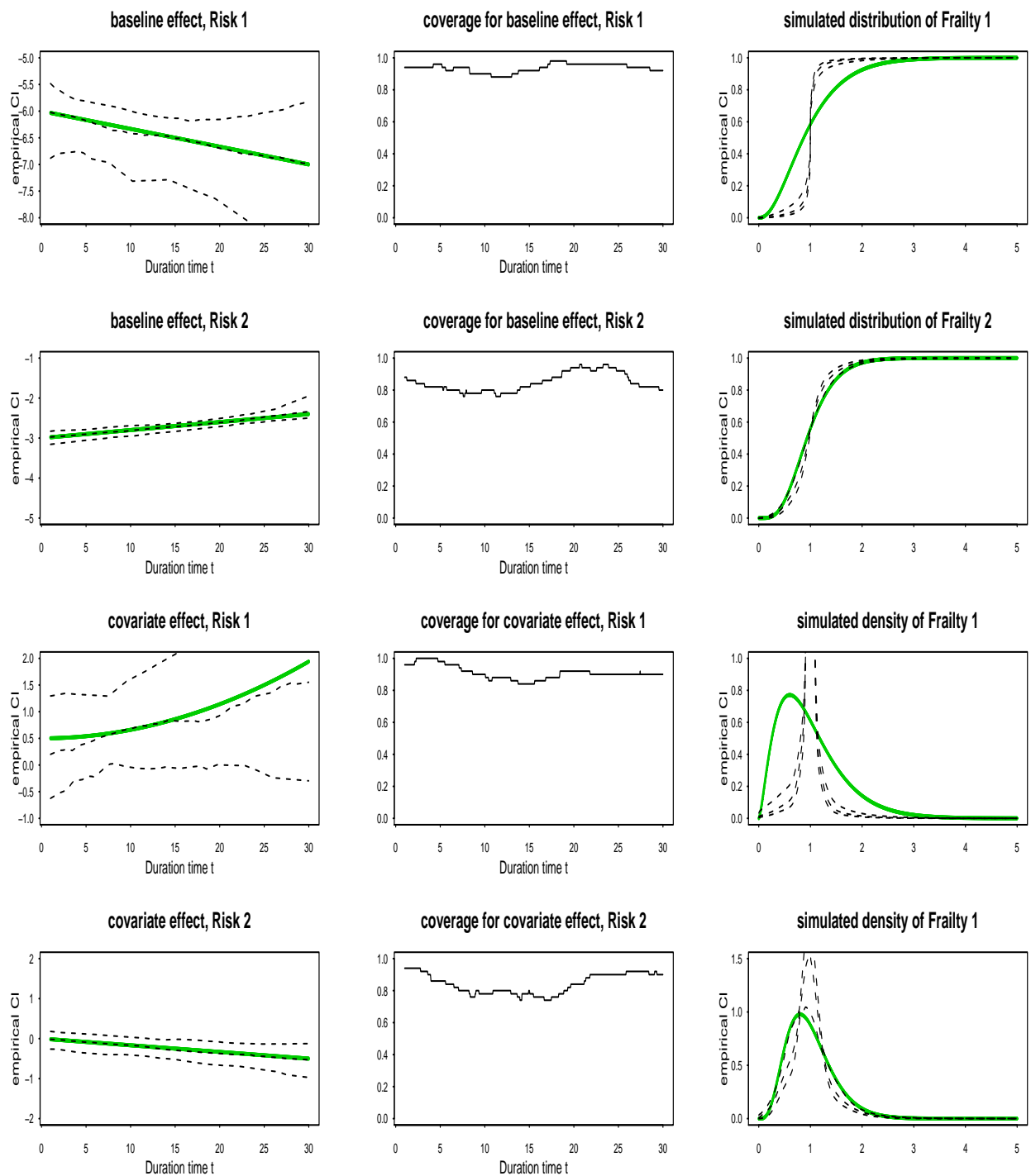


Abbildung B.42: Simulierte glatte Komponenten, coverage probabilities und Frailties. 100 Individuen, davon ca. 20% zensiert, 10 Spells pro Individuum.

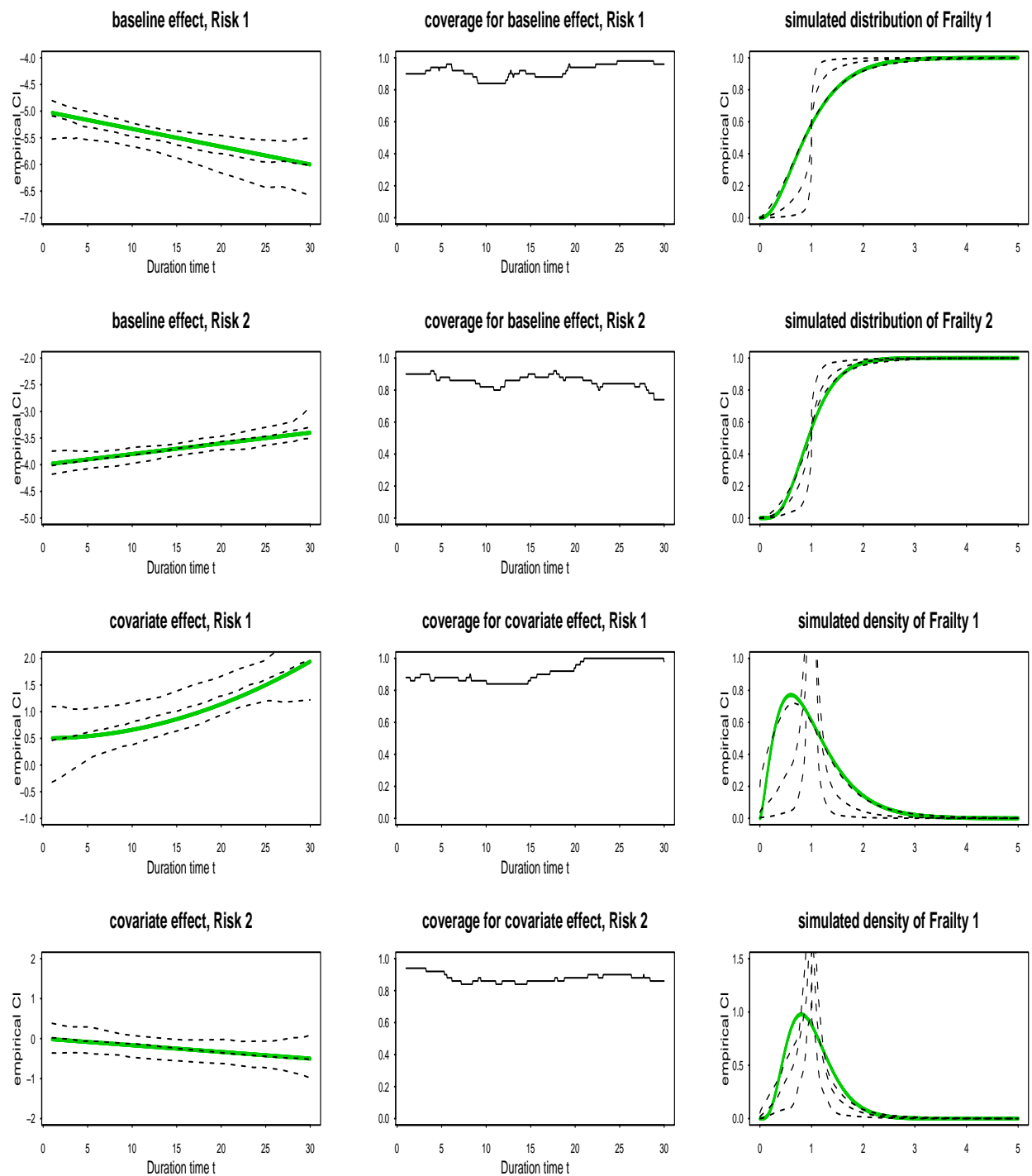


Abbildung B.43: Simulierte glatte Komponenten, coverage probabilities und Frailties. 100 Individuen, davon ca. 50% zensiert, 10 Spells pro Individuum.

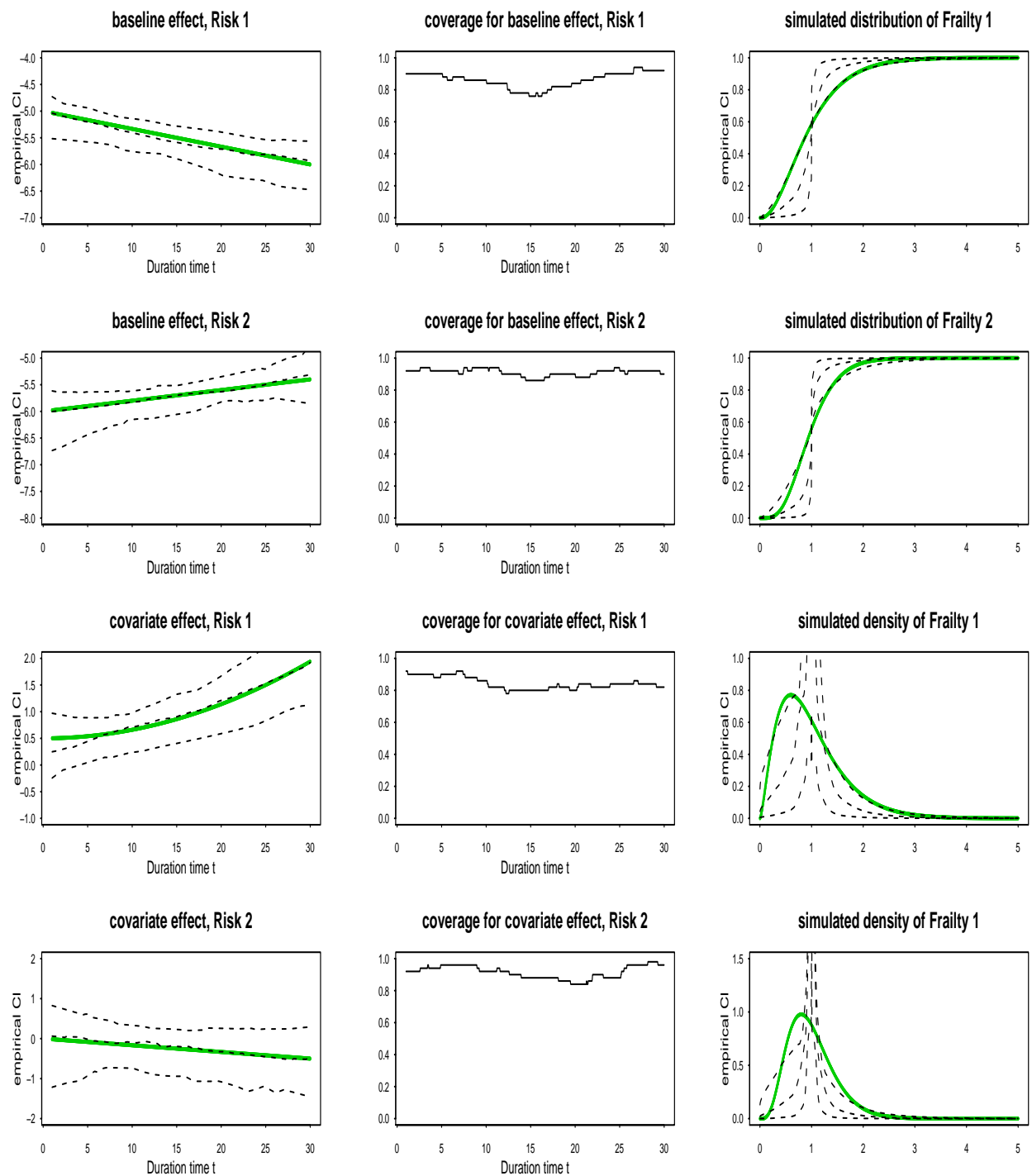


Abbildung B.44: Simulierte glatte Komponenten, coverage probabilities und Frailties. 100 Individuen, davon ca. 75% zensiert, 10 Spells pro Individuum.

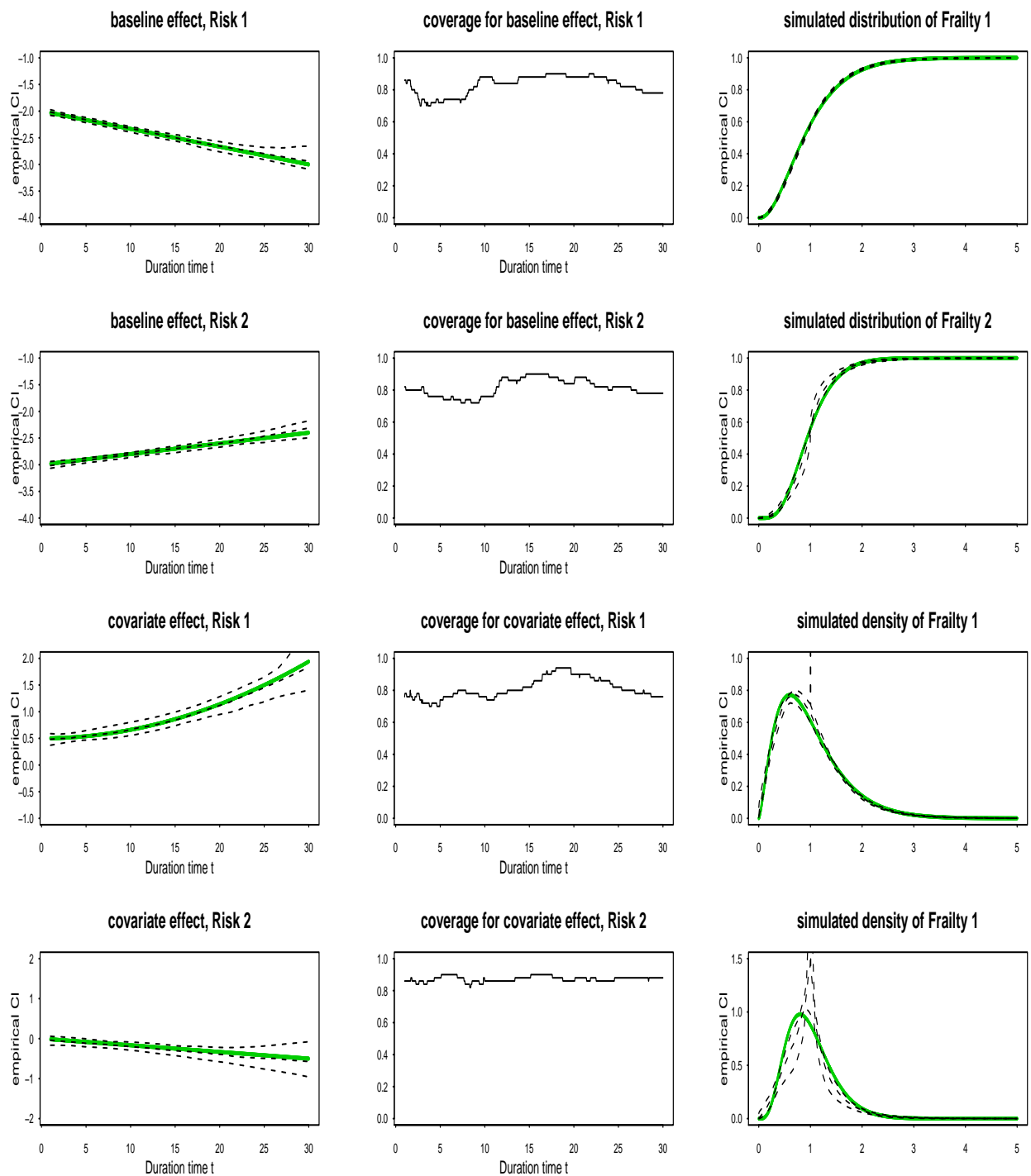


Abbildung B.45: Simulierte glatte Komponenten, coverage probabilities und Frailties. 1000 Individuen, davon ca. 5% zensiert, 10 Spells pro Individuum.

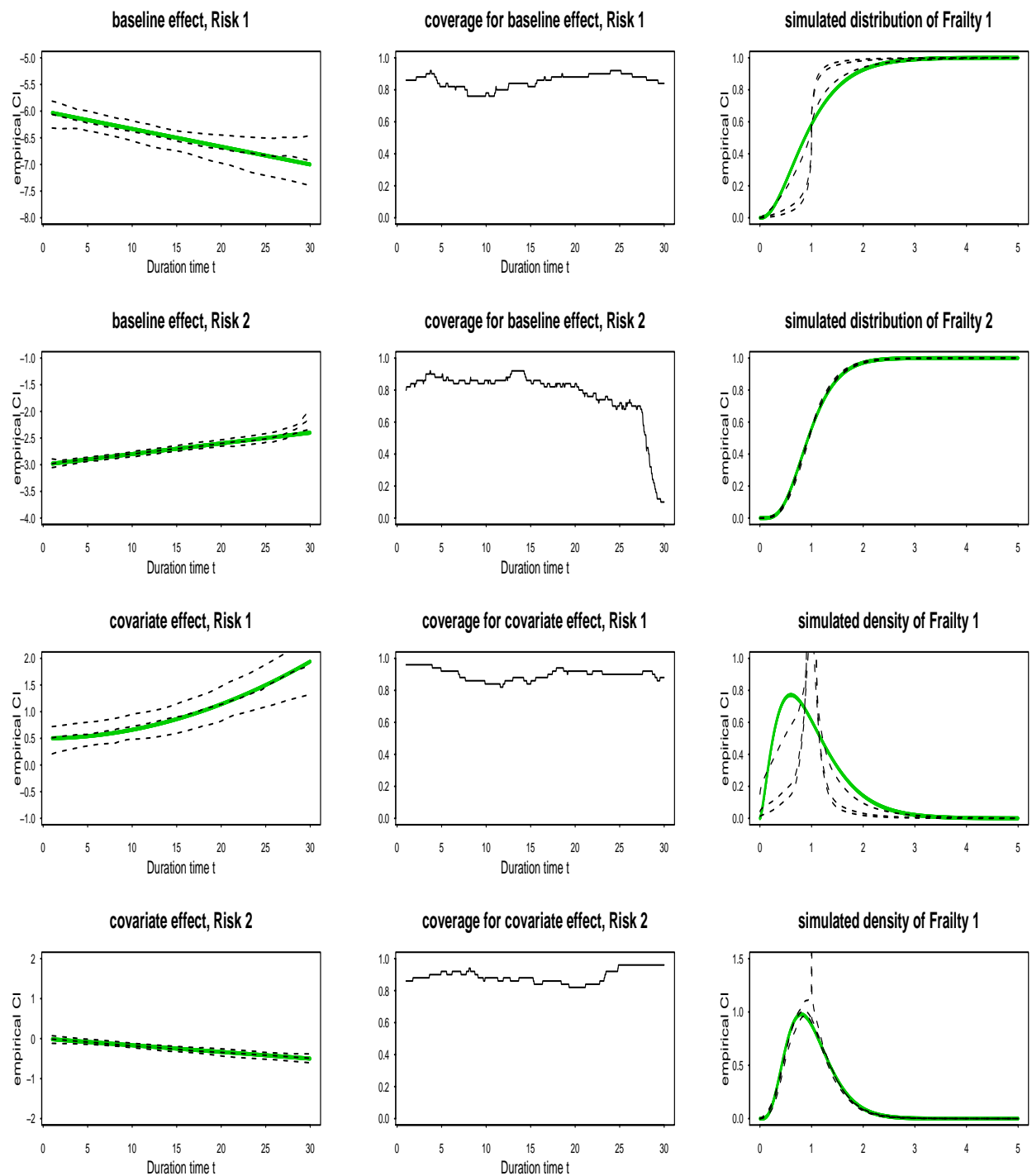


Abbildung B.46: Simulierte glatte Komponenten, coverage probabilities und Frailties. 1000 Individuen, davon ca. 20% zensiert, 10 Spells pro Individuum.

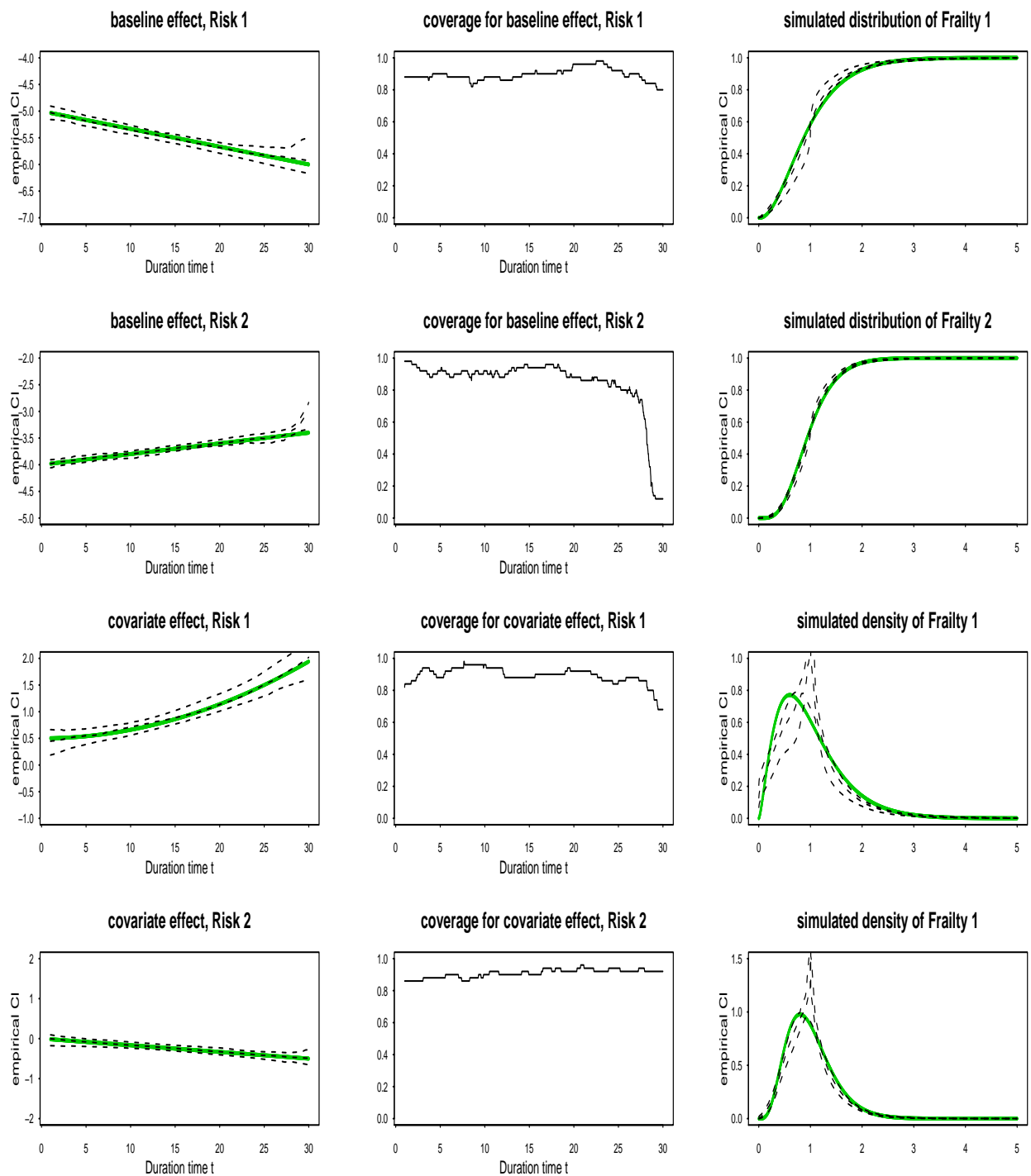


Abbildung B.47: Simulierte glatte Komponenten, coverage probabilities und Frailties. 1000 Individuen, davon ca. 50% zensiert, 10 Spells pro Individuum.

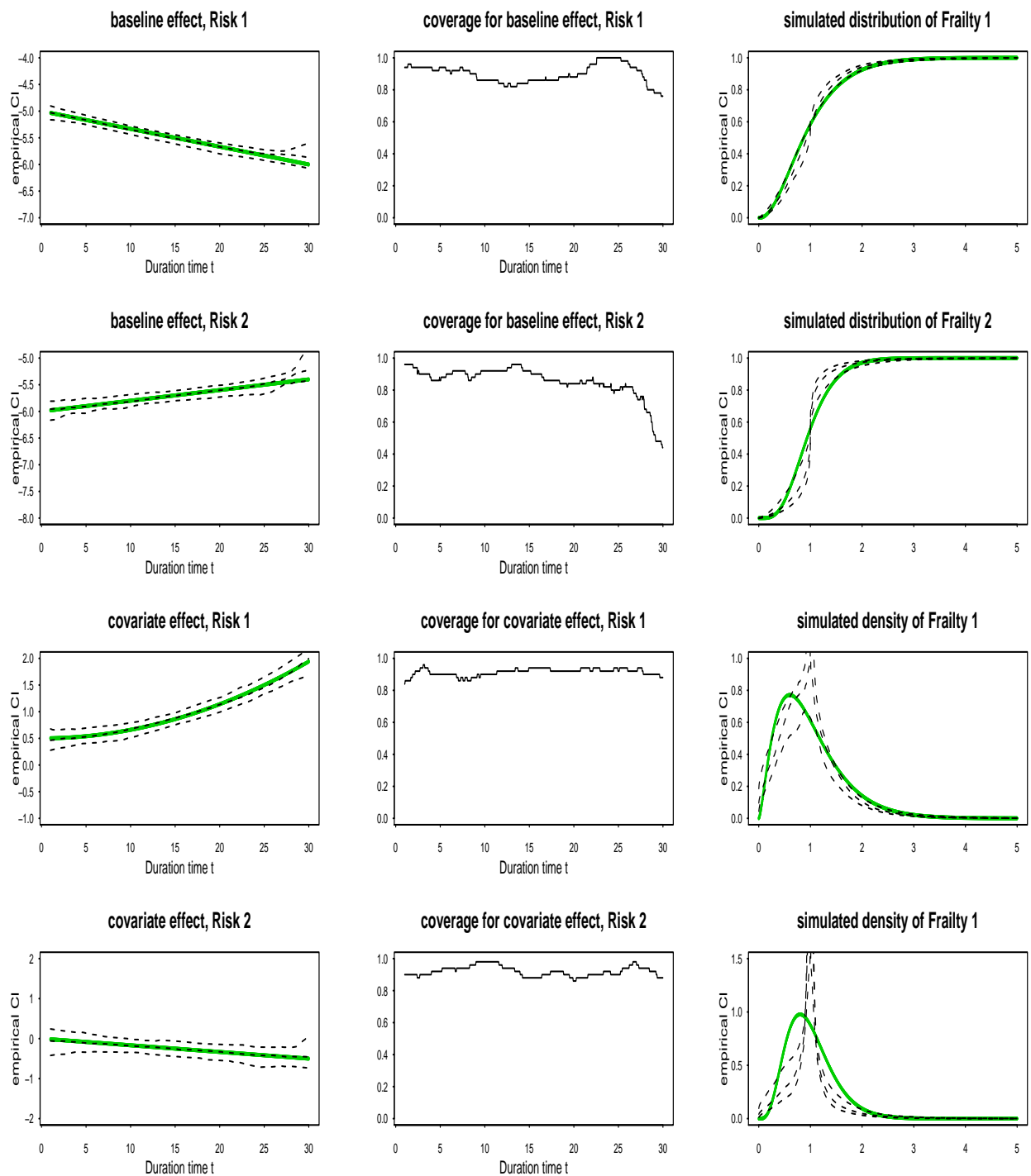


Abbildung B.48: Simulierte glatte Komponenten, coverage probabilities und Frailties. 1000 Individuen, davon ca. 75% zensiert, 10 Spells pro Individuum.

### B.3.2 Simulation auf Tagesbasis

Hier simulieren wir die diskreten ganzzahligen Überlebenszeiten aus dem Intervall  $[1, 900]$  und interpretieren diese Werte als Tage, d. h. wir verfeinern das Intervall  $[1, 30]$ , interpretiert als Monate; Jeder Monat besteht dabei aus 30 Tagen. Für  $L = 2$  Competing Risks ist die Wahl der zu simulierenden glatten Komponenten wie folgt festgelegt:

$$\alpha_{01}^t(t) = \eta_1 - \frac{t/30}{30}, \quad \alpha_{02}^t(t) = \eta_2 + \frac{t/30}{50},$$

$$\alpha_{x1}^t(t) = 0.5 + \left(\frac{t/30}{25}\right)^2, \quad \alpha_{x2}^t(t) = -\frac{t/30}{60},$$

wobei die Parameter  $\eta_1, \eta_2$  im Ausdruck für die logarithmierte Baseline-Hazard gerade die relative Häufigkeit der zensierten Beobachtungen steuern. Die Intensitäten jeweiliger zu simulierender Poisson-Daten sind  $\xi_1 = \frac{1}{30} \exp\{\alpha_{01}^t(t) + x \alpha_{x1}^t(t)\}w_1$  und  $\xi_2 = \frac{1}{30} \exp\{\alpha_{02}^t(t) + x \alpha_{x2}^t(t)\}w_2$  mit simulierten Frailty-Werten  $w_1$  und  $w_2$ . Die folgenden eingesetzten Werte für  $\eta_1, \eta_2$  entsprechen den Zensierungsmustern  $h_\delta$ :

$$\begin{aligned} \eta_1 = -2, \eta_2 = -3 &\rightarrow h_\delta \simeq 5\%, \\ \eta_1 = -6, \eta_2 = -3 &\rightarrow h_\delta \simeq 20\%, \\ \eta_1 = -5, \eta_2 = -4 &\rightarrow h_\delta \simeq 50\%, \\ \eta_1 = -5, \eta_2 = -6 &\rightarrow h_\delta \simeq 75\%. \end{aligned}$$

Weitere Simulationsparameter sind:

- *Frailty-Verteilung.* Für  $L=2$  unabhängige Gamma-Frailties mit jeweils *shape*-Parameter =2.5= *scale*-Parameter (für Risk 1) und *shape*-Parameter =5=*scale*-Parameter (für Risk 2) werden simuliert.
- *Anzahl der Individuen  $N$ ,* jeweils 100 oder 1000.
- *Anzahl der Spells pro Individuum.* Zwei Muster stehen zur Wahl: jeweils 10 Spells pro Individuum und Spell-Verteilung ähnlich der wie im Anwendungsbeispiel, z. B. für 1000 Individuen (vgl. mit Tabelle 8.1):

number of spells	1	2	3	4	5
number of individuals	720	170	70	30	10
proportion	0.72	0.17	0.07	0.03	0.01

Auf der Tagesbasis geschätzt, werden in den darauffolgenden Abbildungen die glatten Komponenten auf der (gröberen) Monatsskala gezeichnet, zum Vergleich mit den jeweiligen Schätzern aus B.3.1, durchgeführt auf Monatsbasis. Näher ist das Simulationsverfahren in Abschnitt 8.9 erläutert. Als Ergebnis jeweiliger Simulationsserie resultieren die geschätzten empirischen 5%-, 50%-, 95%- Quantile sowie *coverage probabilities*. Außerdem sind für Frailty-Verteilung und -Dichte die empirischen 95%- Konfidenzbänder und medialer Wert gezeichnet (vgl. außerdem Kauermann & Khomski, 2007, für weitere Simulationen).



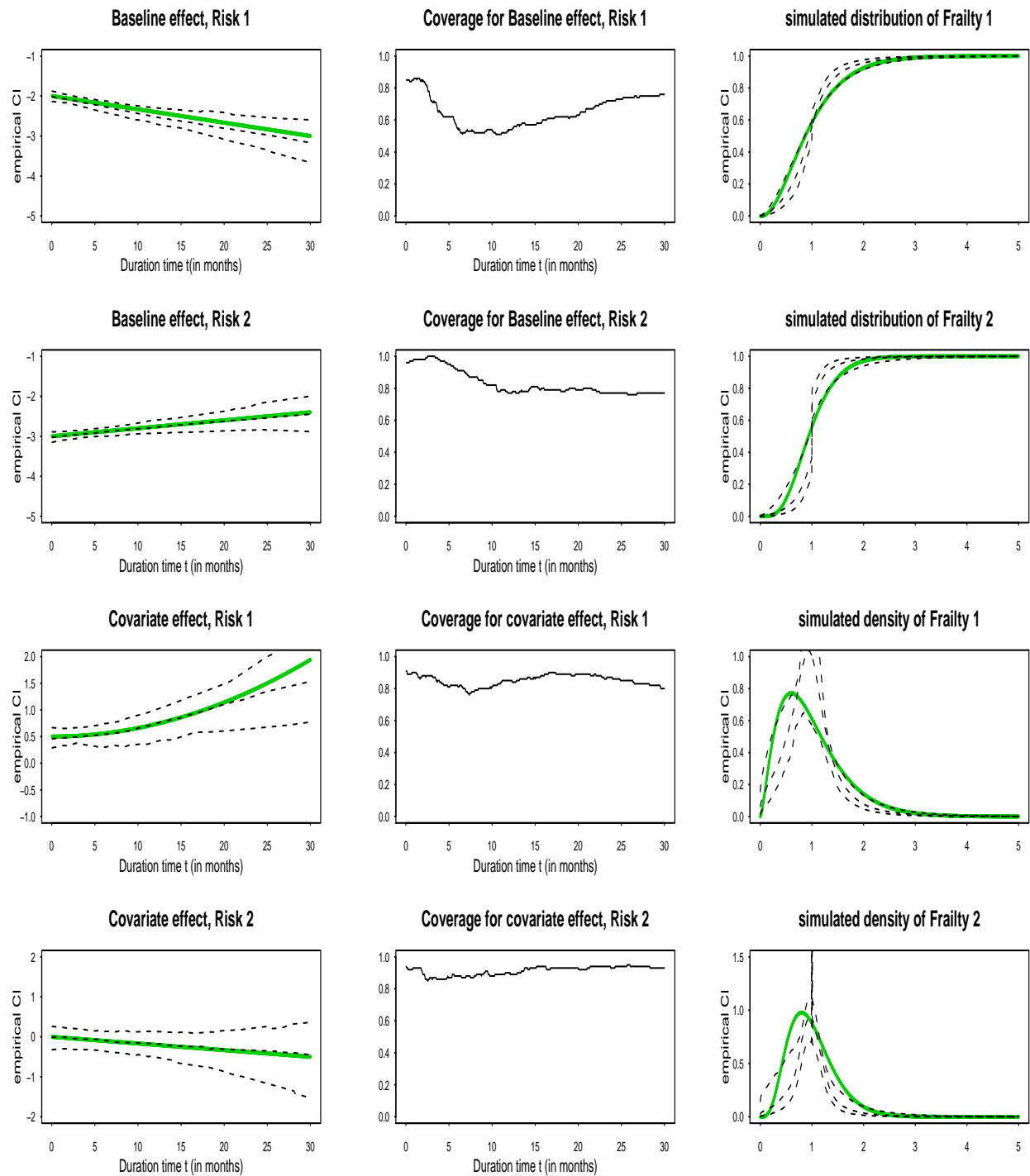


Abbildung B.49: Simulierte glatte Komponenten (geschätzt auf Tagesbasis, dargestellt in Monaten), coverage probabilities und Frailties. 1000 Individuen, davon ca. 5% zensiert. Häufigkeiten der Spells wie im Anwendungsbeispiel, Kapitel 8.

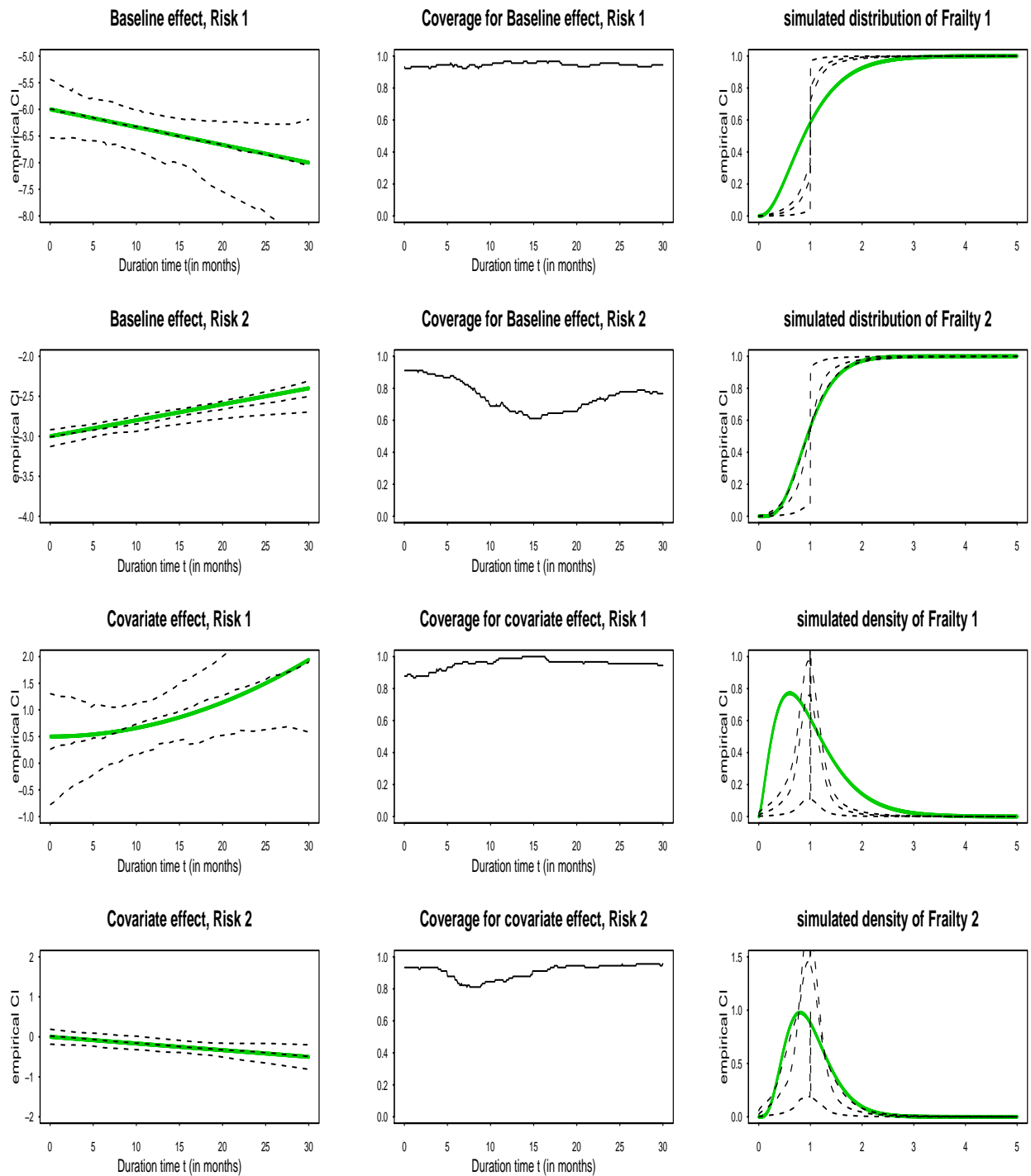


Abbildung B.50: Simulierte glatte Komponenten (geschätzt auf Tagesbasis, dargestellt in Monaten), coverage probabilities und Frailties. 1000 Individuen, davon ca. 20% zensiert. Häufigkeiten der Spells wie im Anwendungsbeispiel, Kapitel 8.

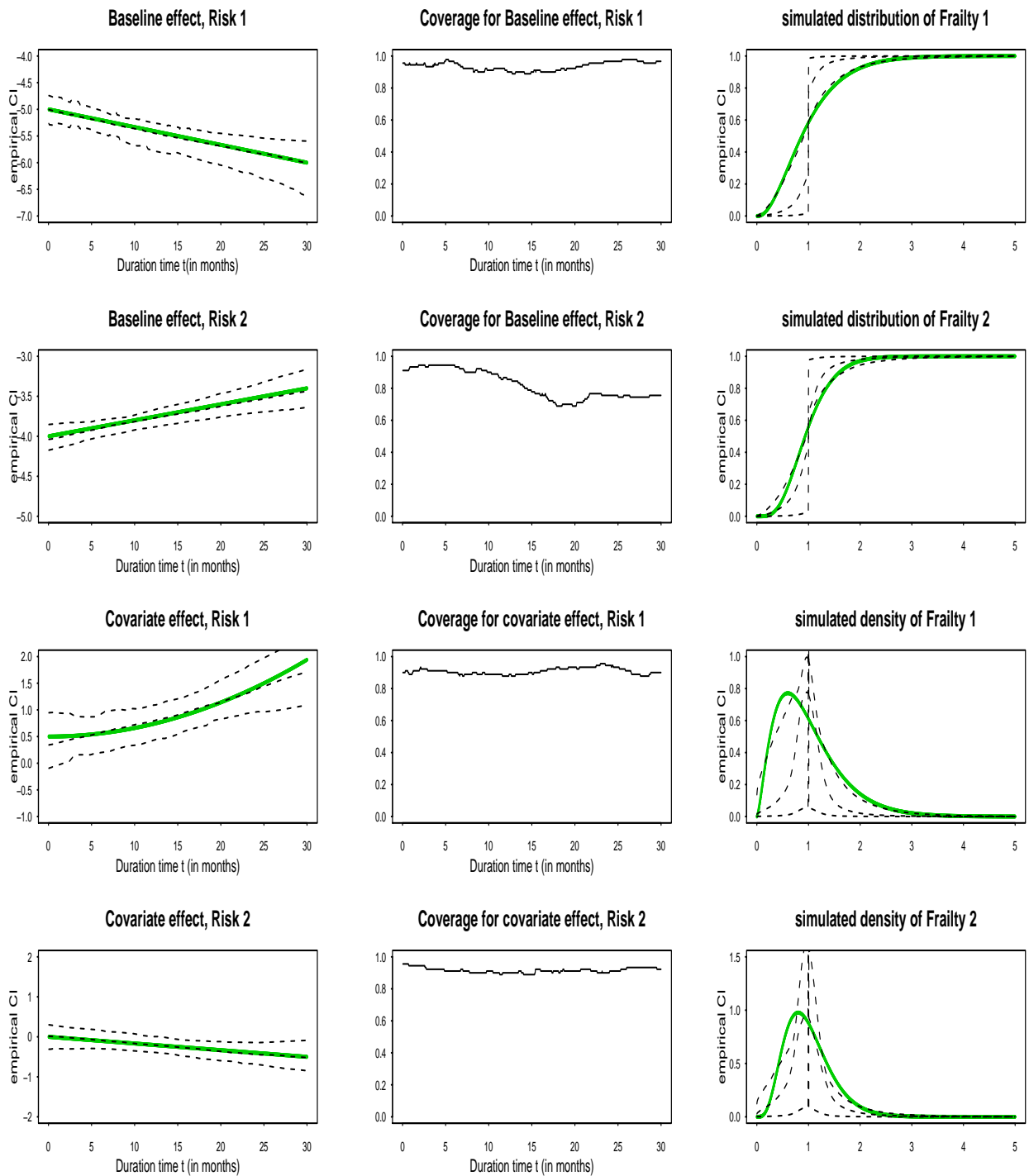


Abbildung B.51: Simulierte glatte Komponenten (geschätzt auf Tagesbasis, dargestellt in Monaten), coverage probabilities und Frailties. 1000 Individuen, davon ca. 50% zensiert. Häufigkeiten der Spells wie im Anwendungsbeispiel, Kapitel 8.

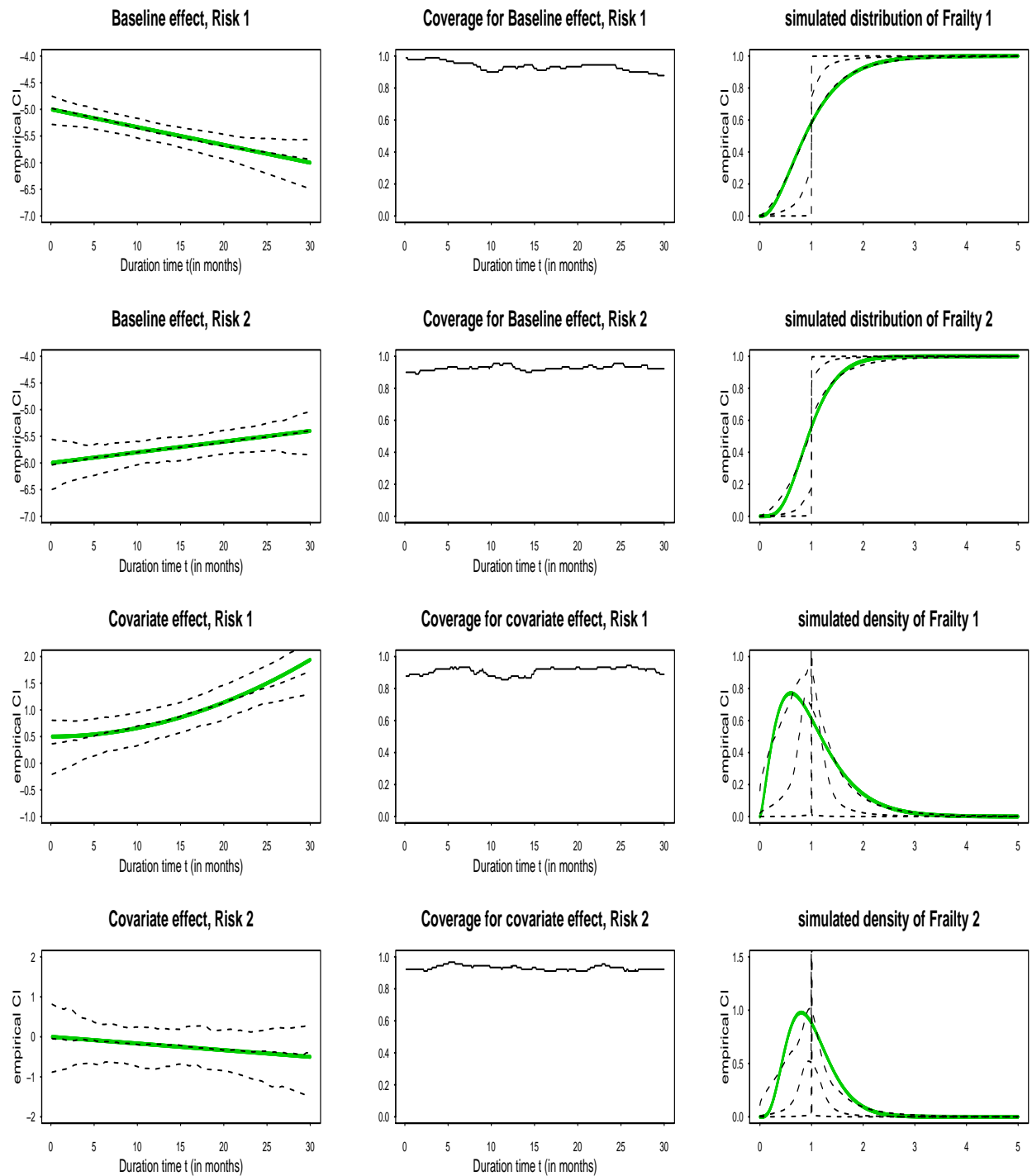


Abbildung B.52: Simulierte glatte Komponenten (geschätzt auf Tagesbasis, dargestellt in Monaten), coverage probabilities und Frailties. 1000 Individuen, davon ca. 75% zensiert. Häufigkeiten der Spells wie im Anwendungsbeispiel, Kapitel 8.

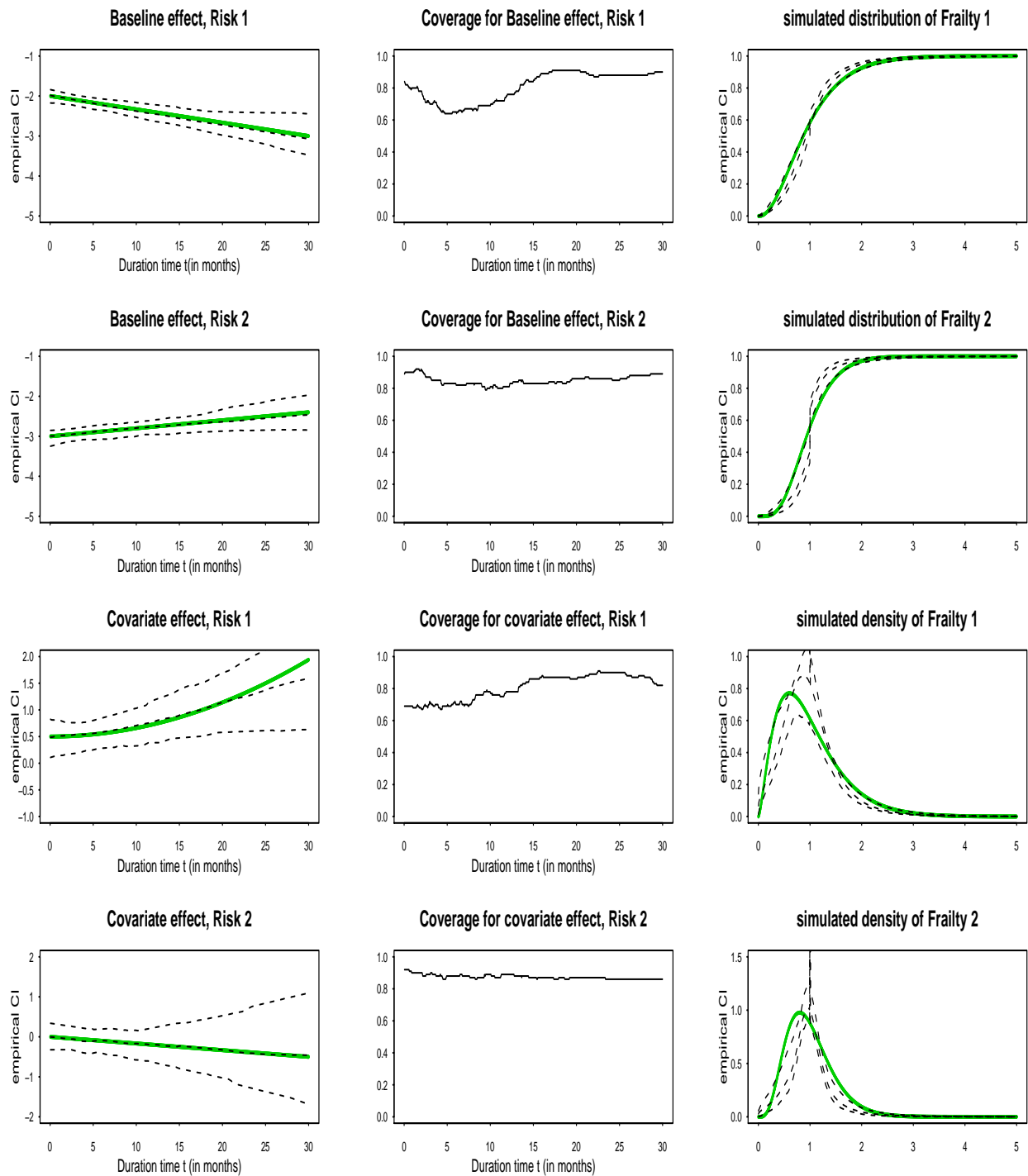


Abbildung B.53: Simulierte glatte Komponenten (geschätzt auf Tagesbasis, dargestellt in Monaten), coverage probabilities und Frailties. 100 Individuen, davon ca. 5% zensiert, 10 Spells pro Individuum.

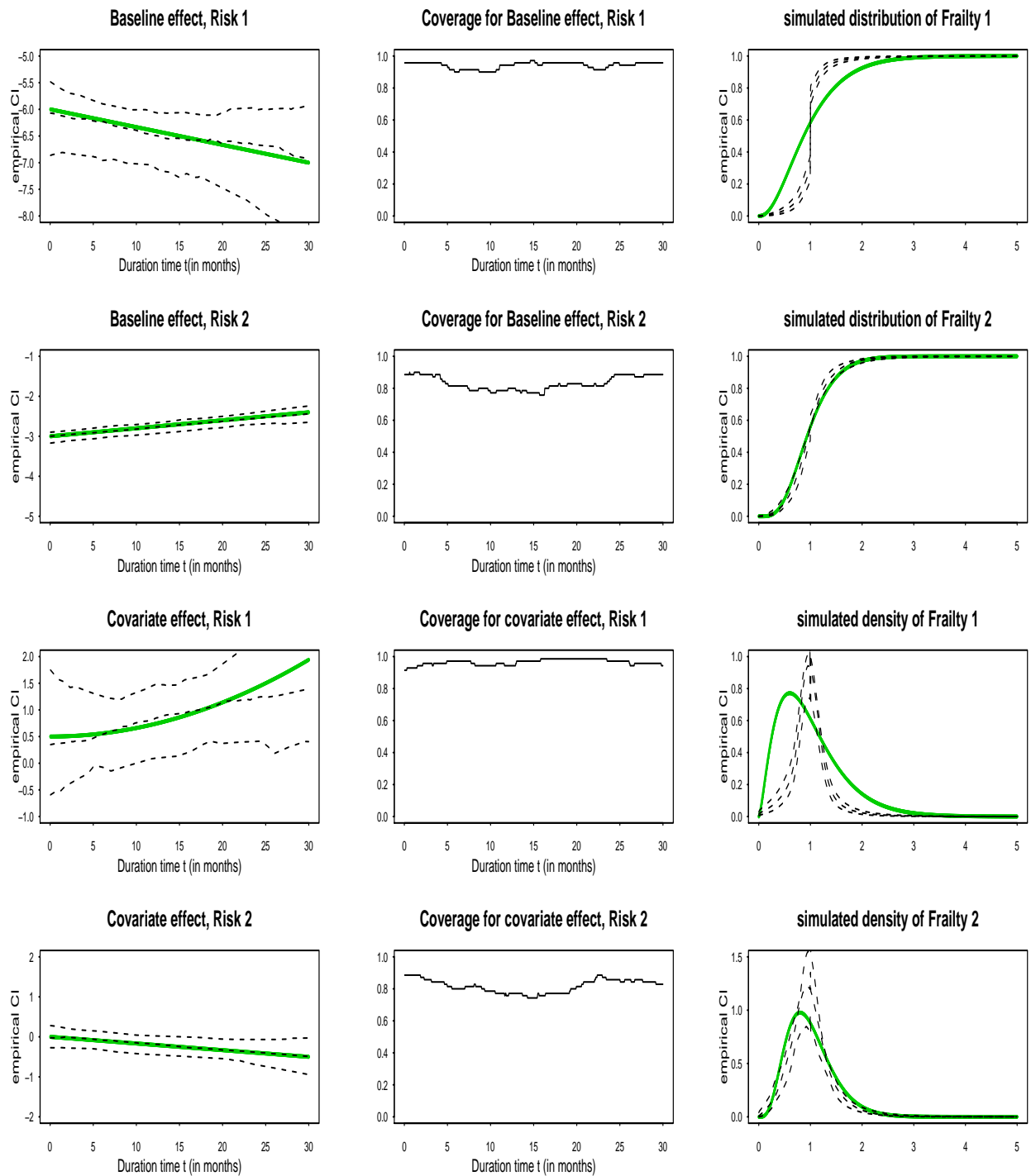


Abbildung B.54: Simulierte glatte Komponenten (geschätzt auf Tagesbasis, dargestellt in Monaten), coverage probabilities und Frailties. 100 Individuen, davon ca. 20% zensiert, 10 Spells pro Individuum.

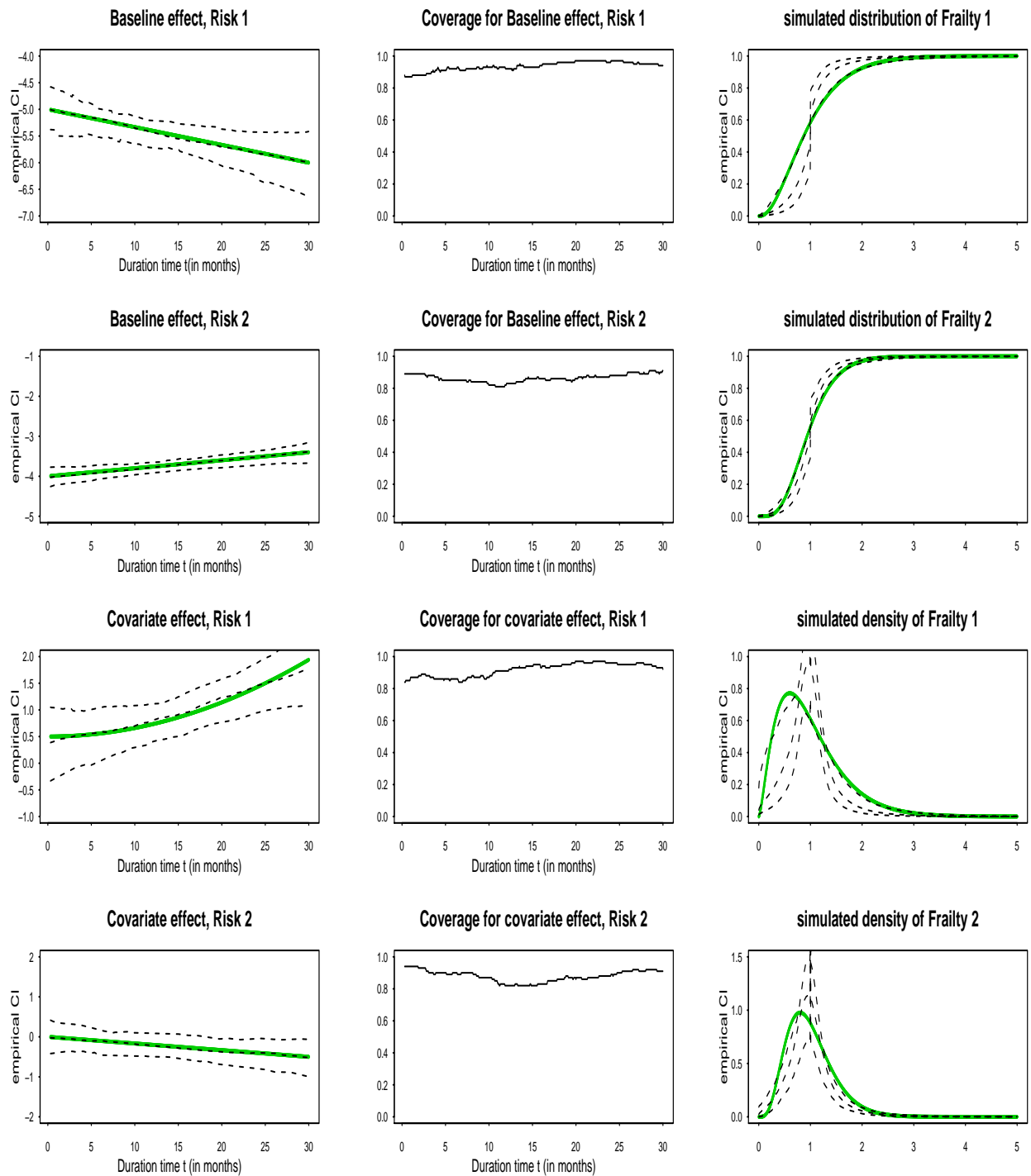


Abbildung B.55: Simulierte glatte Komponenten (geschätzt auf Tagesbasis, dargestellt in Monaten), coverage probabilities und Frailties. 100 Individuen, davon ca. 50% zensiert, 10 Spells pro Individuum.

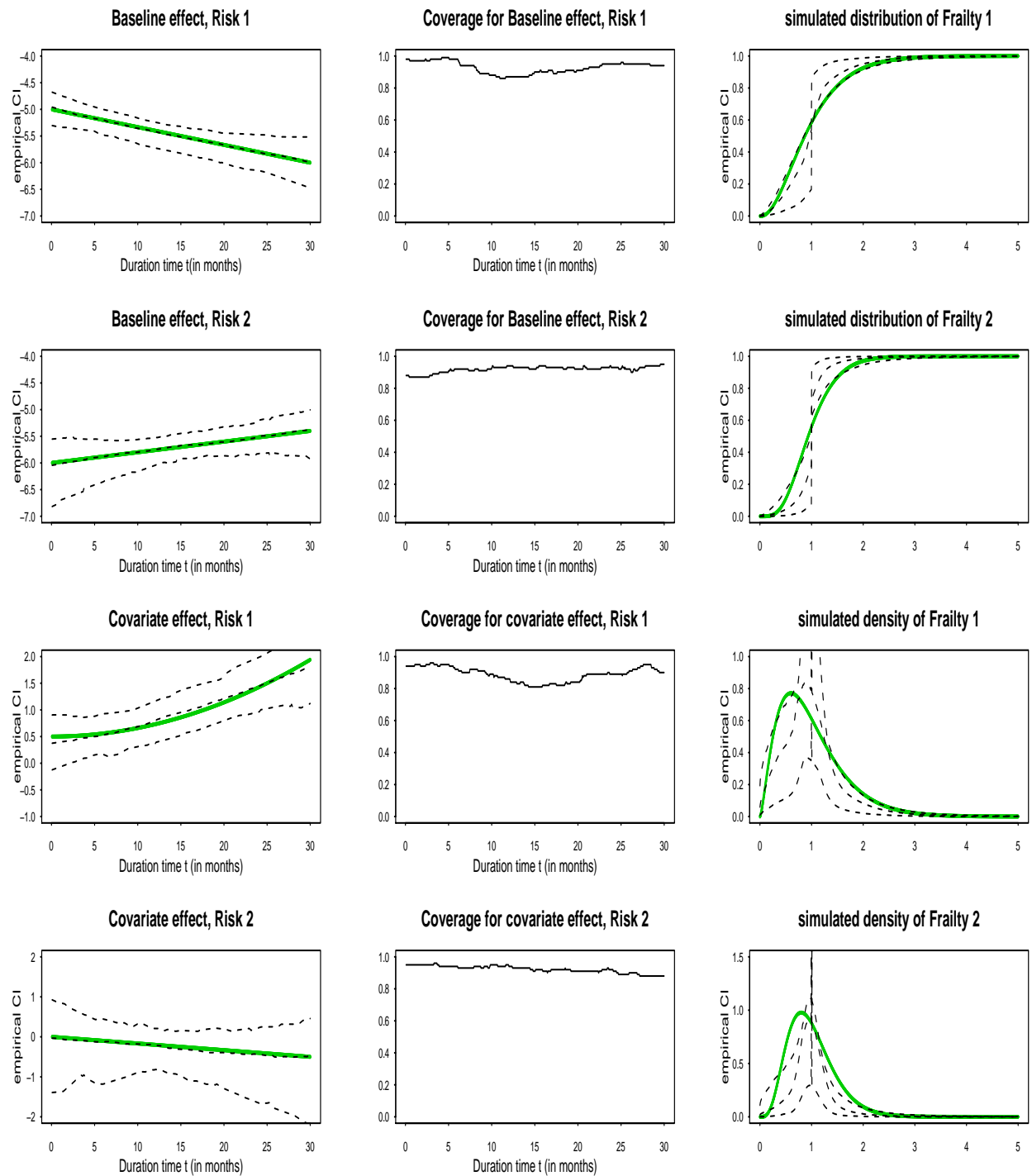


Abbildung B.56: Simulierte glatte Komponenten (geschätzt auf Tagesbasis, dargestellt in Monaten), coverage probabilities und Frailties. 100 Individuen, davon ca. 75% zensiert, 10 Spells pro Individuum.



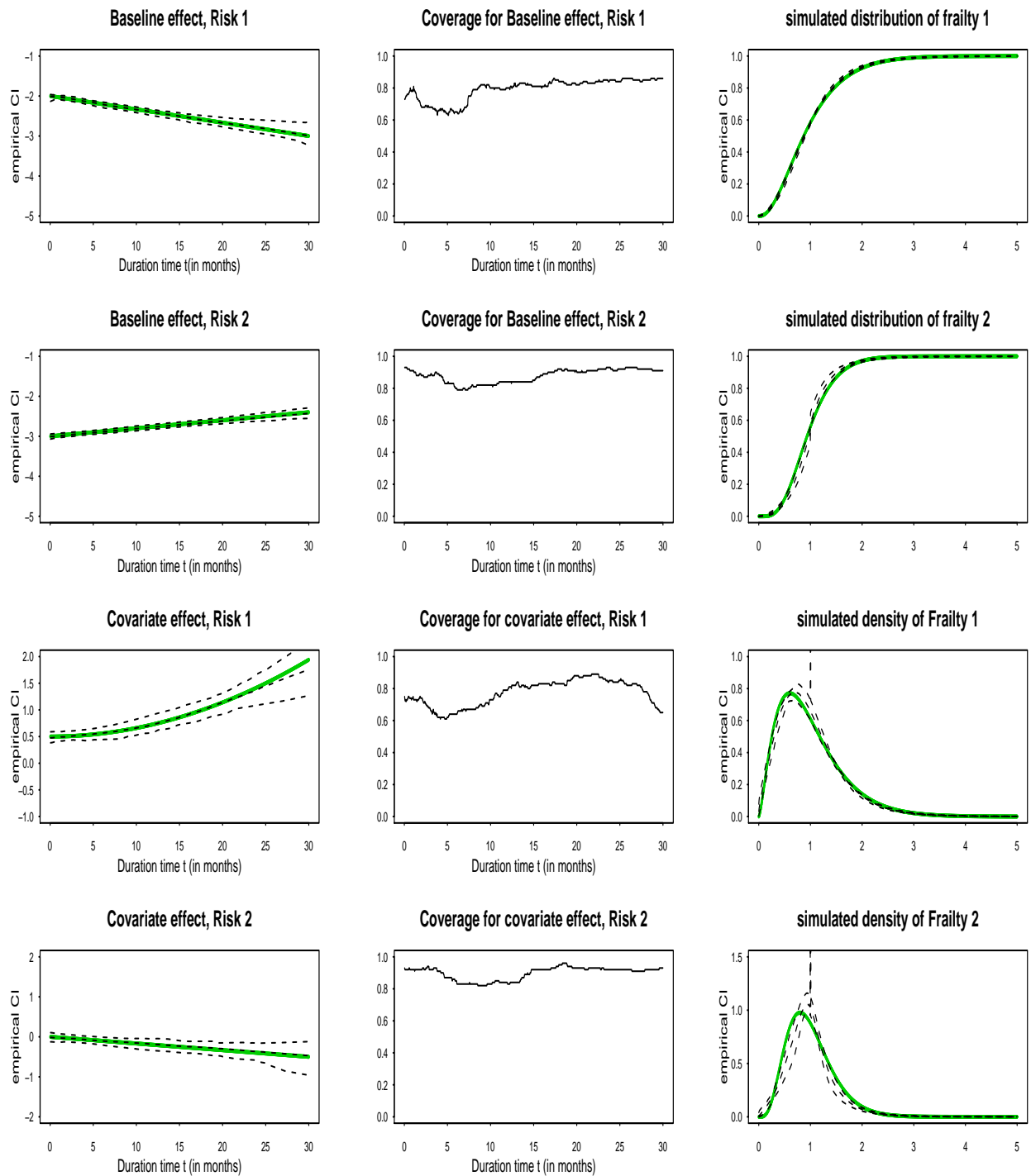


Abbildung B.57: Simulierte glatte Komponenten (geschätzt auf Tagesbasis, dargestellt in Monaten), coverage probabilities und Frailties. 1000 Individuen, davon ca. 5% zensiert, 10 Spells pro Individuum.

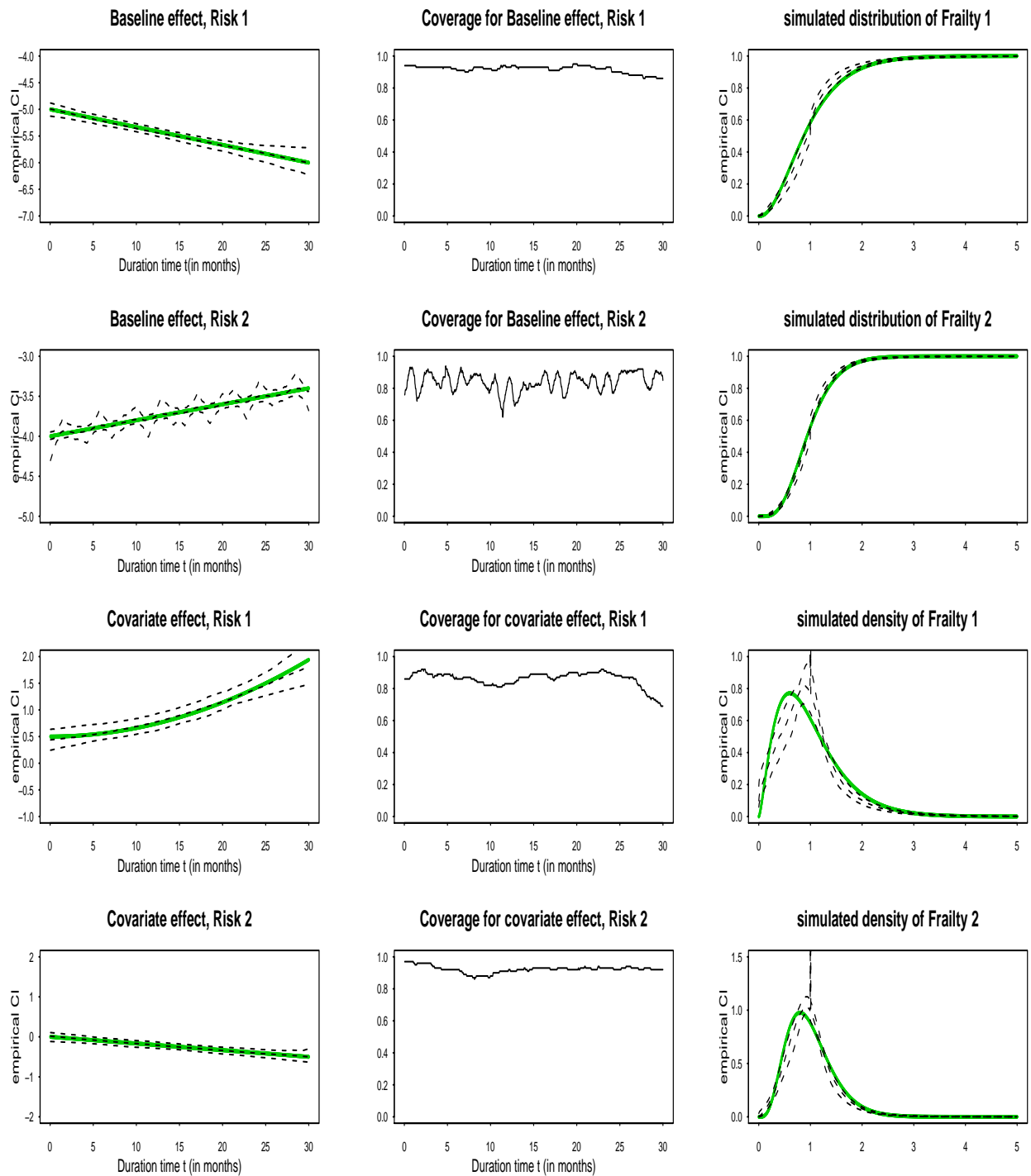


Abbildung B.58: Simulierte glatte Komponenten (geschätzt auf Tagesbasis, dargestellt in Monaten), coverage probabilities und Frailties. 1000 Individuen, davon ca. 50% zensiert, 10 Spells pro Individuum.

# Literatur

- Aalen, O. (1988). Heterogeneity in survival analysis. *Statistics in Medicine* **7**, 1121–1137.
- Aalen, O. (1994). Effects of frailty in survival analysis. *Statistical Methods in Medical Research* **3**, 227–243.
- Abramovitz, M. und Stegun, I. A. (1965). *Handbook of mathematical functions*. New York: Dover Publications.
- Aerts, M., Claeskens, G., und Wand, M. P. (2002). Some theory for penalized spline generalized additive models. *Journal of Statistical Planning and Inference* **103**, 455–470.
- Akaike, H. (1974). A new look at the statistical model identification. *IEEE Transactions on Automatic Control* **AC 19**, 716–723.
- Aldrin, M. (2006). Improved predictions penalizing both slope and curvature in additive models. *Computational Statistics and Data Analysis* **50**, 267–284.
- Andersen, P. K., Borgan, O., Gill, R. D., und Keiding, N. (1993). *Statistical Models Based on Counting Processes*. New York: Springer.
- Becher, H., Kauermann, G., und Khomski, P. (2007). Using spline regression to model age- and season-of-birth-dependent effects of childhood mortality risk factors in rural Burkina Faso. *submitted for publication*.
- Becher, H. und Kouyatè, B. (Hrsg.) (2005). *Health Research in Developing Countries. A collaboration between Burkina Faso and Germany*. Heidelberg: Springer.
- Becher, H., Müller, O., Jahn, A., Gbangou, A., Kynast-Wolf, G., und Kouyaté, B. (2004). Risk factors of infant and child mortality in rural Burkina Faso. *Bulletin of the World Health Organization* **82**, 265–273.
- Blossfeld, H.-P. und Hamerle, A. (1987). Unobserved heterogeneity in hazard rate models: A test and an illustration from a study of career mobility. Forschungsbericht 253, J.W. Goethe-Universität Frankfurt und Universität Mannheim. Sonderforschungsbereich 3: Mikroanalytische Grundlagen der Gesellschaftspolitik.
- de Boor, C. (2001). *A Practical Guide to Splines*. New York: Springer-Verlag.
- Boyd, S. und Vandenberghe, L. (2004). *Convex Optimization*. Cambridge University Press.
- Breslow, N. E. (1972). Comment on "regression and life tables" by D. R. Cox. *Journal of the Royal Statistical Society, Series B* **34**, 216–217.

- Breslow, N. E. (1974). Covariance analysis of censored survival data. *Biometrics* **30**, 89–99.
- Breslow, N. E. und Clayton, D. G. (1993). Approximate inference in generalized linear mixed model. *Journal of the American Statistical Association*. **88**, 9–25.
- Brüderl, J., Preisendörfer, P., und Ziegler, R. (1998). *Der Erfolg neugegründeter Betriebe* (2 Aufl.). Berlin: Duncker & Humboldt.
- Brüderl, J., Preisendörfer, P., und Ziegler, R. (1992). Survival chances of newly founded business organizations. *American Sociological Review* **57**, 227–242.
- Brumback, B. A., Ruppert, D., und Wand, M. P. (1999). Comment on Shively, Kohn and Wood. *Journal of the American Statistical Association*. **94**, 794–797.
- Burnham, K. P. und Anderson, D. (2002). *Model Selection and Multi-Model Inference* (2 Aufl.). New York: Springer-Verlag.
- Cai, T., Hyndman, R., und Wand, M. (2002). Mixed model-based hazard estimation. *Journal of Computational and Graphical Statistics* **11**, 784 – 798.
- Carlin, B. P. und Louis, T. A. (2000). *Bayes and Empirical Bayes Methods for Data Analysis* (2 Aufl.). London: Chapman & Hall/CRC.
- Cox, D. R. (1972). Regression models and life tables (with discussion). *Journal of the Royal Statistical Society, Series B* **34**, 187–220.
- Cox, D. R. (1975). Partial likelihood. *Biometrika* **62**, 187–220.
- Cox, D. R. und Oakes, D. (1984). *Analysis of Survival Data*. London: Chapman and Hall.
- Crainiceanu, C. M., Ruppert, D., Claeskens, G., und Wand, M. P. (2005). Exact likelihood ratio tests for penalized splines. *Biometrika* **92**, 91–103.
- Crainiceanu, C. M., Ruppert, D., und Wand, M. P. (2005). Bayesian analysis for penalized spline regression using WinBUGS. *Journal of Statistical Software* **14**(11), 1–24.
- Craven, P. und Wahba, G. (1979). Smoothing noisy data with spline functions: Estimating the correct degree of smoothing by the method of generalized cross-validation. *Numerische Mathematik* **31**, 377–403.
- Crowder, M. J. (2001). *Classical Competing Risks*. Chapman & Hall.
- Currie, I. D., Durban, M., und Eilers, P. (2004). Smoothing and forecasting mortality rates. *Statistical Modelling* **4**, 279–298.
- Currie, I. D., Durban, M., und Eilers, P. H. C. (2006). Generalized linear array models with applications to multidimensional smoothing. *Journal of the Royal Statistical Society, Series B* **68**, 259–280.
- David, H. A. und Moeschberger, M. L. (1978). *The theory of competing risks*. London: Griffin's Statistical Monographs & Courses.
- Dempster, A. P., Laird, N. M., und Rubin, D. B. (1977). Maximum likelihood estimation from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society, Series B* **39**, 1–38.
- Dierckx, P. (1993). *Curve and Surface Fitting with Splines*. Oxford Science Publications.
- Dobhammer, G. und Vaupel, J. W. (2001). Lifespan depends on month of

- birth. *Proceedings of the National Academy of Sciences of the United States of America* **98**, 2934–2939.
- Duchesne, T. und Lawless, J. (2000). Alternative time scales and failure time models. *Lifetime Data Analysis* **6**, 157–179.
- Efron, B. (1977). The efficiency of Cox' Likelihood function for censored data. *Journal of the American Statistical Association*. **72**, 557–565.
- Efron, B. (2002). The two-way proportional hazards model. *Journal of the Royal Statistical Society, Series B* **64**, 899–909.
- Eilers, P., Currie, I. D., und Durban, M. (2006). Fast and compact smoothing on large multidimensional grids. *Computational Statistics & Data Analysis* **50**, 61–76.
- Eilers, P. und Marx, B. (1996). Flexible smoothing with B-splines and penalties. *Statistical Science* **11**, 89–121.
- Eilers, P. und Marx, B. (2003). Multivariate calibration with temperature interaction using two-dimensional penalized signal regression. *Chemometrics and Intelligent Laboratory Systems* **66**, 159–174.
- Eilers, P. und Marx, B. (2005). Splines, knots and penalties. *to appear*.
- Etezadi-Amoli, J. und Ciampi, A. (1987). Extended hazard regression for censored survival data with covariates: A spline approximation for the baseline hazard function. *Biometrics* **43**, 181–192.
- Eubank, R. L. (1988). *Spline smoothing and nonparametric regression*. Dekker: New York.
- Fahrmeir, L., Kneib, T., und Lang, S. (2004). Penalized structured additive regression for space-time data; a bayesian perspective. *Statistica Sinica* **14**, 715–745.
- Farewell, V. T. und Cox, D. R. (1979). A note on multiple time scales in life testing. *Journal of the Royal Statistical Society, Series C* **28**, 73–75.
- Fletcher, R. (1987). *Practical Methods of Optimization* (2 Aufl.). John Wiley & Sons.
- Francis, B. und Pritchard, J. (1998). Bertin, Lexis and the graphical representation of event histories. *Bulletin du Comite Francais de Cartographie* **156**, 80–87.
- Gamerman, D. (1991). Dynamic bayesian models for survival data. *Applied Statistics* **40**, 63–79.
- Gavrilov, L. A. und Gavrilova, N. S. (1999). Season of birth and human longevity. *Journal of Anti-Aging Medicine* **2**, 365–366.
- Ghidey, W., Lesaffre, P., und Eilers, P. H. C. (2004). Smooth random effects distribution in a linear mixed model. *Biometrics* **60**, 945–953.
- Gianelli, G. C. (1996). Women's transitions in the labour market: A competing risks analysis on German panel data. *Journal of Population Economics* **9**, 287–300.
- Golub, G. H., Heath, M., und Wahba, G. (1979). Generalized cross validation as a method for choosing a good ridge parameter. *Technometrics* **21**, 215–223.
- Good, I. J. und Gaskins, R. A. (1971). Nonparametric roughness penalties for probability densities. *Biometrika* **58**, 255–277.
- Grambsch, P. M. und Therneau, T. M. (1994). Proportional hazards tests

- and diagnostics based on weighted residuals. *Biometrika* **81**, 515–526.
- Gray, R. J. (1992). Flexible methods for analyzing survival data using splines, with applications to breast cancer prognosis. *Journal of the American Statistical Association*. **87**, 942–951.
- Gray, R. J. (1994). Spline-based tests in survival analysis. *Biometrics* **50**, 640–652.
- Green, D. J. und Silverman, B. W. (1994). *Nonparametric Regression and generalized linear models*. Chapman & Hall.
- Gu, C. (2002). *Smoothing Spline ANOVA Models*. New York: Springer.
- Haight, F. A. (1967). *Handbook of the Poisson distribution*. New York: John Wiley & Sons, Inc.
- Hall, P. und Opsomer, J. D. (2005). Theory for penalized spline regression. *Biometrika* **92**, 105–118.
- Hamerle, A. und Tutz, G. (1989). *Diskrete Modelle zur Analyse von Verweildauer und Lebenszeiten*. Frankfurt / New York: Campus Verlag.
- Harville, D. A. (1997). *Matrix Algebra From A Statistician's Perspective*. Springer.
- Hastie, T. (1996). Pseudosplines. *Journal of the Royal Statistical Society, Series B* **58**, 379–396.
- 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, Series B* **55**, 757–796.
- Heckman, J. und Singer, B. (1984). A method for minimizing the impact of distributional assumptions in econometric models for duration data. *Econometrica* **52**, 271–320.
- Heckman, J. J. und Honore, B. E. (1989). The identifiability of the competing risks model. *Biometrika* **76**, 325–330.
- Hougaard, P. (2000). *Analysis of Multivariate Survival Data*. New York: Springer-Verlag.
- Hüppelshäuser, M., Krafft, M., und Rüger, E. (2006). Hazard-Raten-Modelle im Marketing. *Marketing ZFP* **28**, 197–210.
- Ibrahim, J. G., Chen, M.-H., und Sinha, D. (2001). *Bayesian Survival Analysis*. Springer.
- Jacobson, T. und Carlin, K. (1995). Modeling unemployment duration in a dependent competing risks framework: Identification and estimation. *Life Time Data Analysis* **1**, 111–122.
- Jaffar, S., Leach, A., Greenwood, A., und Greenwood, B. (2000). Season of birth is not associated with delayed childhood mortality in Upper River Division, the Gambia. *Tropical Medicine and International Health* **5**(9), 628–632.
- Johansen, S. (1983). An extension of Cox's regression model. *International Statistical Review* **51**, 165–174.
- Joly, P., Commenges, D., und Letenneur, L. (1998). A penalized likelihood approach for arbitrarily censored and truncated data: application to age-specific incidence of dementia. *Biometrics* **54**, 185–194.
- Kalbfleisch, J. D. und Prentice, R. L. (2002). *The Statistical Analysis of*

- Failure Time Data*. John Wiley & Sons.
- Kauermann, G. (2005a). A note on smoothing parameter selection for penalised spline smoothing. *Journal of Statistical Planning and Inference* **127**, 53–69.
- Kauermann, G. (2005b). Penalised spline fitting in multivariable survival models with varying coefficients. *Computational Statistics and Data Analysis* **49**, 169–186.
- Kauermann, G., Brüderl, J., und Tutz, G. (2005). The survival of newly founded firms: a case-study into varying-coefficient models. *Journal of the Royal Statistical Society, Series A* **168**, 145–158.
- Kauermann, G. und Khomski, P. (2006). Additive two-way hazards model with varying coefficients. *Computational Statistics and Data Analysis* **51**, 1944–1956.
- Kauermann, G. und Khomski, P. (2007). Full time or part time reemployment: A competing risk model with frailties and smooth effects using a penalty based approach. *submitted to Journal of the Computational and Graphical Statistics*.
- Kauermann, G., Ormerod, J. T., und Wand, M. P. (2006). Parsimonious classification via generalized linear mixed models. *submitted for publication*.
- Keiding, N. (1990). Statistical inference in the Lexis diagram. *Philosophical Transactions of the Royal Society, London, A* **332**, 487–509.
- Keiding, N., Andersen, P. K., und Klein, J. (1997). The role of frailty models and accelerated failure time models in describing heterogeneity due to omitted covariates. *Statistics in Medicine* **16**, 215–224.
- Kingman, J. F. C. (1993). *Poisson Processes*. Oxford: Oxford University Press.
- Klein, J. P. (1992). Semiparametric estimation of random effects using the Cox model based on the EM algorithm. *Biometrics* **48**, 795–805.
- Klein, J. P. und Moeschberger, M. L. (1997). *Survival Analysis*. New York: Springer.
- Kneib, T. (2006). *Mixed model based inference in structured additive regression*. Dissertation, Fakultät für Mathematik, Informatik und Statistik der Ludwig-Maximilian-Universität München.
- Komarek, A., Lesaffre, E., und Hilton, J. F. (2005). Accelerated failure time model for arbitrary censored data with smoothed error distribution. *Journal of Computational and Graphical Statistics* **14**, 726–745.
- Kooperberg, C., Stone, C., und Troung, Y. (1995). Hazard regression. *Journal of the American Statistical Association*. **90**, 78–94.
- Krafft, M. (2002). *Kundenbindung und Kundenwert*. Heidelberg: Physica-Verlag.
- Lang, S. und Brezger, A. (2004). Bayesian P-splines. *Journal of Computational & Graphical Statistics* **13**, 183–212.
- Lawless, J. F. (2003). *Statistical models and methods for lifetime data*. New York: John Wiley & Sons.
- Lin, X. (1997). Variance component testing in generalized linear models with random effects. *Biometrika* **84**, 309–326.

- Louis, T. A. (1982). Finding the observed information matrix when using the EM algorithm. *Journal of the Royal Statistical Society, Series B* **44**, 226–233.
- Mardia, K. V., Kent, J. T., und Bibby, J. M. (1979). *Multivariate analysis*. Acad. Press, London.
- Marx, B. und Eilers, P. (1998). Direct generalized additive modelling with penalized likelihood. *Computational Statistics and Data Analysis* **28**, 193–209.
- Marx, B. und Eilers, P. (2002). Generalized linear additive smooth structures. *Journal of Computational and Graphical Statistics* **11**, 758–783.
- Marx, B. und Eilers, P. (2005). Multidimensional penalized signal regression. *Technometrics* **47**, 1–21.
- McCullagh, P. und Nelder, J. A. (1989). *Generalized Linear Models* (2 Aufl.). New York: Chapman and Hall.
- McCulloch, C. E. und Searle, S. R. (2001). *Generalized, Linear, and Mixed Models*. New York: John Wiley & Sons, Inc.
- McGilchrist, C. A. (1993). REML estimation for survival models with frailty. *Biometrics* **49**, 221–225.
- McGilchrist, C. A. und Aisbett, C. W. (1991). Regression with frailty in survival analysis. *Biometrics* **47**, 461–466.
- McLachlan, G. J. und Krishnan, T. (1997). *The EM Algorithm and Extensions*. New York: Wiley.
- McLachlan, G. J. und Peel, D. (2000). *Finite Mixture Models*. New York: Wiley.
- Molenberghs, G. und Verbeke, G. (2005). *Models for Discrete Longitudinal Data*. New York: Springer-Verlag.
- Moore, S., Cole, T., und Poskit, E. (1997). Season of birth predicts mortality in rural Gambia. *Nature* **388**, 434.
- Narendranathan, W. und Stewart, M. B. (1993). Modeling the probability of leaving unemployment: competing risks models with flexible baseline hazards. *Applied Statistics* **42**, 63–83.
- Ngo, L. und Wand, M. P. (2004). Smoothing with mixed models software. *Journal of Statistical Software* **9**, 1–54.
- Nielsen, G. G., Gill, R. D., Andersen, P. K., und Sørensen, T. I. A. (1992). Counting processes approach to maximum likelihood estimation in frailty models. *Scandinavian Journal of Statistics* **19**, 25–43.
- Nielsen, J. P. und Sperlich, S. (2005). Smooth backfitting in practice. *Journal of the Royal Statistical Society, Series B* **67**(1), 43–61.
- Nocedal, J. und Wright, S. J. (1999). *Numerical Optimization*. New York: Springer.
- Oakes, D. (1995). Multiple time scales in survival analysis. *Lifetime Data Analysis* **1**, 7–18.
- O’Sullivan, F. (1986). A statistical perspective on ill-posed inverse problems (c/r: P519-527). *Statistical Science* **1**, 502–518.
- O’Sullivan, F. (1988a). Fast computation of fully automated log-density and log-hazard estimators. *SIAM, Journal of Scientific and Statistical Computing* **9**, 363–379.



- O'Sullivan, F. (1988b). Nonparametric estimation of relative risk using splines and cross-validation. *SIAM, Journal of Scientific and Statistical Computing* **9**, 531–542.
- Parker, R. L. und Rice, J. A. (1985). Discussion of "some aspects of the spline smoothing approach to nonparametric curve fitting" by B.W. Silverman. *Journal of the Royal Statistical Society, Series B* **47**, 40–42.
- Pearce, N. D. und Wand, M. P. (2006). Penalized splines and reproducing kernel methods. *American Statistician* **60**, 233–240.
- Pinheiro, J. C. und Bates, D. M. (2000). *Mixed-Effects Models in S and S-PLUS*. New York: Springer-Verlag.
- R Development Core Team (2006). *R: A language and environment for statistical computing*. URL <http://www.R-project.org>: R Development Core Team. R Foundation for Statistical Computing, Vienna, Austria.
- Ramsay, J. und Silverman, B. (2005). *Functional Data Analysis* (2 Aufl.). New York: Springer.
- Rao, C. R. und Mitra, S. K. (1974). *Generalized Inverse of Matrices and its Applications*. New York: Wiley.
- Rao, C. R. und Rao, M. B. (1998). *Matrix Algebra and its Applications to Statistics and Econometrics*. World Scientific.
- Richards, S. J., Kirkby, J. G., und Currie, I. D. (2005). The importance of year of birth in two-dimensional mortality data. *British Actuarial Journal*, to appear.
- Ripatti, S. und Palmgren, J. (2000). Estimation of multivariate frailty models using penalized partial likelihood. *Biometrics* **56**, 1016–1022.
- Rogers, D. F. (2001). *An Introduction to NURBS: with historical perspective*. Morgan Kaufmann Publishers.
- Rogers, G. S. (1980). *Matrix derivatives*. Marcel Dekker.
- Rosholm, M. (1997). Forschungsbericht 97-14, Centre for Labour Market and Social Research (CLS), Gustav Wieds Vej 10C, 8000 Århus C, Denmark.
- Ruppert, D. (2002). Selecting the number of knots for penalized splines. *Journal of Computational and Graphical Statistics* **11**, 735–757.
- Ruppert, R., Wand, M. P., und Carroll, R. J. (2003). *Semiparametric Regression*. Cambridge University Press.
- Schoenfeld, D. (1980). Chi-squared goodness-of-fit test for the proportional hazards regression model. *Biometrika* **67**, 145–153.
- Searle, S. R., Casella, G., , und McCulloch, C. E. (1992). *Variance Components*. New York: John Wiley & Sons.
- Shumway, T. (2001). Forecasting bankruptcy more accurately: a simple hazard model. *Journal of Business* **74**, 101–124.
- Silverman, B. W. (1985). Some aspects of the spline smoothing approach to non-parametric regression curve fitting. *Journal of the Royal Statistical Society, Series B* **47**, 1–52.
- Stepanova, M. und Thomas, L. (2002). Survival analysis methods for personal loan data. *Operations Research* **50**(2), 277–289.
- Therneau, T. M. und Grambsch, P. M. (2000). *Modelling Survival Data: Extending the Cox Model*. New York: Springer-Verlag.

- Therneau, T. M., Grambsch, P. M., und Pankratz, V. S. (2003). Penalized survival models and frailties. *Journal of Computational and Graphical Statistics* **12**, 156–175.
- Tierney, L. und Kadane, J. B. (1986). Accurate approximations for posterior moments and marginal densities. *Journal of the American Statistical Association*. **81**, 82–86.
- Tikhonov, A. (1963). Solution of incorrectly formulated problems and the regularization method. *Soviet Math. Dokl.* **5**, 1035–1038.
- Titterton, D. M., Smith, A. F. M., und Makov, U. E. (1985). *Statistical Analysis of Finite Mixture Distributions*. New York: John Wiley & Sons.
- Tsiatis, A. (1975). A nonidentifiability aspect of the problem of competing risks. *Proceedings of the National Academy of Sciences* **72**, 20–22.
- Tutz, G. und Binder, H. (2004). Flexible modelling of discrete failure time including time-varying smooth effects. *Statistics in Medicine* **23**, 2445–2461.
- Tysse, T. I. und Vaage, K. (1999). Unemployment of older Norwegian workers: a competing risk analysis. Forschungsbericht 2298, Department of Economics, University of Bergen.
- Vaida, F. und Xu, R. (2000). Proportional hazards model with random effects. *Statistics in Medicine* **19**, 3309–3324.
- van den Berg, G. J. (2001). Duration models: Specification, identification, and multiple durations. In J. J. Heckman & E. Leamer (Hrsg.), *Handbook of Econometrics*, Volume V. Amsterdam: North-Holland.
- van den Berg, G. J. und Lindeboom, M. (1998). Attrition in panel survey data and the estimation of multi-state labor market models. *The Journal of Human Resources* **33**, 459–478.
- Vaupel, J. W., Manton, K. G., und Stallard, E. (1979). The impact of heterogeneity in individual frailty on the dynamics of mortality. *Demography* **16**, 439–454.
- Verbeke, G. und Molenberghs, G. (2000). *Linear Mixed Models for Longitudinal Data*. New York: Springer-Verlag.
- Wager, C., Vaida, F., und Kauermann, G. (2005). Model selection for P-spline smoothing using Akaike information criteria. *Australian and New Zealand Journal of Statistics*, submitted for publication.
- Wahba, G. (1990). *Spline models for observational data*. Philadelphia: SIAM, Society for Industrial and Applied Mathematics.
- Wand, M. P. (2003). Smoothing and mixed models. *Computational Statistics* **18**, 223–249.
- Wand, M. P. und Ormerod, J. T. (2007). On semiparametric regression with O’Sullivan penalized splines. *to appear*.
- Wolfinger, R. und O’Connell, M. (1993). Generalized linear mixed models: A pseudo-likelihood approach. *Journal of Statistical Computation and Simulation* **48**, 233–243.
- Wong, W. (1986). Theory of partial likelihood. *The Annals of Statistics* **14**, 88–123.
- Wood, S. N. (2006). *Generalized Additive Models: an Introduction with R*. Chapman & Hall/CRC.

- Zhang, D. und Lin, X. (2003). Hypothesis testing in semiparametric additive mixed models. *Biostatistics* **4**, 57–74.
- Zucker, D. M. und Karr, A. F. (1990). Nonparametric survival analysis with time-dependent covariate effects: a penalized partial likelihood approach. *The Annals of Statistics* **18**, 329–352.

Gedruckt auf alterungsbeständigem Papier °° ISO 9706