



universität
wien

MAGISTERARBEIT

Titel der Magisterarbeit

„Die Schätzung von nichtlinearen Interaktionen in der
Cox-Regression mittels Fraktionaler Polynome“

Verfasserin

Maria Kohl, Bakk. rer. soc. oec.

angestrebter akademischer Grad

Magistra der Sozial- und Wirtschaftswissenschaften

(Mag. rer. soc. oec.)

Wien, im Mai 2011

Studienkennzahl lt. Studienblatt: A 066 951

Studienrichtung lt. Studienblatt: Statistik

Betreuer: Ao.Univ.-Prof.Mag.Dr. Georg Heinze

Dank

An dieser Stelle möchte ich mich besonders bei Professor Georg Heinze bedanken, der mich bei meiner Magisterarbeit umfangreich unterstützt hat, indem er hilfreiche Anregungen gab und bei Fragen Rede und Antwort stand. Außerdem möchte ich mich herzlich bei meinen Eltern Magareta und Stefan sowie bei meinem Freund Stephan bedanken, die mich bei meinem Studium unterstützt haben.

Vielen Dank auch an meine Tante Karin für das Korrekturlesen und dafür, dass sie und mein Onkel Ossi mir in Wien des Öfteren ein Dach über dem Kopf boten.

Bei meiner Arbeitskollegin Daniela Dunkler möchte ich mich auch bedanken, da sie ebenfalls meine Magisterarbeit korrekturgelesen hat.

Inhaltsverzeichnis

Einleitung	7
1 Methoden	9
1.1 Das Cox-Modell	9
1.1.1 Zeitabhängige Effekte in der Cox-Regression	11
1.2 Monotone Likelihood	12
1.2.1 Monotone Likelihood im Cox-Modell	12
1.2.2 Die „Firth-Korrektur“ im Cox-Modell	13
1.3 Nichtlineare Modellierung	16
1.3.1 Fraktionale Polynome	16
1.3.1.1 Definition	16
1.3.1.2 Kurvenformen von Fraktionalen Polynomen	17
1.3.1.3 Modellselektierung	20
1.3.1.4 Verschiebung, Skalierung und Zentrierung	21
1.3.1.5 Multiple Kovariablen	22
1.3.1.6 Fehlerrate 1.Art für RA2-Algorithmus	23
1.3.1.7 Interaktionen	23
1.3.1.8 Grafische Darstellung der Fraktionalen Polynome	26
1.3.1.9 FP-Analyse in R	27
1.3.2 Restricted Cubic Splines	29
1.3.2.1 Definition	29
1.3.2.2 Anzahl und Position der Knoten	29
1.3.2.3 Restricted Cubic Splines im Cox-Modell	30
1.3.2.4 RCS-Analyse in SAS	31
1.3.3 Der Vergleich von FP und Restricted Cubic Splines	33
1.4 Konfidenzintervalle	35
1.4.1 Das parametrische (Bootstrap-) Konfidenzintervall	35
1.4.2 Das nichtparametrische (Bootstrap-) Konfidenzintervall	35
1.4.3 Das modellbasierte Konfidenzintervall	36

2	Simulation	37
2.1	Einleitung	37
2.2	Datengenerierung	37
2.3	Überblick über Simulation	40
2.4	Szenarien	41
2.5	Punktschätzer	45
2.5.1	Selektierte Modelle	46
2.5.2	Fixiertes Modell	48
2.6	Abweichungen der Schätzer	50
2.6.1	Der mediane Bias und der MAD	50
2.6.2	Berechnung der Gewichte	51
2.7	Results	53
2.7.1	Monotone Likelihood	53
2.7.2	Grafische Darstellung der Punktschätzer	54
2.7.3	Grafische Darstellung der Abweichungen der Schätzer	82
2.8	Interpretation	132
2.8.1	Die Punktschätzer	132
2.8.2	Die Abweichungen der Schätzer	133
3	Fallstudie zur Venenthrombose	135
3.1	Einleitung	135
3.2	Der HCT Datensatz	136
3.2.1	Univariate Betrachtung der HCT Daten	136
3.2.2	Kaplan-Meier-Kurven	138
3.2.3	Proportionalitätsannahme	139
3.2.4	Cox-Analyse am HCT Datensatz	140
3.2.5	Analyse der Konfidenzintervalle	146
3.2.6	Fazit	161
	Diskussion	162
	Zusammenfassung	168
	Abstract	170
	Lebenslauf	171
	Elektronischer Anhang	172

Einleitung

In medizinischen Studien werden häufig Überlebensdaten betrachtet. Dabei ist die Zeit bis zum Auftreten eines bestimmten Ereignisses von Interesse, zum Beispiel die Zeit von einer Nierentransplantation bis zum Abstoß der Niere. Dadurch erhält man Daten mit unterschiedlichen Beobachtungsdauern pro Patient und eine Zensierungsvariable, die angibt, ob das zu beobachtende Ereignis eingetreten ist. Patienten, bei denen das Ereignis bis zum Ende der Studie nicht eingetreten ist, werden als rechtszensiert betrachtet. Um diese Informationen nicht zu vergeuden, gibt es für die Analyse von Überlebensdaten („*Time-to-event data*“) spezielle statistische Verfahren.

Das Modell nach Cox ist eines der bekanntesten Regressionmodelle für die Analyse von Überlebensdaten. Es wurde 1972 durch die Veröffentlichung des Papers „Regression Models and Life-Tables“ von D. R. Cox im „Journal of the Royal Statistic Society“ erstmals bekannt.

Häufig weisen Datensätze aus medizinischen Studien kleine Stichprobengrößen auf und bei selten beobachtbaren Ereignissen wie zum Beispiel einer seltenen Krankheit kommt es zu hohen Zensierungsanteilen. Dies sind Faktoren, die die Wahrscheinlichkeit für das Auftreten von unendlichen Schätzern mit unendlichen Standardfehlern, die als Folge undefinierte Konfidenzintervalle erzeugen, erhöhen. Dieses Phänomen ist bekannt unter dem Namen Monotone Likelihood. Eine Lösung für das Auftreten von Monotoner Likelihood stellt die Schätzung der Parameter mit der Firth-Korrektur von David Firth (1993) dar. Dabei wird vom Maximum-Likelihood-Schätzer ein Strafterm abgezogen, wodurch die Schätzer konvergieren (Firth-Methode). Somit scheint es sinnvoll bei der Schätzung im Cox-Modell die Firth-Korrektur zu verwenden.

In der praktischen Anwendung gehen die meisten Anwender immer noch davon aus, ausschließlich lineare Modelle zu betrachten, obwohl mittlerweile eine Vielzahl an Literatur zur nichtlinearen Modellierung angeboten wird. Ein moderner Ansatz für die Modellierung von Nichtlinearität sind die *Fraktionale Polynome* (FP) (Royston und Sauerbrei, 1994), aber auch die *Restricted Cubic Splines* (Durrleman und Simon, 1989) werden häufig dafür verwendet.

In dieser Masterarbeit werden im Speziellen nichtlineare Interaktionen und deren Modellierung durch *Fraktionale Polynome* und *Restricted Cubic Splines* im Cox-Modell betrachtet. Hier werden ausschließlich Interaktionen einer kontinuierlichen und einer binären Variable berücksichtigt, d.h. es wird angenommen, dass der Effekt der kontinuierlichen Variable von der Ausprägung der binären Variable abhängt. Die Schätzungen erfolgen sowohl mit der Maximum-Likelihood-Methode als auch mit der Firth-Methode.

Ziel ist es mittels einer Simulationsstudie herauszufinden, ob die Schätzung zuzüglich der Firth-Korrektur außer der Vermeidung von Monotoner Likelihood noch andere Vorteile gegenüber der Maximum-Likelihood-Methode aufweist. Dafür werden die Abweichungen der Punktschätzer der beiden Schätzverfahren miteinander verglichen.

Ein weiteres Ziel ist, anhand der Punktschätzer der Log Relative Hazards zu überprüfen, ob die Fraktionalen Polynome die wahren zugrunde liegenden Funktionen gut genug abbilden. Dabei werden die Schätzer der Fraktionalen Polynome mit Schätzern verglichen, die aus einem Modell mit vorgegebenen Transformationen der kontinuierlichen Variable der Interaktion stammen. Dabei basieren die Transformationen auf den Potenzen der zugrunde liegenden wahren Log Relative Hazards der simulierten Daten.

In beiden Fragestellungen ist der Einfluss bezüglich der Stichprobengröße und des Zensierunganteils der simulierten Daten von Interesse. Ausgehend von der Tatsache, dass die Firth-Methode eine Lösung für Monotone Likelihood ist, scheint es sinnvoll das Auftreten von dieses Phänomens in der Simulationsstudie zu dokumentieren.

Zusätzlich werden in einer Fallstudie über Venenthrombosen die Punkt- und Intervallschätzer der *Fraktionalen Polynomen* und *Restricted Cubic Splines* miteinander verglichen.

Das erste Kapitel dieser Masterarbeit beinhaltet die Beschreibung des Cox-Modells, die Schätzung mit der Firth-Korrektur und das Auftreten des Phänomens der Monotonen Likelihood. Ebenso wird die nichtlineare Modellierung mittels *Fraktionaler Polynome* und *Restricted Cubic Splines* sowie die Berechnung derer Konfidenzintervalle erklärt.

Im Kapitel zwei wird die Simulation medizinischer Daten von der Generierung bis hin zu den Ergebnissen und Interpretationen beschrieben. Dabei werden alle Methoden, die in Kapitel eins getrennt voneinander betrachtet wurden, kombiniert.

Das Kapitel drei beschäftigt sich ausschließlich mit der Fallstudie zur Venenthrombose. Die anschließende Diskussion beinhaltet den Überblick über alle Kapitel und Schlussfolgerungen aus dieser Arbeit sowie einen Ausblick auf offene und interessante Fragestellungen in diesem Bereich.

Kapitel 1

Methoden

1.1 Das Cox-Modell

Das Modell nach Cox (1972) findet Anwendung in der Schätzung von Effekten verschiedener Einflussfaktoren (Kovariablen) auf die Überlebenszeit (abhängige Variable). Von Bedeutung dabei ist, dass wir keine parametrische Annahme über die Verteilung der Überlebenszeit festlegen müssen. Bei Überlebensdaten gehört zu der abhängigen Variable (der Überlebenszeit) noch die sogenannte Zensierungsvariable. Diese ist dichotom und gibt an, ob das zu betrachtende Ereignis (zum Beispiel: Tod, Wiederauftreten einer Krankheit, ...) während der Beobachtungszeit eingetreten ist oder nicht. Eine weitere wichtige Eigenschaft dieses Modells ist die Annahme von konstanten Hazardraten, d.h. diese seien von der Zeit unabhängig. Bei zwei oder mehr Gruppen bedeuten konstante Hazardraten über die Zeit, dass die Hazardwerte proportional sind, daher ist dieses Modell auch weit verbreitet unter dem Namen proportionales Hazard-Modell (engl.: proportional hazards model). Wie wir später sehen werden, gilt diese Annahme nicht, wenn wir zeitabhängige Effekte von unabhängigen Variablen betrachten.

Das Cox-Modell mit k Kovariablen kann durch die Hazard-Funktion folgendermaßen definiert werden:

$$h(t, X) = h_0(t) \exp\{\beta_1 X_1 + \beta_2 X_2 + \dots + \beta_k X_k\} \quad (1.1)$$

Die Hazard-Funktion $h(t, X)$ gibt die Wahrscheinlichkeit pro Zeiteinheit an, dass für ein Individuum innerhalb eines kleinen Zeitintervalls das Zielereignis eintritt, gegeben, das Individuum hat bis zu diesem Zeitpunkt überlebt. Unter $h_0(t)$ versteht man den Baseline-Hazard. Er gibt den Hazard für das Eintreten eines Ereignisses, wenn alle Kovariablen gleich null sind, zum Zeitpunkt t an.

Die Werte der unabhängigen Variablen sind gegeben durch X_1, X_2, \dots, X_k und deren zu schätzende Regressionskoeffizienten sind $\beta_1, \beta_2, \dots, \beta_k$. Verändert sich der Wert einer unabhängigen Variable um eine Einheit, so verändert sich der Wert des log Hazards um den Regressionskoeffizienten dieser Variable. Der Baseline-Hazard kann jede beliebige Funktion der Zeit annehmen und wird dabei als nichtparametrisch betrachtet, während der *Hazardquotient* $\exp(\beta)$ den parametrischen Teil des Modells darstellt. Folglich handelt es sich bei der Cox-Regression um ein semiparametrisches Modell.

Durch Logarithmierung kann das Modell als lineare Gleichung dargestellt werden:

$$\log h(t, x) = \log h_0(t) + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_k x_k \quad (1.2)$$

Der daraus resultierende log Hazardratio wird geschrieben als:

$$\text{LHR}(x) = \log \left(\frac{h(t, x)}{h_0(t)} \right) \quad (1.3)$$

In einer Stichprobe von n Individuen können wir m verschiedene und unzensierte Überlebenszeiten $t_{(j)}$ ($1 \leq j \leq m$) aus n möglichen Überlebenszeiten t_i ($1 \leq i \leq n$) beobachten. Der Kovariablenvektor $x_i = (x_{i1}, x_{i2}, \dots, x_{ik})$ bezieht sich auf jedes Individuum. Die Wahrscheinlichkeit, dass das i -te Individuum zum Zeitpunkt $t_{(i)}$ stirbt, ist $\exp(x_i \beta) / \sum_{l \in \mathbf{R}_{t_{(i)}}} \exp(x_l \beta)$. Das Produkt dieser Wahrscheinlichkeiten ergibt die partielle Likelihood, welche nach Cox (1975) wie folgt definiert ist:

$$L(\beta) = \prod_{i=1}^n \left[\frac{\exp(x_i \beta)}{\sum_{l \in \mathbf{R}_{t_{(i)}}} \exp(x_l \beta)} \right]^{d_i}$$

$$\log L(\beta) = \sum_{i=1}^n \left[x_i \beta - \log \sum_{l \in \mathbf{R}_i} \exp(x_l \beta) \right]^{d_i}$$

wobei $\beta = (\beta_1, \dots, \beta_r, \dots, \beta_k)$ den Vektor der Regressionskoeffizienten bezeichnet. Das *Riskset* wird bezeichnet mit \mathbf{R}_i und beinhaltet die Individuen, die zum Zeitpunkt $t_{(i)}$ noch am Leben sind. Da d_i die Anzahl der Sterbefälle zum Zeitpunkt $t_{(i)}$ ist, reichen die zensierten Beobachtungen zwar das Riskset an, Produkt-Terme gibt es für sie allerdings nicht.

Durch die Bindungskorrektur nach Breslow (1974) erhalten wir eine Likelihood (mit Bindungen in t) wie folgt:

$$L(\beta) = \prod_{j=1}^m \frac{\exp(\beta s_j)}{[\sum_{h \in \mathbf{R}_j} \exp(x_h \beta)]^{d_j}}$$

und eine Log-Likelihood mit

$$\log L(\beta) = \sum_{j=1}^m \left[\beta s_j - d_j \log \left\{ \sum_{h \in \mathbf{R}_j} \exp(x_h \beta) \right\} \right].$$

Der Vektor β und die Vektoren x_1, \dots, x_k , sowie das *Riskset* \mathbf{R}_j sind definiert wie zuvor, zusätzlich ist d_j die Anzahl der Sterbefälle zum Zeitpunkt $t_{(j)}$ und s_j der Vektor der Summen der Kovariablen über die betroffenen Individuen.

Leiten wir die Log-Likelihood einmal ab und setzen sie Null, so erhalten wir die *Score-Gleichung* $\partial \log L(\beta_r) / \partial \beta_r \equiv U(\beta_r) = 0$, deren Lösungen die Maximum-Likelihood-Schätzer $\hat{\beta}_r$ der Regressionsparameter β_r mit $1 \leq r \leq k$ sind:

$$\frac{\partial \log L(\beta)}{\partial \beta_r} = \sum_{j=1}^m \left[s_{jr} - \frac{d_j \sum_{h \in \mathbf{R}_j} x_{hr} \exp(x_h \beta)}{\sum_{h \in \mathbf{R}_j} \exp(x_h \beta)} \right]$$

Der Maximum-Likelihood-Schätzer ist asymptotisch normalverteilt mit $\hat{\beta} \sim N_k(\beta, I^{-1}(\beta))$ wobei

$$I(\beta) = [I_{rs}(\beta)]_{k \times k} \quad \text{mit} \quad I_{rs}(\beta) = \frac{\partial^2 \log L(\beta)}{\partial \beta_r \partial \beta_s}$$

und

$$\begin{aligned} \frac{\partial^2 \log L(\beta)}{\partial \beta_r \partial \beta_s} = & - \sum_{j=1}^m d_j \left[\frac{\sum_{h \in \mathbf{R}_j} x_{hr} x_{hs} \exp(x_h \beta)}{\sum_{h \in \mathbf{R}_j} \exp(x_h \beta)} - \right. \\ & \left. - \frac{\left\{ \sum_{h \in \mathbf{R}_j} x_{hr} \exp(x_h \beta) \right\} \left\{ \sum_{h \in \mathbf{R}_j} x_{hs} \exp(x_h \beta) \right\}}{\left\{ \sum_{h \in \mathbf{R}_j} \exp(x_h \beta) \right\}^2} \right]. \end{aligned}$$

1.1.1 Zeitabhängige Effekte in der Cox-Regression

Für nichtkonstante Effekte müssen die Log-Likelihood und die Hazard-Funktion neu definiert werden. Wir nehmen, wie schon zuvor an, dass in einer Stichprobe mit n Individuen m verschiedene und unzensierte Überlebenszeiten $t_{(j)}$ ($1 \leq j \leq m$) aus n möglichen Überlebenszeiten t_i ($1 \leq i \leq n$) beobachtet werden. Für jedes Individuum i aus dieser Stichprobe gibt es einen Vektor mit den Kovariablenwerten $(x_{i1}, \dots, x_{ir}, \dots, x_{ik})$ und x_{jr} ist der r -te Kovariablenwert des Individuums, das zum Zeitpunkt $t_{(j)}$ stirbt. Für die Kovariable x_i sei der Log-Hazardquotient von der Zeit abhängig, was bedeutet, dass x_i in Interaktion mit der Funktion $g(t)$ steht, sodass der Log-Hazardquotient für

x_i mit $\beta_1 + \gamma g(t)$ definiert ist. Zur Vereinfachung des Modells nehmen wir an, dass keine Bindungen in t vorliegen, dann sieht die partielle Log-Likelihood für zeitabhängige Effekte aus wie folgt:

$$\log L(\beta, \gamma) = \sum_{j=1}^m \left[\sum_{r=1}^k x_{jr} \beta_r + x_{j1} g(t_{(j)}) \gamma - \log \left\{ \sum_{h \in \mathbf{R}_j} \exp \left(\sum_{r=1}^k x_{hr} \beta_r + x_{h1} g(t_{(j)}) \gamma \right) \right\} \right] \quad (1.4)$$

β ist der Vektor der k Regressionskoeffizienten, welcher durch die k *Score-Gleichungen* $\partial \log L(\beta, \gamma) / \partial \beta_r \equiv U(\beta_r) = 0$ ($1 \leq r \leq k$) geschätzt wird. Die Gleichung $\partial \log L(\beta, \gamma) / \partial \gamma \equiv U(\gamma) = 0$ hat als Lösung $\hat{\gamma}$, den Schätzer für den Regressionsparameter γ . Bei den Schätzern handelt es sich um partielle Maximum-Likelihood-Schätzer. Die Hazard-Funktion mit einem zeitabhängigen Effekt der Kovariable x_r ist somit gegeben durch

$$h(t; x) = h_0(t) \exp(\beta' x + \gamma x_r g(t)).$$

1.2 Monotone Likelihood

Das Phänomen der Monotonen Likelihood wurde als erstes von Bryson and Johnson (1981) angesprochen.

1.2.1 Monotone Likelihood im Cox-Modell

Das Problem der Monotonen Likelihood wird während des Anpassungsverfahrens im Cox-Modell beobachtet, wenn die Log-Likelihood gegen einen endlichen Wert konvergiert, hingegen ein Parameterschätzer gegen $\pm\infty$ divergiert. Das Problem wird hierbei durch die fehlende Existenz des Maximum-Likelihood-Schätzers ausgelöst. Dazu kann es kommen, wenn spezielle Bedingungen in einer Stichprobe vorliegen.

Ein kleines Beispiel dafür: Bei einer einzelnen Kovariate kann Monotone Likelihood auftreten, wenn für jeden Ausfallszeitpunkt der Wert der Kovariable des ausgefallenen Individuums der größte Wert unter den Kovariablenwerten aller Individuen, die sich zu diesem Zeitpunkt im *Riskset* befinden, ist. Selbiges gilt, wenn es sich dabei immer um den kleinsten Wert der Kovariable handelt. Wenn ausschließlich unzensierte Daten betrachtet werden, so tritt monotone Likelihood auf, wenn die nach der Lebensdauer geordneten Kovariablenwerte monoton sind. Daraus folgt, dass die partielle Likelihood monoton ist und die Schätzer der Parameter divergieren.

Monotone Likelihood ist ein Problem der Stichprobe und nicht der zugrundeliegenden Population. Für jede endliche Stichprobe ist die Wahrscheinlichkeit für das Auftreten des Problems der monotonen Likelihood größer Null. Diese Wahrscheinlichkeit erhöht sich laut Bryson und Johnson (1981) mit steigendem Zensierungsgrad. Das Hinzufügen von Kovariablen zu dem Modell führt ebenfalls zu einer erhöhten Wahrscheinlichkeit für Monotone Likelihood, da die Möglichkeit des Auftretens von Monotonität durch jede Kovariable steigt. Heinze und Schemper (2001) bestätigten das gehäufte Auftreten von Monotoner Likelihood bei zusätzlicher Aufnahme von dichotomen Kovariablen in das Modell und bei der Erhöhung des Anteils der zensierten Überlebenszeiten. Dieses Problem kann noch auf weitere Bedingungen zurückgeführt werden, wie die Simulationsstudie von Heinze und Schemper zeigt. Die Stichprobengröße, das Ausmaß des relativen Risikos im Zusammenhang mit dichotomen Kovariablen sowie deren Balance in der Verteilung, sind ausschlaggebend für das häufigere Auftreten von Monotoner Likelihood. (vgl. Heinze u. Schemper, 2001)

Monotone Likelihood tritt nicht nur im Cox-Modell auf, sondern auch in anderen Modellen wie zum Beispiel in der Logistischen Regression (Heinze u. Schemper, 2002). Die Firth-Korrektur bietet eine Lösung für dieses Problem (vgl. Heinze u. Schemper, 2001 und 2002).

1.2.2 Die „Firth-Korrektur“ im Cox-Modell

Die Firth-Methode wurde ursprünglich von David Firth (1993) entwickelt, um den Bias des Maximum-Likelihood-Schätzers in verschiedenen Regressionsmodellen zu minimieren. Dabei sollte die Reduktion des Bias nicht von der Existenz eines endlichen Parameterschätzers $\hat{\beta}$ abhängen. Deshalb entwickelte Firth eine systematische Korrektur für die Score-Gleichungen, über die die Maximum-Likelihood-Schätzer berechnet werden. Bis zu diesem Zeitpunkt waren ausschließlich Methoden bekannt, die für die Minimierung des Bias einen endlichen Parameterschätzer voraussetzten, wie die *Jackknife Methode* (Efron u. Tibshirani, 1993, S.133 und 138) und die *Bias-korrigierten Schätzer (BC-Schätzer)* (Cordeiro u. Cribari-Neto, 1998). Heinze und Schemper (2001) verwendeten die Firth-Prozedur, um das Problem der monotonen Likelihood in der Cox-Regression zu eliminieren.

Die Maximum-Likelihood-Schätzer $\hat{\beta}$ der Regressionsparameter β_r ($1 \leq r \leq k$) erhalten wir als Lösungen der Score-Gleichung $\partial \log L(\beta_r) / \partial \beta_r \equiv U(\beta_r) = 0$. Der Bias der Maximum-Likelihood-Schätzer entsteht dabei durch die Kombination der Erwartungstreue der Scoregleichung, $E[U(\beta)] = 0$ für den wahren Wert von β und die Krümmung der Score-Gleichung $U'(\beta) \neq 0$ von β . Durch die Hinzugabe eines passenden Korrekturterms zur Score-Gleichung

konnte Firth den Bias von $\hat{\beta}$ reduzieren. Die Cox-Regression kann als ein Modell der Exponentialfamilie mit dem kanonischen Parameter β formuliert und interpretiert werden. Die modifizierte Score-Gleichung für Modelle der Exponentialfamilie sieht nun folgendermaßen aus:

$$U(\beta_r)^* \equiv U(\beta_r) + a_r = 0 \quad 1 \leq r \leq k$$

mit dem Strafterm

$$a_r = 0.5 \cdot \text{Spur}[I(\beta)^{-1}\{\partial I(\beta)/\partial\beta_r\}] \quad (1.5)$$

wobei $I(\beta)^{-1}$ die an der Stelle β ausgewertete Inverse der Informationsmatrix darstellt, die ebenso bekannt ist als die geschätzte Kovarianzmatrix an der Stelle $\hat{\beta}$. Der Term $\partial I(\beta)/\partial\beta_r$ ist die Ableitung der Informationsmatrix an der Stelle β_r :

$$\begin{aligned} \frac{\partial I_{rs}(\beta)}{\partial\beta_t} &= -\frac{\partial^3 \log L(\beta)}{\partial\beta_r \partial\beta_s \partial\beta_t} \\ &= -\sum_{j=1}^m d_j \left\{ \left(\frac{S_{j,rst}}{S_{j,0}} - \frac{S_{j,rs}S_{j,t}}{S_{j,0}^2} \right) - \frac{S_{j,s}}{S_{j,0}} \left(\frac{S_{j,rt}}{S_{j,0}} - \frac{S_{j,r}S_{j,t}}{S_{j,0}^2} \right) - \right. \\ &\quad \left. - \frac{S_{j,r}}{S_{j,0}} \left(\frac{S_{j,st}}{S_{j,0}} - \frac{S_{j,s}S_{j,t}}{S_{j,0}^2} \right) \right\}, \quad 1 \leq r, s, t \leq k, \end{aligned}$$

mit

$$\begin{aligned} S_{j,0} &= \sum_{h \in R_j} \exp(x_h \beta) & S_{j,r} &= \sum_{h \in R_j} x_{hr} \exp(x_h \beta) \\ S_{j,s} &= \sum_{h \in R_j} x_{hs} \exp(x_h \beta) & S_{j,t} &= \sum_{h \in R_j} x_{ht} \exp(x_h \beta) \\ S_{j,rs} &= \sum_{h \in R_j} x_{hr} x_{hs} \exp(x_h \beta) & S_{j,rt} &= \sum_{h \in R_j} x_{hr} x_{ht} \exp(x_h \beta) \\ S_{j,st} &= \sum_{h \in R_j} x_{hs} x_{ht} \exp(x_h \beta) & S_{j,rst} &= \sum_{h \in R_j} x_{hr} x_{hs} x_{ht} \exp(x_h \beta) \end{aligned}$$

Zusammengefasst lässt sich daher sagen, dass man für die Anwendung der Firth-Korrektur im Cox-Modell ausschließlich die Ableitung der Informationsmatrix benötigt. Die penalisierte Log-Likelihood und die Likelihood sind gegeben durch $\log L(\beta)^* = \log L(\beta) + 0.5 \log |I(\beta)|$ und $L(\beta)^* = L(\beta)|I(\beta)|^{0.5}$. Die Straffunktion $|I(\beta)|^{0.5}$ hat einen verschwindenden asymptotischen Einfluss und ist auch bekannt unter dem Namen *Jeffreys invariant prior*. Firth (1993) zeigte, dass diese Modifikation den $O(n^{-1})$ Bias der Maximum-Likelihood-Schätzer $\hat{\beta}$ beseitigt. Dies ist zwar in großen Stichproben unbedeutend, jedoch in kleinen oder spärlichen Datensätzen von großer Bedeutung.

Im Allgemeinen kann gezeigt werden, dass die Cox-Regression mit der Firth-Korrektur k endliche Schätzer liefert, solange für jedes Riskset von mindestens k verschiedenen Ausfallzeiten eine Kovariable mit einer Varianz ungleich

Null existiert und jede Kovariable in mindestens einem Riskset eine Varianz ungleich Null hat (Heinze u. Schemper, 2001). In Folge muss man die Informationsmatrix $I(\beta)$ betrachten,

$$I_{rs}(\beta) = \sum_{j=1}^m d_j \left[\sum_{h \in R_j} x_{hr} x_{hs} w_h - \left(\sum_{h \in R_j} x_{hr} w_h \right) \left(\sum_{h \in R_j} x_{hs} w_h \right) \right] \quad (1.6)$$

mit $1 \leq r, s \leq k$ und $w_h = \exp(x_h \beta) / \sum_{l \in R_j} \exp(x_l \beta)$. Die m Summanden der Informationsmatrix stellen m Risksets dar. Jedes dieser Risksets kann als Kovarianzmatrix der Kovariaten aller Individuen, die sich zum Zeitpunkt j im Riskset befinden und durch w_h gewichtet sind, gesehen werden. Wenn nun β_r gegen ∞ geht, so erhält der höchste (bzw. niedrigste) beobachtete Wert x_r einer Kovariable in jedem Riskset mehr Gewicht, im Vergleich zu den anderen Kovariatenwerten. In Folge nähert sich die Varianz von x_r in jedem Riskset sowie die Determinante der Informationsmatrix dem Wert Null an, sodass, selbst wenn die Likelihood monoton in β ist, die bestrafte Likelihood $L(\beta)^*$ ihr Maximum an einem endlichen Wert von $\hat{\beta}$ erreicht. Daraus folgt, dass das Cox-Modell mit der Annäherung von Firth das Auftreten von Monotoner Likelihood vollkommen eliminiert. Laut Heinze und Schemper (2001) sind die Standardfehler von $\hat{\beta}$ in der Cox-Regression mit der Firth-Korrektur kleiner und immer endlich, im Gegensatz zu den Standardfehlern der normalen Cox-Regression. Der Parameterschätzer wird durch den Strafterm zu einem Wert verschoben, an dem die Determinante der inversen Kovarianzmatrix maximiert wird, d.h. die Varianz der Parameterschätzer wird minimiert. (vgl. Heinze u. Schemper, 2001)

Die penalisierte Likelihood Ratio Statistik (PLR) ist gegeben durch:

$$PLR = 2 \times [\log L(\hat{\beta})^* - \log L(\beta_0)^*] \quad (1.7)$$

wobei $\hat{\beta}$ den Vektor der penalisierten Maximum-Likelihood-Schätzer darstellt und β_0 als ein Nullvektor definiert ist. Es macht in Folge einen Unterschied, ob alle Koeffizienten auf Null gesetzt werden oder ob sie einfach weggelassen werden, da die Informationsmatrix an der Stelle 0 nicht Null ist. Somit bleibt der Bestrafungsterm auch bei Koeffizient gleich Null erhalten. Die PLR-Statistik ist approximativ χ^2 -verteilt mit so vielen Freiheitsgraden (df) wie Parameter im Modell sind. Heinze (2006) stellte fest das im Logistischen Modell die Annäherung an die χ^2 -Verteilung der PLR-Statistik besser ist als die der Likelihood Ratio Statistik.

1.3 Nichtlineare Modellierung

In den meisten Studien betrachtet man eine oder mehrere kontinuierliche Kovariablen, jedoch ist es immer noch Routine, ausschließlich lineare Terme bei der Modellierung zu inkludieren. Ist die Beziehung der abhängigen Variable und einer metrischen Kovariable kurvenförmig, so wird oftmals noch ein quadratischer Term in das Modell aufgenommen. In der Anwendung von statistischen Verfahren bleibt es großteils bei der Überprüfung von linearen oder quadratischen Beziehungen. Kubische oder höhergradige Polynome werden nur selten verwendet. In den folgenden beiden untergeordneten Kapiteln (1.3.1) und (1.3.2) werden zwei Möglichkeiten der nichtlinearen Modellierung näher vorgestellt, die *Fraktionalen Polynome* (FP) und die *Restricted Cubic Splines* (RCS).

1.3.1 Fraktionale Polynome

Polynome niedrigen Grades reichen aufgrund ihrer wenigen Kurvenformen oftmals nicht aus, um Daten ausreichend anzupassen, während Polynome höheren Grades die Daten zwar besser anpassen, es damit aber nicht immer gelingt, die Datenendpunkte gut genug abzubilden. Aus diesem Grund haben Royston und Altman (1994) ein kleines, fest definiertes Set von Potenzen, bestehend aus ganzzahligen und nicht ganzzahligen Werten, vorgeschlagen, welches eine ausreichende Vielfalt von Kurvenformen für die Datenmodellierung liefert. Aus diesem Set werden ein oder zwei Potenzen selektiert, somit werden nur Teile (engl.: fractionals) eines Polynoms bei der Modellierung verwendet. Deshalb ist diese Familie von Kurven auch besser bekannt unter dem Namen *Fraktionale Polynome* (FP).

1.3.1.1 Definition

Es wird nun eine Familie von Modellfunktionen für eine Kovariable x eingeführt, mit der Restriktion dass $x > 0$. (Für $x \leq 0$ siehe etwas später 1.3.1.4.) Ein Polynommodell m -ten Grades wird geschrieben als $\sum_{j=1}^m \beta_j x^j$, während ein fraktionales Polynom m -ten Grades geschrieben wird als $\sum_{j=1}^m \beta_j x^{p_j}$ (ohne Konstante β_0) für $p_1 < p_2 < \dots < p_m$. Ein FP ist ein einfaches Polynom mit Potenzen aus dem Set $S = \{-2, -1, -0.5, 0, 0.5, 1, 2, 3, \dots, \max(3, m)\}$, wobei 0 als der natürliche Logarithmus von x ($x^0 = \ln x$) definiert ist. Weiters gilt, dass ein herkömmliches Polynom m -ten Grades auch ein FP m -ten Grades ist, was umgekehrt nicht gelten muss. Für FP mit $p_1 \leq p_2 \leq \dots \leq p_m$ kann die allgemeine Form geschrieben werden als:

$$\varphi_m(x; p) = \sum_{j=1}^m \beta_j h_j(x) \quad (1.8)$$

mit

$$h_j(x) = \begin{cases} x^{p_j} & \text{wenn } p_j \neq p_{j-1} \\ h_{j-1}(x) \log(x) & \text{wenn } p_j = p_{j-1} \end{cases}$$

Zum Verständnis sei noch erwähnt, dass unter einer *Transformation* x^p mit $p \in S$ gemeint ist, während unter einer *Funktion* oder einem Modell $\varphi_m(x; p)$ gemeint ist. Zur Notation sei hinzugefügt, dass $\varphi_m(x; p)$ ebenso geschrieben werden kann als FPm oder $FPm(p_1, \dots, p_m)$.

Da in der Praxis Modelle mit $m > 2$ selten sind und $\varphi_1(x; p)$ und $\varphi_2(x; p)$ eine ausreichend große Vielfalt an Funktionen liefert, wird in weiterer Folge nur mehr auf diese eingegangen. Man arbeitet daher mit dem fixen Set $S = \{-2, -1, -0.5, 0, 0.5, 1, 2, 3\}$ und den beiden Modellen $\beta_1 x^{p_1}$ und $\beta_1 x^{p_1} + \beta_2 x^{p_2}$, wobei bei $p_1 = p_2$ das Modell $\beta_1 x^{p_1} + \beta_2 x^{p_1} \log x$ verwendet wird.

(vgl. Royston u. Altman, 1994; Royston u. Sauerbrei, 2008, S.71-75)

1.3.1.2 Kurvenformen von Fraktionalen Polynomen

Für das Set S liefert das Modell $\varphi_1(x; p)$ 8 monotone Funktionen, d.h. das diese Funktionen einen positiven oder negativen Anstieg haben. Hinzu kommt, dass jene Funktionen, wo $p < 0$ ist, Asymptoten mit $x \rightarrow \infty$ haben. Im Gegensatz dazu kann das Modell $\varphi_2(x; p)$ 36 verschiedene Funktionen annehmen und diese können sowohl monoton als auch unimodal sein, d.h. sie haben ein Minimum oder Maximum. Wenn bei $FP2$ p_1 und p_2 negativ sind, dann besitzen diese eine Asymptote mit $x \rightarrow \infty$. Bei $p_1 \neq p_2$ ist die Funktion monoton, wenn $sign(\beta_1 \beta_2) sign(p_2) = sign(p_1)$ gilt, ansonsten sind diese unimodal. Wenn $p_1 = p_2$, dann liegen immer unimodale Funktionen vor.

Weiters ist zu berücksichtigen, dass wenn bei einer unimodalen Funktion das Maximum oder Minimum außerhalb der Spannweite der Variable x liegt, diese Funktion für den Bereich von x monoton ist. Eine Extrapolation außerhalb von x ist daher nicht zu empfehlen.

Die Abbildung (1.1) zeigt die Vielfalt an möglichen Kurven, die mit nur wenigen Potenzen wie $p_1 \in \{-2, -1\}$ und $p_2 \in \{\pm 2, \pm 1\}$ möglich sind. Ein $FP2(-1, -1)$ kann unter der Verwendung von verschiedenen Koeffizientenvektoren β ebenfalls sehr unterschiedliche Kurvenformen liefern, wie in der Abbildung (1.1) sichtbar wird. Erwähnenswert ist noch, dass bei $FP2$ Modellen aus teilweise unterschiedlichen p_1 und p_2 Kombinationen sehr ähnliche

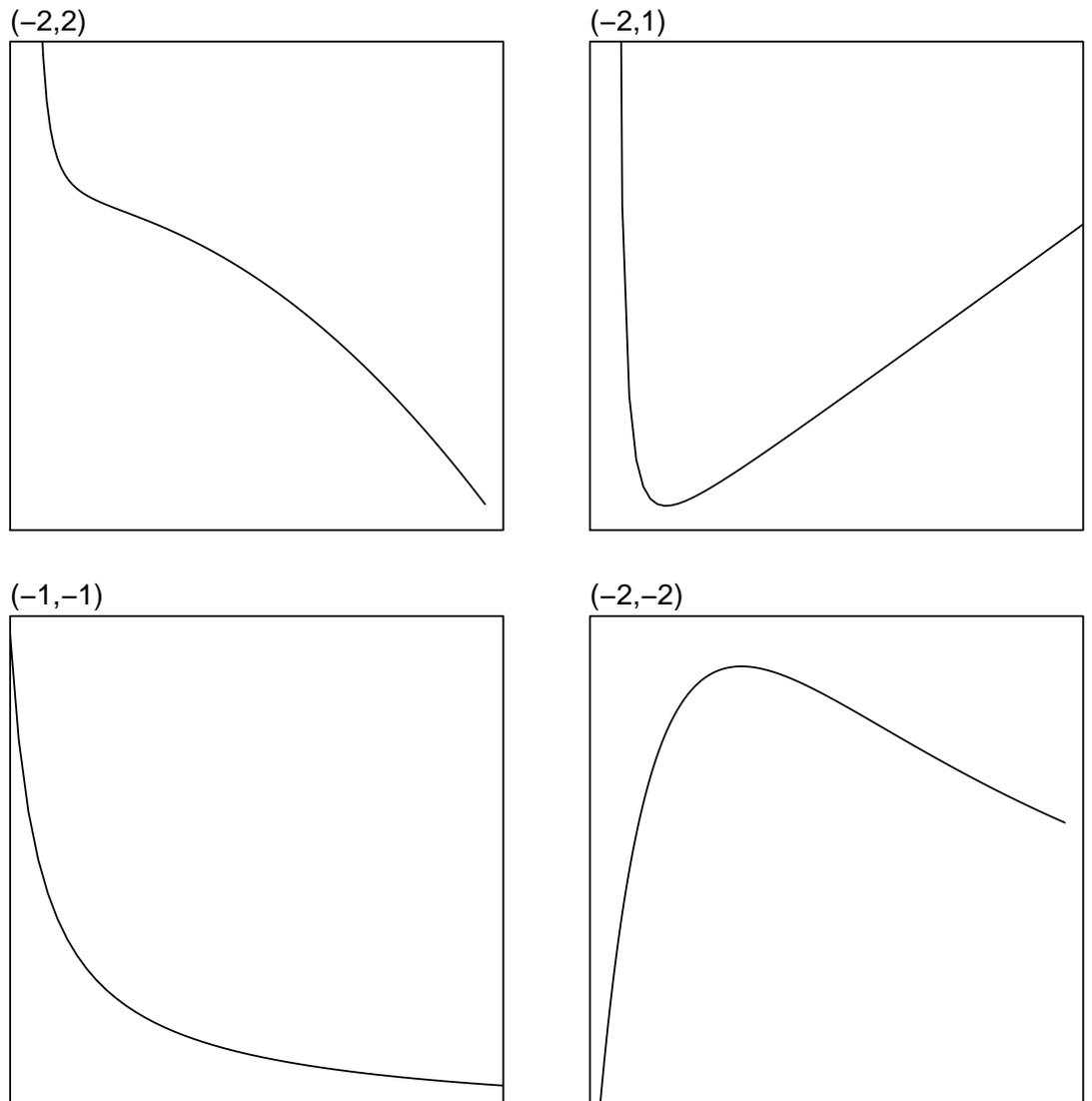


Abbildung 1.1: Beispiele für FP2 mit $p = (-2, 2)$, $(-2, 1)$, $(-1, -1)$ und $(-2, -2)$. (Adaptiert von Royston und Altman, 1994)

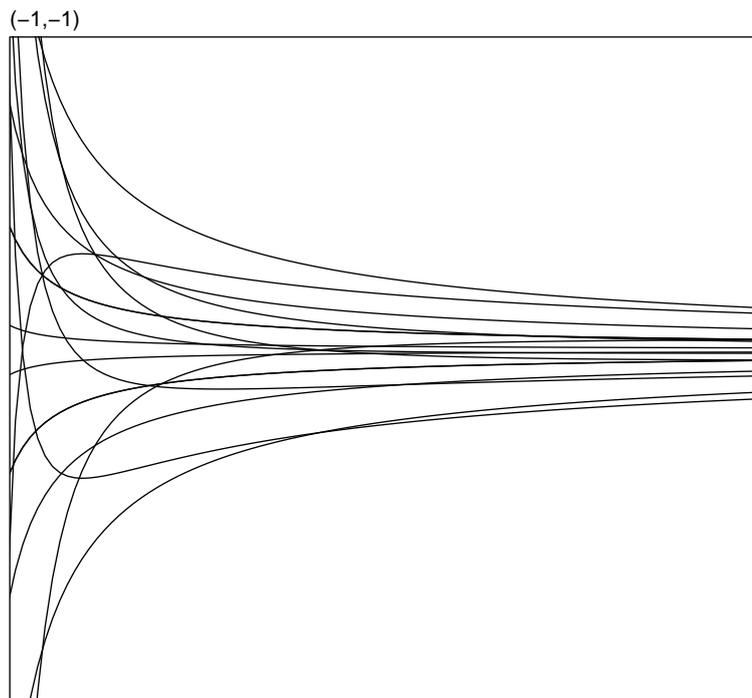


Abbildung 1.2: Eine Auswahl von 17 Kurven mit $p = (-1, -1)$ unter der Verwendung von verschiedenen Koeffizientenvektoren β . (Adaptiert von Royston und Altman, 1994)

Kurven resultieren.

(vgl. Royston u. Altman, 1994; Royston u. Sauerbrei, 2008, S.75-78)

1.3.1.3 Modellselektierung

Für das Auffinden einer passenden Funktion für die Variable x wird der sogenannte *RA2-Algorithmus* verwendet (Ambler u. Royston, 2001).

Der Algorithmus läuft wie folgt:

1. Das beste FP2 Modell für x wird mit einem α_1 - Level und 4 Freiheitsgraden gegen das Nullmodell getestet. Ist das Ergebnis nicht signifikant, so wird der Algorithmus hier angehalten. Bei einem signifikanten Ergebnis wird fortgefahren.
2. Das beste FP2 Modell für x wird mit einem Signifikanzniveau α_2 gegen das lineare Modell (eine Gerade) getestet, unter der Verwendung von 3 Freiheitsgraden. Ist der Test nicht signifikant, so wird hier der Algorithmus abgebrochen und das Endmodell ist das lineare Modell, d.h. der Effekt von x ist linear. Sonst wird fortgefahren.
3. Das beste FP2 Modell von x wird mit einem Signifikanzniveau α_3 gegen das beste FP1 Modell von x getestet, dafür werden 2 Freiheitsgrade verwendet. Ist der Test nicht signifikant, so ist das Endmodell ein FP1 Modell. Bei einem signifikanten Ergebnis ist das Endmodell ein FP2 Modell.
Ende der Prozedur.

Im ersten Schritt wird überprüft, ob die Variable x überhaupt in das Modell aufgenommen werden soll. Steht außer Diskussion, dass die Variable x im Modell enthalten sein soll, so setzt man $\alpha_1 = 1$. Im zweiten Schritt wird das Modell auf einen Nachweis von Nichtlinearität überprüft. Im Schritt drei gehen wir von einem nichtlinearen Zusammenhang aus und entscheiden nur noch, ob es sich um ein einfaches oder komplexeres nichtlineares Modell handelt. Welche Werte für α_1 , α_2 und α_3 verwendet werden, hängt von der jeweiligen Modellierungsstrategie ab. (vgl. Ambler u. Royston, 2001)

In der Praxis wird häufig das *Akaike Informationskriterium* (AIC) für die Modellauswahl bei konkurrierenden, auch nicht genesteten Modellen verwendet. Daher scheint ein Signifikanzniveau von 0.157 für α_1 und α_2 äußerst sinnvoll, da dieses approximativ einer Modellauswahl über das $AIC = -2l + 2 \dim(M)$ (engl. Abk. für *Akaike information criterion*) entspricht (Royston u. Sauerbrei, 2008, S.32,33,95), wobei M das vorgegebene Modell ist, l die maximierte

Log-Likelihood und $\dim(M)$ die Anzahl der zu schätzenden Parameter.

Die Schätzung der einzelnen Modelle erfolgt mittels Maximum-Likelihood (ML). Das beste FP m Modell ist jenes Modell mit der höchsten Likelihood oder der niedrigsten Devianz (D), wobei $D = -2 \times \text{Log-Likelihood}$ ist. Nehmen wir an, dass $\hat{\mathbf{p}}$ der Maximum-Likelihood-Schätzer (MLE) von \mathbf{p} ist und $\mathbf{p} \in (-\infty, +\infty)$. Dann sei $\tilde{\mathbf{p}}$ ebenso ein MLE von \mathbf{p} , jedoch mit der Restriktion $\tilde{\mathbf{p}} \in S$. Die Differenz $D(m, \mathbf{p}) - D(m, \hat{\mathbf{p}})$ ist approximativ χ^2 -verteilt mit m Freiheitsgraden (DF). Da $D(m, \tilde{\mathbf{p}}) \geq D(m, \hat{\mathbf{p}})$ und für m generell die Fläche der Likelihood in der Nähe von $\hat{\mathbf{p}}$ sehr flach ist, bedeutet das, dass es mehrere FP Modelle mit ähnlichen Devianzen gibt. Daraus folgt, dass $D(m, \mathbf{p}) - D(m, \tilde{\mathbf{p}}) < \chi_{m,1-\alpha}^2$ ein asymptotischer konservativer Test für \mathbf{p} ist, der uns das beste FP m Modell liefert.

Um ein FP Modell m -ten Grades mit einem FP Modell mit $m+1$ zu vergleichen, müssen zwei Parameter geschätzt werden, einer für die Potenz und einer für den Regressionskoeffizienten. Somit ist $D(m, \hat{\mathbf{p}}) - D(m+1, \hat{\mathbf{p}})$ wieder χ^2 -verteilt mit 2 df. Folglich liefert uns $D(m, \tilde{\mathbf{p}}) - D(m+1, \tilde{\mathbf{p}}) > \chi_{2,1-\alpha}^2$ einen Test, um zwischen FP m und FP($m+1$) zu entscheiden, was auch für FP1 und FP2 im Schritt drei des *RA2-Algorithmus* geschieht.

Im Schritt 2 der *RA2-Prozedur* ist es notwendig, das lineare Modell gegen das beste FP2 Modell zu testen. Das lineare Modell kann geschrieben werden als $\varphi_1(x; 1)$ und hat somit die Devianz $D(1, 1)$, welche gegen $D(2, \tilde{\mathbf{p}})$ mit drei Freiheitsgraden getestet wird.

Zuletzt soll gezeigt werden, wie das Nullmodell gegen das beste FP m Modell getestet wird. Da $D(0) - D(m, \hat{\mathbf{p}})$ einer χ^2 -Verteilung mit $2m$ Freiheitsgraden entspricht, erhalten wir einen konservativen Test mit $D(0) - D(m, \tilde{\mathbf{p}}) > \chi_{2m,1-\alpha}^2$, der die wahren Werte überschätzt. (vgl. Royston u. Altman, 1994; Royston u. Sauerbrei, 2008, S.79-80)

1.3.1.4 Verschiebung, Skalierung und Zentrierung

Da die Kovariable positiv sein muss (aufgrund der Berechnung des Logarithmus und der Wurzel), ist es oftmals notwendig, eine Variable x mit nichtpositiven Werten vor der Anwendung von FP zu transformieren. Folglich muss die Variable so verschoben werden, dass ihre Werte größer Null sind. Dies bewerkstelligen Royston und Altman (1994) mit einer einfachen Transformation von x . Diese lautet $x^* = x - x_{(1)} + \xi$ mit $\xi = \min\{x_{(i)} - x_{(i-1)}\}$ und $x_{(i)}$ ist der eindeutig geordnete x -Wert an der i -ten Stelle. Es können auch andere Transformationen verwendet werden, jedoch können schlecht gewählte Verschiebungen den *Fehler 1.Art* beeinflussen und zu einer Verringerung der Power, die für das Auffinden von Nichtlinearität notwendig ist, führen.

Es ergeben sich verschiedene Vorteile, wenn die Variable x zusätzlich skaliert und zentriert wird. Die Zentrierung dient dazu, dass die Baseline Hazard im Cox-Modell sinnvoll interpretiert werden kann, während eine ausbleibende Skalierung dazu führt, dass die Schätzung des Modells unmöglich wird oder es zu Ungenauigkeiten in der Schätzung kommt. Die Skalierung und Zentrierung ist bei den Softwares bereits im Algorithmus implementiert als:

$$\begin{aligned} \text{logrange} &= \log_{10}[\max(x) - \min(x)] \\ \text{scale} &= 10^{\text{sign}(\text{logrange})\text{int}(|\text{logrange}|)} \\ x^* &= \frac{x}{\text{scale}} \end{aligned}$$

mit $\text{int}(x)$ als den nächsten ganzzahligen Wert unterhalb von x . Nach der Skalierung kommt die Schätzung der Potenzen der FP und zuletzt folgt die Zentrierung, indem von x^{p^*} der Mittelwert $\bar{x}^{p^*} = \frac{1}{n} \sum_{i=1}^n x_i^{p^*}$ abgezogen wird. Während die Skalierung keinen Einfluß auf die Schätzung ausübt, führt eine Zentrierung vor dem Schätzprozess zu einer anderen Auswahl an Potenzen. (vgl. Royston u. Sauerbrei, 2008, S.79, 84-85)

1.3.1.5 Multiple Kovariablen

In der Praxis wird meist mehr als nur eine kontinuierliche Variable in ein Modell aufgenommen, von denen mehrere nichtlinear sein könnten. Betrachtet man ein Modell mit k kontinuierlichen Variablen x und c kategoriellen Variablen z , so ist das Ziel, eine Kombination von Potenzen $\tilde{\mathbf{p}}_1, \tilde{\mathbf{p}}_2, \dots, \tilde{\mathbf{p}}_k$ und Funktionsgraden m_1, m_2, \dots, m_k zu finden. Dazu wird ein iterativer Prozess benötigt, der zuerst für das Modell (hier ohne Konstante)

$$\eta_1 = \sum_{j=1}^{m_1} \beta_{1j} h_{1j}(x_1) + \sum_{j=2}^k \delta_j x_j + \sum_{j=1}^c \gamma_j z_j$$

das passende $m_1, \tilde{\mathbf{p}}_1$ und $h_1(x_1)$ schätzt, wobei β_{1j}, δ_j und γ_j die zu schätzenden Regressionskoeffizienten des Modells sind. Im nächsten Schritt werden die Schätzer von $m_2, \tilde{\mathbf{p}}_2$ und $h_2(x_2)$ für das Modell

$$\eta_2 = \sum_{j=1}^{m_2} \beta_{2j} h_{2j}(x_2) + \sum_{j=1}^{m_1} \beta_{1j} h_{1j}(x_1) + \sum_{j=3}^k \delta_j x_j + \sum_{j=1}^c \gamma_j z_j$$

optimiert und β_{1j}, γ_j und δ_j erneut geschätzt. Dieses Prozedere wiederholt sich bis η_k , dann beginnt der Prozess erneut mit der Schätzung von $h_1(x_1)$, wobei $h_2(x_2)$ bis $h_k(x_k)$ und z fixiert sind, und läuft wiederum bis zum k -ten Schritt. Der Prozess wird solange immer wieder von Neuem gestartet, bis sich

die Polynomfunktionen $h_1(x_1), \dots, h_k(x_k)$ nicht mehr ändern. (vgl. Royston u. Altman, 1994)

Dieser Prozess lässt sich mit einem Rückwertseliminations-Verfahren kombinieren und wird von Sauerbrei und Royston als der MFP-Algorithmus beschrieben (für mehr Details siehe Royston u. Sauerbrei, 2008, S.115-150).

1.3.1.6 Fehlerrate 1.Art für RA2-Algorithmus

Der *RA2-Algorithmus* ist eine Weiterentwicklung des *RA1-Algorithmus* von Royston und Altman (1994). Der *RA1-Algorithmus* besteht ebenfalls aus drei Schritten. Bei ihm findet jedoch gleich im ersten Schritt der Vergleich der beiden Fraktionalen Polynome FP1 und FP2 statt. Wenn *RA1* sich für das FP Modell ersten Grades entscheidet, wird im zweiten Schritt dieses mit dem linearen Modell verglichen. Bei einer wiederholten Entscheidung für das FP1 Modell wird im letzten Schritt dieses mit dem Nullmodell verglichen. In einer Simulationsstudie verglichen Ambler und Royston (2001) die beiden Algorithmen bezüglich ihres *Fehlerrate 1.Art* (engl.: *Type I error*) und kamen zu dem Schluss, dass der *RA1-Algorithmus* im Gegensatz zum *RA2-Algorithmus* den Fehler 1.Art bei einem Signifikanzniveau von α nicht einhalten kann, deshalb wird nur mehr der *RA2-Algorithmus* in weiterführender Literatur erwähnt.

1.3.1.7 Interaktionen

In diesem Kapitel wird auf die Modellierung Fraktionaler Polynome mit einer nichtlinearen Interaktion im Cox-Modell eingegangen. Der Vollständigkeit wegen ist es ebenfalls sinnvoll grafisch und theoretisch zu zeigen, wie ein Modell ohne Interaktion und ein Modell mit einer linearen Interaktion aussehen. Wenn man von einer kontinuierlichen Variable x und einer binären Variable z ausgeht, zeigt die Gleichung 1.9 das Cox-Modell ohne Interaktion der Variablen x und z .

$$h(t; x, z) = h_0(t) \exp(\beta_1 x + \beta_2 z) \quad (1.9)$$

$$h(t; x, z) = h_0(t) \exp(\beta_1 x + \beta_2 z + \beta_3 x \cdot z) \quad (1.10)$$

$$h(t; x, z) = h_0(t) \exp(\beta_1 f(x) + \beta_2 z + \beta_3 f(x) \cdot z) \quad (1.11)$$

Eine Interaktion zwischen den Variablen x und z ist dann vorhanden, wenn z die Beziehung zwischen der Variablen x und dem Outcome modifiziert, d.h. dass der Effekt von x für die verschiedenen Untergruppen von z verschieden ist. Dabei können wir zwischen einer linearen und einer nichtlinearen Interaktion unterscheiden. Bei der linearen Interaktion werden die Variablen x und

z einfach miteinander multipliziert (siehe Gleichung 1.10), während bei der nichtlinearen Interaktion eine nichtlineare Funktion von x ($f(x)$) modelliert wird, die mit z multipliziert wird (siehe Gleichung 1.11).

Es sei nun die kontinuierliche Variable zum Beispiel das Alter („Age“) der Probanden, die zwei unterschiedliche Behandlungen (A und B) bekommen („Treatment“). In Abbildungen 1.3(a) ist keinerlei Interaktion sichtbar, da zwei parallele Linien pro Behandlung vorhanden sind, wobei der Anstieg der Geraden durch β_1 und der Abstand der Linien durch β_2 definiert ist. Die zweite Abbildung 1.3(b) zeigt eine nichtlineare Interaktion. Diese ist erkennbar, da wieder zwei Geraden vorliegen, jedoch jetzt mit unterschiedlichen Anstiegen. In der Abbildung 1.3(c) ist eine Interaktion mit zwei sehr unterschiedlichen Funktionen für jede Behandlung erkennbar.

Ein Beispiel: Eine nichtlineare Interaktion zwischen einer Behandlung und dem Alter der Probanden kann dann vorliegen, wenn die Behandlung A die Überlebensdauer der jüngeren Probanden gegenüber den jüngeren Probanden der Gruppe B deutlich anhebt, während für ältere Probanden die Überlebensdauer gleich ist.

In weiterer Folge wird ausschließlich auf nichtlineare Interaktionen, die aus einer kontinuierlichen Variablen x und einer binären Variablen z bestehen, eingegangen. Die Verallgemeinerung auf Interaktionen mit einer kontinuierlichen Variable und einer kategoriellen Variable mit mehr als zwei Ausprägungen ist unter der Beachtung von globalen Tests und paarweisen Vergleichen möglich. Hingegen wird hier auf Interaktionen mit zwei kontinuierlichen Variablen nicht eingegangen. (vgl. Royston u. Sauerbrei, 2008, S.152)

Die herkömmliche Modellierung von Interaktionen kategorisiert die Variable x in eine gewisse Anzahl von Gruppen gemäß einem oder mehrerer Schnittpunkte und analysiert die Interaktionen in einem Modell mit Haupteffekten und vielfachen Interaktionstermen. Anhand eines Trendtests werden die Effekte von z über die geordneten Kategorien von x eruiert. Diese Art der Betrachtung der Interaktionen wirft einige Schwierigkeiten auf, zum Beispiel die komplizierte Interpretation der Ergebnisse sowie die Anzahl und Position der Schnittpunkte und die Abhängigkeit der statistischen Signifikanz von den Schnittpunkten. Laut Royston et al. (2006) soll die Dichotomisierung einer kontinuierlichen Variable generell vermieden werden. Die Modellierung von Interaktionen sollte so einfach wie möglich sein, daher verwenden wir in Folge FP ersten und zweiten Grades, welche über eine Vielfalt von einfachen, aber ausreichend komplexen Funktionen verfügen und gleichzeitig mathematische Einfachheit aufweisen, sowie einfach zu interpretieren sind. (vgl. Royston u. Sauerbrei, 2004)

Wird eine Interaktion einer kontinuierlichen Variable mit einer kategoriellen-

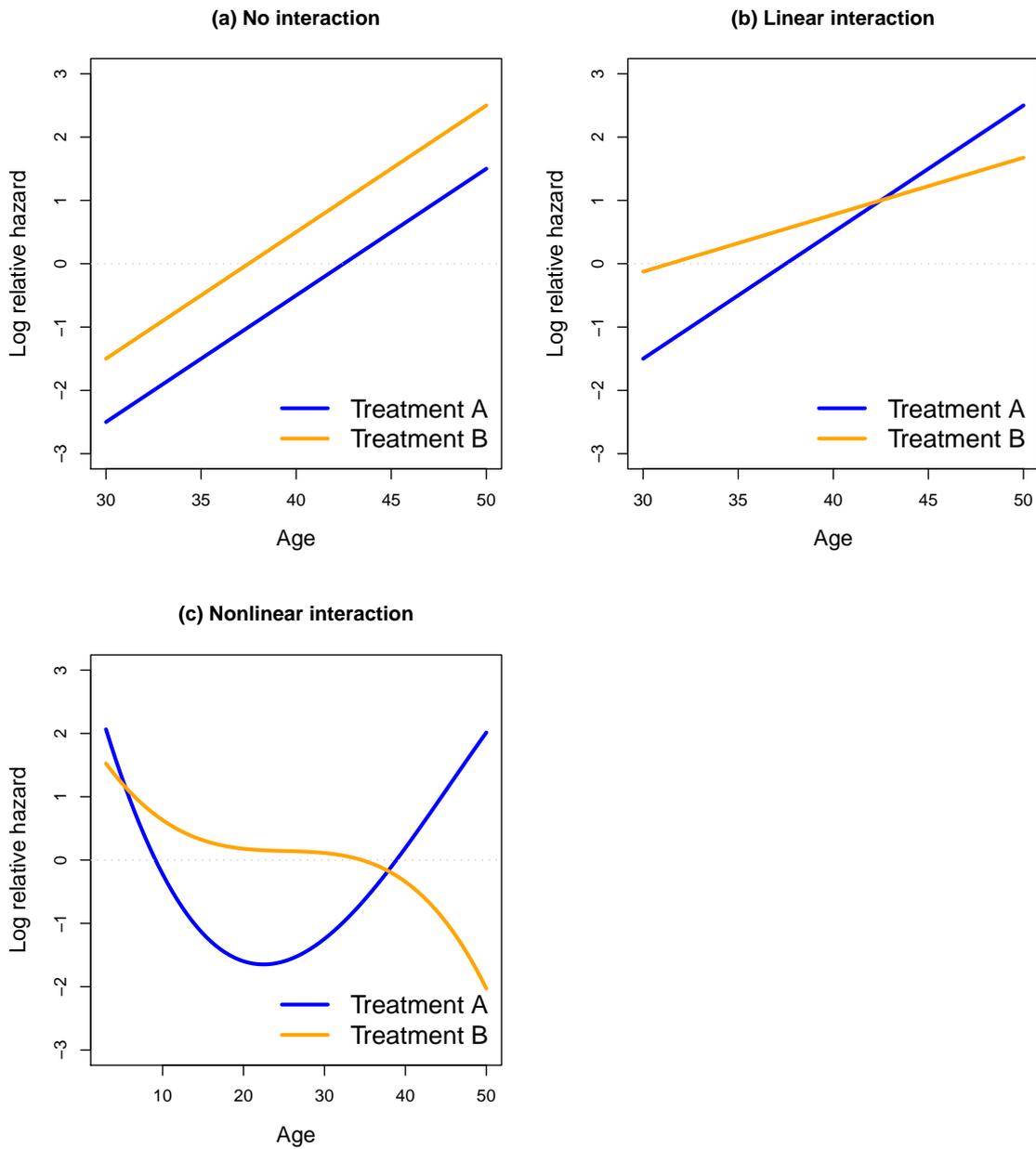


Abbildung 1.3: (a) Modell ohne Interaktion; (b) Modell mit linearer Interaktion; (c) Modell mit nichtlinearer Interaktion, der Variablen *Age* und *Treatment*.

len Variable vermutet, so kann der *MFP-Algorithmus* nicht mehr verwendet werden. Royston und Sauerbrei haben deshalb den *MFP-Algorithmus* zum *MFPI-Algorithmus* erweitert, der es ermöglicht auch Interaktionen zu testen (siehe Royston u. Sauerbrei, 2008, S.156)

Die Beziehung zwischen dem Outcome und der Variable x wird von FP mit denselben Potenzen modelliert, jedoch mit unterschiedlichen Koeffizienten β für jede Ausprägung von z . Für ein FP ersten und zweiten Grades sehen die Hazardfunktionen daher wie folgt aus:

$$\begin{aligned} h(t; x, z) &= h_0(t) \exp\{\beta_1 f_1(x) + \beta_2(z) + \beta_3 f_1(x) \cdot z\} \\ h(t; x, z) &= h_0(t) \exp\{\beta_1 f_1(x) + \beta_2 f_2(x) + \beta_3(z) + \\ &\quad + \beta_4 f_1(x) \cdot z + \beta_5 f_2(x) \cdot z\} \end{aligned}$$

Viele Studien sind so geplant, dass genügend statistische Mächtigkeit vorhanden ist, um einen Haupteffekt zu finden; die Power, um eine starke Interaktion zu finden, ist jedoch oftmals zu gering. Daher ist eine wichtige Voraussetzung für eine sinnvolle Analyse von Interaktionen eine ausreichende Stichprobengröße. (vgl. Royston u. Sauerbrei, 2008)

1.3.1.8 Grafische Darstellung der Fraktionalen Polynome

Der nichtlineare Effekt einer kontinuierlichen Variable x kann grafisch in einem Plot gezeigt werden, indem auf der x-Achse die Spanne der Variable x aufgetragen wird und auf der y-Achse die geschätzten Prädiktoren. Im Fall der Cox-Regression würde auf der y-Achse der sogenannte relative Hazard oder dessen Logarithmus aufgetragen werden.

Für die grafische Darstellung der Gruppeneffekte wird der sogenannte *Treatment-Effect Plot* verwendet; dieser zeigt, wie sich der Gruppeneffekt kontinuierlich über x ändert. Dafür wird der Gruppeneffekt $t(x) = \hat{f}_1(x) - \hat{f}_0(x)$ und dessen 95% Konfidenzintervall gegen die Variable x aufgetragen, wobei \hat{f}_j die geschätzte Funktion der Fraktionalen Polynome für die Gruppe j ($j = 0, 1$) darstellt. Eine signifikante Interaktion ist qualitativ, wenn $t(x)$ die Nulllinie kreuzt, denn dann ist nicht nur der Wert, sondern auch die Richtung des Gruppeneffekts von der Variable x abhängig. Eine quantitative Interaktion liegt dann vor, wenn sich der Gruppeneffekt über x ändert, aber für alle Werte von x in die selbe Richtung geht. (vgl. Royston u. Sauerbrei, 2004 und 2008, S.154)

1.3.1.9 FP-Analyse in R

In der Software R (R Development Core Team, 2009) ist die Cox-Regression mit der Schätzung von nichtlinearen Effekten mit Hilfe der Fraktionalen Polynome und zusätzlich mit der Firth-Korrektur im Package `coxphf.beta` (Version 1.51) möglich. Es besteht die Möglichkeit dieses Package unter http://www.meduniwien.ac.at/msi/biometrie/programme/frameset_programme.htm (Stand vom 25.03.2011) herunterzuladen.

Die Cox-Regression wird mit dem Befehl `coxphf.beta` aufgerufen. In den Aufruf kommt bei `data` der Name des Datensatzes und bei `formula` die Modellformel. Die meisten der anderen Elemente haben bereits einen Defaultwert, das Signifikanzniveau ist mit `alpha=0.05` festgelegt, `alpha.fp = c(1,0.2,0.2)` legt die Signifikanzniveaus für die Tests im *RA2-Algorithmus* fest. Dabei ist `alpha.fp[1]` der α -Wert, der für die Testung von FP2 vs. dem Nullmodell verwendet wird, `alpha.fp[2]` ist das Signifikanzniveau für den Test zwischen dem FP2 und dem linearen Modell und `alpha.fp[3]` wird für die Testung von FP2 gegen FP1 benützt. Durch `fp.max` wird der maximale Grad der FP festgelegt, ausgewählt werden kann zwischen 1 und 2 (Defaultwert). Mit `print=TRUE` (Defaulteinstellung) lässt man sich die Ergebnisse für alle gerechneten Modelle ausgeben, d.h. wenn man `print=FALSE` setzt, erhält man nur die Ergebnisse für das vom *RA2-Algorithmus* ausgewählte Endmodell. Sollte es Probleme bei der Konvergenz der Schätzer geben, so kann man die maximale Anzahl an Iterationen (`maxit=50` default) erhöhen. Mit `selection.by` kann man Einfluss auf die Selektierung nehmen; setzt man den Befehl gleich `"plr"` (Penalized Likelihood Ratio Statistik, Defaultwert), so wird das beste FP Modell anhand der PLR Statistik selektiert, während bei `"score"` die Score-Statistik dafür verwendet wird. Diese ist schneller, da keine Iterationen benötigt werden. Möchte man, dass die Schätzung mit der Firth-Korrektur durchgeführt wird, so schreibt man `firth=TRUE`; für `firth=FALSE` wird der MLE berechnet. Mit `penalty` kann man die Stärke des Strafterms verändern, dies funktioniert nur in Kombination mit `firth=TRUE`. Der Defaultwert ist auf 0.5 gesetzt, dies entspricht der Firth-Korrektur; bei Verwendung von größeren Zahlenwerten wird der Strafterm größer. Für die Berechnungsmethode der Konfidenzintervalle kann man zwischen der Profile Penalized Log Likelihood (`p1=TRUE`, default) und der Waldstatistik (`p1=FALSE`) wählen. Zur Veranschaulichung soll folgendes Beispiel mit den Kovariablen *BMI*, *Alter*, *Raucherstatus* und dem *Geschlecht* dienen:

```
coxphf.beta(data=mydata, Surv(survival,censur) ~ fp(BMI) +
Geschlecht + Alter + Raucherstatus, alpha.fp=c(1,0.157,0.157),
firth=TRUE, maxit=500).
```

Da es sich um ein Cox-Modell handelt, muss die Modellformel ein `Survival Object` sein, d.h. die linke Seite der Formel enthält die Überlebenszeit (hier `survival`) und die Zensierungsvariable (hier `censur`). Möchte man für eine Variable nichtlineare Effekte zulassen, wie zum Beispiel für die Variable `BMI`, so muss man diese Variable als `fp(BMI)` schreiben. Um noch zusätzlich eine Interaktion zwischen der kontinuierlichen Variable `BMI` und der binären Variable `Geschlecht` einzufügen, wird die Modellformel wie folgt geschrieben:

$$\text{Surv}(\text{survival}, \text{censur}) \sim \text{fp}(\text{BMI}) * \text{Geschlecht} + \text{Alter} + \text{Raucherstatus}$$

Weitere Optionen können in der R-Hilfe, die man mit `?coxphf.beta` erhält, nachgelesen werden.

Es ist noch zu beachten, dass der *RA2-Algorithmus* im R-Befehl `coxphf.beta` so programmiert ist, dass ein Modell mit einem Interaktionsterm der kontinuierlichen Variablen x und der binären Variable z nicht gegen ein Modell getestet wird, das z und $fp(x)$ enthält. D.h., die kontinuierliche Variable x kann als nichtlineare Interaktion mit der binären Variable z oder linear (lineares Modell) modelliert werden. Sie kann aber auch nicht im Modell inkludiert sein (Nullmodell).

Zur Rechenzeit (mit einem Intel(R) Core(TM)2 Quad CPU Prozessor mit 2.66GHz (4 CPUs)) : Diese verlängert sich eindeutig, wenn eine Variable mit FP modelliert wird (von < 0.01 Sekunden auf 0.41 für obiges Beispiel mit $n = 619$). Kommt noch eine Interaktion hinzu, so nimmt die Rechenzeit wiederum deutlich zu (0.64 Sekunden für obiges Beispiel). Die Laufzeit für das Cox-Modell mit der ML-Methode ist im Vergleich zur Firth-Methode um ein Vielfaches kürzer (von 0.64 auf 5.81 Sekunden für obiges Beispiel mit Interaktion und mit FP). Wird die Stichprobengröße erhöht, so steigt die Rechenzeit exponentiell an. Ebenso nimmt die Rechenzeit zu, je mehr Kovariablen in das Modell aufgenommen werden.

Alle folgenden Berechnungen in R wurden unter der Version 2.9.2 durchgeführt (R Development Core Team, 2009).

1.3.2 Restricted Cubic Splines

Für die Modellierung von nichtlinearen Effekten können neben den FP auch Splines verwendet werden. Diese produzieren geglättete Kurven und bieten daher eine flexible Möglichkeit zur Identifizierung von nichtlinearen Beziehungen. Splines n -ten Grades bestehen aus mehreren einzelnen Polynomfunktionen, deren maximaler Grad n ist. Die Knoten sind jene Punkte, an denen die Polynomfunktionen zusammengesetzt werden. In Folge betrachten wir ausschließlich eine Untergruppe der Familie der Splines, die *Restricted Cubic Splines* (RCS).

1.3.2.1 Definition

Restricted Cubic Splines gehören zu den nichtparametrischen Methoden in der Statistik. Sie setzen sich aus (maximal) kubischen Polynomfunktionen zusammen. Bei k Knoten t_1, t_2, \dots, t_k ist die Splinefunktion gegeben durch:

$$C(x) = \beta_0 + \beta_1 x + \sum_{j=1}^{k-2} \theta_j C_j(x) \quad \text{mit } j = 1, \dots, k-2 \quad (1.12)$$

wobei die kubischen Terme gegeben sind durch $C_1(x), \dots, C_{k-2}(x)$ und

$$C_j(x) = (x - t_j)_+^3 - \frac{(x - t_{k-1})_+^3 [t_k - t_j]}{[t_k - t_{k-1}]} + \frac{(x - t_k)_+^3 [t_{k-1} - t_j]}{[t_k - t_{k-1}]} \quad (1.13)$$

Dabei ist $(k)_+ = \max(0, k)$ und

$$(x - k)_+^3 = \begin{cases} (x - k)^3 & \text{wenn } x \geq k \\ 0 & \text{wenn } x < k. \end{cases}$$

Splines sind so konstruiert, dass sie $k - 1$ mal stetig differenzierbar sind, d.h. bei den RCS existiert eine erste und zweite Ableitung. Für die Knoten spielt dies eine wichtige Rolle, da ihre Übergänge dadurch geglättet werden. $C(x)$ ist linear in x für $x < t_1$ und $x \geq t_k$, d.h. die Splinefunktion ist in den Enden linear. Durch diese Restriktion (engl.: *restriction*) kam diese Untergruppe der Splines zu dem Namen *Restricted Cubic Splines*. (vgl. Heinzl et al., 1996; Heinzl u. Kaider, 1997)

1.3.2.2 Anzahl und Position der Knoten

Die Anzahl und Position der Knoten muss festgelegt werden, wobei die Wahl der Knoten unabhängig von der abhängigen Variable sein soll. Für die in der Medizin vorkommenden Stichprobengrößen reichen meist 3 – 5 Knoten.

Wird die Anzahl der Knoten erhöht, so steigt die Varianz der Schätzer der Splines und die Gefahr einer Überanpassung an die Daten nimmt zu. Laut Durrleman und Simon (1989) sollen die Knoten nahe den Rändern gesetzt werden, da die Enden als linear definiert sind, zum Beispiel können die Knoten dem 5-ten oder 10-ten Perzentil entsprechen. Die Knoten direkt auf die Ränder zu setzen scheint nicht sinnvoll, da sonst die Splinefunktionen von Ausreißern extrem beeinflusst werden können. Die restlichen Knoten sollten über die Quantile der Variable x gleichverteilt werden, dann sind die Ergebnisse unempfindlich bezüglich der Knotenpositionen. Von Harrell (2001, S.23) werden für $k=3, 4$ und 5 als Empfehlung die Perzentile $\{10, 50, 90\}$, $\{5, 35, 65, 95\}$ und $\{5, 27.5, 50, 72.5, 95\}$ angegeben. Laut Durrleman und Simon (1989) kann ein Plot der Residuen ebenfalls hilfreich für die Setzung der Knoten sein.

1.3.2.3 Restricted Cubic Splines im Cox-Modell

Verwendet man nichtlineare Funktionen zur Modellierung von Beziehungen im Cox-Modell, so erhalten wir statt (1.1) folgende Hazardfunktion:

$$h(t, x) = h_0(t) \exp(f(x)). \quad (1.14)$$

Der Term $x\beta$ wird durch die nichtlineare Funktion $f(x)$ ersetzt. Für die log Hazardratio folgt daher:

$$\text{LHR}(x) = \log(h(t, x)/h_0(t)) = f(x). \quad (1.15)$$

Wenn $x = 0$, dann muss auch $f(x)$ gleich Null sein. Für die Approximation von $f(x)$ werden die RCS verwendet:

$$\begin{aligned} \text{LHR}(x) &= \log(h(t, x)/h_0(t)) \approx C(x) - C(0) \\ &= \beta_1 x + \sum_{j=1}^{k-2} \theta_j (C_j(x) - C_j(0)) \end{aligned} \quad (1.16)$$

wobei $C_j(0) = 0$ für $j = 1, \dots, k-2$, wenn $t_1 \geq 0$, d.h. jeder Hazardwert wird in Relation zu $x = 0$ interpretiert. Oftmals ist es jedoch sinnvoll die Hazardwerte in Relation zu einem anderen Wert (Referenzwert) zu sehen, zum Beispiel dem Median oder einem Wert, der sich in der Mitte der Variablenstrecke befindet. Betrachtet man zum Beispiel den systolischen Blutdruck, so macht es wenig Sinn, die LHR eines Patienten mit einem LHR bei einem systolischen Blutdruck von 0 zu vergleichen, während ein Vergleich mit einem Normalwert von zum Beispiel 120 für die Interpretation einfacher und interessanter wäre.

Für einen Referenzwert von m erhält man somit eine LHR von:

$$\begin{aligned} \text{LHR}_m(x) &= \log(h(t, x)/h_0(t; m)) = f(x) - f(m) \approx C(x) - C(m) \\ &= \beta_1(x - m) + \sum_{j=1}^{k-2} \theta_j(C_j(x) - C_j(m)). \end{aligned} \quad (1.17)$$

Die grafische Darstellung der Splinefunktionen ist ident mit der grafischen Darstellung der Fraktionalen Polynome. Auf der x-Achse werden die Werte der Variable x aufgetragen, während auf der y-Achse die geschätzte Funktion $\hat{C}(x)$ aufgetragen wird. Bei den Splines ist die grafische Darstellung sehr wichtig, da die geschätzten Parameter alleine kaum interpretierbar sind. (vgl. Heinzl u. Kaider, 1997)

1.3.2.4 RCS-Analyse in SAS

Mit dem SAS Makro RCS (Heinzl und Kaider, 1997) ist es möglich nichtlineare Effekte anhand von RCS zu schätzen. Für die Schätzung verwenden wir dasselbe Modell, welches schon bei den FP als Beispiel diente, mit der Variable *survival* für die Überlebenszeit, der Zensierungsvariable *censur* und den beiden metrischen Kovariablen *BMI* und *Alter* sowie den beiden binären Kovariablen *Geschlecht* und *Raucherstatus*. Somit erhält man folgenden SAS-Aufruf für das Makro RCS:

```
%MACRO RCS(TITLE=%STR(Schätzung nichtlinearer Effekt mit RCS),
DATA=mydata, TIME=survival, STATUS=censur, COV1=BMI, COV2=Alter,
COV3=Geschlecht, COV4=Raucherstatus, WHAT1=0, KNOTS1=17 25 40,
GRAPH=LR, TIMEUNIT=MONTHS, DIRDATA= %STR(C: \Maria\DATA\), PROGRAM=
%STR(C: \Maria\Results\File.sas));
```

Durch die Option `TITLE` kann ein Titel für die Analyse eingegeben werden, `DATA` definiert den Datensatz, der für die Berechnungen verwendet wird, wobei durch die Parameter `TIME` und `STATUS` zusätzlich definiert wird, welche der Variablen die Überlebenszeit und den Status der Patienten widerspiegelt. Bei der Statusvariable ist es notwendig, dass die zensierten Beobachtungen mit 0 codiert sind, während die Ereignisse jede beliebige Zahl außer Null annehmen dürfen. Mit den Optionen `COV1`, ..., `COV20` können bis zu 20 Kovariablen in das Modell aufgenommen werden. Zu jeder Kovariable i ($i = 1, \dots, 20$) gehört eine `WHATi` Option. Diese teilt dem Makro mit, wie die Kovariable verwendet werden soll. Wird `WHATi=0` gesetzt, wie im obigen Beispielcode, so wird für diese Variable ein nichtlinearer Effekt unter der Verwendung von RCS (nach Durrleman u. Simon, 1989) geschätzt. Bei `WHATi=1` wird mit

der Kovariable i und der Zeit eine Interaktion (nach Hess, 1994) modelliert. Wenn `WHATi=2`, so wird mit Hilfe der RCS ein Interaktionseffekt von der Zeit und einer zeitabhängigen binären Kovariable errechnet, wie es bei Heinzl et al. (1996) beschrieben ist. Wird kein Wert angegeben, wird die Kovariable wie gewöhnlich geschätzt. Wurde `WHATi` ein Wert zugeordnet, so kann man zusätzlich die Anzahl und Position der Knoten für jede Kovariable i mit `KNOTSi` festlegen. Mit der Option `GRAPH` kann man zusätzlich die Art des grafischen Outputs bestimmen, `GRAPH=HR` produziert einen Plot mit der relativen Hazardfunktion, `GRAPH=LR` (Defaulteinstellung) zeichnet in die Grafik die logarithmierte relative Hazardfunktion und `GRAPH=BOTH` zeichnet beide Plots. Die Grafiken beinhalten auch immer das 95% Konfidenzintervall. Wenn `WHATi` gleich 1 oder 2 ist, wird mit der Option `TIMEUNIT` die Zeiteinheit der x-Achse festgelegt (default=`MONTHS`). Die Option `DIRDATA` gibt das Verzeichnis an, welches den Datensatz aus `DATA` enthält, `PROGRAM` gibt den Filenamen an, der für die Generierung der SAS Files benötigt wird, beide müssen vom User spezifiziert werden.

Das Makro kann unter http://www.meduniwien.ac.at/msi/biometrie/programme/frameset_programme.htm (Stand vom 25.03.2011) herunter geladen werden. (vgl. Heinzl u. Kaider, 1997 und 2006)

Für die Analyse der Fallstudie (siehe Kapitel 3) wurde das unveröffentlichte Makro `showinteractionv2` (siehe elektronischer Anhang) verwendet. Dieses Makro baut zwar auf dem Makro `RCS` auf, kann jedoch zusätzlich eine nichtlineare Interaktion einer kontinuierlichen Variable und einer binären Variable berechnen. Nimmt man das Beispiel von zuvor und möchte man eine Interaktion der kontinuierlichen Variable *BMI* und der binären Variable *Geschlecht* schätzen, so lautet der Befehl in SAS:

```
%showinteractionv2(data=mydata, time=survival, cens=censur,
treat=Geschlecht, var=BMI, covar=Alter Raucherstatus, knots=17 25
40, options=firth, nknots=3)
```

Bei `data` setzt man wie zuvor den Namen des Datensatzes ein, bei `time` die Überlebenszeit oder Beginn und Ende (`start, stop`) der Beobachtungszeit und bei `cens` die Zensierungsvariable. Bei `treat` wird jene binäre Variable eingefügt, die Teil der Interaktion ist und bei `var` jene kontinuierliche Variable, die zur Interaktion gehört. Weitere Variablen, die zur Adjustierung des Modells dienen, können durch die Option `covar` hinzugefügt werden. Die Anzahl und Position der Knoten wird durch `nknots` und `knots` festgelegt. Die Schätzer, die das Modell auswirft, sind MLE; durch die Option `options=firth` erhält man Schätzer, die mit der Firth-Methode geschätzt wurden (für weitere

Möglichkeiten von `options` siehe `proc phreg`). Es besteht außerdem die Möglichkeit, mit `estcov` eine Kovarianzmatrix für das Modell anzugeben, wobei die Datenmatrix dann nicht für die Schätzung verwendet wird. Die Anzahl der Iterationen für die Evaluierung ist mit `steps=101` festgelegt, kann jedoch bei Bedarf geändert werden. Bei der Option `by` wird eine Variable eingesetzt, die Subgruppen definiert, wodurch separate Analysen für die Subgruppen erfolgen. Mit `where` kann man eine einzelne Subgruppenanalyse durchführen, zum Beispiel `where= %str(Raucherstatus=1)`). Als Output erhält man drei Grafiken. Die erste Abbildung zeigt die Hazardfunktion für die Werte der Variable `var` relativ zum Referenzwert und das dazugehörige modellbasierte 95% KI für die Gruppe 0 der binären Variable `treat`. Die zweite Abbildung zeigt selbiges für die Gruppe 1 der binären Variable `treat` und die dritte Abbildung zeigt die eigentliche Interaktion der Gruppen (= *Treatment-Effect Plot*), die Hazardratio zwischen der Gruppe 0 und 1 sowie wieder das dazugehörige modellbasierte 95% KI. Durch die Option `label` kann man die Beschriftung der y-Achse in den Plots festlegen. (vgl. Heinze, 2011) Für die weitere Evaluierung wurde neben der Software R auch das SAS System (Version 9.2, SAS Institute, 2010 und 2009) verwendet.

1.3.3 Der Vergleich von FP und Restricted Cubic Splines

Da für die nichtlineare Schätzung von Effekten sowohl FP als auch RCS verwendet werden können, liegt es am Anwender, für welche Modellierung er/sie sich entscheidet. Folgende Betrachtungsweisen können bei der Entscheidung dienlich sein:

FP sind globale „influence smoother“, während RCS lokale „influence smoother“ sind. Ein Ausreißer wirkt sich daher bei den FP auf die gesamte Schätzung des nichtlinearen Effekts aus, während bei den RCS nur jene Schätzung zwischen zwei Knoten betroffen ist, wo sich auch der Ausreißer befindet. (vgl. Royston et al., 1999)

Die Funktion für FP ist sehr simpel und ihre Schätzer bei FP1 leicht zu interpretieren, im Gegensatz dazu ist die Funktion für die RCS komplexer und ihre Schätzer sind nur schwer interpretierbar.

Im *RA2-Algorithmus* werden 45 Modelle betrachtet, was auch dementsprechend zeitaufwendig sein kann. Bei den RCS betrachten wir ein Modell, das schnell gerechnet werden kann. Dieser Vorteil der RCS geht jedoch verloren, wenn man für sie ebenso einen Selektionsalgorithmus verwendet (siehe Royston u. Sauerbrei, 2008, S.205). Bei den FP besteht die Möglichkeit einer linearen Modellierung. Anders ist es bei den RCS, die immer aus Polynomfunktionen

bestehen (außer unter der Verwendung eines Selektionsalgorithmus).

Die Erfahrungen von Royston und Sauerbrei (2008) zeigen, dass in Regionen mit wenigen Datenpunkten die Unterschiede der Schätzungen von FP und Splines stark unterschiedlich sein können, hingegen bei genügend Daten die geschätzten Funktionen und ihre Interpretationen sehr ähnlich sind. Sie bevorzugen trotzdem die Ergebnisse einer MFP-Analyse, da diese ihren Kenntnissen nach stabiler sind und ihrer Meinung nach die Ergebnisse einfacher zu interpretieren sind.

1.4 Konfidenzintervalle

1.4.1 Das parametrische (Bootstrap-) Konfidenzintervall

Das parametrische 95% Bootstrap-Konfidenzintervall (KI) für die relative Hazardfunktion der Variable x wird wie folgt berechnet:

1. Es werden b Bootstrapsamples angelegt, indem von einem Originaldatensatz jeweils n Beobachtungen mit Zurücklegen gezogen werden. Dabei ist n frei wählbar, meist ist n jedoch die Anzahl der Beobachtungen des Originaldatensatzes. Von diesen Bootstrapsamples werden jeweils die relativen Hazardfunktionen berechnet.
2. Danach wird von jedem einzelnen Wert der Variable x die Standardabweichung (Sd) \hat{s}_x der b relativen Hazardfunktionen errechnet. D.h. man erhält einen Vektor \hat{s} von Standardabweichungen für die Variable x , wobei sich jeder Eintrag in \hat{s} auf einen Punkt von x bezieht.
3. Das parametrische 95% Bootstrap-KI ist gegeben durch:

$$[h(t; x; m) \pm 1,96 \cdot \hat{s}] \quad (1.18)$$

Bootstrap-KI benötigen ein vielfaches der Rechenzeit, da b -mal so viele Modelle geschätzt werden müssen.

1.4.2 Das nichtparametrische (Bootstrap-) Konfidenzintervall

Das nichtparametrische Bootstrap-KI zum Niveau $(1-\alpha)$, wobei meist $\alpha = 5\%$, ist gegeben durch:

$$[q_{\frac{\alpha}{2}}; q_{1-\frac{\alpha}{2}}] \quad (1.19)$$

wobei $q_{\frac{\alpha}{2}}$ und $q_{1-\frac{\alpha}{2}}$ Quantile der Bootstrap-Verteilung von $h(t; x; m)$ sind.

Man erhält dieses Konfidenzintervall, indem man

1. für jedes der b Bootstrapsamples die relative Hazardfunktion $h(x; t; m)$ berechnet und
2. für jeden Wert der Variable x das 0,025-te und das 0,975-te Quantil der b relativen Hazardfunktionen errechnet.

Die obere nichtparametrische Konfidenzgrenze ist dann der Vektor $q_{\frac{\alpha}{2}}$ der 0,025-Quantile und die untere nichtparametrische Konfidenzgrenze ist der Vektor $q_{1-\frac{\alpha}{2}}$ der 0,975-Quantile.

Bei den nichtparametrischen Bootstrap-KI ist von Vorteil, dass die Parameter nicht normalverteilt sein müssen.

1.4.3 Das modellbasierte Konfidenzintervall

Modellbasierte (engl.: modelbased) Konfidenzintervalle (KI) werden hauptsächlich bei RCS verwendet, sie können jedoch auch auf FP angewandt werden. Möchte man für eine geschätzte Splinefunktion $\hat{C}(x)$ modellbasierte KI konstruieren, dann nimmt man zunächst einen fixen Wert x_0 , für diesen kann man dann $\hat{C}(x_0)$ schreiben als

$$\hat{C}(x_0) = \beta' \hat{X}_0 \quad (1.20)$$

wobei $\hat{\beta} = (\hat{\beta}_0, \hat{\beta}_1, \hat{\theta}_1, \dots, \hat{\theta}_{k-2})^t$ und $X_0 = (1, x_0, C_1(x_0), \dots, C_{k-2}(x_0))^t$. Das $(1 - \alpha)$ Konfidenzintervall für $\hat{C}(x_0)$ ist dann gegeben durch

$$[\hat{C}_{lower}(x_0), \hat{C}_{upper}(x_0)] = [\hat{C}(x_0) \pm (\gamma X_0' V X_0)^{\frac{1}{2}}], \quad (1.21)$$

mit der Kovarianzmatrix V für $\hat{\beta}$ und mit $\gamma = \chi_{p, 1-\alpha}^2$, wobei γ das $(1 - \alpha)$ -Quantil einer χ^2 -Verteilung mit p Freiheitsgraden ist. Wenn $p = 1$, dann entspricht das KI dem Standard KI für einen Parameter; wenn p gleich der Anzahl der Kovariablen ist, dann entspricht das KI einem *Scheffé*-Konfidenzintervall, wobei das KI für alle p Parameter simultan eingehalten wird.

Für die Fraktionalen Polynome sieht das modellbasierte KI wie folgt aus:

$$[\hat{C}_{lower}(x_0), \hat{C}_{upper}(x_0)] = [\hat{C}(x_0) \pm (\gamma X_0' V X_0)^{\frac{1}{2}}], \quad (1.22)$$

wobei $\hat{\beta} = (\hat{\beta}_1, \dots, \hat{\beta}_m)^t$ und $X_0 = (\varphi_m(x_0, p))^t$.
(vgl. Heinzl et al., 1996)

Durch den Parameter γ kann man somit für die zwei Freiheitsgrade, die durch die Auswahl der Potenzen bei FP2 verbraucht werden, korrigieren. Dabei wird $p = 3$ gesetzt und γ entspricht dem $\chi_{0.95;3}^2$ -Quantil.

Bei den modellbasierten KI geht das gerechnete Modell in Form der Modellauswahl in die Berechnung der KI mitein, dadurch kommt es zu überoptimistischen KI. In Folge weisen modellbasierte KI oftmals eine geringe Breite auf.

Kapitel 2

Simulation

2.1 Einleitung

Die Simulationsstudie dient zur Gewinnung weiterer Erkenntnisse über die Firth-Korrektur und über Fraktionale Polynome. Royston und Altman (1994) untersuchten bereits nichtlineare Effekte von kontinuierlichen Variablen im Cox-Modell anhand von Fraktionalen Polynomen (FP). In dieser Simulation werden im Speziellen nichtlineare Effekte von Interaktionen einer kontinuierlichen und einer dichotomen Variable, die durch Fraktionale Polynome modelliert werden, untersucht. Die Parameter werden sowohl mit der Firth-Korrektur als auch mit der Maximum-Likelihood-Methode geschätzt. Somit können im Anschluß die beiden Schätzverfahren miteinander verglichen werden. Für diesen Vergleich wurden die Abweichungen der Schätzer, genauer der *mediane Bias* und die mittlere absolute Abweichung (engl.: *median absolute deviation* (MAD)) der Punktschätzer berechnet. Die Firth-Korrektur wurde schon eingehend von Heinze und Schemper (2001) und Heinze und Dunkler (2008) im Cox-Modell untersucht. In Kombination mit Fraktionalen Polynomen stellt die vorliegende Arbeit jedoch die erste Untersuchung dar.

2.2 Datengenerierung

Die Daten der Simulationsstudie wurden in Anlehnung an eine Nierentransplantationsstudie aus dem Paper „*Mortality in renal transplant recipients given erythropoietins to increase haemoglobin concentration*“ (Heinze et al., 2009) generiert.

Für die Simulationsstudie wurden 7 Kovariablen (Tabelle 2.1) erzeugt. Dabei stellt die Variable *HB* die Hämoglobinkonzentration (g/l) der Probanden dar, *SAP* gibt Auskunft über den systolische Blutdruck (mm Hg) und *Cholesterol*

(mmol/l) simuliert die Cholesterinkonzentration der Patienten im Blut. Die Variable *eryth* gibt an, ob ein Patienten mit dem Medikament *Erythropoietin*, besser bekannt unter *EPO*, behandelt wurde, *diab* steht für eine Diabeteserkrankung, *heart.fail* gibt, an ob die Probanden bisher an einem Herzleiden litten und *vasc.disease* spiegelt eine Gefäßerkrankung wieder. Bei den drei Variablen *HB*, *SAP* und *Cholesterol* handelt es sich um kontinuierliche Variablen, bei den Variablen *eryth*, *diab*, *heart.fail* und *vasc.disease* um vier binäre Variablen.

Die abhängigen Variablen sind der Zensierungsindikator *censur*, welcher angibt, ob ein Event, in unserem Falle der Tod, eingetreten ist oder nicht, und die Zeitvariable *time*, welche die Zeitdauer bis zum beobachteten Ereignis oder dem Ende der Nachbeobachtung misst.

Tabelle 2.1: Kodierung der Variablen

Variable	Variablenlabel	Erläuterung	
HB	Hämoglobinwert	gemessen in g/l	
SAP	systolischer Blutdruck	gemessen in mm Hg	
Cholesterol	Cholesterinkonzentration	gemessen in mmol/l	
eryth	Erythropoietin	0 nein	1 ja
diab	Diabetes	0 nein	1 ja
heart.fail	Herzversagen	0 nein	1 ja
Vasc.disease	Gefäßerkrankung	0 nein	1 ja
censur	Zensierungsindikator	0 zensiert	1 Event
time	Beobachtungszeitraum	in Monaten	

Bei der Datengenerierung wurde der Datensatz zuerst in zwei gleich große Teile geteilt. Ein Teil wurde danach *eryth* = 1 zugeordnet, entsprechend den mit EPO behandelten Patienten, der zweite Teil, die unbehandelten Patienten, *eryth* = 0. Die kontinuierlichen Variablen stammen aus einer Normalverteilung, die binären Variablen aus einer Binomialverteilung, die jeweiligen Parameter sind in Tabelle 2.2 ersichtlich. Die Effekte der Variablen wurden zusätzlich festgelegt, *SAP* bekam den kleinsten Effekt mit 0.01, *diab*, *heart.fail* und *vasc.disease* bekamen einen Effekt von 0.69 und *Cholesterol* erhielt einen Effekt von 0.42. Durch zwei komplexe Funktionen konnte für die Variable *HB* ein nichtlinearer Effekt erzeugt werden. Zusätzlich wurde eine Interaktion mit der Variable *eryth* festgelegt, indem eine der komplexen Funktionen der Variable *HB*, wo *eryth* = 1 ist, zugeordnet wurde und die zweite Funktion der Gruppe *eryth* = 0. Da unterschiedliche Funktionen gewählt wurden, auf

die im übernächsten Kapitel (2.4) noch genauer eingegangen wird, konnten verschiedene Szenarien betrachtet werden.

Tabelle 2.2: Parameter der Verteilungen der Variablen

Normalverteilte Variablen			
Variable	eryth	Mittelwert	Standardabw.
HB	1	90	15
	0	110	16
SAP	1	110.3	51.1
	0	88	52
Cholesterol	1	5.26	1.64
	0	5.10	1.46
Binomialverteilte Variablen			
Variable	eryth	Wert	Wahrscheinlichkeit
diab	1	1	0.4
	0	1	0.25
Heart.fail	1	1	0.186
	0	1	0.170
Vasc.disease	1	1	0.342
	0	1	0.299

Unzensierte Sterbezeiten wurden aus einer Weibullverteilung mit den Parametern $\alpha = 1$ und $\beta = 2$ gezogen. Die Beobachtungsdauer jedes Patienten ergab sich aus einer Gleichverteilung zwischen 0 und τ . Zensierte Lebensdauern ergaben sich aus dem Minimum von Sterbe- und Beobachtungszeit jedes Patienten, wobei der Parameter τ so gewählt wurde, dass sich 0%, 33% und 67% zensierte Lebensdauern ergaben.

2.3 Überblick über Simulation

Zum besseren Verständnis des Ablaufs der Simulationsstudie dient ein Flowchart (Abbildung 2.1) der die verschiedenen Ebenen der Analyse widerspiegelt.

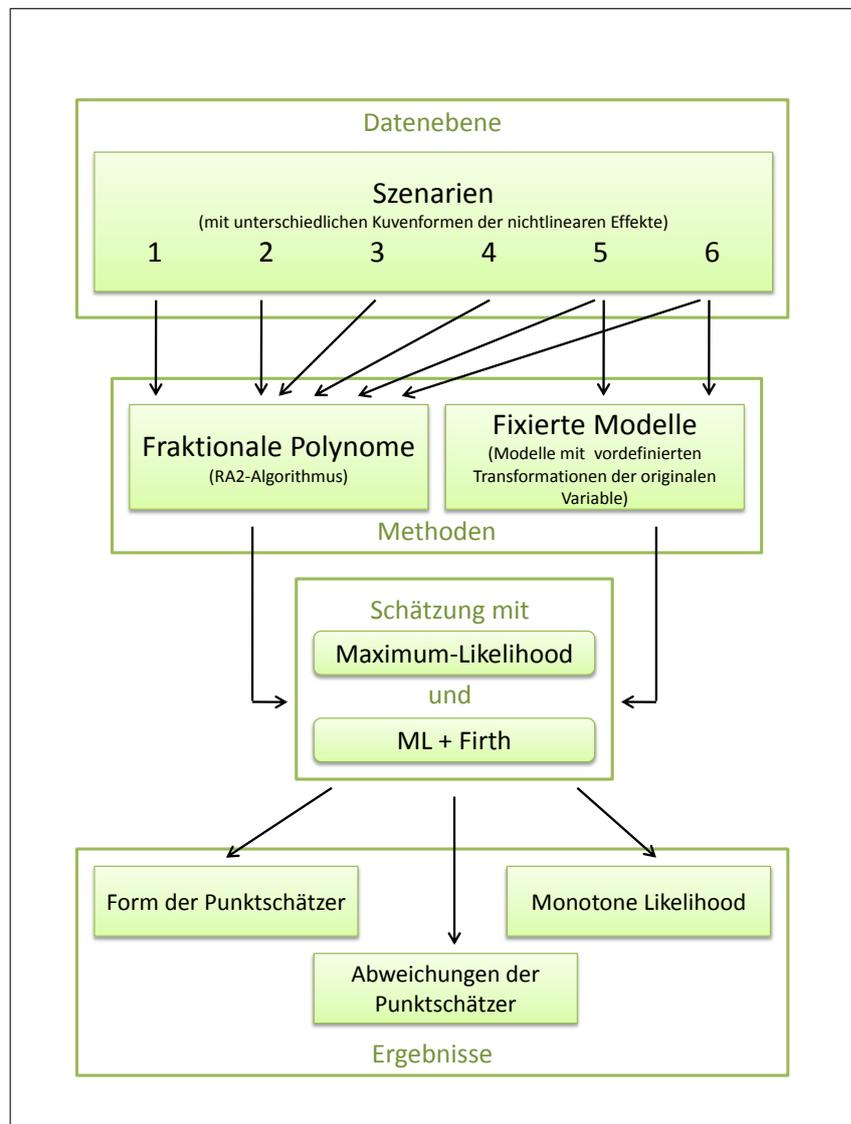


Abbildung 2.1: Flowchart von Simulation

2.4 Szenarien

Durch 12 komplexe Funktionen erhalten wir 6 verschiedene Szenarien. Pro Szenario benötigen wir zwei Funktionen, die den nichtlinearen Effekt der Variable HB widerspiegeln. Eine Funktion $f_1(x)$ für jene Gruppe, die das Medikament *Erythropoietin* erhalten hat ($eryth = 1$) und die zweite Funktion $f_2(x)$ für die unbehandelte Gruppe ($eryth = 0$). Die Wechselwirkung ergibt sich aus der Differenz der beiden Funktionen ($f_3(x) = f_1(x) - f_2(x)$). Durch sie lässt sich eine Beobachtung aus der ersten Gruppe mit einer Beobachtung aus der zweiten Gruppe bei gleichem Hämoglobingehalt vergleichen. In Folge sind alle Funktionen der sechs Szenarien aufgeführt:

Szenario 1:

$$f_1(x) = \mathbf{x}^3(-4.028775 \cdot 10^{-06}) + \mathbf{x}^2 0.001636204 + \mathbf{x}(-0.2138003) + \log(\mathbf{x}/2 - 5)1.152772 + 3.732537 \quad (2.1)$$

$$f_2(x) = \mathbf{x}^3(-4.116858 \cdot 10^{-05}) + \mathbf{x}^2 0.01314295 + \mathbf{x}(-1.599217) + \log(\mathbf{x})20.26735 - 23.71787 \quad (2.2)$$

Szenario 2:

$$f_1(x) = \mathbf{x}^3(-2.578416 \cdot 10^{-05}) + \mathbf{x}^2 0.01047171 + \mathbf{x}(-1.368322) + \log(\mathbf{x}/2 - 5)7.377741 + 28.28824 \quad (2.3)$$

$$f_2(x) = \mathbf{x}^3(-3.216295 \cdot 10^{-05}) + \mathbf{x}^2 0.01026793 + \mathbf{x}(-1.249388) + \log(\mathbf{x})15.83387 - 18.35771 \quad (2.4)$$

Szenario 3:

$$f_1(x) = \mathbf{x}^3 4.028775 \cdot 10^{-06} + \mathbf{x}^2(-0.001636204) + \mathbf{x}(0.2138003) + \log(\mathbf{x}/2 - 5)(-1.152772) - 2.432537 \quad (2.5)$$

$$f_2(x) = \mathbf{x}^3(-4.116858 \cdot 10^{-05}) + \mathbf{x}^2 0.01314295 + \mathbf{x}(-1.599217) + \log(\mathbf{x})20.26735 - 23.11787 \quad (2.6)$$

Szenario 4:

$$f_1(x) = \mathbf{x}^3 2.578416 \cdot 10^{-05} + \mathbf{x}^2(-0.01047171) + \mathbf{x}(1.368322) + \log(\mathbf{x}/2 - 5)(-7.377741) - 27.78824 \quad (2.7)$$

$$f_2(x) = \mathbf{x}^3(-4.116858 \cdot 10^{-05}) + \mathbf{x}^2 0.01314295 + \mathbf{x}(-1.599217) + \log(\mathbf{x})20.26735 - 23.11787 \quad (2.8)$$

Szenario 5:

$$f_1(x) = \mathbf{x}^3 7.977153 \cdot 10^{-07} + \mathbf{x}^2 (-8.90214 \cdot 10^{-05}) - 0.5840333 \quad (2.9)$$

$$f_2(x) = \mathbf{x}^3 (-4.557718 \cdot 10^{-06}) + \mathbf{x}^2 0.0005717053 - 0.26 \quad (2.10)$$

Szenario 6:

$$f_1(x) = \mathbf{x}^3 (-2.642726 \cdot 10^{-06}) + \mathbf{x}^{-2} (-24239.83) + 6.762096 \quad (2.11)$$

$$f_2(x) = \mathbf{x}^3 (-8.77397 \cdot 10^{-07}) + \mathbf{x}^{-2} (-12328.42) + 2.8 \quad (2.12)$$

Interaktion:

$$f_3(x) = f_1(x) - f_2(x) \quad (2.13)$$

Wenn man diese Polynomfunktionen genauer betrachtet, so fällt auf, dass in Szenario 1 bis 4 Polynomfunktionen dritten Grades mit den Potenzen 3, 2 und 1 sowie ein Logarithmus vorliegen. Hingegen haben die Funktionen der Szenarien 5 und 6 Polynomfunktionen dritten Grades mit ausschließlich zwei Polynomgliedern mit den Potenzen 3 und 2. Fraktionale Polynome bestehen ebenso nur aus zwei Polynomgliedern und beinhalten in ihrem Set S die Potenzen 3 und 2 bzw. -2 . Folglich ist es sinnvoll zu überprüfen, ob die Schätzer der Szenarien 5 und 6 mit den Fraktionalen Polynomen besser abgebildet werden können als die Schätzer der Szenarien 1 bis 4.

Für jedes Szenario wurden sechs unterschiedliche Datenzellen mit darin je 200 Datensätzen angelegt. Diese Datenzellen unterscheiden sich durch ihren Zensierungsanteil und durch die Stichprobengröße der darin befindlichen Datensätze. Für die Simulationsstudie wurden die drei unterschiedlichen Zensierungsanteile 0%, 33% und 67% gewählt, d.h. bei einem Zensierungsanteil von 0% liegen ausschließlich Events vor, bei einem Zensierungsanteil von 33% sind in den Daten in 67% der Fälle Events aufgetreten, usw. Die Stichprobengröße wurde entweder mit 100 Beobachtungen oder 400 Beobachtungen pro Datensatz festgelegt. Schlussendlich erhalten wir für die Kombination aus Zensierungsanteil und Stichprobengröße sechs Datenzellen pro Szenario, das sind in Summe 36 unterschiedliche Datenzellen, die untersucht wurden.

Die Tabelle 2.3 veranschaulicht die Unterschiede der 36 Datenzellen nochmals überblicksmäßig. Jede Datenzelle erhält zur einfacheren Benennung eine Nummer, welche sich aus der Nummer des Szenarios, dem Zensierungsanteil und der Stichprobengröße zusammensetzt (siehe Tabelle 2.3, Spalte 5).

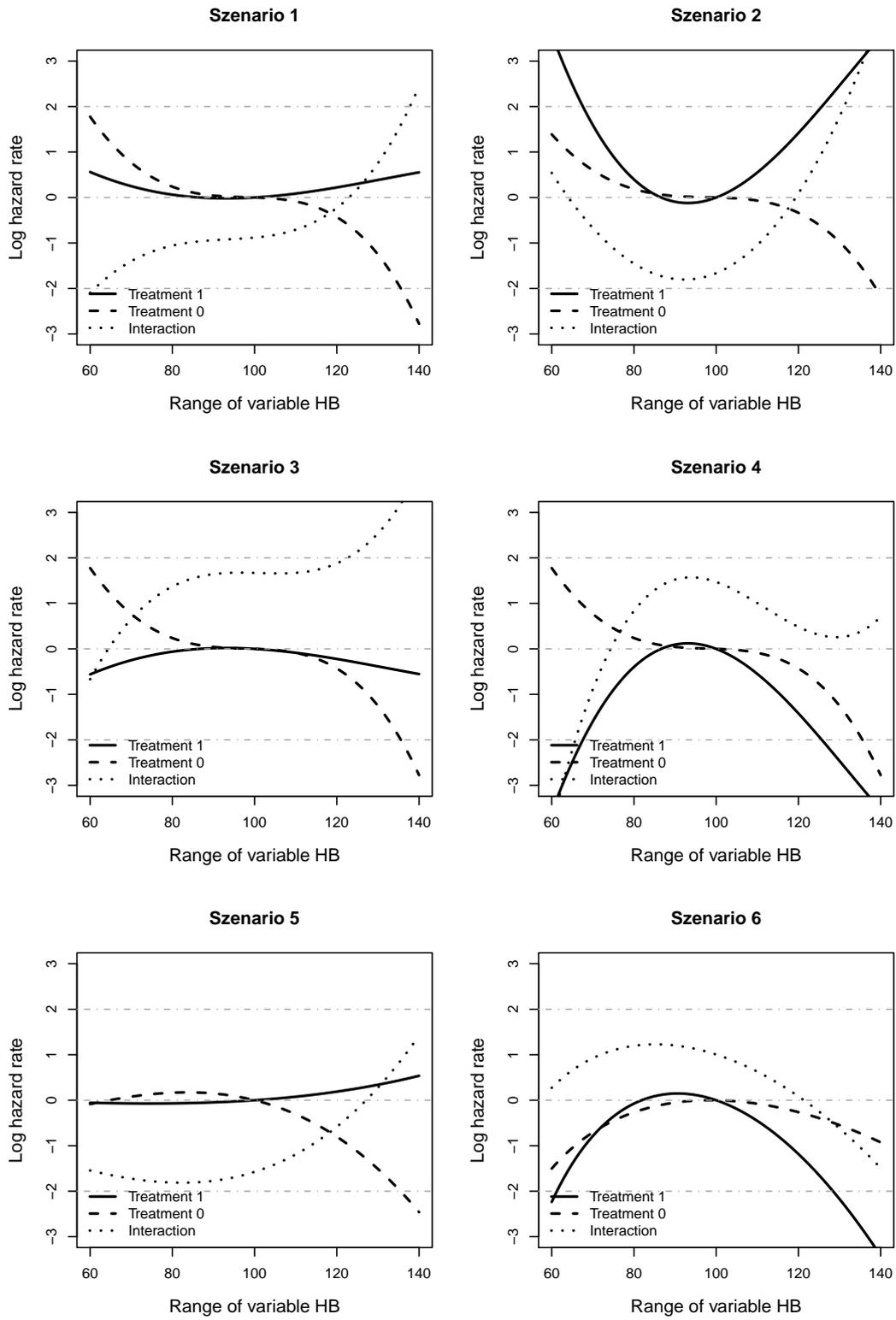


Abbildung 2.2: Wahre Funktionen

Tabelle 2.3: Übersicht der Datenzellen

Datenzellename	Szenario	zens.%	N	Nummer
Y1_e0_n100	1	0%	100	101
Y1_e0.33_n100	1	33%	100	1331
Y1_e0.67_n100	1	67%	100	1671
Y1_e0_n400	1	0%	400	104
Y1_e0.33_n400	1	33%	400	1334
Y1_e0.67_n400	1	67%	400	1674
Y2_e0_n100	2	0%	100	201
Y2_e0.33_n100	2	33%	100	2331
Y2_e0.67_n100	2	67%	100	2671
Y2_e0_n400	2	0%	400	204
Y2_e0.33_n400	2	33%	400	2334
Y2_e0.67_n400	2	67%	400	2674
Y3_e0_n100	3	0%	100	301
Y3_e0.33_n100	3	33%	100	3331
Y3_e0.67_n100	3	67%	100	3671
Y3_e0_n400	3	0%	400	304
Y3_e0.33_n400	3	33%	400	3334
Y3_e0.67_n400	3	67%	400	3674
Y4_e0_n100	4	0%	100	401
Y4_e0.33_n100	4	33%	100	4331
Y4_e0.67_n100	4	67%	100	4671
Y4_e0_n400	4	0%	400	404
Y4_e0.33_n400	4	33%	400	4334
Y4_e0.67_n400	4	67%	400	4674
Y5_e0_n100	5	0%	100	501
Y5_e0.33_n100	5	33%	100	5331
Y5_e0.67_n100	5	67%	100	5671
Y5_e0_n400	5	0%	400	504
Y5_e0.33_n400	5	33%	400	5334
Y5_e0.67_n400	5	67%	400	5674
Y6_e0_n100	6	0%	100	601
Y6_e0.33_n100	6	33%	100	6331
Y6_e0.67_n100	6	67%	100	6671
Y6_e0_n400	6	0%	400	604
Y6_e0.33_n400	6	33%	400	6334
Y6_e0.67_n400	6	67%	400	6674

2.5 Punktschätzer

Es wurden die Punktschätzer für zwei unterschiedliche Modelle berechnet: für das selektierte Modell, welches der *RA2-Algorithmus* zurück gibt, und für das fixierte Modell mit fixen Termen der kontinuierlichen Variable der Interaktion.

Tabelle 2.4: PLR-Statistiken der FP1 und FP2 Modelle des Beispieldatensatzes (geschätzt mittels ML-Methode)

FP1		FP2					
Power	PLR	Power		PLR	Power		PLR
	p_1	p_1	p_2		p_1	p_2	
-2	111.23	1	1	131.40	-1	0.5	132.82
-1	111.75	1	-2	133.46	-1	2	132.29
-0.5	112.45	1	-1	132.63	-1	3	132.14
0	113.35	1	-0.5	132.17	-0.5	-0.5	132.92
0.5	114.37	1	0	131.79	-0.5	0	132.68
1	115.46	1	0.5	131.54	-0.5	0.5	132.41
2	117.75	1	2	131.40	-0.5	2	131.85
3	120.02	1	-2	131.61	-0.5	3	131.78
		-2	-2	130.81	0	0	132.34
		-2	-1	132.28	0	0.5	132.03
		-2	-0.5	132.79	0	2	131.56
		-2	0	133.14	0	3	131.60
		-2	0.5	133.36	0.5	0.5	131.73
		-2	2	133.47	0.5	2	131.43
		-2	3	133.40	0.5	3	131.56
		-1	-1	132.98	2	2	131.57
		-1	-0.5	133.05	2	3	131.89
		-1	0	132.98	3	3	132.26

2.5.1 Selektierte Modelle

Das selektierte Modell (hier in R Schreibweise)

$$\text{Surv}(time, censur) \sim \text{fp}(HB) \cdot \text{eryth} + \text{diab} + \text{sap} + \text{Cholesterol} + \\ + \text{heart.fail} + \text{vasc.disease}$$

enthält die vom *RA2-Algorithmus* ausgewählten Terme für die Variable *HB*. Um die Ergebnisfindung und die Funktionsweise des *RA2-Algorithmus* besser veranschaulichen zu können, geben die Tabellen 2.4 bis 2.6 einen Einblick in den Ablauf dieser Prozedur. Die vorgeführten Ergebnisse stammen von einem Beispieldatensatz aus der Datenzelle 1674 (entspricht Szenario 1 mit einem Zensierungsanteil von 67% und $n = 400$; siehe Tabelle 2.3), der mit Hilfe der Maximum-Likelihood-Methode ausgewertet wurde. In Tabelle 2.4 sind die 8 FP1 und die 36 FP2 Modelle des *RA2-Algorithmus* mit ihren *Penalized Likelihood Ratios* (PLR) abgebildet. Das beste FP1 Modell ist jenes Modell mit der größten PLR, in diesem Beispiel ist es das Modell mit $p_1 = 3$, welches eine PLR von 120,02 aufweist. Bei den FP2 Modellen ist jenes Modell mit $p_1 = -2$ und $p_2 = 2$ das beste Modell, da es die größte PLR mit 133,47 hat.

Tabelle 2.5: Test Prozedur des *RA2-Algorithmus* (anhand des Beispieldatensatzes)

Modell	PLR	Powers	Vergleich	PLR.Diff	df	α	p-Wert
Null	102.61	-	FP2 vs. Null	30.86	6	1	< 0.001
Linear	115.46	1	FP2 vs. linear	18.01	4	0.157	0.001
FP1	120.02	3	FP2 vs. FP1	13.46	3	0.157	0.004
FP2	133.47	-2, 2	-	-	-	-	-

In der Tabelle 2.5 werden die Schritte des *RA2-Algorithmus* dargestellt. Zuerst wird das Nullmodell, das die Variable *HB* nicht enthält, gegen das beste FP2 Modell getestet. Dies geschieht mit einem χ^2 -Test mit nun 6 statt 4 Freiheitsgraden, da wir im FP2 Modell zusätzlich eine Interaktion haben. Da der p-Wert < 0.001 ist, gehen wir zum zweiten Schritt über. In diesem wird das beste FP2 Modell gegen das lineare Modell getestet, unter der Verwendung von 4 Freiheitsgraden. Da das Ergebnis erneut signifikant ist (p-Wert = 0.001), landet man im letzten Schritt dieser Prozedur. Hier wird das beste FP2 Modell gegen das beste FP1 Modell getestet. Da diese bereits

eruiert wurden, können wir die Differenz der PLR-Statistiken von FP1(3) gegen FP2(-2, 2) unter der Verwendung von 3 Freiheitsgraden testen. Ein p-Wert von 0.004 bedeutet, dass das FP2 Modell das bessere Modell ist.

Tabelle 2.6: Ergebnisse der Schätzung mit der Maximum-Likelihood-Methode

Variable	$\hat{\beta}$	HR	95% KI		p-Wert
			lower	upper	
powM2(HB/100)	-3.94	0.02	0.00	0.34	0.007
pow2(HB/100)	-4.46	0.01	0.00	0.14	0.000
powM2(HB/100):eryth	4.98	145.56	7.62	2781.10	0.001
pow2(HB/100):eryth	5.97	393.01	24.70	6252.11	0.000
eryth	-12.14	0.00	0.00	0.00	0.000
diab	0.46	1.58	1.11	2.26	0.012
sap	0.01	1.01	1.01	1.01	0.000
Cholesterol	0.39	1.48	1.32	1.66	0.000
Heart.fail	0.47	1.60	1.00	2.56	0.050
Vasc.disease	0.68	1.98	1.37	2.86	0.000

Das selektierte Modell, das der *RA2-Algorithmus* für den Beispieldatensatz liefert, lautet daher:

$$\begin{aligned} \log h(t, x) = & \log h_0(t) + \beta_1 HB^{-2} + \beta_2 HB^2 + \beta_3 HB^{-2} \cdot eryth + \\ & \beta_4 HB^2 \cdot eryth + \beta_5 eryth + \beta_6 diab + \beta_7 sap + \beta_8 Cholesterol + \\ & \beta_9 Hear.fail + \beta_{10} Vasc.disease \end{aligned} \quad (2.14)$$

Die Koeffizienten β_1 bis β_{10} , die Hazardratios und deren dazugehöriges 95% Konfidenzintervall sowie die p-Werte sind in der Tabelle 2.6 angegeben.

Wird derselbe Datensatz mit der Firth-Methode ausgewertet, so gibt der *RA2-Algorithmus* als Ergebnismodell das FP2 Modell mit $p_1 = 3$ und $p_2 = 3$ zurück. Die Ergebnisse sind in der Tabelle 2.7 zu finden. Durch dieses Beispiel kann man erkennen, dass der *RA2-Algorithmus* für beide Schätzmethoden nicht automatisch das gleiche Modell zurück gibt.

Die Log Relative Hazards für die beiden Gruppen und für deren Interaktion können formal niedergeschrieben werden: die Gleichungen (2.15) und (2.16) tun dies mit einem Referenzwert von 0. Bei den Gleichungen (2.18) und (2.19) liegt der Referenzwert entsprechend der klinischen Interpretation bei 100, wie auch in allen Abbildungen. Für die Interaktion (2.17) spielt der Referenzwert keine Rolle.

$$\log \left[\frac{h(t, HB | eryth = 1)}{h_0(t)} \right] = (\beta_1 + \beta_3) \left(\frac{HB}{100} \right)^{p_1} + (\beta_2 + \beta_4) \left(\frac{HB}{100} \right)^{p_2} + \beta_5 \quad (2.15)$$

$$\log \left[\frac{h(t, HB | eryth = 0)}{h_0(t)} \right] = \beta_1 \left(\frac{HB}{100} \right)^{p_1} + \beta_2 \left(\frac{HB}{100} \right)^{p_2} \quad (2.16)$$

$$\log \left[\frac{h(t, HB | eryth = 1)}{h(t, HB | eryth = 0)} \right] = \beta_3 \left(\frac{HB}{100} \right)^{p_1} + \beta_4 \left(\frac{HB}{100} \right)^{p_2} + \beta_5 \quad (2.17)$$

$$\log \left[\frac{h(t, HB | eryth = 1)}{h(t | eryth = 1, HB = 100)} \right] = (\beta_1 + \beta_3) \left[\left(\frac{HB}{100} \right)^{p_1} - 1 \right] + (\beta_2 + \beta_4) \left[\left(\frac{HB}{100} \right)^{p_2} - 1 \right] \quad (2.18)$$

$$\log \left[\frac{h(t, HB | eryth = 0)}{h(t | eryth = 0, HB = 100)} \right] = \beta_1 \left[\left(\frac{HB}{100} \right)^{p_1} - 1 \right] + \beta_2 \left[\left(\frac{HB}{100} \right)^{p_2} - 1 \right] \quad (2.19)$$

Der Unterschied zwischen den beiden Schätzverfahren ist grafisch in der Abbildung 2.3 ersichtlich.

Jedes Modell wurde mit den zwei unterschiedlichen Alphakombinationen $\text{Alpha}=\text{c}(1, 0.157, 0.157)$ und $\text{Alpha}=\text{c}(1, 1, 1)$ berechnet, um in späterer Folge (Kapitel 2.6) zu sehen, inwieweit sich die Wahl der Alpha-Werte auf die Schätzung auswirkt.

2.5.2 Fixiertes Modell

Zusätzlich zu diesen Modellen wurde für die Szenarien 5 und 6 ein weiteres Modell mit fixen Termen der Variable HB , welche den vorgegebenen Exponenten der Gleichungen (2.9) (bzw. (2.10)) und (2.11) (bzw. (2.12)) entsprechen, berechnet. Somit erhält man für die beiden Szenarien 5 und 6 das fixierte Modell (R Schreibweise):

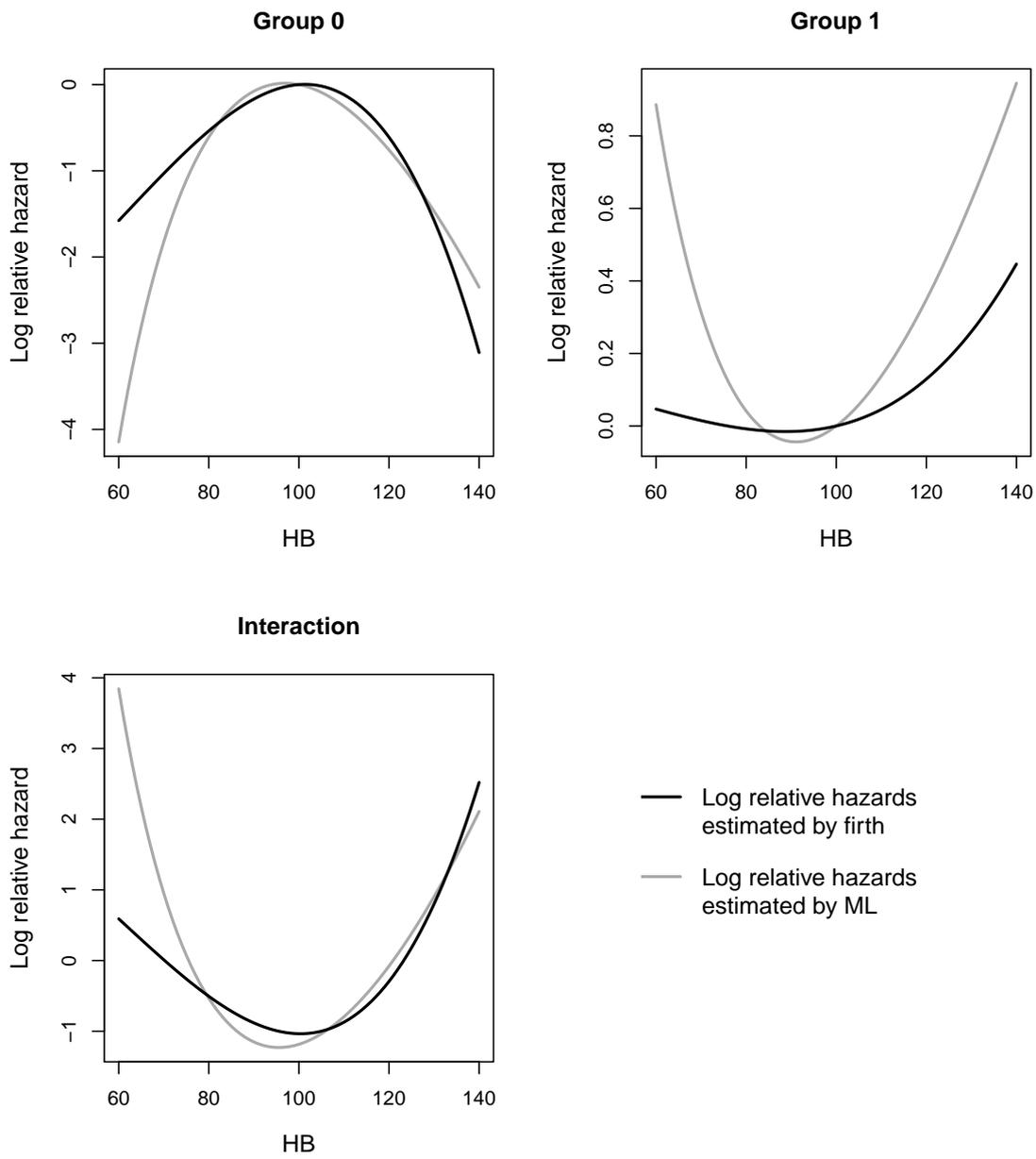


Abbildung 2.3: Vergleich der Punktschätzer

Tabelle 2.7: Ergebnisse der Schätzung mit der Firth-Methode

Variable	$\hat{\beta}$	HR	95% KI		p-Wert
			lower	upper	
pow3(HB/100)	3.39	29.61	1.70	517.00	0.020
Rpow3(HB/100)	-9.77	0.00	0.00	0.05	0.005
eryth	2.53	12.51	0.55	285.94	0.114
diab	0.48	1.62	1.13	2.31	0.008
sap	0.01	1.01	1.01	1.01	0.000
Cholesterol	0.38	1.47	1.31	1.64	0.000
Heart.fail	0.54	1.71	1.07	2.72	0.024
Vasc.disease	0.67	1.96	1.36	2.82	0.000
pow3(HB/100):eryth	-3.56	0.03	0.00	0.79	0.036
Rpow3(HB/100):eryth	10.58	$3.9 \cdot 10^4$	12.99	$1.19 \cdot 10^8$	0.010

$$\text{Surv}(time, censor) \sim HB_1 + HB_2 + HB_3 + HB_4 + eryth + diab + sap + \\ + Cholesterol + heart.fail + vasc.disease$$

mit

$$\begin{aligned} HB_1 &= HB^{p_1} \\ HB_2 &= HB^{p_2} \\ HB_3 &= HB_1 \cdot eryth \\ HB_4 &= HB_2 \cdot eryth. \end{aligned}$$

Dieses Modell soll dazu dienen, zu überprüfen, wie gut die Fraktionalen Polynome die wahren Effekte abbilden und ob die Ergebnisse stark von der Selektion abhängen.

2.6 Abweichungen der Schätzer

2.6.1 Der mediane Bias und der MAD

Um ein Maß für die Abweichung der Schätzer von den wahren Funktionen zu bekommen, wurden der *mediane Bias* und die *Median Absolute Deviation*

(MAD) berechnet. Damit sollte es im Anschluss möglich sein, die Maximum-Likelihood-Methode mit der Firth-Methode zu vergleichen. Wie man in den Gleichungen (2.20) und (2.21) sehen kann, wird in beiden Fällen im ersten Schritt die Differenz der Schätzer (\hat{Y}) und der vorgegebenen wahren Funktionen (Y) berechnet. Im Anschluss wird beim medianen Bias der Median der Differenzen berechnet, während beim MAD der Median der *absoluten* Differenzen errechnet wird. Als Ergebnis erhält man für jede Datenzelle einen Vektor von Abweichungen, dessen Länge der Spannweite der untersuchten Variable entspricht. In der Simulationsstudie ist dies die Spannweite der Variable HB mit den Werten von 60 bis 140. Bei beiden Maßen handelt es sich um nichtparametrische Berechnungen.

$$Bias = median(\hat{Y} - Y) \quad (2.20)$$

$$MAD = median|\hat{Y} - Y| \quad (2.21)$$

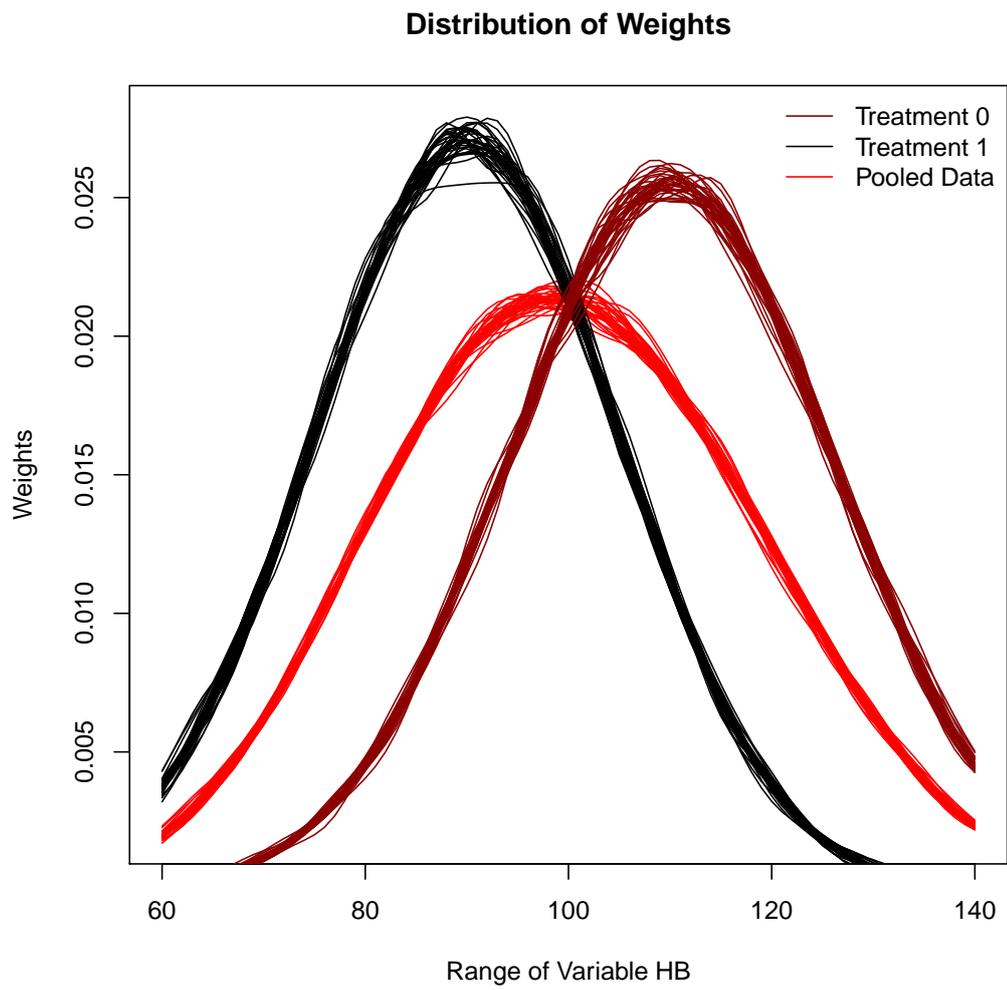
Da Vektoren nur schwer vergleichbar sind, wird für die Untersuchung der Schätzgenauigkeit die gewichtete Fläche unter der Funktion der Abweichungen berechnet, indem der gewichtete Mittelwert des Vektors der Abweichungen errechnet wird. Die Berechnung der Gewichte wird im Anschluss erläutert (2.6.2).

2.6.2 Berechnung der Gewichte

Für die Berechnung der Gewichte wird zuerst die Variable HB aller 200 Datensätze einer Datenzelle zu einer Variable mit der Länge $200 \times n$ zusammengefügt und von dieser die Dichte für die Datenpunkte 60 bis 140 berechnet. Dieser Dichtevektor mit der Länge 81 wird durch die Summe des Dichtevektors dividiert, somit erhält man einen Vektor mit den Gewichten für die Datenpunkte 60 bis 140. Diese Gewichtung gilt jedoch nur für die Interaktion; für die Gewichte der Gruppe 1 bzw. der Gruppe 0 verwendet man ausschließlich die Variable HB , wo $eryth = 1$ bzw. wo $eryth = 0$.

Die Abbildung 2.4 zeigt die Verteilung der Gewichte der beiden Gruppen und der Interaktion für die 36 Datenzellen.

Abbildung 2.4: Verteilung der Gewichte



2.7 Results

2.7.1 Monotone Likelihood

Sowohl für die generierten Datensätze der Simulation als auch für ihre Bootstrapsamples wurde überprüft, wie oft Monotone Likelihood bei der Maximum-Likelihood-Schätzung auftritt.

Aus Tabelle 2.8 kann dies für die 36 generierten Datenzellen, jeweils bestehend aus 200 Datensätzen, abgelesen werden. Bei einer Stichprobengröße von 400 kommt es in den Datenzellen kein einziges Mal zum Auftreten des Phänomens der Monotonen Likelihood, während bei einer Stichprobengröße von 100 in 6 von 18 Datenzellen Monotone Likelihood auftritt. Dabei zeigt sich, dass speziell bei einem Zensierungsanteil von 67% ein gehäuftes Auftreten feststellbar ist.

Somit können die schon bekannten Annahmen, dass Monotone Likelihood umso häufiger auftritt, je kleiner die Stichprobengröße bzw. je höher der Zensierungsanteil ist (Bryson und Johnson, 1981), bekräftigt werden.

Tabelle 2.8: Monotone Likelihood in den generierten Daten

Szenario	n=100			n=400		
	0%	33%	67%	0%	33%	67%
1	0/200	0/200	0/200	0/200	0/200	0/200
2	4/200	7/200	3/200	0/200	0/200	0/200
3	0/200	0/200	4/200	0/200	0/200	0/200
4	0/200	0/200	2/200	0/200	0/200	0/200
5	0/200	0/200	0/200	0/200	0/200	0/200
6	0/200	0/200	2/200	0/200	0/200	0/200

Pro Datenzelle wurden 200 Datensätze generiert, wobei in x aus 200 Monotone Likelihood auftritt ($x/200$).

In den Bootstrapsamples (Tabelle 2.9) ist der Trend bezüglich der Stichprobengröße und des Zensierungsanteils noch deutlicher zu erkennen. Während bei einer Stichprobengröße von 100 in allen Datenzellen Monotone Likelihood auftritt, kommt es bei $n = 400$ nur in neun Datenzellen dazu. Bei $n = 100$ kommt es dabei in 0.006% bis 13.628% der Fälle zu Monotoner Likelihood, bei $n = 400$ hingegen nur in 0.002% bis 1.024% der Fälle. Im Mittel tritt Monotone Likelihood bei $n = 100$ und einem Zensierungsanteil von 0% in 0.41% der Datensätze einer Datenzelle auf. Steigt der Zensierungsanteil auf

33% bzw. 67%, so steigt auch das Mittel auf 0.68% bzw. 5.60%. Bei $n = 400$ nimmt das Mittel der Häufigkeiten von Monotoner Likelihood ebenfalls mit einem steigenden Zensierungsanteil zu.

Zusätzlich zeigt der Vergleich der beiden Tabellen (2.8 und 2.9), dass Monotone Likelihood in Bootstrapsamples häufiger auftritt als in den zugrundeliegenden Datensätzen (Loughin, 1995).

Tabelle 2.9: Monotone Likelihood in den Bootstrapsamples

Szenario	n=100			n=400		
	0%	33%	67%	0%	33%	67%
1	0.006%	0.272%	1.492%	0.002%	0.002%	0.094%
2	2.182%	3.648%	3.656%	0.230%	0.466%	1.024%
3	0.008%	0.032%	13.628%	0%	0%	0.024%
4	0.258%	0.038%	6.358%	0%	0%	0.050%
5	0.008%	0.032%	2.070%	0%	0.004%	0%
6	0.012%	0.070%	6.414%	0%	0%	0%

Pro Datenzeile gibt es für die 200 generierten Datensätze insgesamt 50000 Bootstrapsamples, dabei tritt pro Datenzeile in $x\%$ der Fälle Monotone Likelihood auf.

2.7.2 Grafische Darstellung der Punktschätzer

Jede der Abbildungen 2.5 bis 2.40 beinhaltet außer den konvergierten Schätzfunktionen (mit der Firth-Methode) noch deren Median und die der Simulation zugrundeliegende wahre Funktion. Jede Zeile einer Abbildung, d.h. jeweils drei Grafiken, stellt die Schätzfunktionen einer Datenzeile dar. Jede linke Grafik zeigt die geschätzten Log Relative Hazards der Variable HB , wo $eryth = 1$ (Gruppe 1), jeder mittlere Plot zeigt selbiges für $eryth = 0$ (Gruppe 0) und jede rechte Grafik die Schätzfunktionen der Interaktion der beiden Gruppen. Die roten Linien sind die wahren Funktionen der Schätzer und die grünen Linien stellen die Mediane der Schätzfunktionen dar. Da nicht immer alle Schätzer der 200 Datensätze konvergierten, ist zusätzlich in jeder Grafik die Anzahl der konvergierten Schätzer angegeben.

In den Bildern der Abbildungen 2.5 bis 2.6 kann man deutlich sehen, dass die Schätzer des selektierten Modells die Originallinie der Gruppe 1 gut wieder spiegeln, jedoch die der Gruppe 0 nicht; folglich wird die Wechselwirkung

ebenso schlecht angepasst. Wird die Stichprobengröße auf 400 erhöht (Abbildungen 2.8 bis 2.10), so nimmt die Varianz der Schätzer deutlich ab.

Ähnliches gilt für die Abbildungen (2.11 bis 2.16) des Szenarios 2, mit dem Unterschied, dass vom Median ausgehend trotz der schlechten Anpassung der Schätzer der Gruppe 0 die Wechselwirkung gut geschätzt werden kann. Es fällt aber auch auf, dass speziell in diesem Szenario einige Male das lineare Modell den komplexeren Funktionen der Fraktionalen Polynome vorgezogen wird.

Die geschätzten Funktionen der Szenarien 3 (Abbildungen 2.20 bis 2.22) und 4 (Abbildungen 2.26 bis 2.28) weichen stark von den vorgegebenen Originalfunktionen ab, sowohl in der Gruppe 0 als auch in der Gruppe 1 und folglich auch bei der Wechselwirkung. Es hat den Anschein, dass die Fraktionalen Polynome die komplexeren Funktionen nicht ausreichend gut abbilden können. Im Szenario 5 stimmen bei einer Datensatzstichprobengröße von 100 die Schätzer der Gruppe 1 ideal mit dem Originalschätzer überein (Abbildungen 2.29 bis 2.31). Die Schätzer der Gruppe 0 weichen im Hämoglobinbereich von 60 bis 100 ein wenig von dem vorgegebenen Effekt ab, was sich auch auf die Wechselwirkung in diesem Bereich auswirkt. Die Anpassung im Bereich von 100 bis 140 ist gut. Die Schätzer der Datenzellen 504, 5334 und 5674 (Abbildungen 2.32 bis 2.34) stimmen ausgezeichnet mit den vorgegebenen Effekten überein und ihre Varianz ist aufgrund der gestiegenen Stichprobengröße dieser drei Datenzellen wesentlich kleiner.

Die Bilder des letzten Szenarios (Abbildung 2.35 bis 2.40) liefern stark unterschiedliche Ergebnisse: die Punktschätzer der Datenzellen 601, 6331, 604 und 6674 liefern eine sehr gute bis perfekte Anpassung, während die Mediane der geschätzten Effekte in den Datenzellen 6671 und 6334 stark von den wahren Funktionen abweichen.

Abbildung 2.5: Punktschätzer des selektierten Modells der Zelle 101

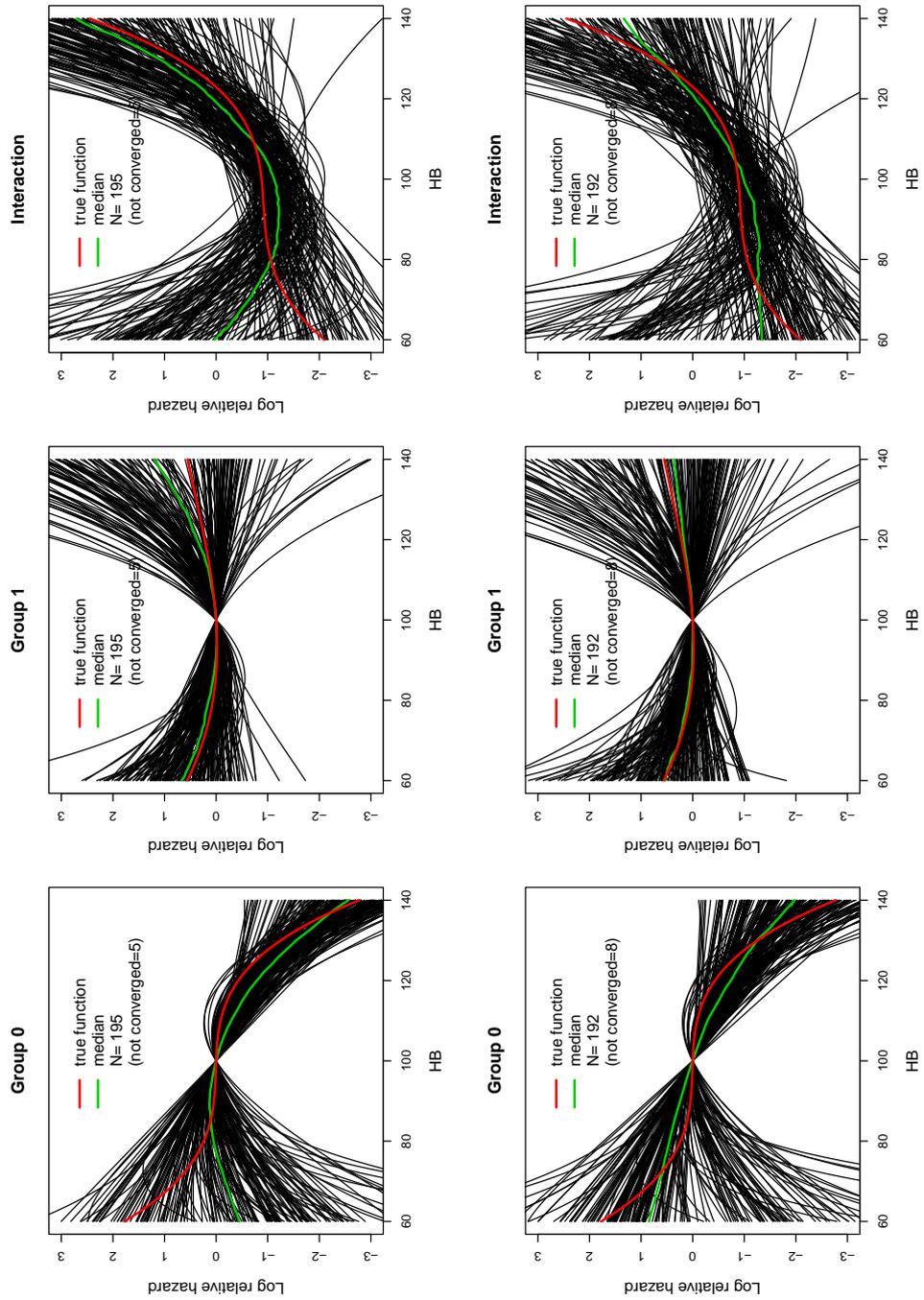


Abbildung 2.6: Punktschätzer des selektierten Modells der Datenzelle 1331

Abbildung 2.7: Punktschätzer des selektierten Modells der Zelle 1671

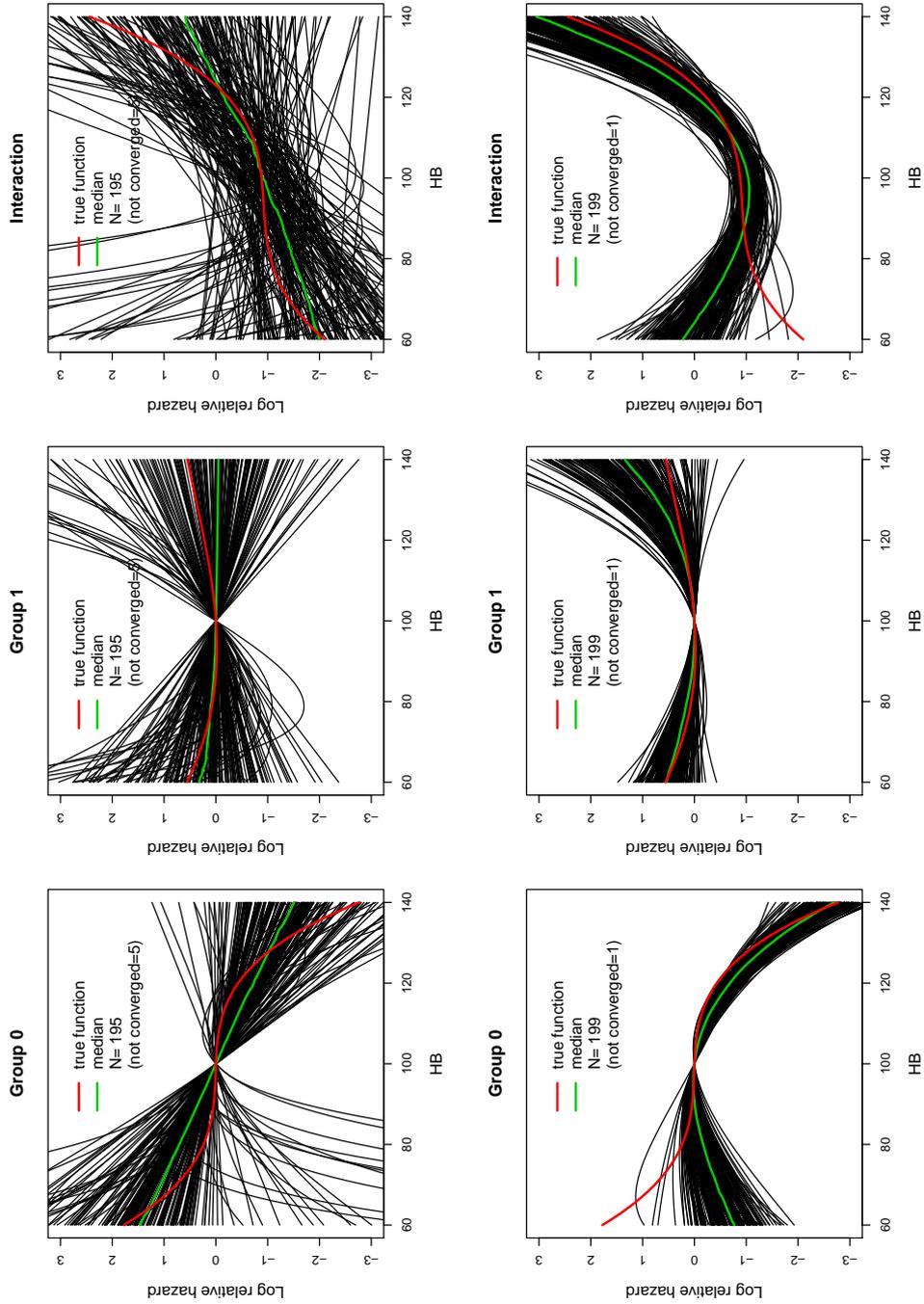


Abbildung 2.8: Punktschätzer des selektierten Modells der Datenzelle 104

Abbildung 2.9: Punktschätzer des selektierten Modells der Zelle 1334

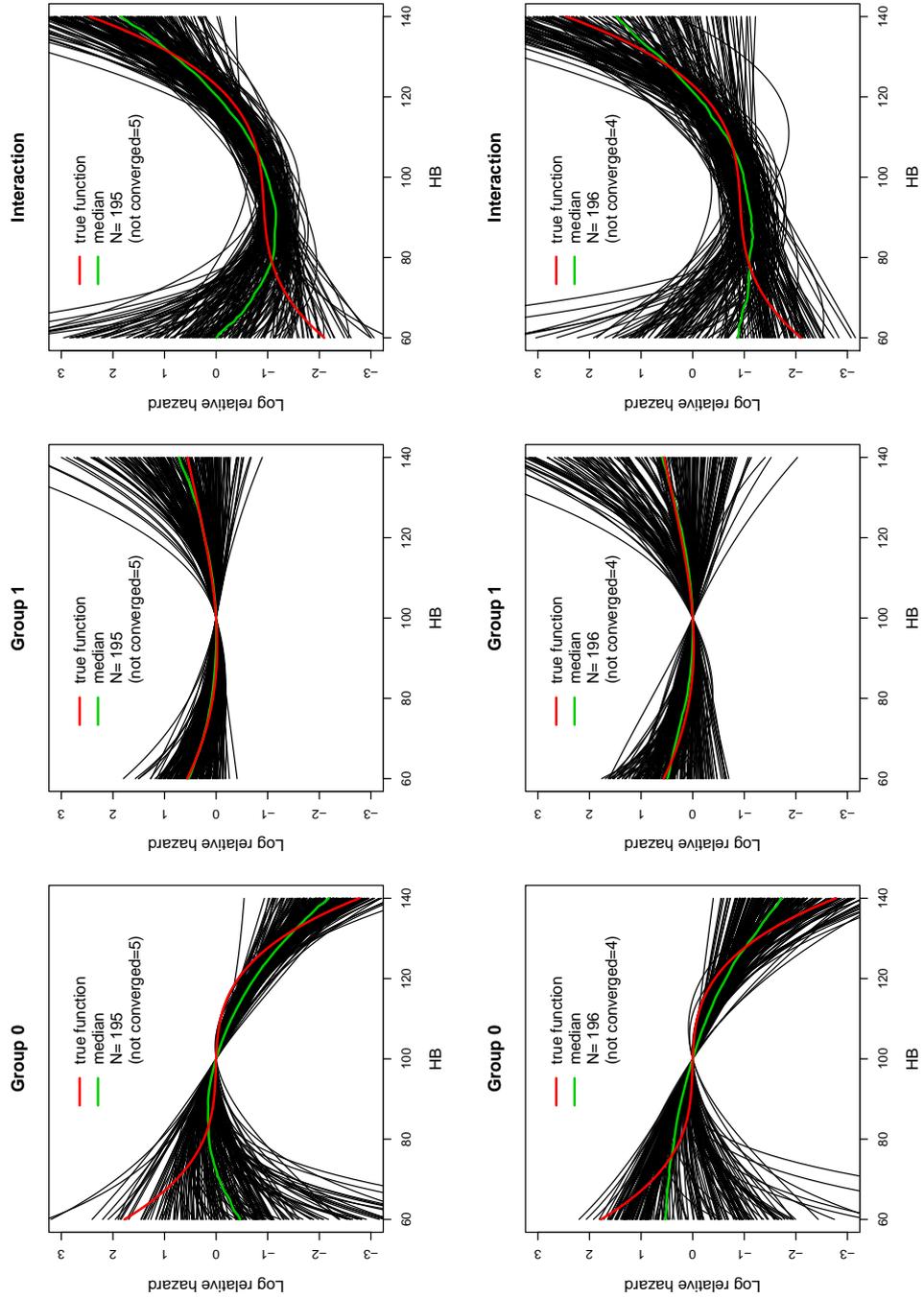


Abbildung 2.10: Punktschätzer des selektierten Modells der Datenzelle 1674

Abbildung 2.11: Punktschätzer des selektierten Modells der Zelle 201

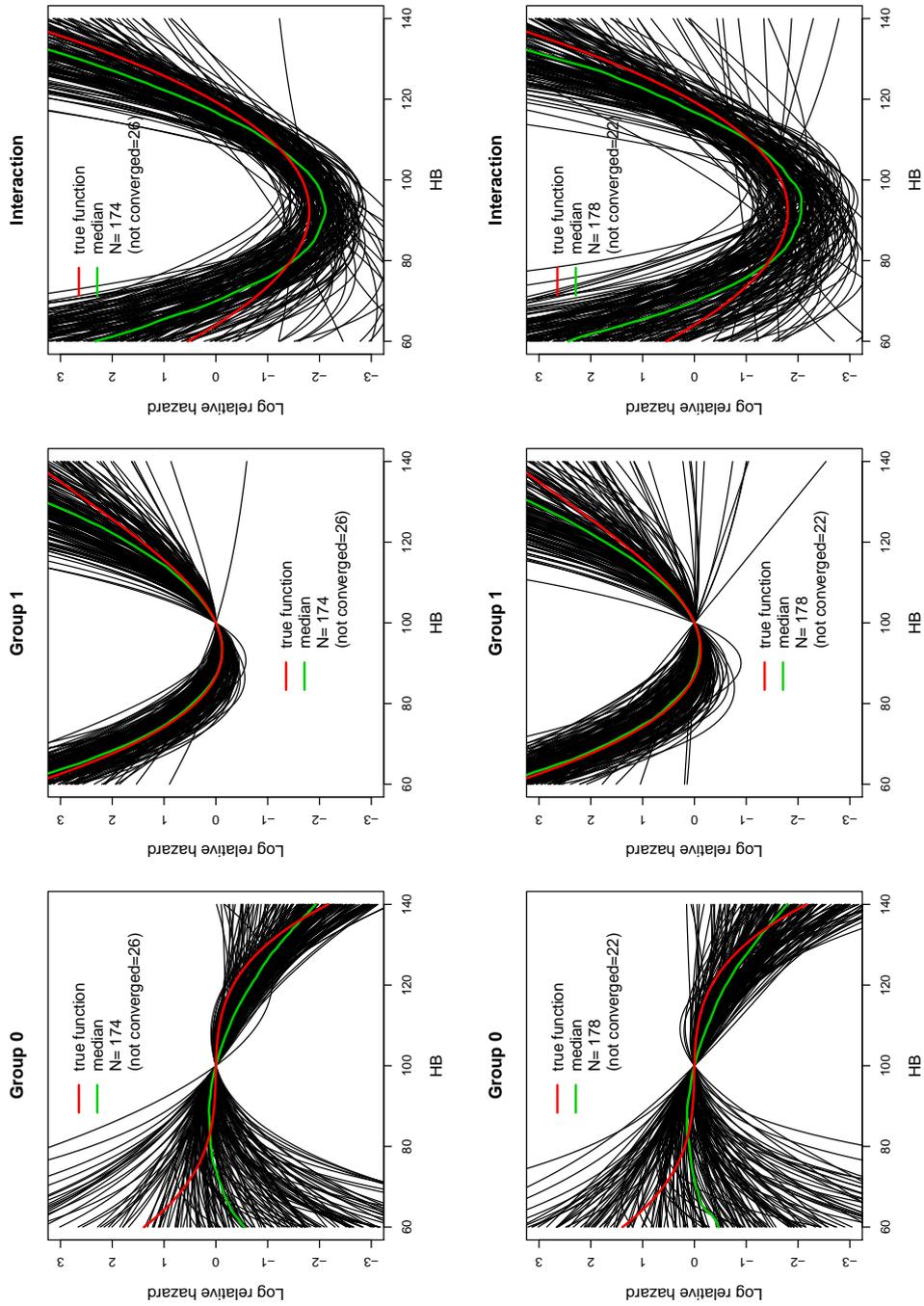


Abbildung 2.12: Punktschätzer des selektierten Modells der Datenzelle 2331

Abbildung 2.13: Punktschätzer des selektierten Modells der Zelle 2671

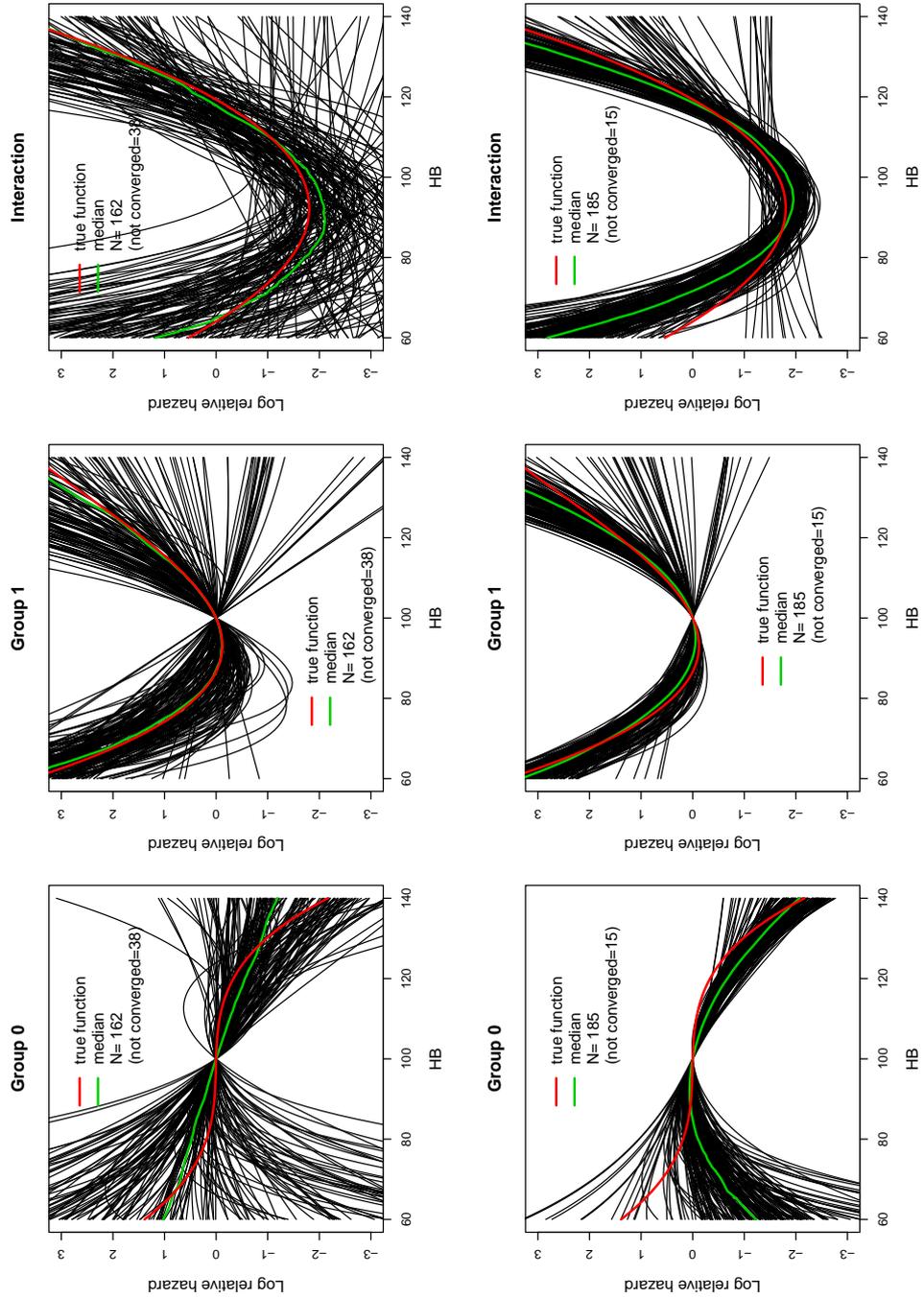


Abbildung 2.14: Punktschätzer des selektierten Modells der Datenzelle 204

Abbildung 2.15: Punktschätzer des selektierten Modells der Zelle 2334

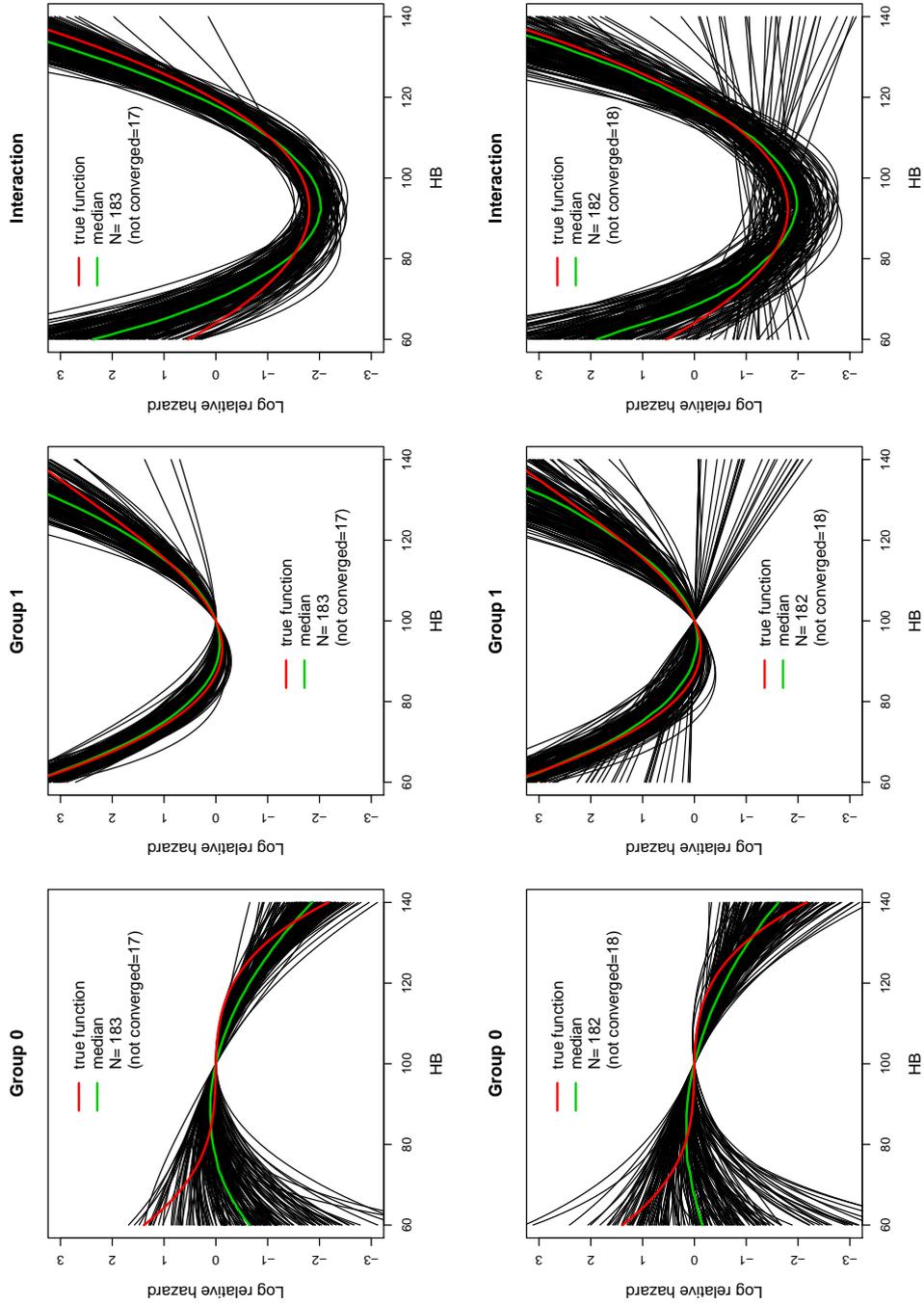


Abbildung 2.16: Punktschätzer des selektierten Modells der Datenzelle 2674

Abbildung 2.17: Punktschätzer des selektierten Modells der Zelle 301

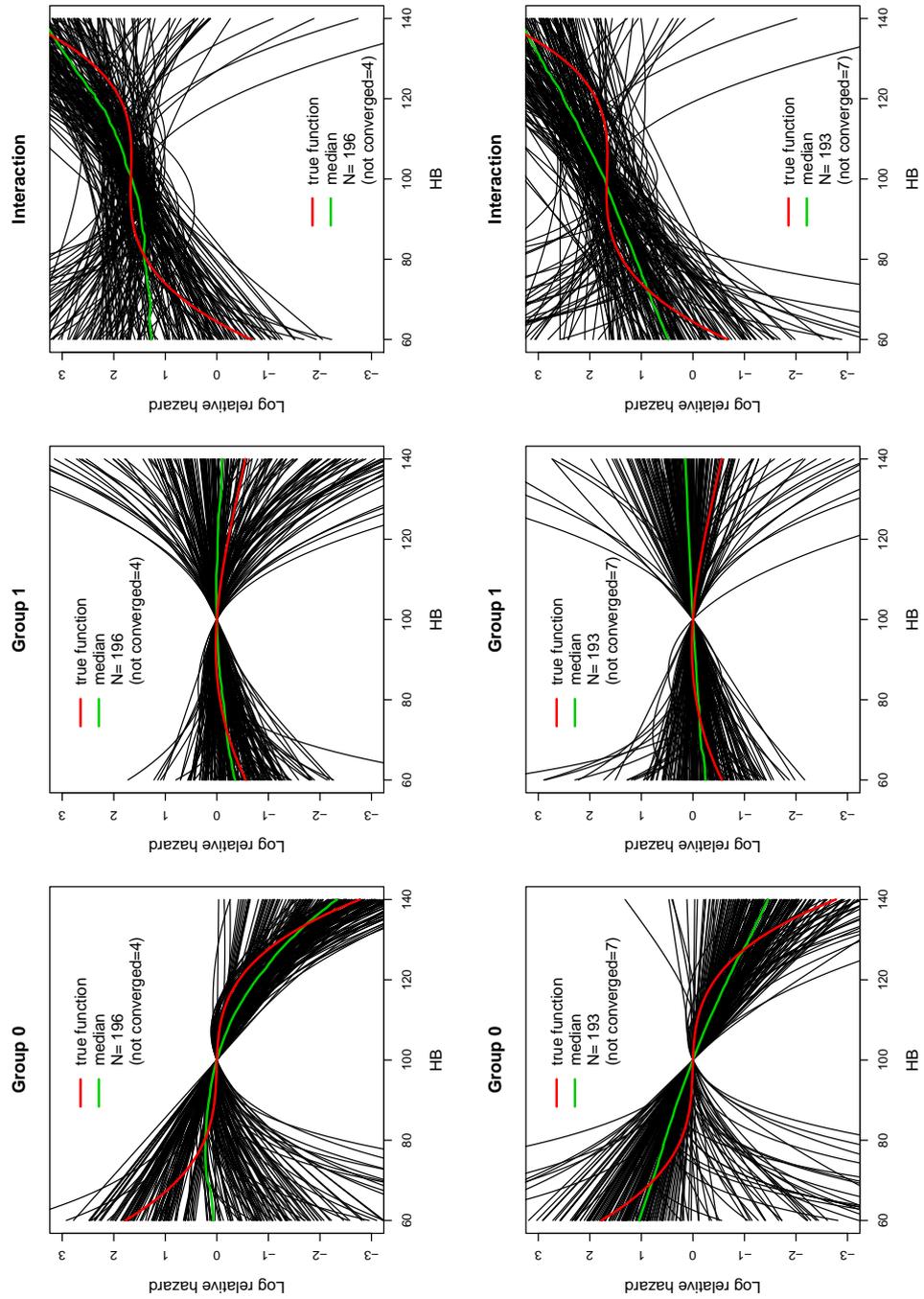


Abbildung 2.18: Punktschätzer des selektierten Modells der Datenzelle 3331

Abbildung 2.19: Punktschätzer des selektierten Modells der Zelle 3671

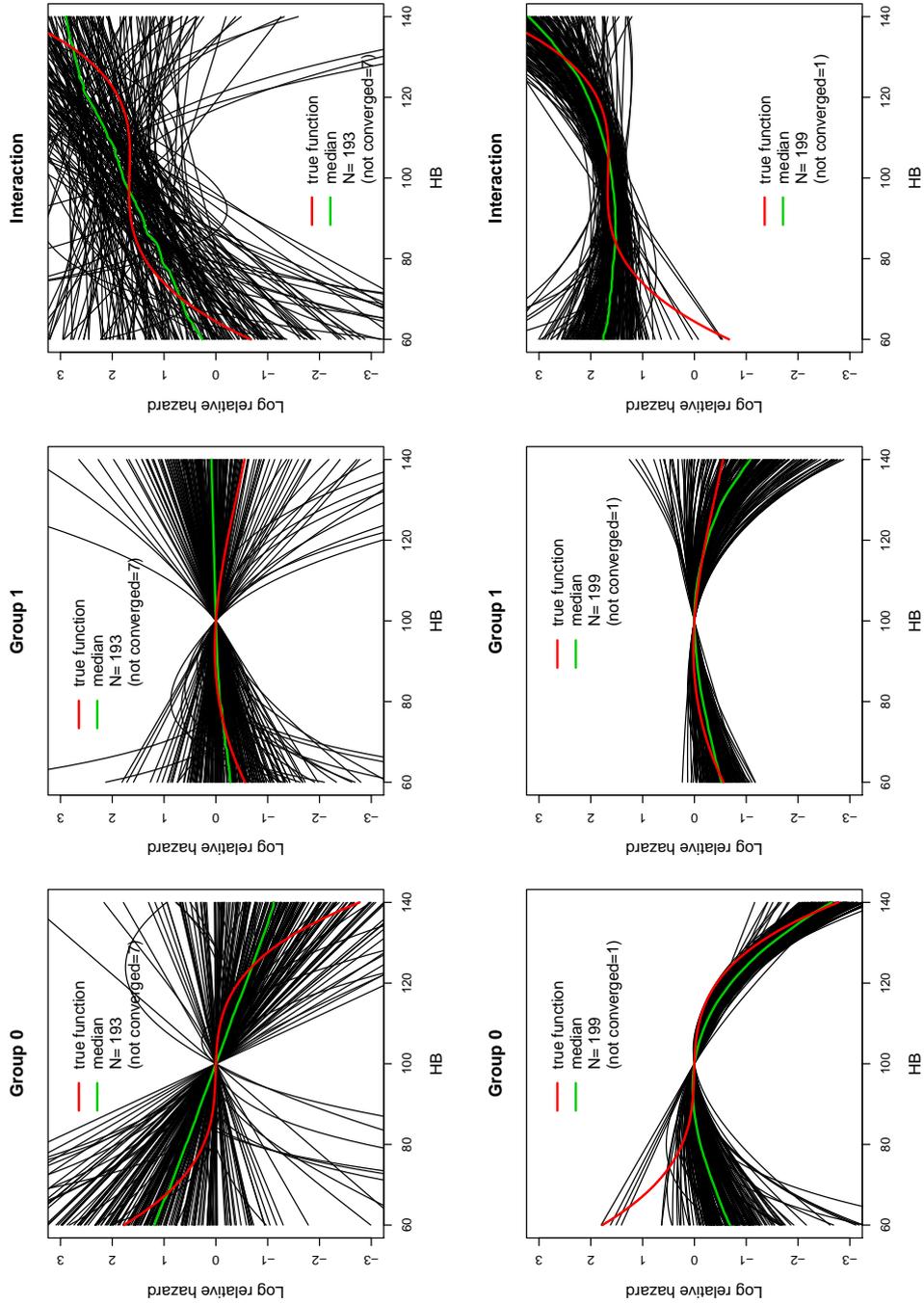


Abbildung 2.20: Punktschätzer des selektierten Modells der Datenzelle 304

Abbildung 2.21: Punktschätzer des selektierten Modells der Zelle 3334

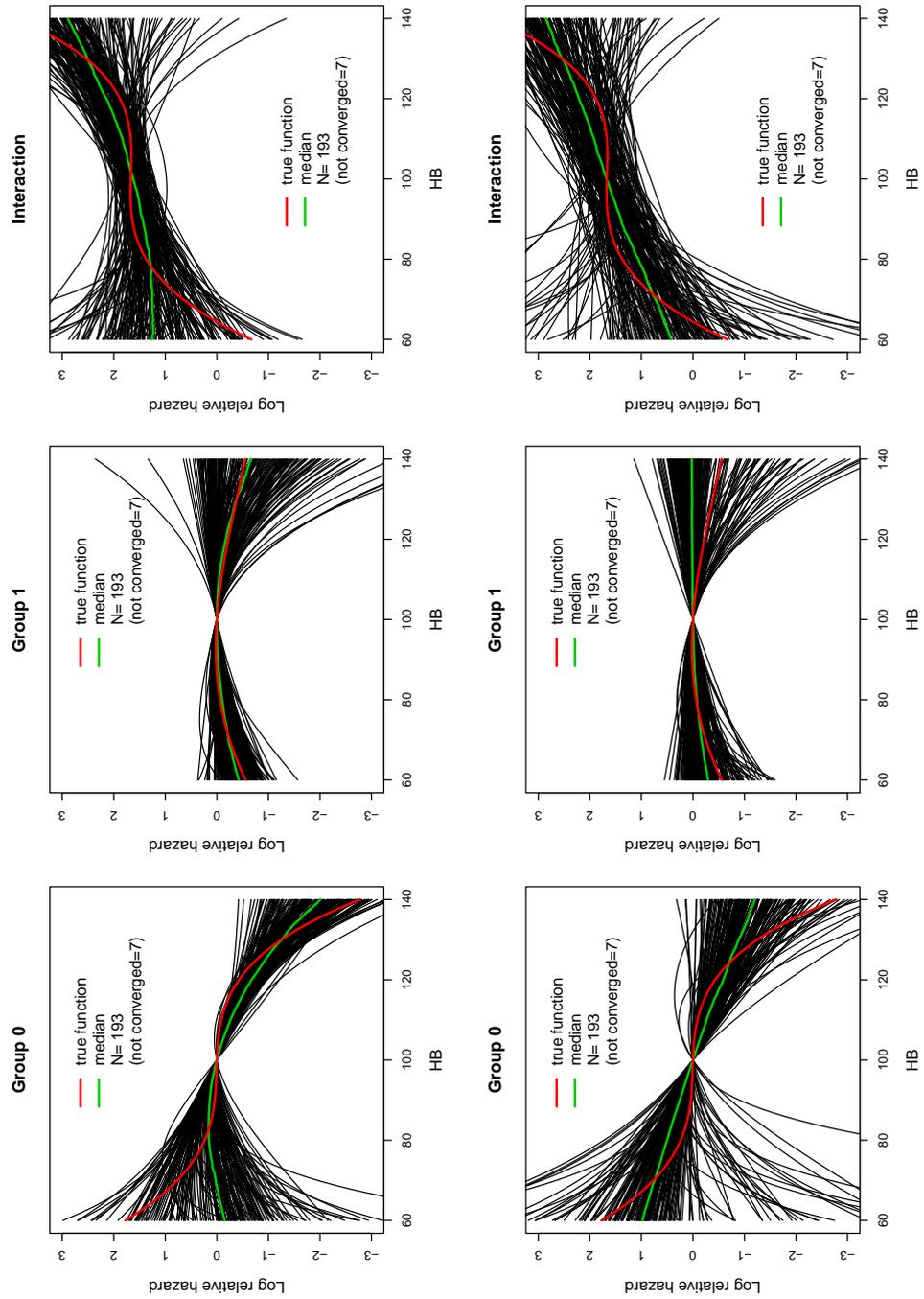


Abbildung 2.22: Punktschätzer des selektierten Modells der Datenzelle 3674

Abbildung 2.23: Punktschätzer des selektierten Modells der Zelle 401

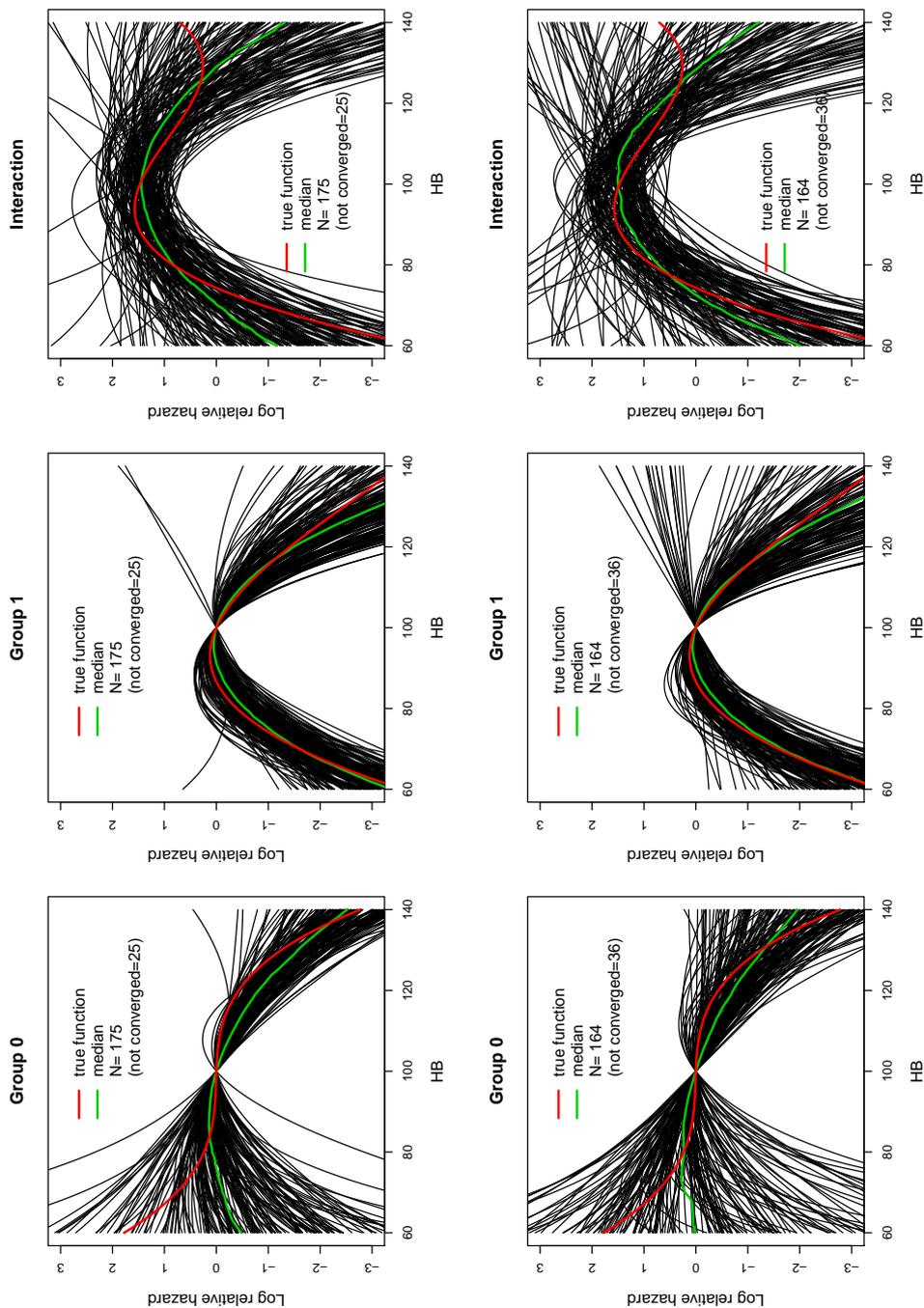


Abbildung 2.24: Punktschätzer des selektierten Modells der Datenzelle 4331

Abbildung 2.25: Punktschätzer des selektierten Modells der Zelle 4671

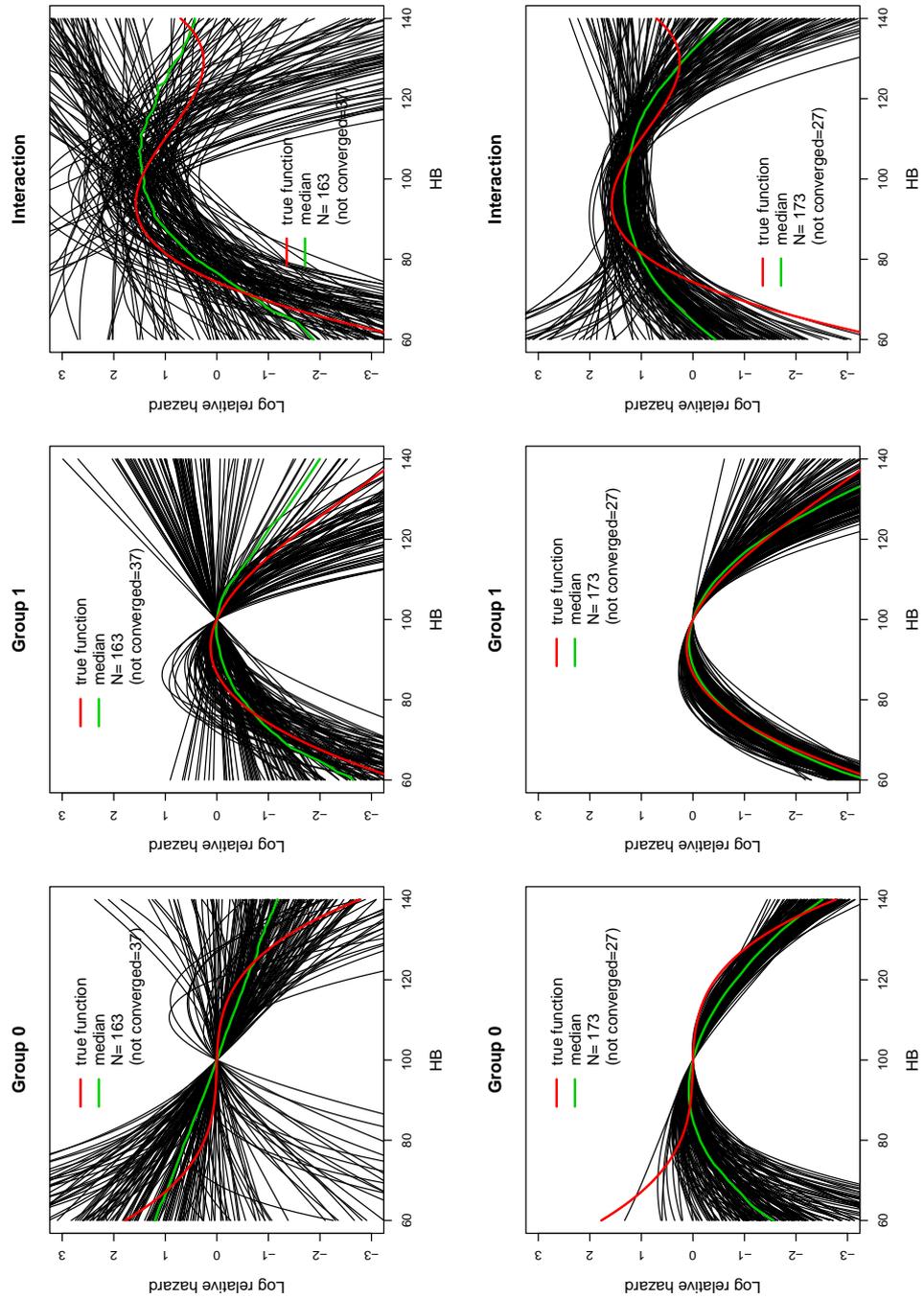


Abbildung 2.26: Punktschätzer des selektierten Modells der Datenzeile 404

Abbildung 2.27: Punktschätzer des selektierten Modells der Zelle 4334

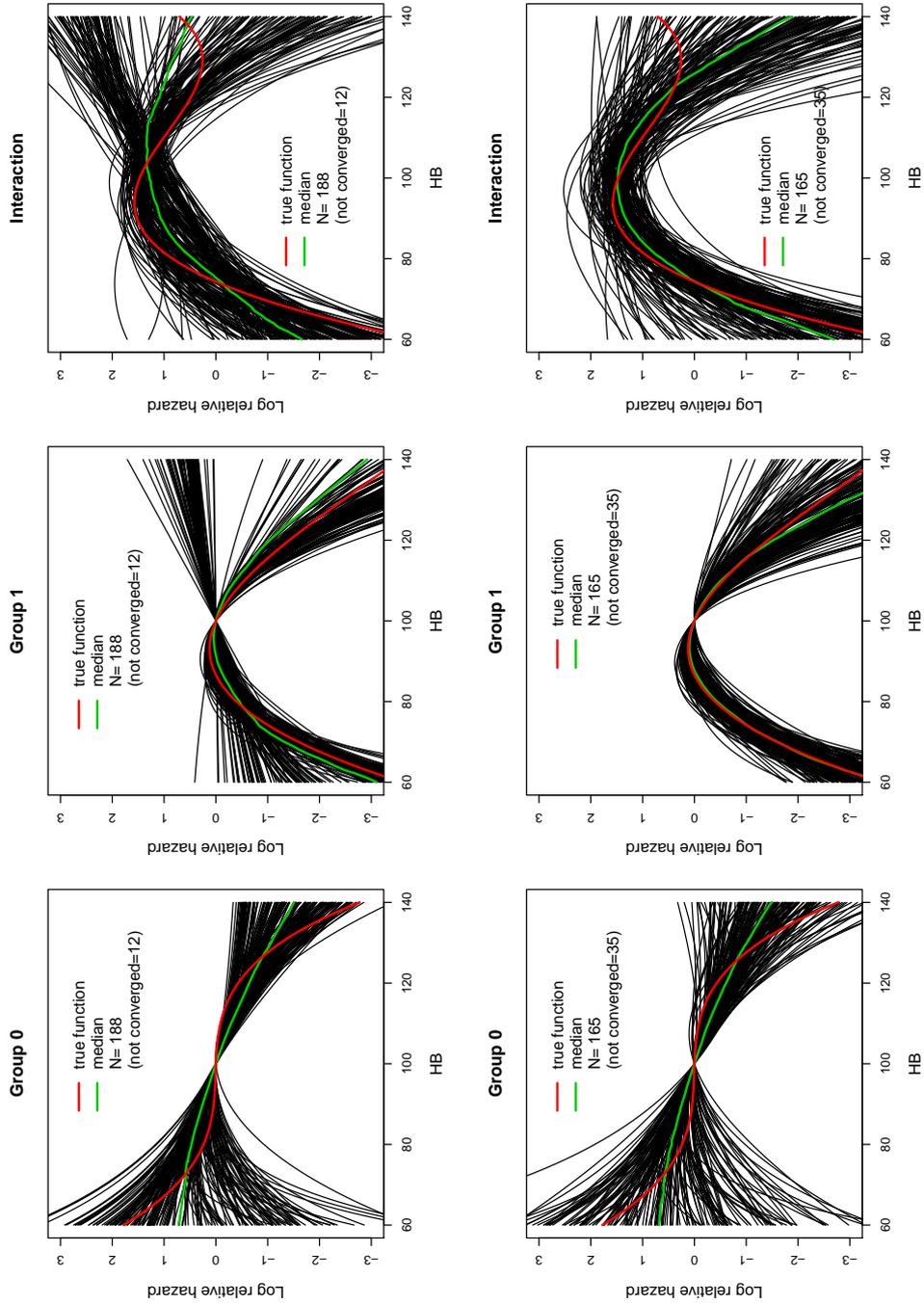


Abbildung 2.28: Punktschätzer des selektierten Modells der Datenzelle 4674

Abbildung 2.29: Punktschätzer des selektierten Modells der Zelle 501

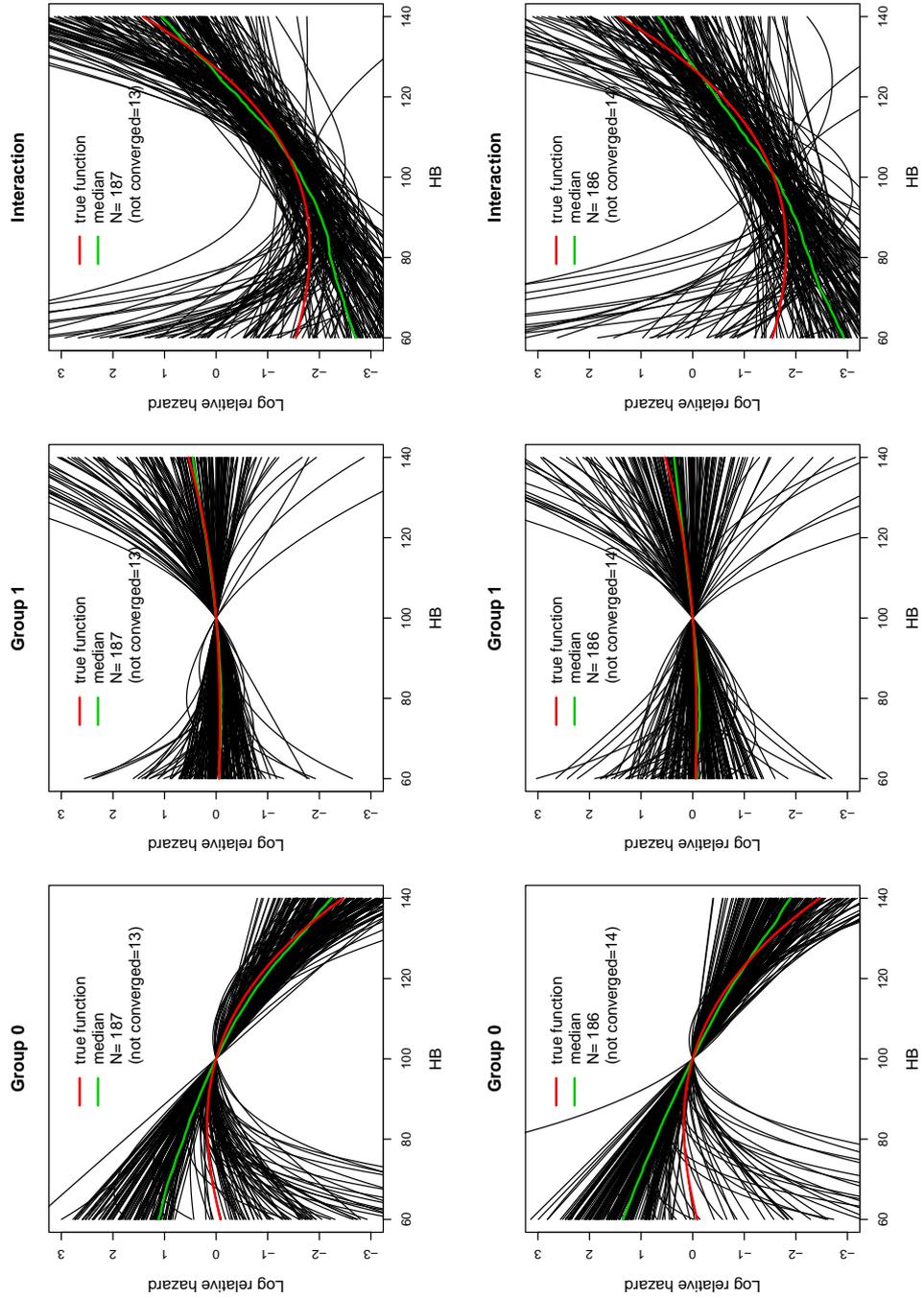


Abbildung 2.30: Punktschätzer des selektierten Modells der Datenzelle 5331

Abbildung 2.31: Punktschätzer des selektierten Modells der Zelle 5671

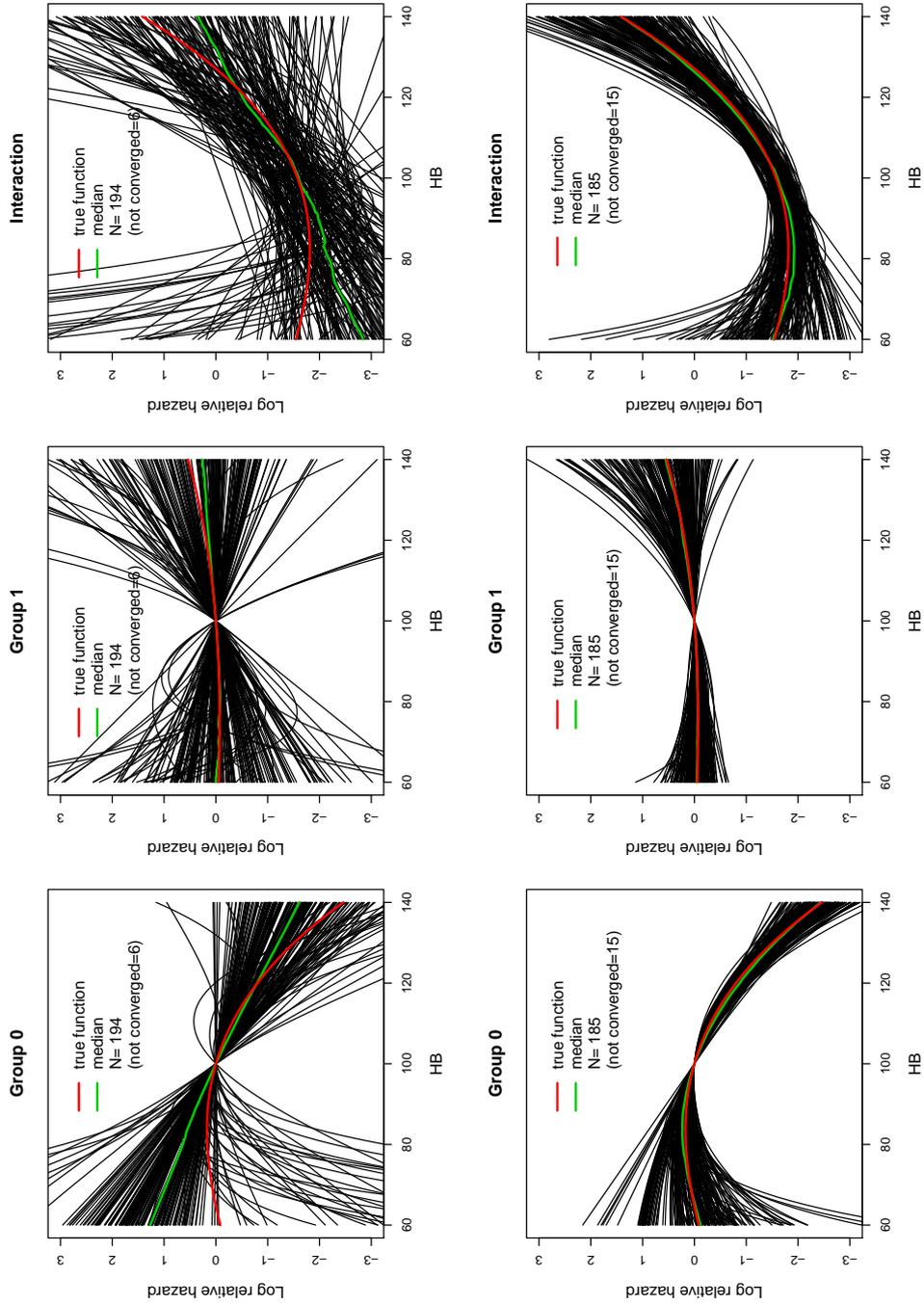


Abbildung 2.32: Punktschätzer des selektierten Modells der Datenzelle 504

Abbildung 2.33: Punktschätzer des selektierten Modells der Zelle 5334

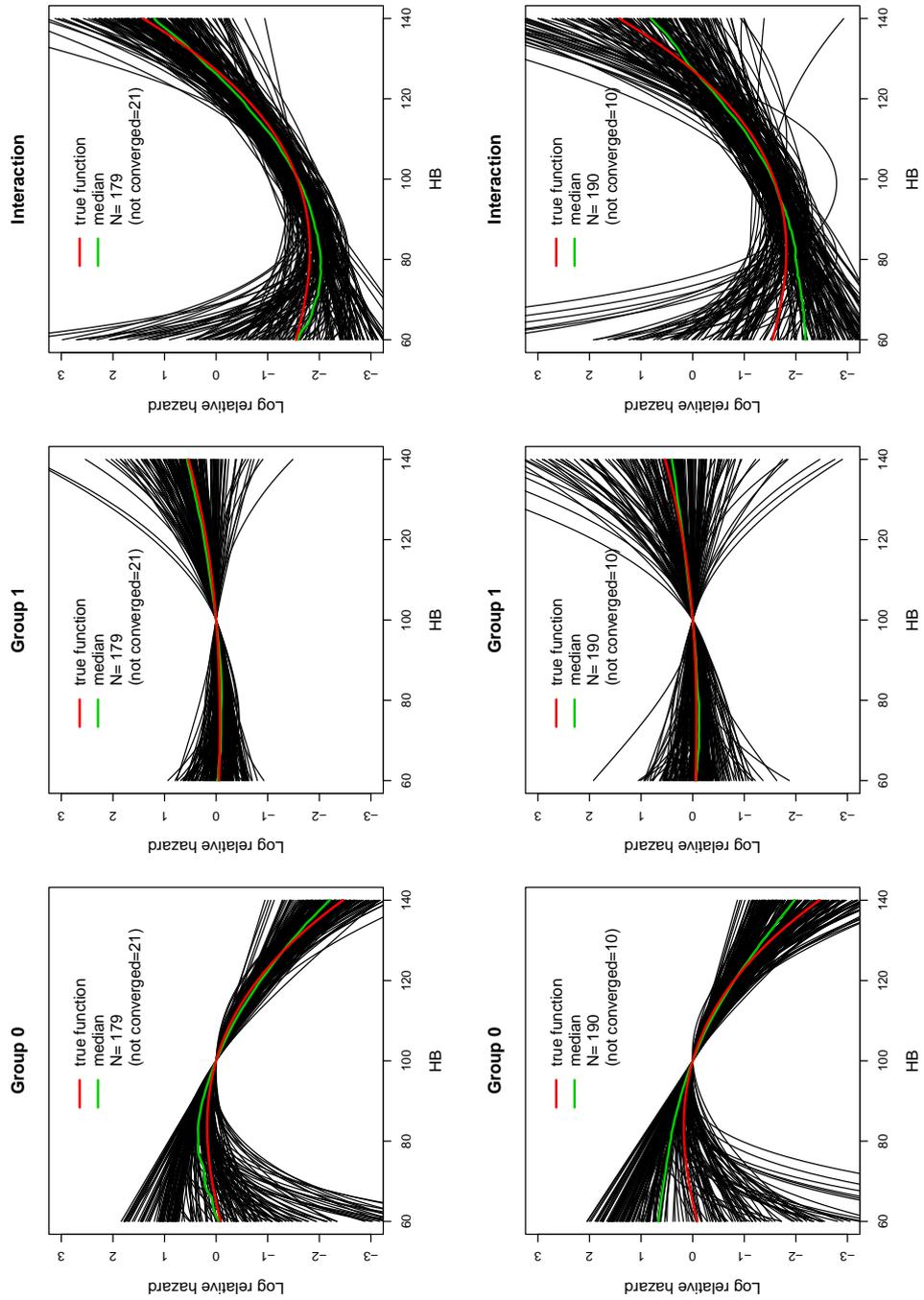


Abbildung 2.34: Punktschätzer des selektierten Modells der Datenzelle 5674

Abbildung 2.35: Punktschätzer des selektierten Modells der Zelle 601

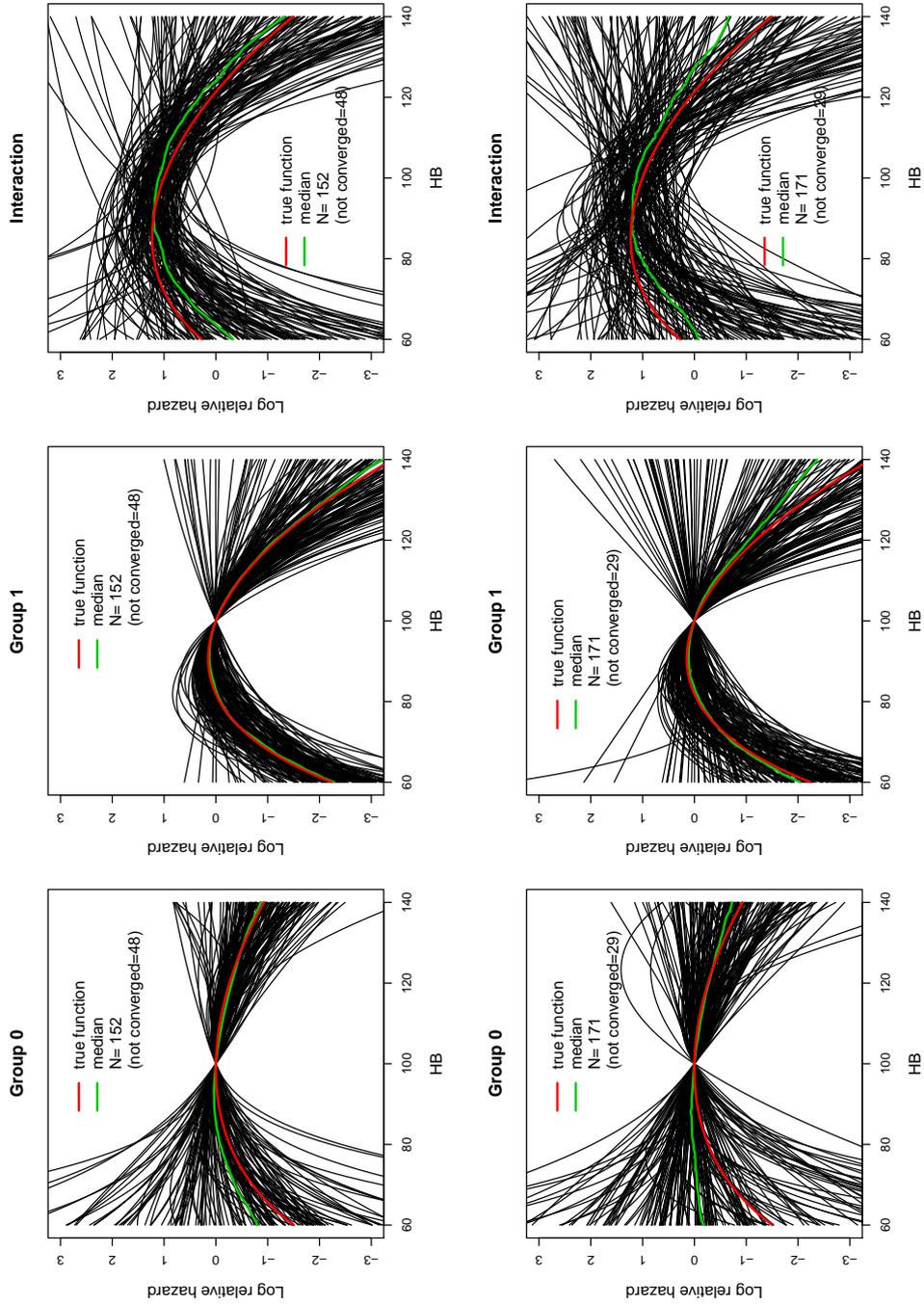


Abbildung 2.36: Punktschätzer des selektierten Modells der Datenzelle 6331

Abbildung 2.37: Punktschätzer des selektierten Modells der Zelle 6671

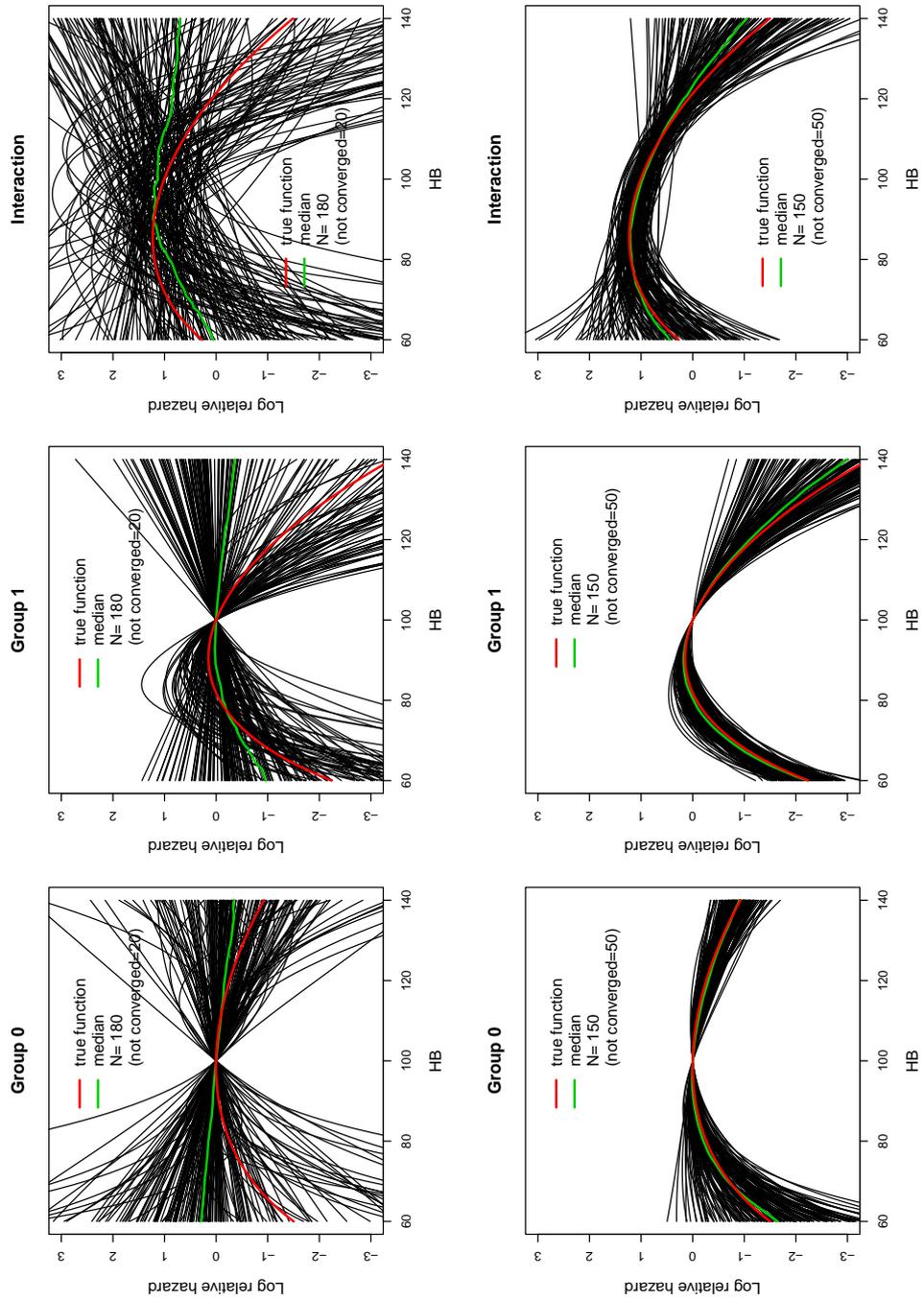


Abbildung 2.38: Punktschätzer des selektierten Modells der Datenzelle 604

Abbildung 2.39: Punktschätzer des selektierten Modells der Zelle 6334

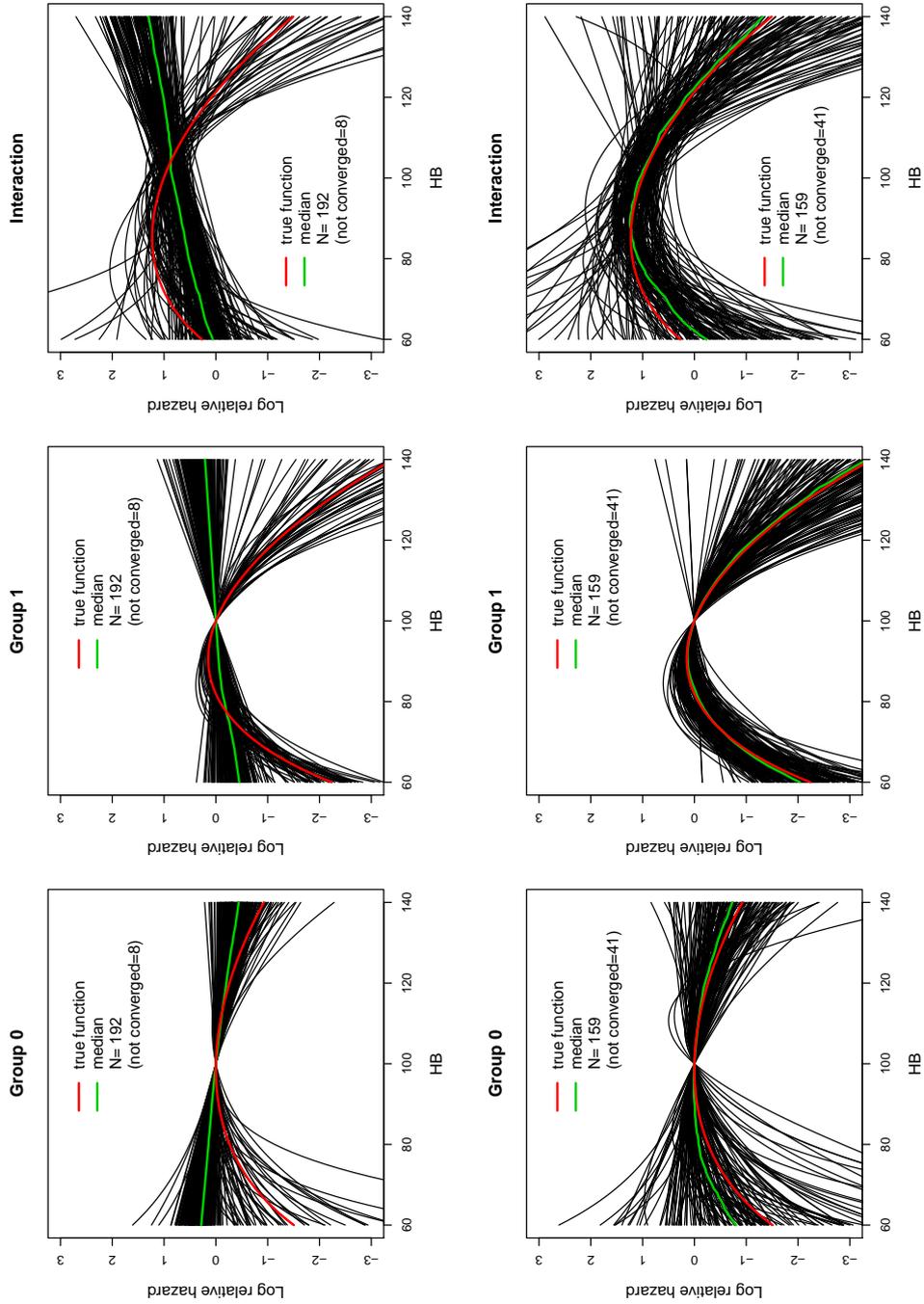


Abbildung 2.40: Punktschätzer des selektierten Modells der Datenzelle 6674

Betrachtet man nun die geschätzten Log Relative Hazards der fixierten Modelle, so kann man in Szenario 5 sehen, dass in den Datenzellen 501, 5331 und 5671 (Abbildungen 2.41 bis 2.43) der Median und die Originalfunktion der Log Relative Hazards die gleiche Form wiedergeben, d.h. die Schätzer liefern eine gute Anpassung. Etwas genauer ausgedrückt decken sich die Linien des Medians und der Originalfunktion in Gruppe 1 im Bereich von 60 bis 110 nahezu perfekt ab, während sie im Bereich von 110 bis 140 auseinanderdriften. Im Falle der Gruppe 0 erhalten wir ein ähnliches Ergebnis, nur in umgekehrter Richtung, d.h. im Bereich von 90 bis 140 liegt eine ausgezeichnete Übereinstimmung von Median und Originalfunktion vor, während im Bereich von 60 bis 90 die Linien geringfügig voneinander abweichen. Die Interaktion liefert eine gute, aber nicht perfekte Anpassung.

Die Anpassung der Datenzellen 504, 5334 und 5674 (Abbildungen 2.44 bis 2.46) ist ausgezeichnet. Außerdem wird in diesen drei Abbildungen deutlich, dass der Zensierungsanteil einen Einfluß auf die Varianz der Schätzer der Log Relative Hazards hat. Je größer der Anteil an zensierten Lebensdauern ist, umso mehr nimmt die Streuung der Schätzer zu.

Die Abbildungen 2.47 bis 2.49 zeigen uns die Schätzerplots der Datenzellen 601, 6331 und 6671. Dieses Mal ist die Anpassung in der Gruppe 1 bei allen drei Datenzellen hervorragend, während sie in der Gruppe 0 im Bereich von 60 bis 90 nicht besonders zufriedenstellend ist, obwohl Richtung und Form der Linien ident sind. In Folge ist die Anpassung in der Interaktion im selbigen Bereich schwächer als bei den anderen Hämoglobinwerten.

Wie schon in Szenario 5 liefern auch in Szenario 6 jene Datenzellen, deren Datensätze eine Stichprobengröße von 400 aufweisen, eine perfekte Übereinstimmung des Medians und der Originalfunktion der Log Relative Hazards. Zusätzlich ist wieder erkennbar, dass mit zunehmenden Zensierungsanteil die Varianz größer wird.

Vergleicht man die Selektionsmodelle und die fixierten Modelle von Szenario 5 und 6, so fällt gleich zu Beginn auf, dass bei den fixierten Modellen die Punktschätzer der Log Relative Hazards öfter zu einem Ergebnis kommen als bei den Selektionsmodellen. Die Schätzer der fixierten Modelle von Szenario 5 kommen sogar immer zu einem Ergebnis, während die Schätzung der Selektionsmodelle pro Datenzelle zwischen 6 und 21 nicht konvergierte Punktschätzer liefert. In Szenario 6 konvergierte im fixierten Modell in jeder Datenzelle maximal ein Punktschätzer nicht, im Gegensatz zu den Schätzern der Selektionsmodelle, wo in jeder Datenzelle zwischen 8 und 50 nicht konvergierten.

Aber nicht nur das Konvergenzverhalten war in den fixierten Modellen wesentlich besser, sondern auch die Anpassung an die wahren Effekte verlief großteils

deutlich besser. Betrachtet man Szenario 5, so gibt es bei den Datenzellen mit einer Stichprobengröße von 400 geringere Unterschiede zwischen den Modellen als bei den Datenzellen 501, 5331 und 5671. Im Speziellen weisen die fixierten Modelle in diesen Datenzellen bei der Gruppe 0 und der Interaktion eine deutlich bessere Anpassung auf. Erwähnt werden sollte aber auch, dass die Selektionsmodelle in der Gruppe 1 im Hämoglobinbereich von 115 bis 140 eine geringere Differenz des Medians und des wahren Effekts zeigen. Die fixierten Modelle sind in Szenario 6 bezüglich der Anpassung besser als die Selektionsmodelle. Besonders in den Datenzellen 6671 und 6334 wird dies deutlich, sowie in allen anderen Datenzellen in der Gruppe 1.

Abbildung 2.41: Punktschätzer der fixierten Modelle der Zelle 501

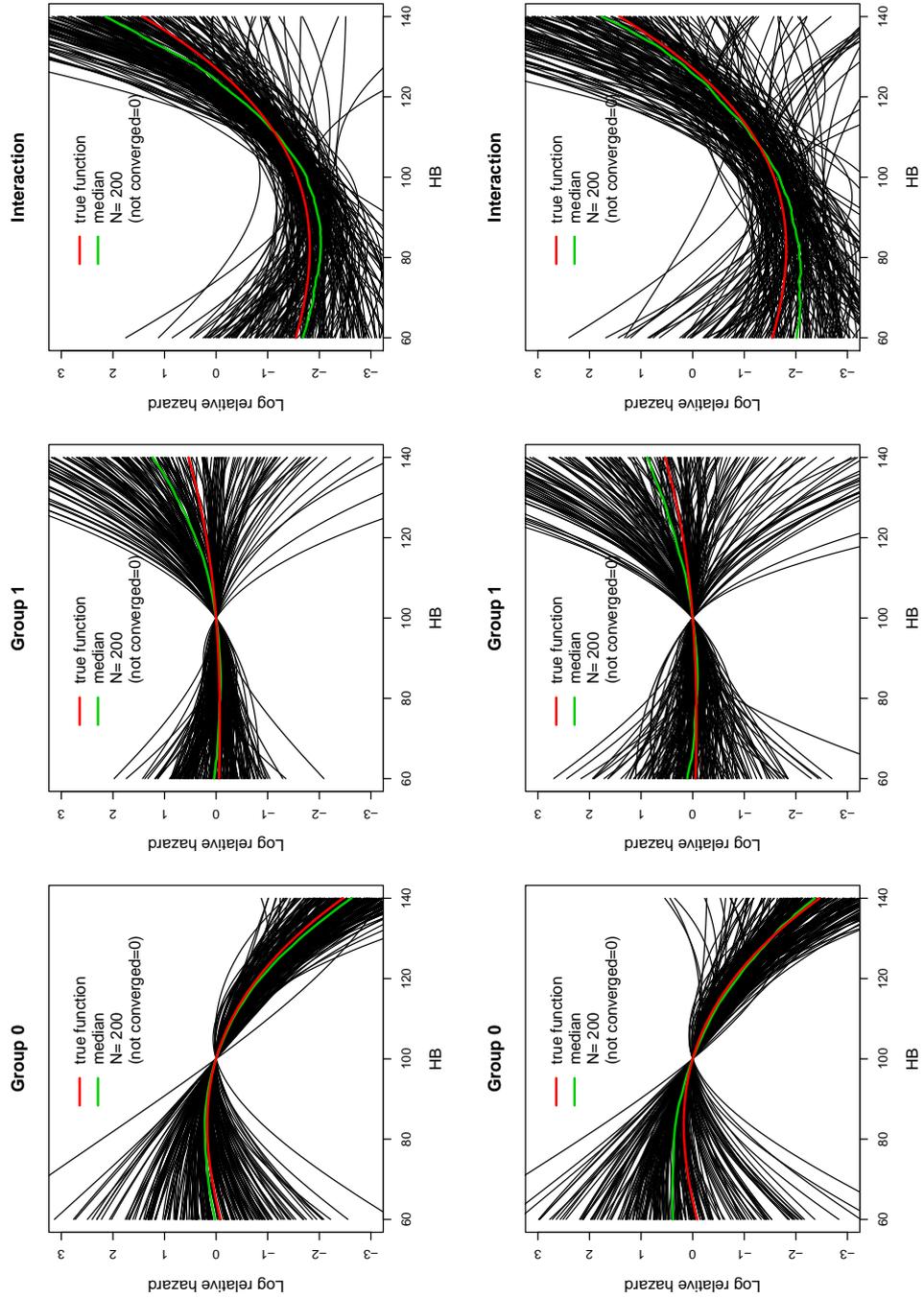


Abbildung 2.42: Punktschätzer der fixierten Modelle der Datenzelle 5331

Abbildung 2.43: Punktschätzer der fixierten Modelle der Zelle 5671

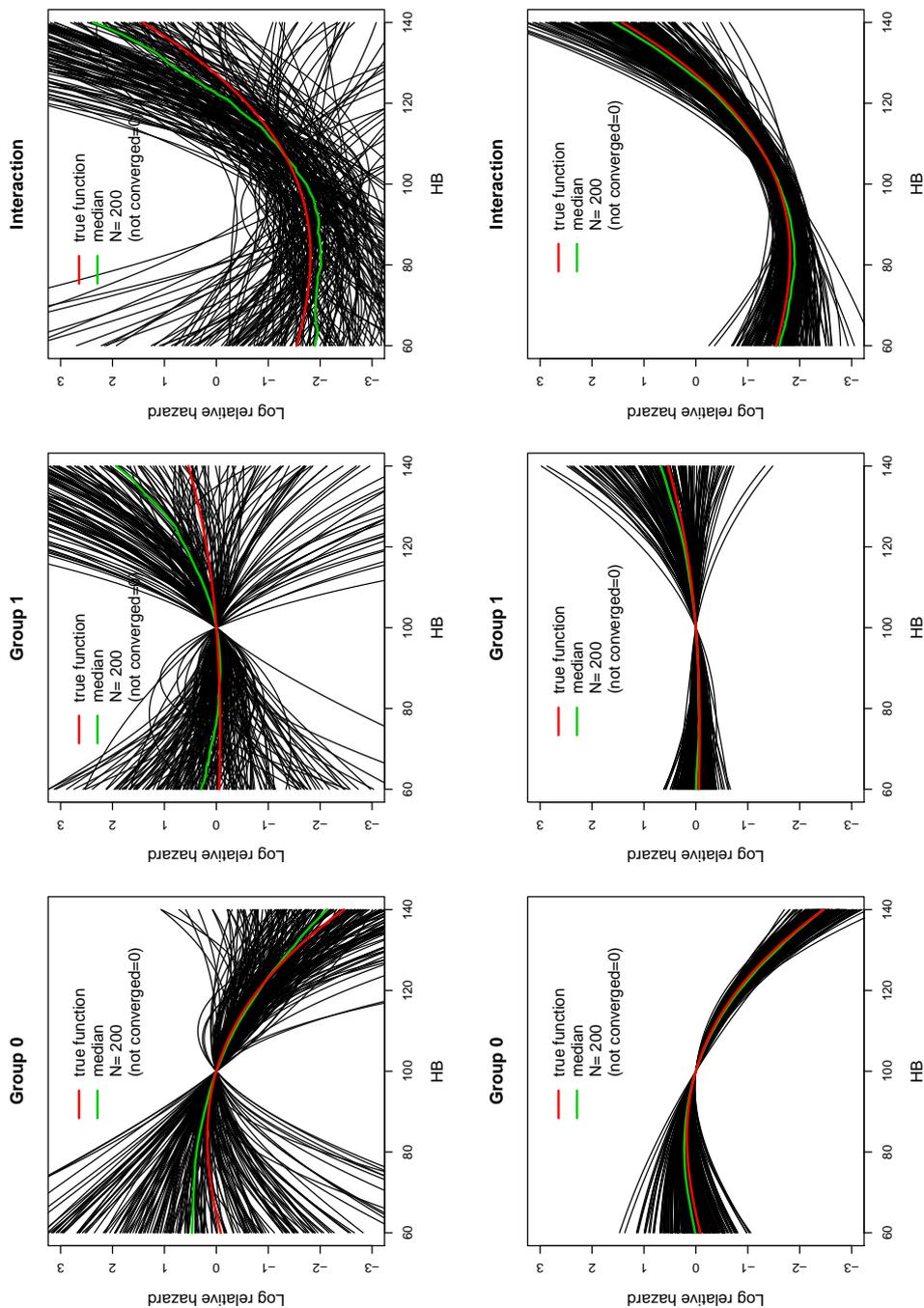


Abbildung 2.44: Punktschätzer der fixierten Modelle der Datenzelle 504

Abbildung 2.45: Punktschätzer der fixierten Modelle der Zelle 5334

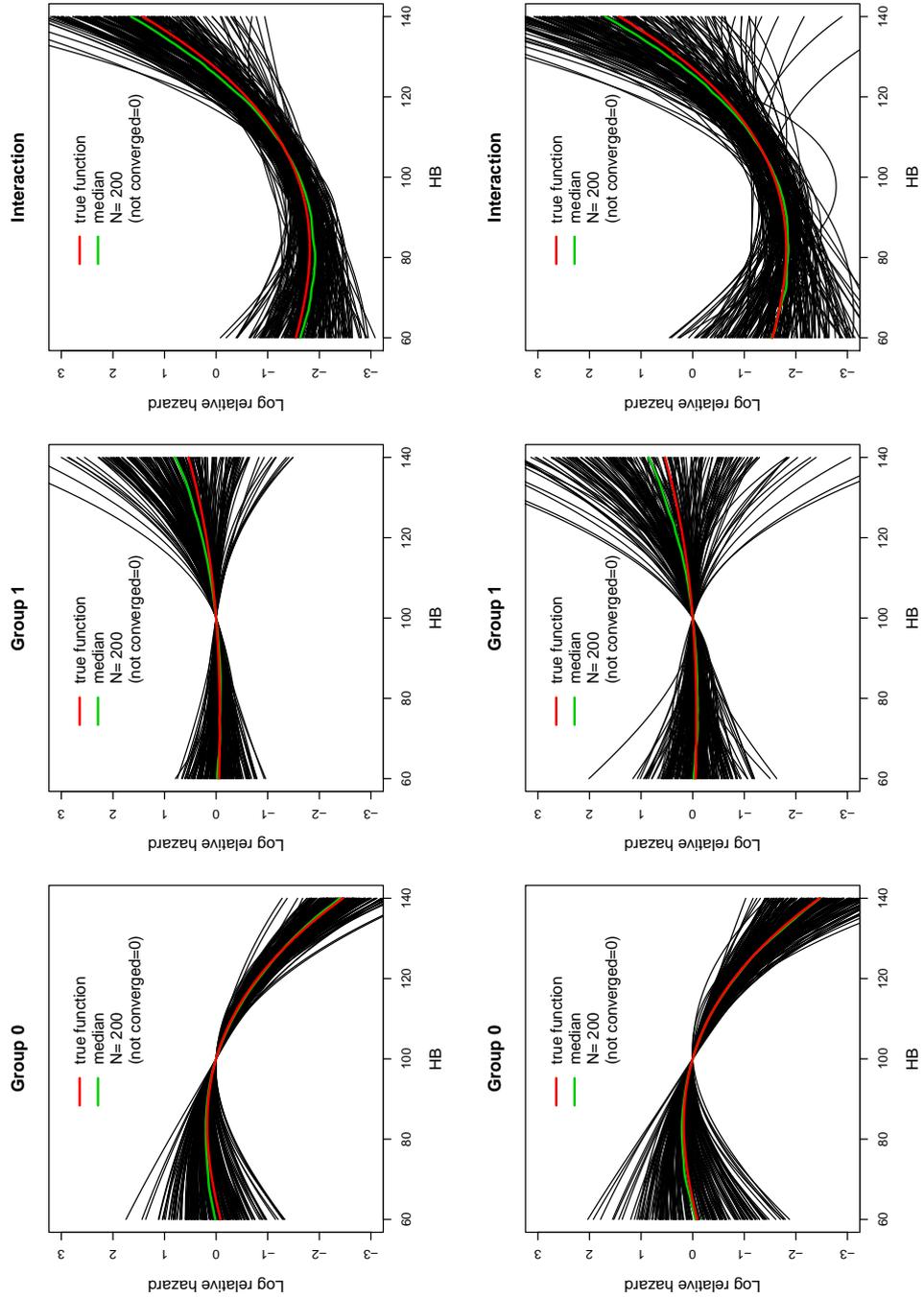


Abbildung 2.46: Punktschätzer der fixierten Modelle der Datenzelle 5674

Abbildung 2.47: Punktschätzer der fixierten Modelle der Zelle 601

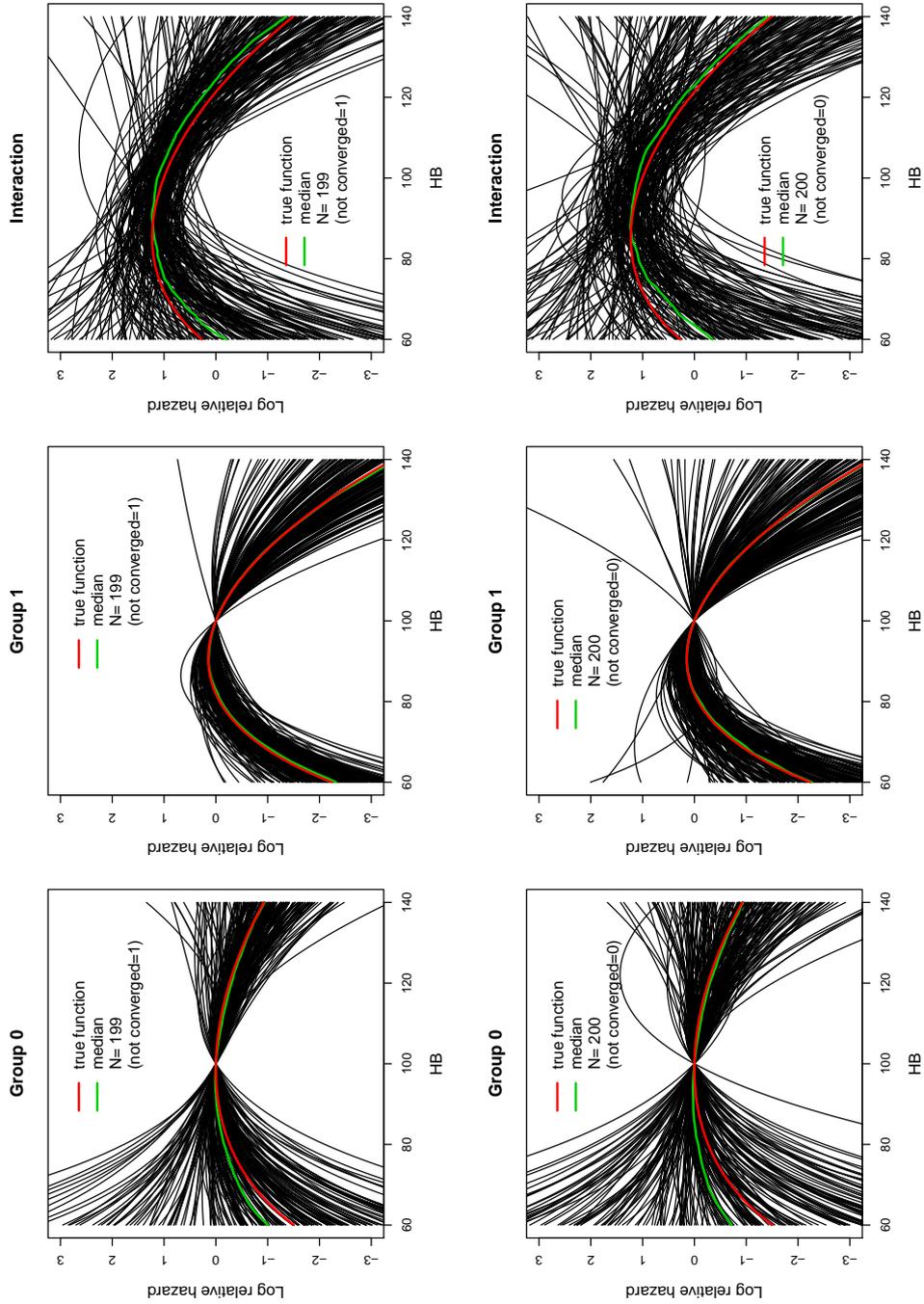


Abbildung 2.48: Punktschätzer der fixierten Modelle der Datenzelle 6331

Abbildung 2.49: Punktschätzer der fixierten Modelle der Zelle 6671

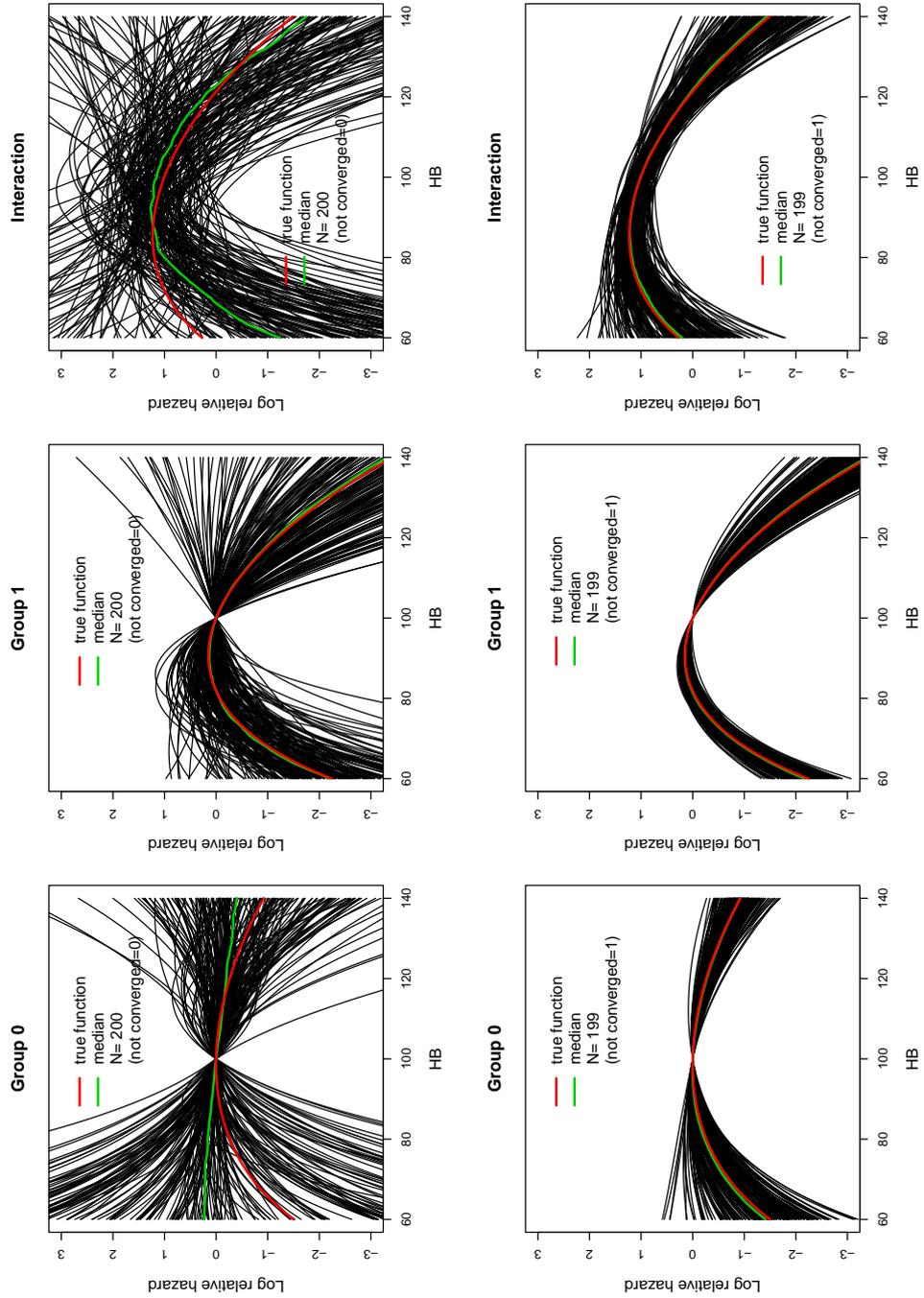


Abbildung 2.50: Punktschätzer der fixierten Modelle der Datenzelle 604

Abbildung 2.51: Punktschätzer der fixierten Modelle der Zelle 6334

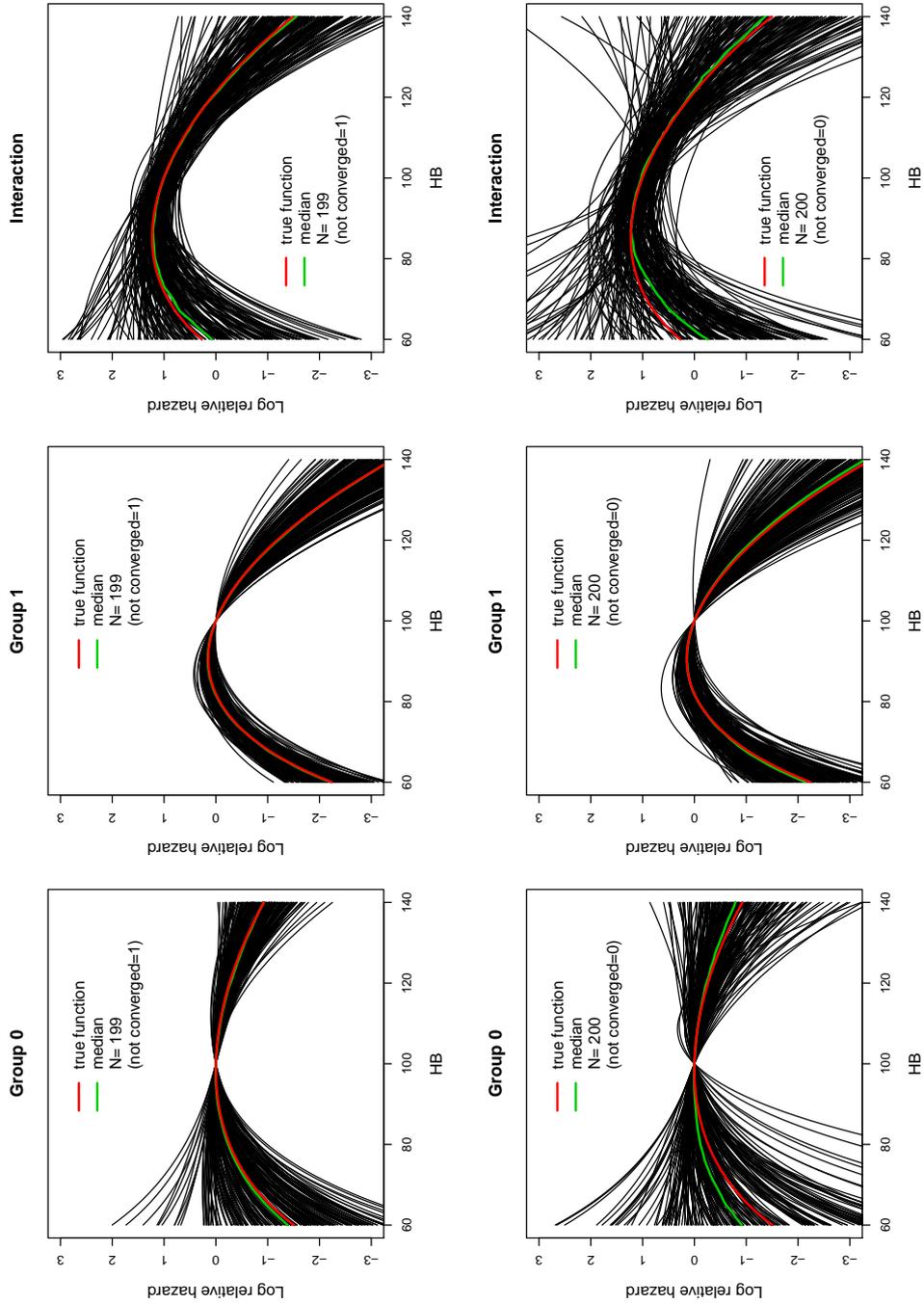


Abbildung 2.52: Punktschätzer der fixierten Modelle der Datenzelle 6674

2.7.3 Grafische Darstellung der Abweichungen der Schätzer

In den Abbildungen 2.53 bis 2.88 wurden der Bias und der MAD für die ML-Schätzer (graue Linie) und die Schätzer der Firth-Methode (schwarze Linie) grafisch festgehalten, außerdem ist in jeder Grafik der gewichtete Mittelwert der Fläche der Abweichungen der Schätzer abgedruckt. Jede Abbildung besteht aus 6 Bildern: die erste Bilderreihe zeigt den medianen Bias, die zweite Reihe den MAD. Jeweils die erste Grafik zeigt die Abweichungen der Gruppe 0, während die Abweichungen der Gruppe 1 bzw. der Interaktion durch die zweite bzw. dritte Grafik wiedergegeben werden.

Eine überblicksmäßige Betrachtung der Grafiken lässt darauf schließen, dass die Maximum-Likelihood-Methode im Schnitt einen kleineren medianen Bias liefert, im Gegensatz dazu jedoch die Firth-Korrektur einen kleineren MAD. Um einen detaillierteren Überblick zu bekommen, ist es sinnvoll, die einzelnen Szenarien zu betrachten.

Betrachtet man den Bias des Szenarios 1 (Abbildungen 2.53 bis 2.58,) so sieht man, dass es hier keine Präferenz für ein bestimmtes Schätzverfahren gibt, während der MAD großteils bei der Firth-Methode ein kleineres Ergebnis liefert. In Szenario 2 (Abbildungen 2.59 bis 2.64) kann die Firth-Methode einen geringeren Bias nachweisen, während beim MAD auffällt, dass die Datenzelle des Szenarios 2 mit einer Stichprobengröße von 100 von der Firth-Methode besser abgebildet wird. Jene Datenzellen mit einer Stichprobengröße von 400 können teilweise mit der ML-Methode besser abgebildet werden. Was den Bias des Szenarios 3 (Abbildungen 2.65 bis 2.70) betrifft, so kann dieser mit der ML-Methode besser erklärt werden. Hingegen liefert beim MAD die Firth-Methode bessere Ergebnisse. Szenario 4 (Abbildungen 2.71 bis 2.76) liefert überwiegend einen kleineren Bias und einen kleineren MAD bei der ML-Methode.

Bei den Szenarien 5 und 6 gibt es jeweils wieder Selektionsmodelle und fixierte Modelle. Die Selektionsmodelle von Szenario 5 (Abbildungen 2.77 bis 2.82) liefern keine eindeutigen Ergebnisse, man kann aber sagen, dass wiederum der Bias größtenteils bei der ML-Methode kleiner ist, während der MAD überwiegend bei der Firth-Methode kleiner ist. Ähnliche Ergebnisse liefern die Abweichungen der Schätzer des Szenarios 5 (Abbildungen 2.89 bis 2.94) für die fixierten Modelle. Die Selektionsmodelle, sowie die fixierten Modelle des Szenarios 6 (Abbildungen 2.83 bis 2.88, sowie 2.95 bis 2.100) weisen bei der ML-Methode einen kleineren Bias auf. Die ML-Methode weist einerseits für das Selektionsmodell einen geringeren Wert des MAD auf, andererseits liefert die Firth-Methode bei den fixierten Modellen bessere Ergebnisse. Eine weitere Erkenntnis ist, dass großteils jene Datenzellen, die Datensätze

mit einer Stichprobengröße von 100 aufweisen, eher kleinere Abweichungen bei der Firth-Methode haben, während jene Datenzellen, die Datensätze mit einer Stichprobengröße von 400 beinhalten, eher kleinere Abweichungen bei der Maximum-Likelihood-Methode aufweisen können.

In allen Szenarien, sowohl in den fixierten Modellen als auch in den Selektionsmodellen, ist ein eindeutiger Einfluss der Zensierungsanteile sichtbar. Nimmt der Anteil der zensierten Daten zu, so wird der gewichtete Mittelwert der MAD größer. Bezüglich der Zensierungsanteile ist für den medianen Bias keinerlei Trend sichtbar.

Erwähnenswert scheint auch die Tatsache, dass oftmals beim medianen Bias die Gruppen 1 und 0 jeweils durch unterschiedliche Schätzmethoden besser abgebildet werden können, wie es zum Beispiel in den Abbildungen 2.54, 2.57, 2.58, und 2.64 der Fall ist. Weitere interessante Ergebnisse liefern uns die Abbildungen 2.77, 2.80, 2.95 und 2.98, in denen jeweils der gewichtete Mittelwert des MAD für beide Gruppen und die Interaktion ident sind, hingegen der gewichtete Mittelwert des Bias unterschiedlich ist. Es ist auch oftmals der Trend ersichtlich, dass der MAD von einer Schätzmethode besser abgebildet wird, während in derselben Abbildung der Bias besser von der jeweiligen konträren Schätzmethode abgebildet wird, wie in den Abbildungen 2.66, 2.70, 2.71 und 2.85 ersichtlich.

Schlussendlich sollte auch darauf hingewiesen werden, dass sich die Abweichungen der Schätzer und die gewichteten Mittelwerte der Abweichungen oftmals nur minimal unterscheiden und das obige Aussagen auf diesen minimalen Unterschieden basieren.

Abbildung 2.53: Abweichungen der Schätzer der Datenzelle 101

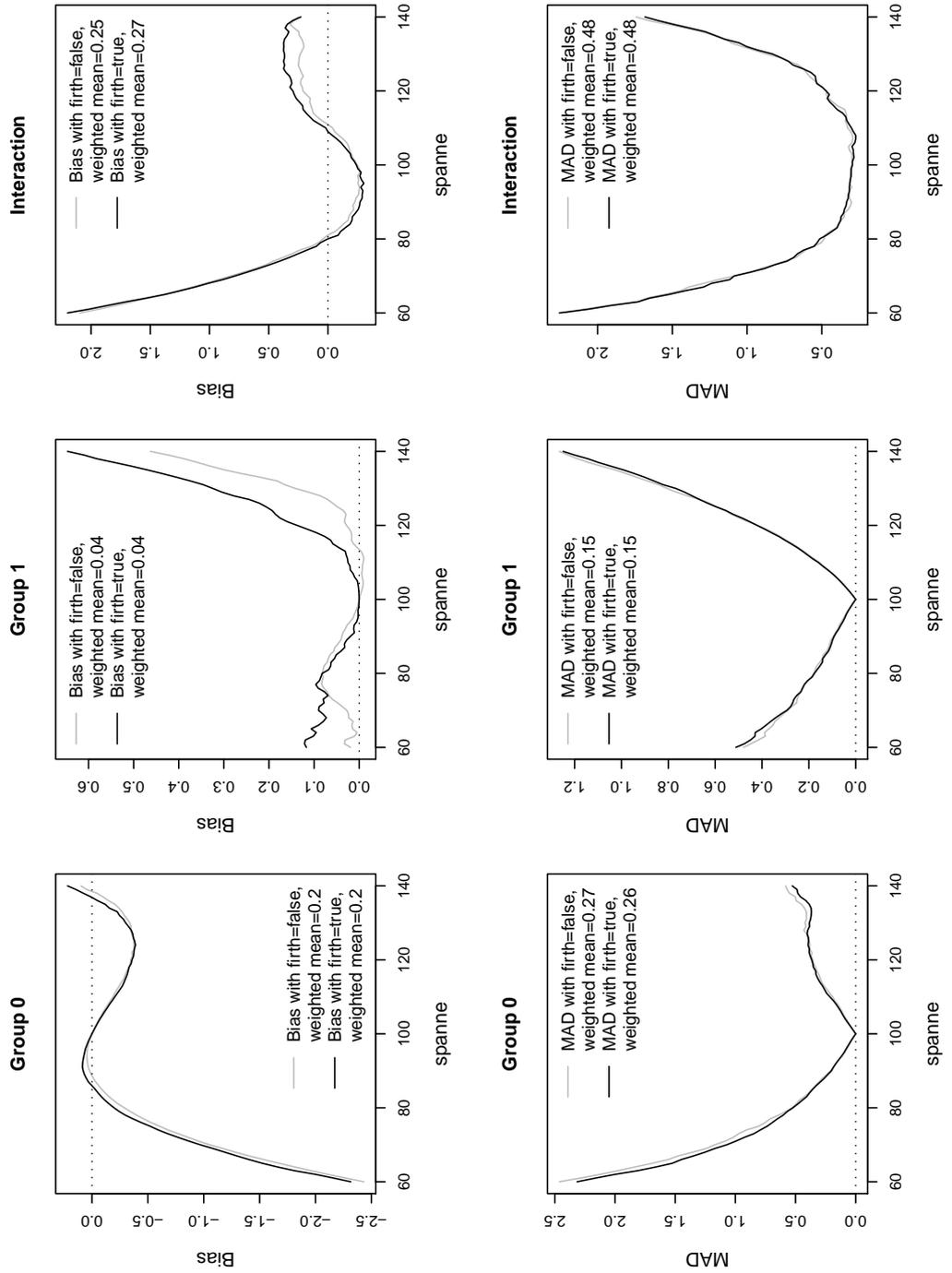


Abbildung 2.54: Abweichungen der Schätzer der Datenzelle 1331

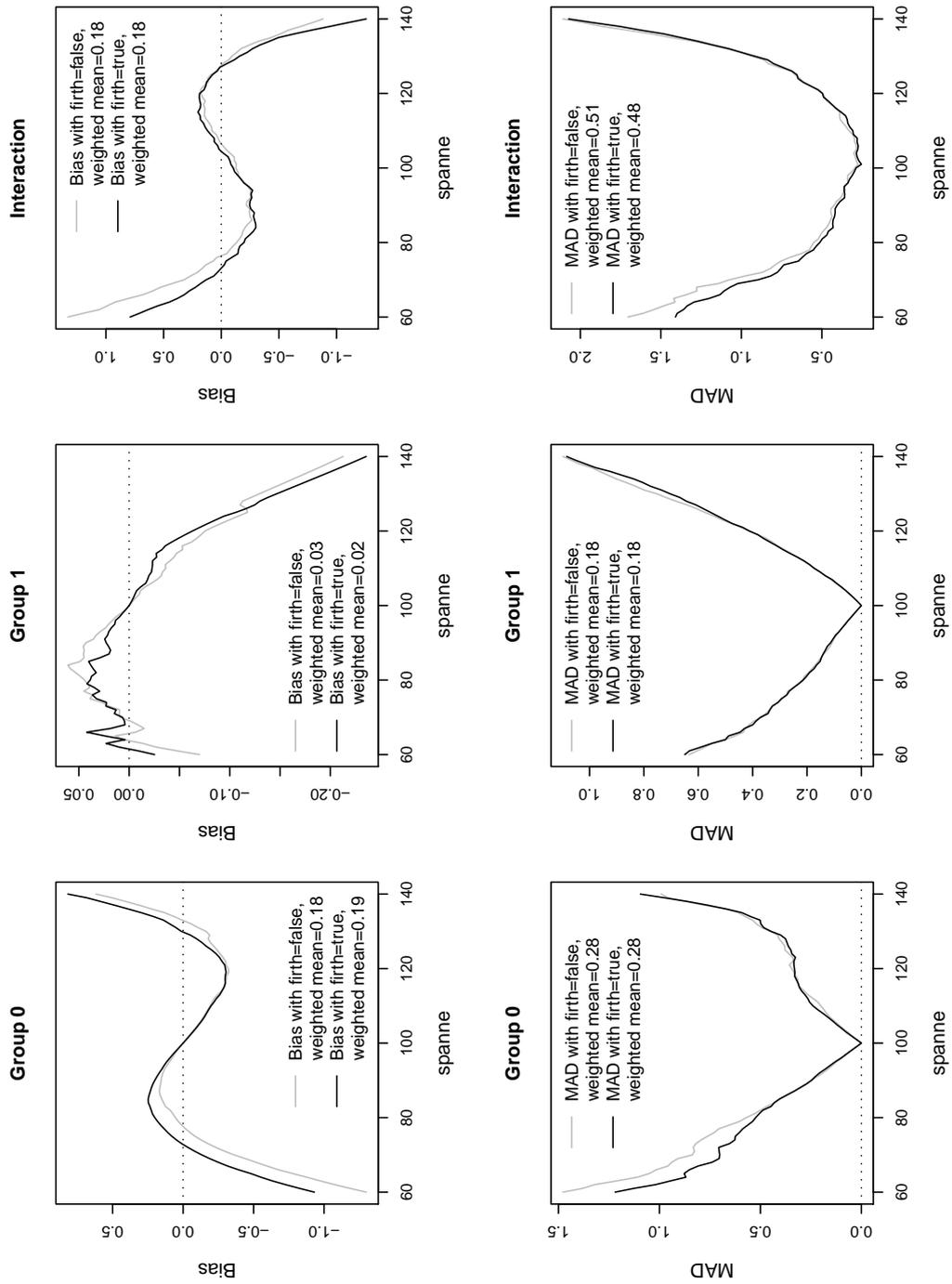


Abbildung 2.55: Abweichungen der Schätzer der Datenzelle 1671

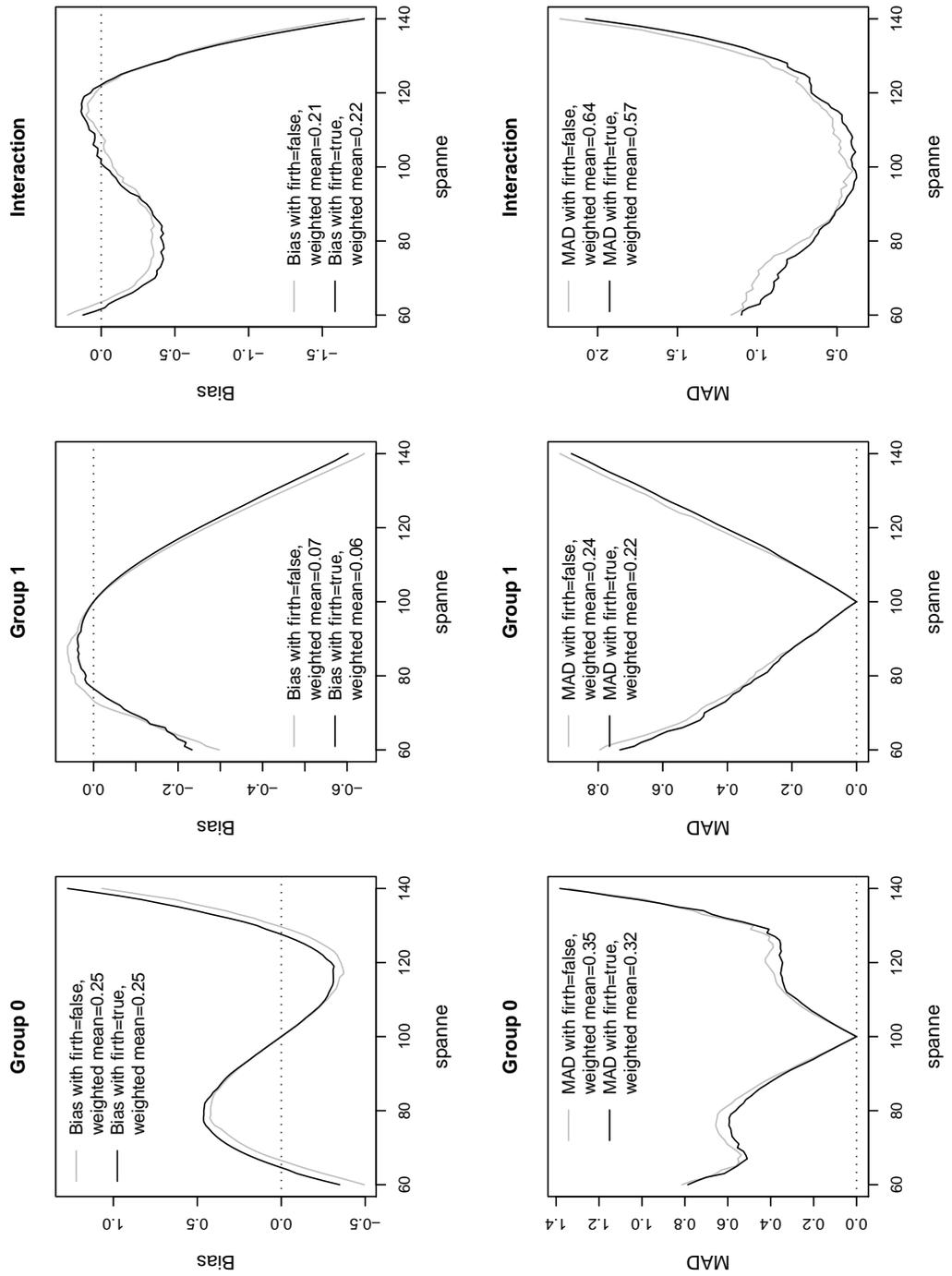


Abbildung 2.56: Abweichungen der Schätzer der Datenzelle 104

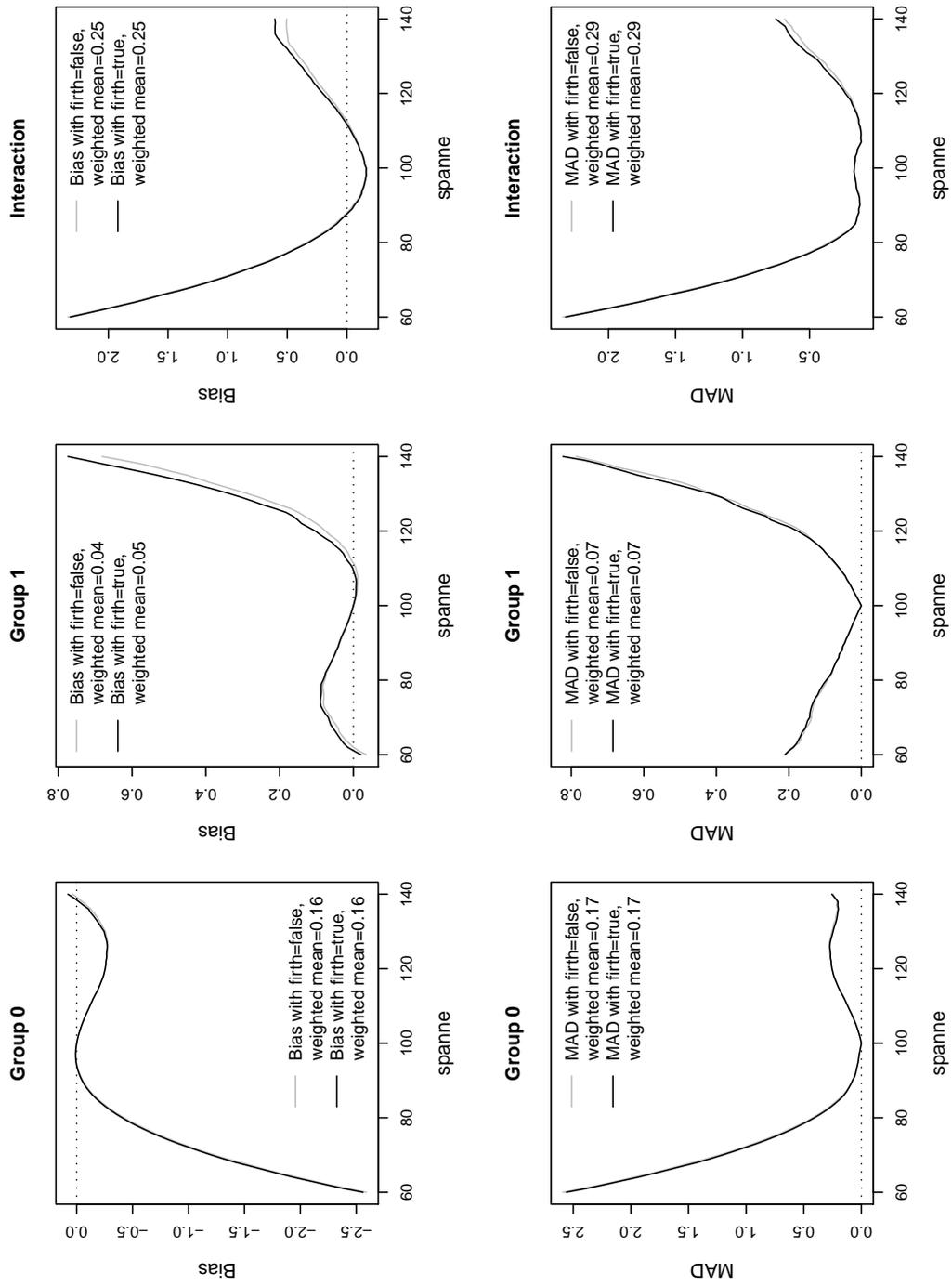


Abbildung 2.57: Abweichungen der Schätzer der Datenzelle 1334

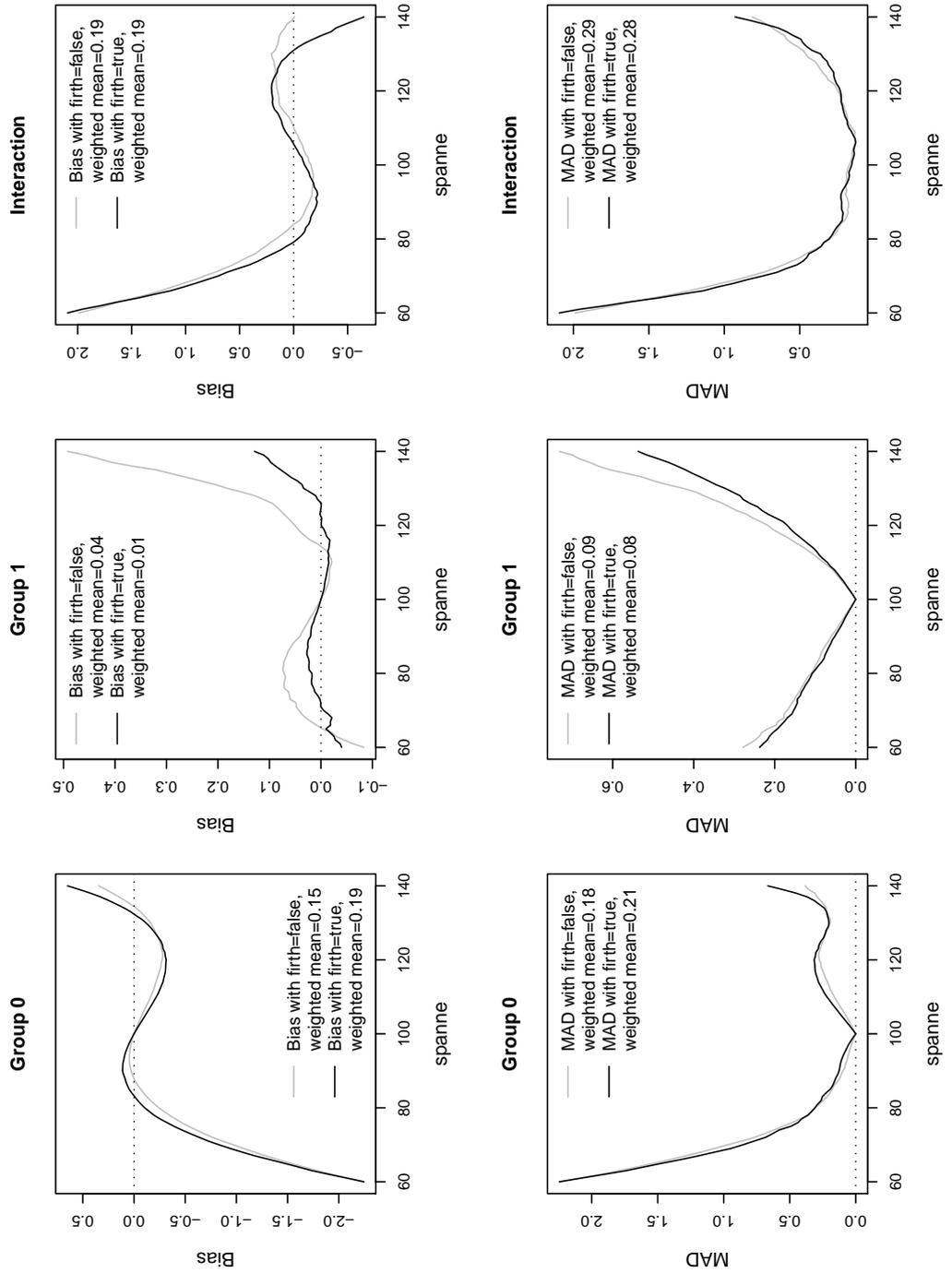


Abbildung 2.58: Abweichungen der Schätzer der Datenzelle 1674

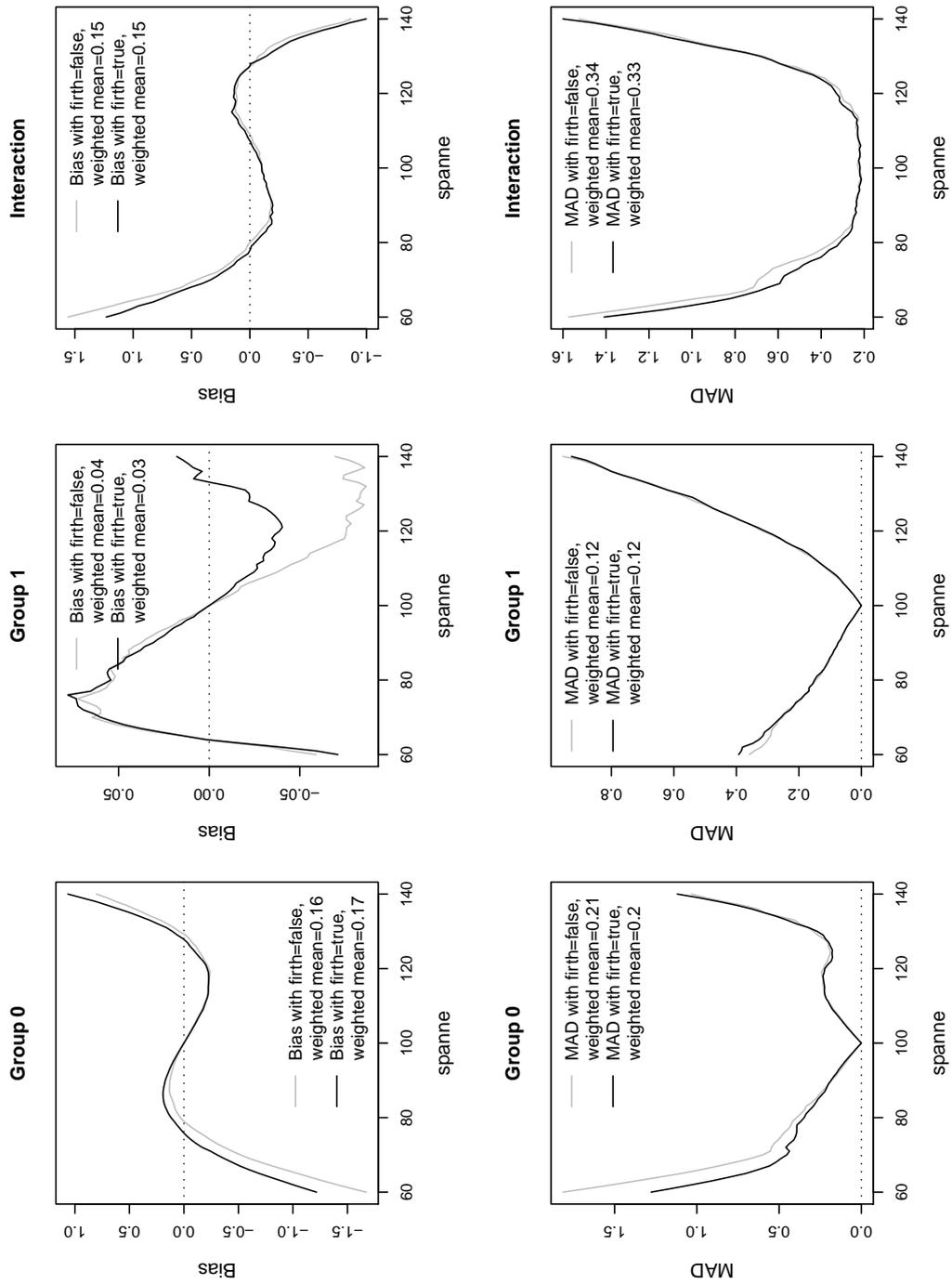


Abbildung 2.59: Abweichungen der Schätzer der Datenzelle 201

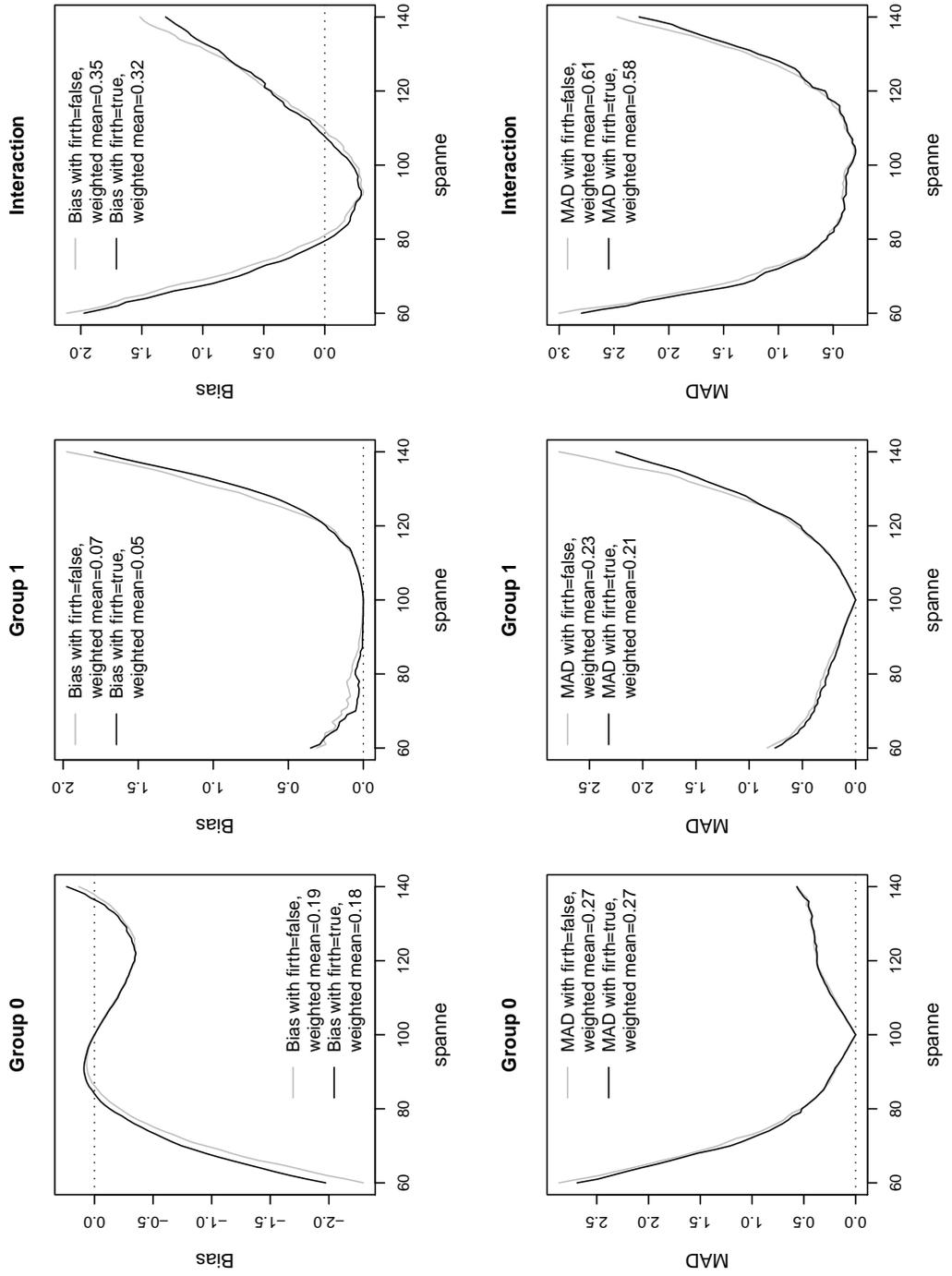


Abbildung 2.60: Abweichungen der Schätzer der Datenzelle 2331

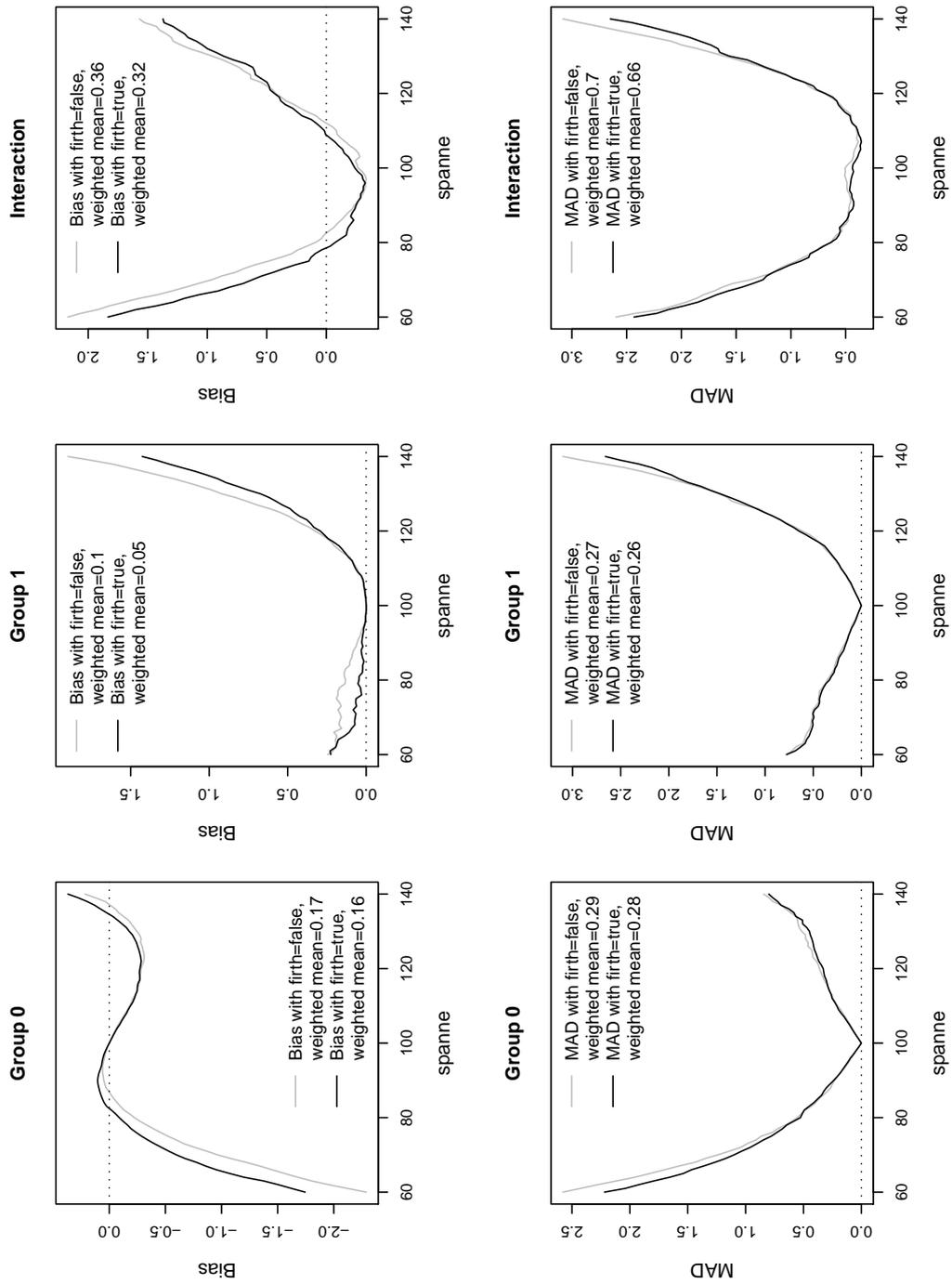


Abbildung 2.61: Abweichungen der Schätzer der Datenzelle 2671

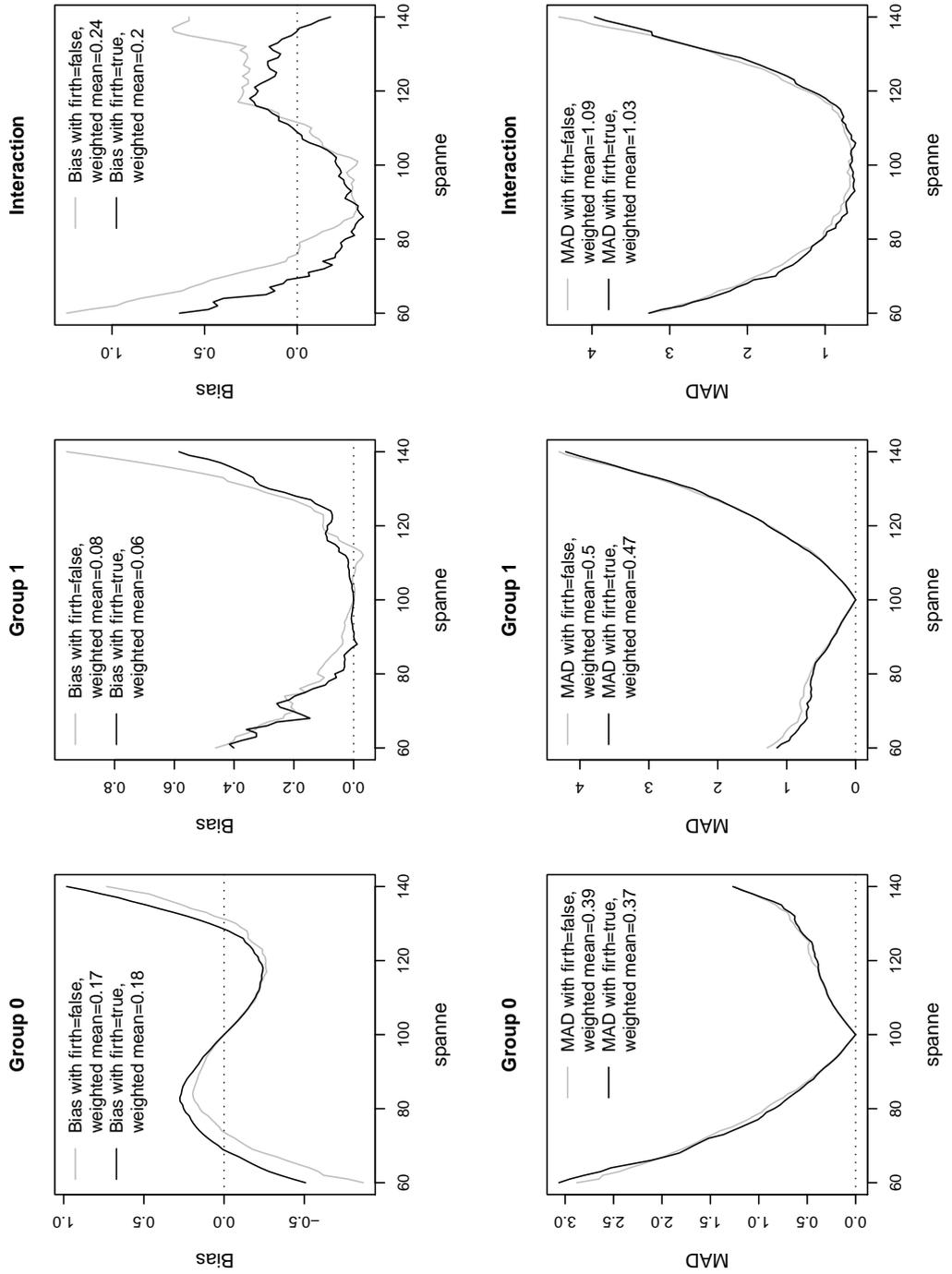


Abbildung 2.62: Abweichungen der Schätzer der Datenzelle 204

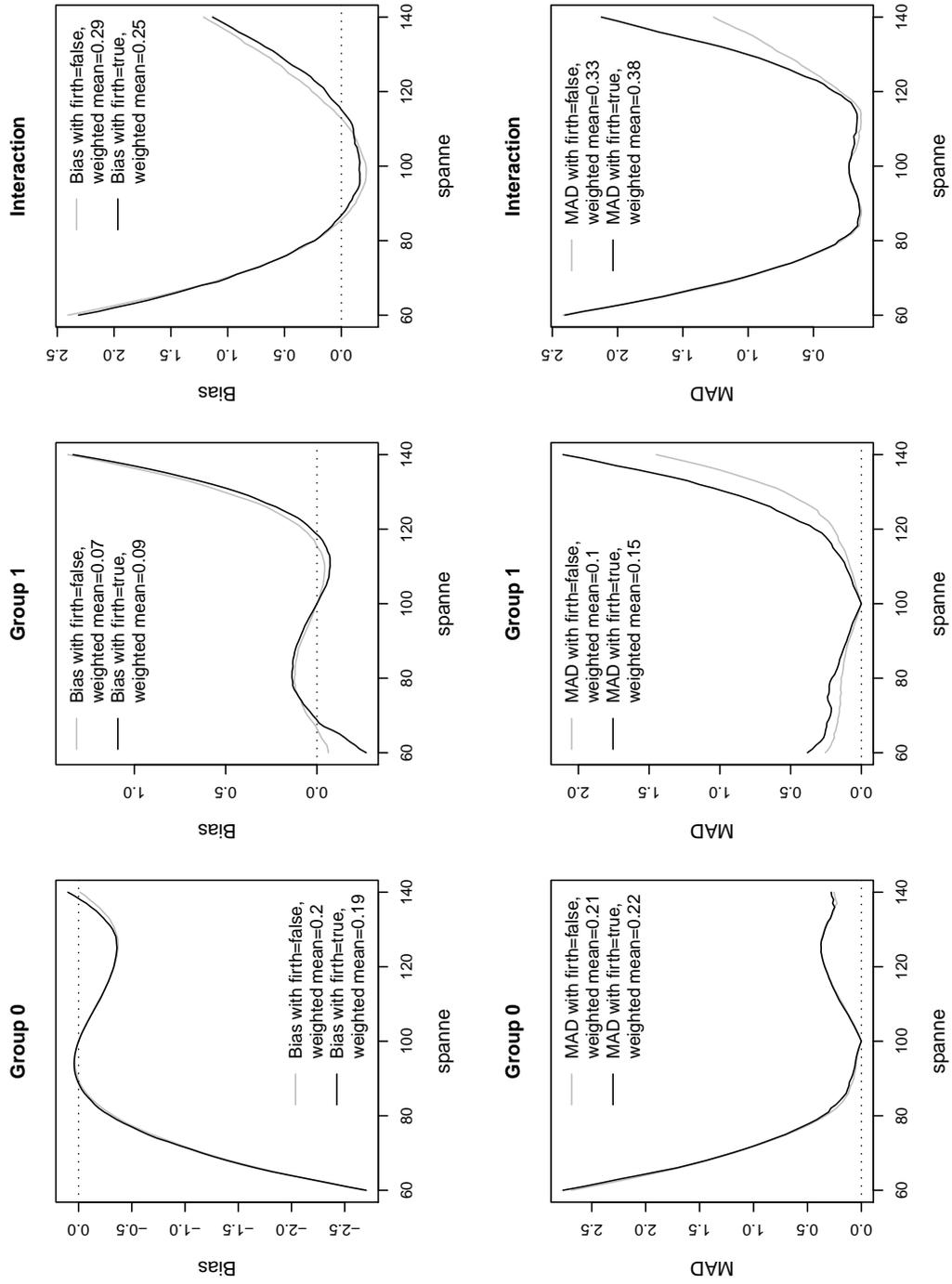


Abbildung 2.63: Abweichungen der Schätzer der Datenzelle 2334

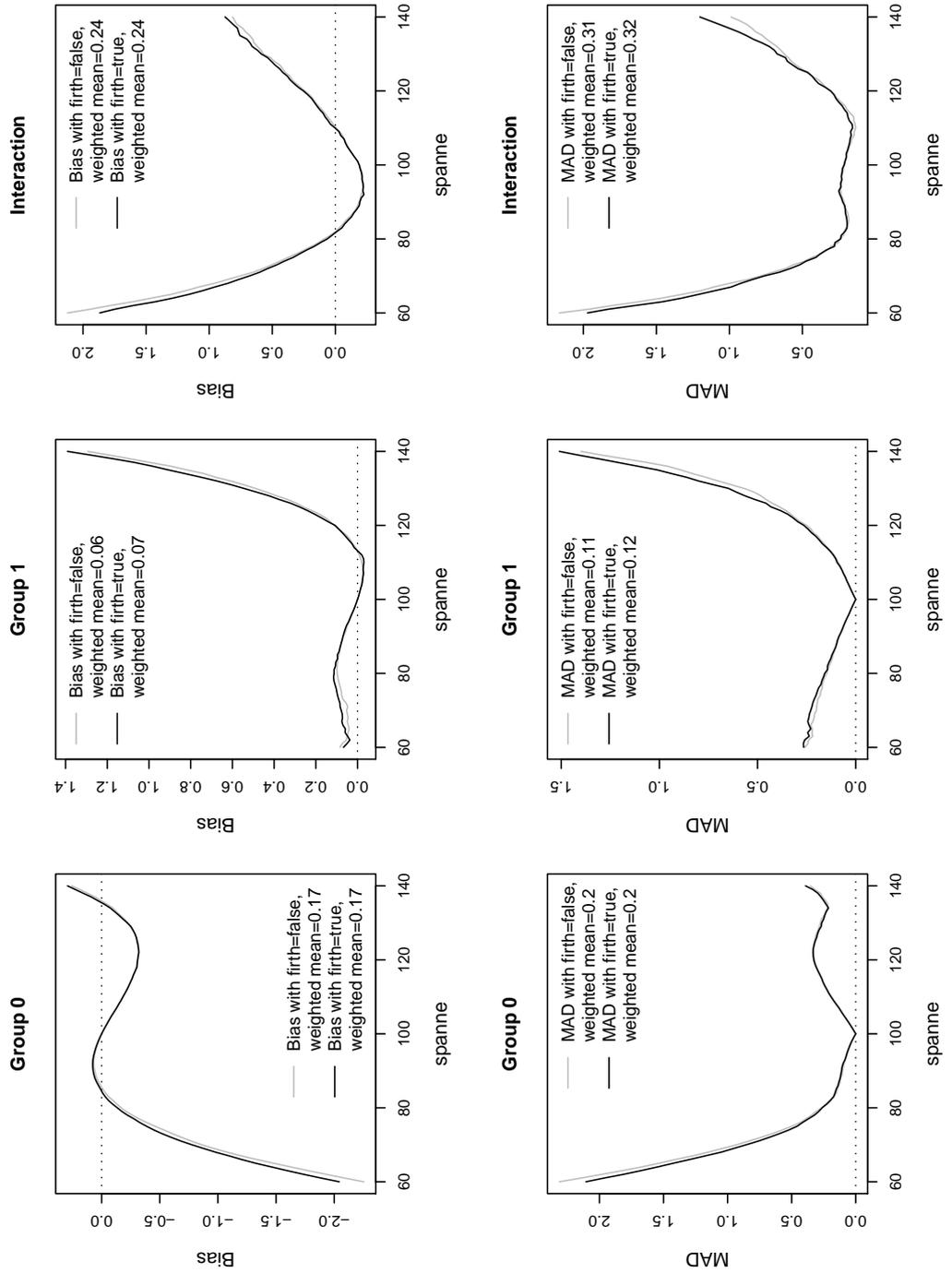


Abbildung 2.64: Abweichungen der Schätzer der Datenzelle 2674

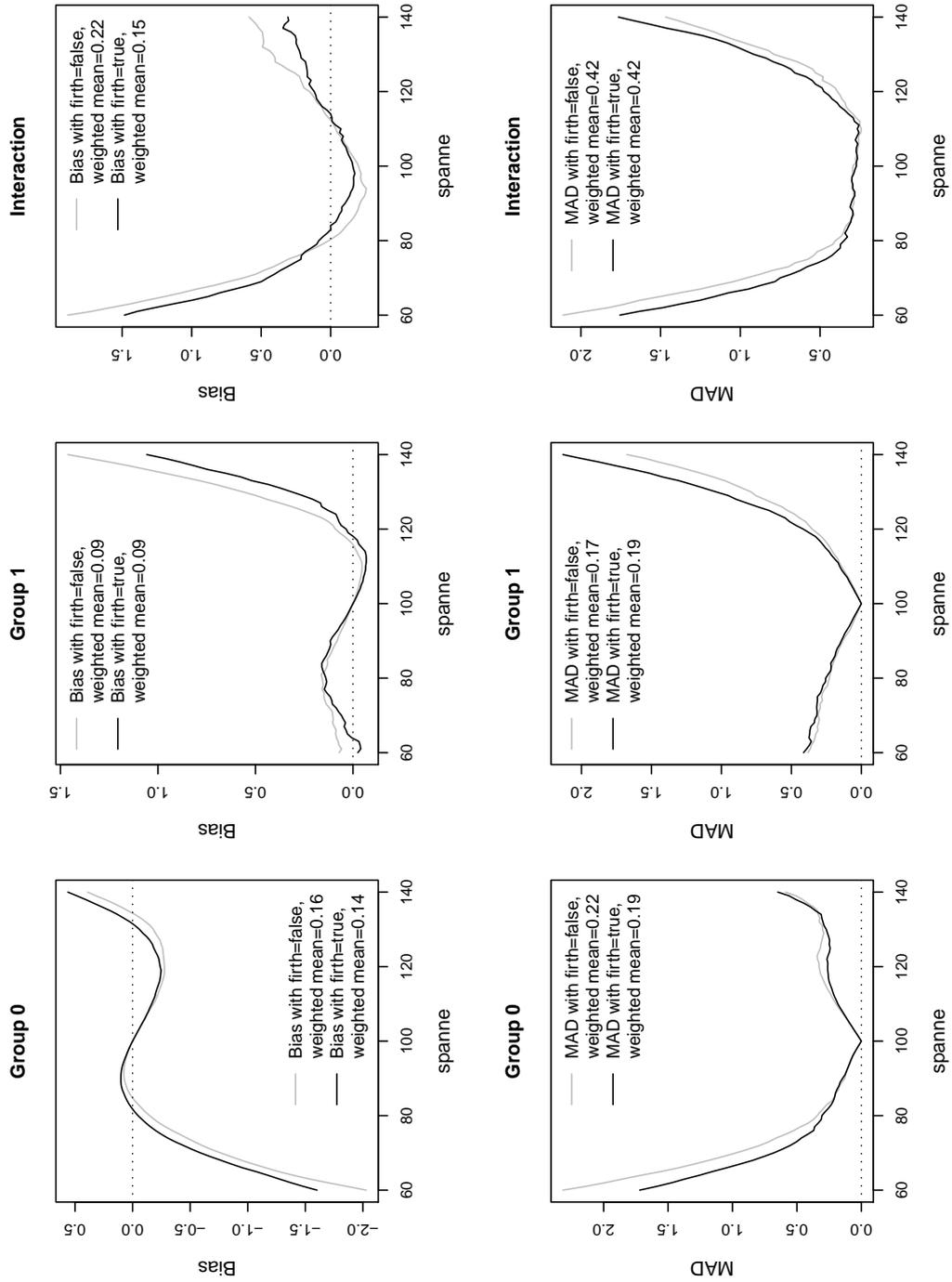


Abbildung 2.65: Abweichungen der Schätzer der Datenzelle 301

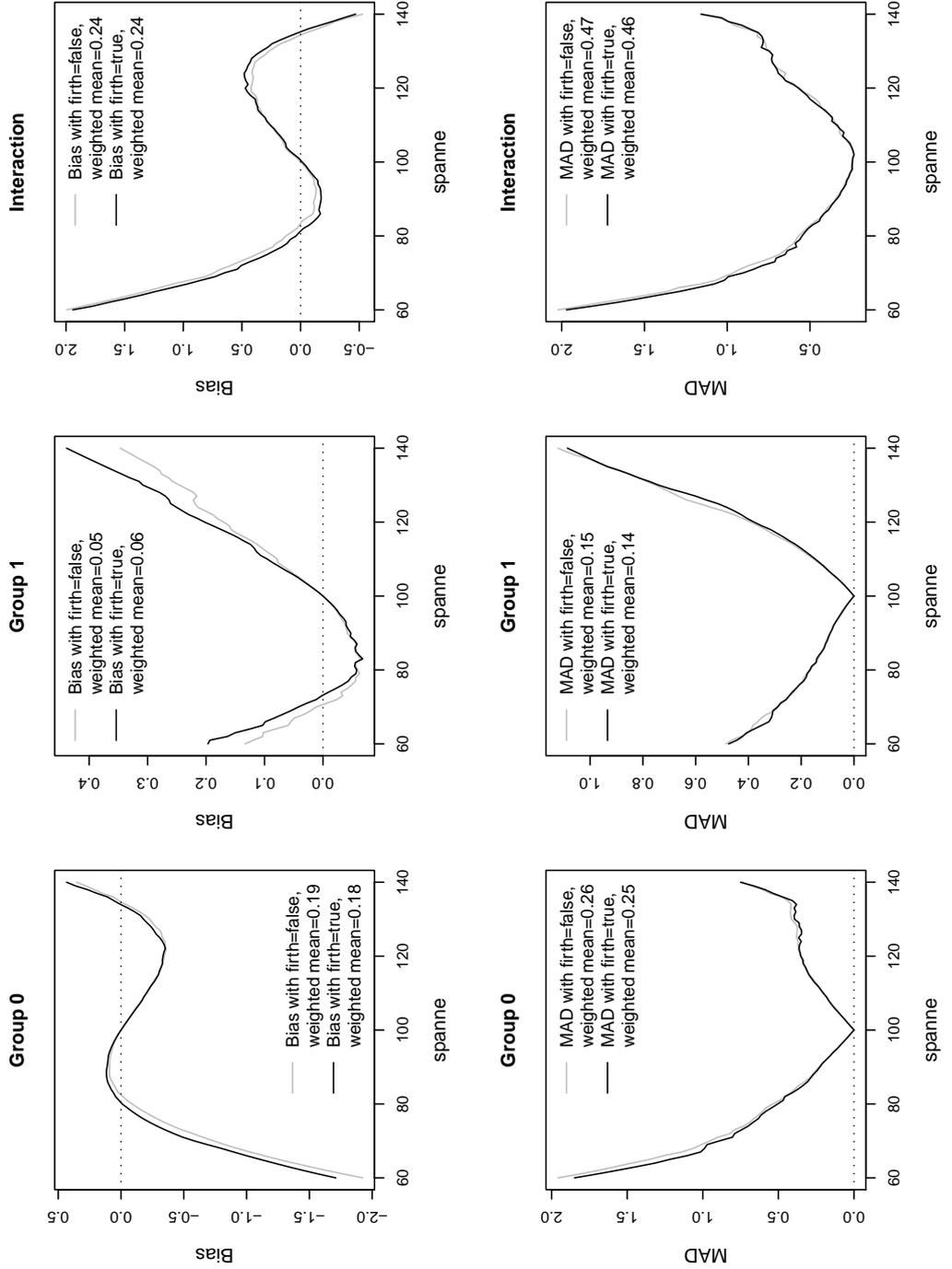


Abbildung 2.66: Abweichungen der Schätzer der Datenzelle 3331

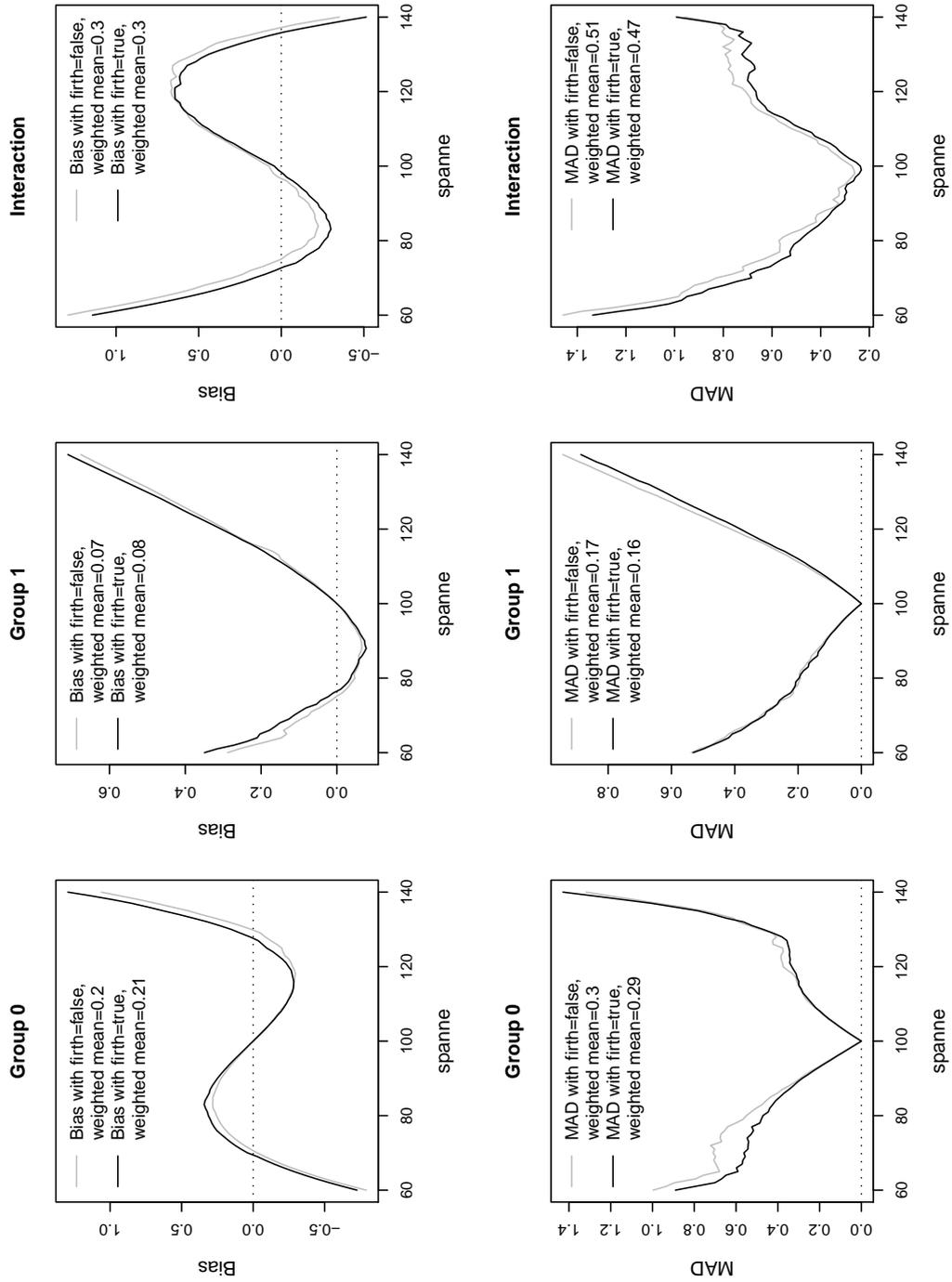


Abbildung 2.67: Abweichungen der Schätzer der Datenzelle 3671

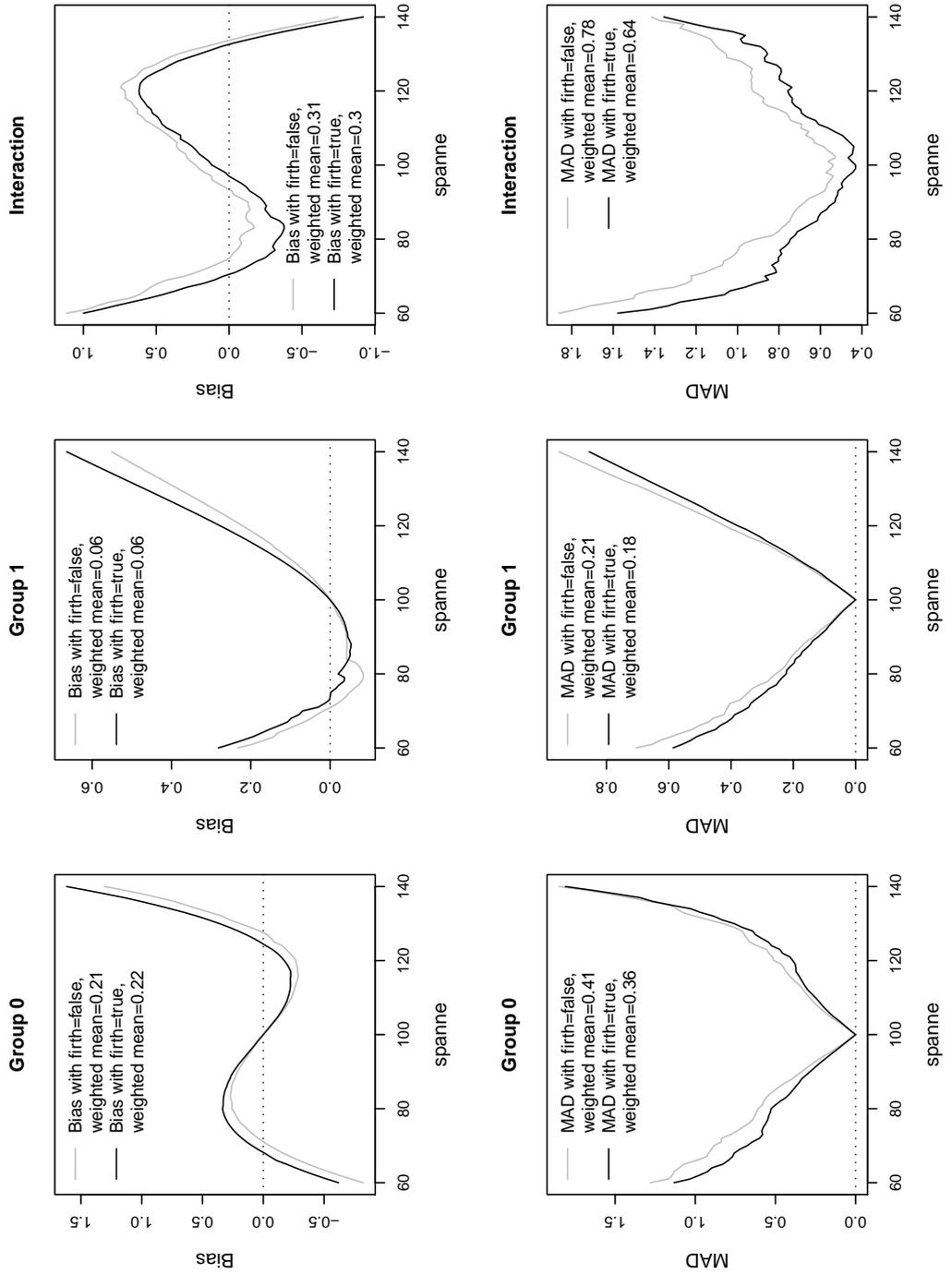


Abbildung 2.68: Abweichungen der Schätzer der Datenzelle 304

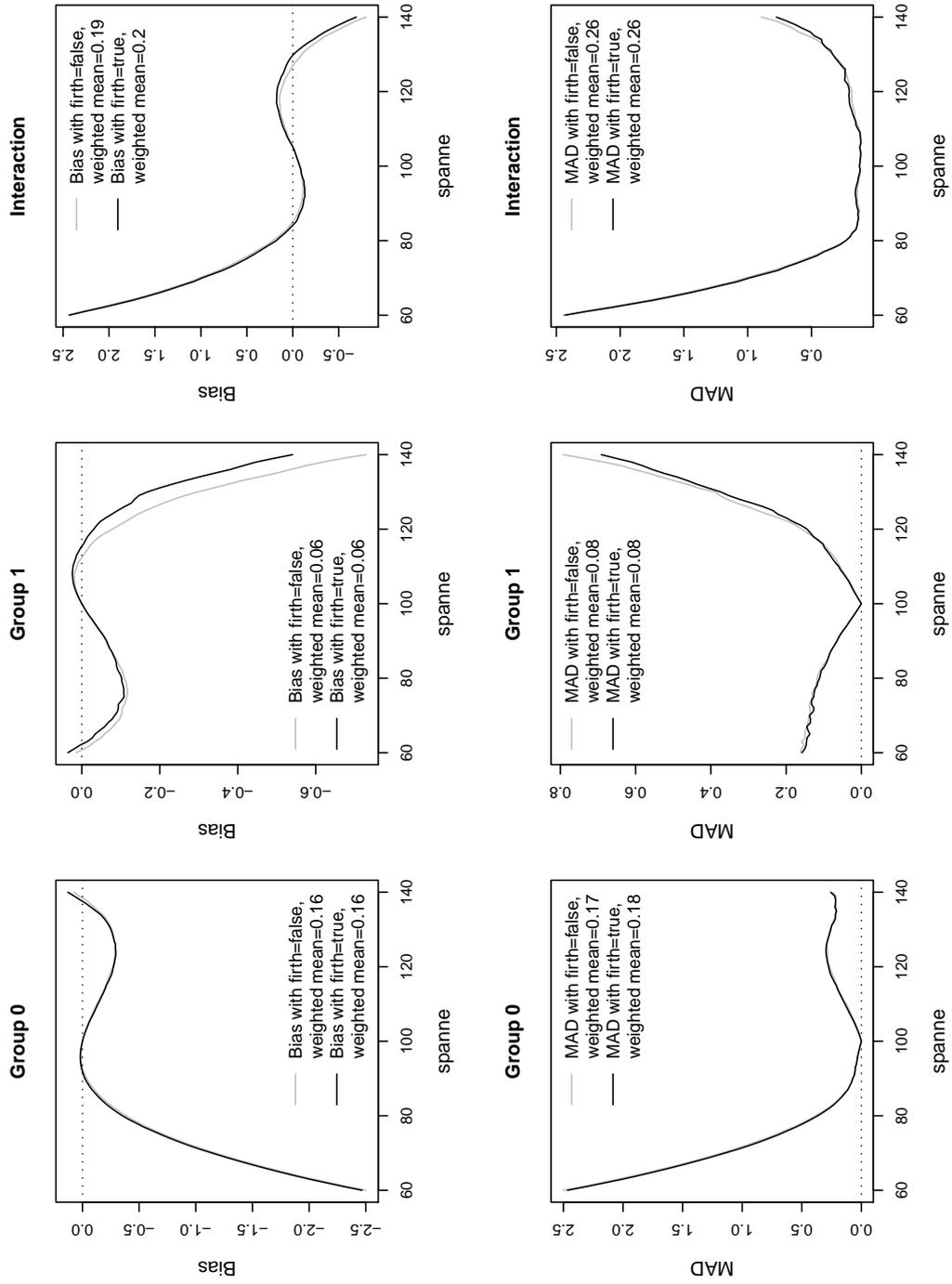


Abbildung 2.69: Abweichungen der Schätzer der Datenzelle 3334

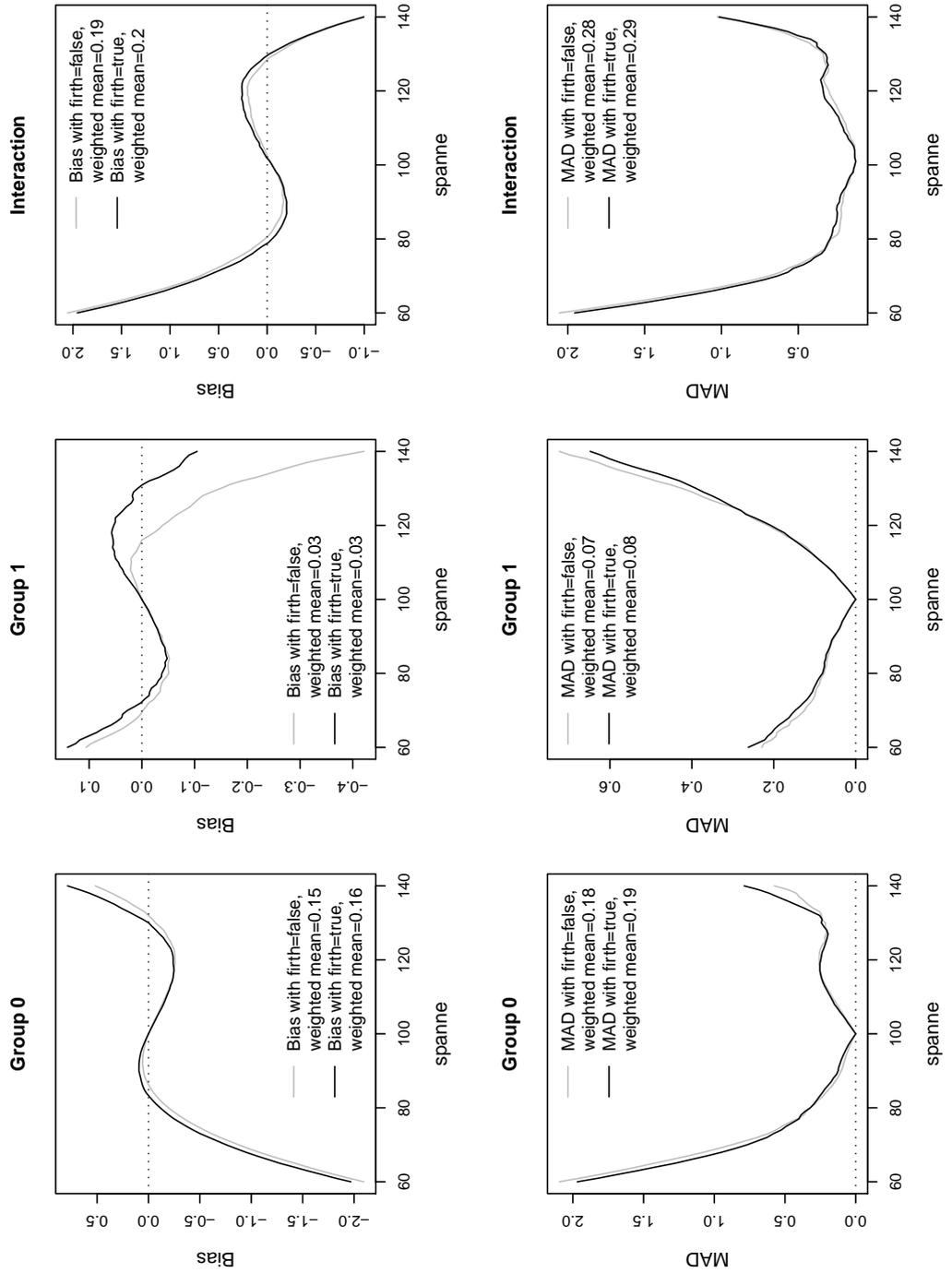


Abbildung 2.70: Abweichungen der Schätzer der Datenzelle 3674

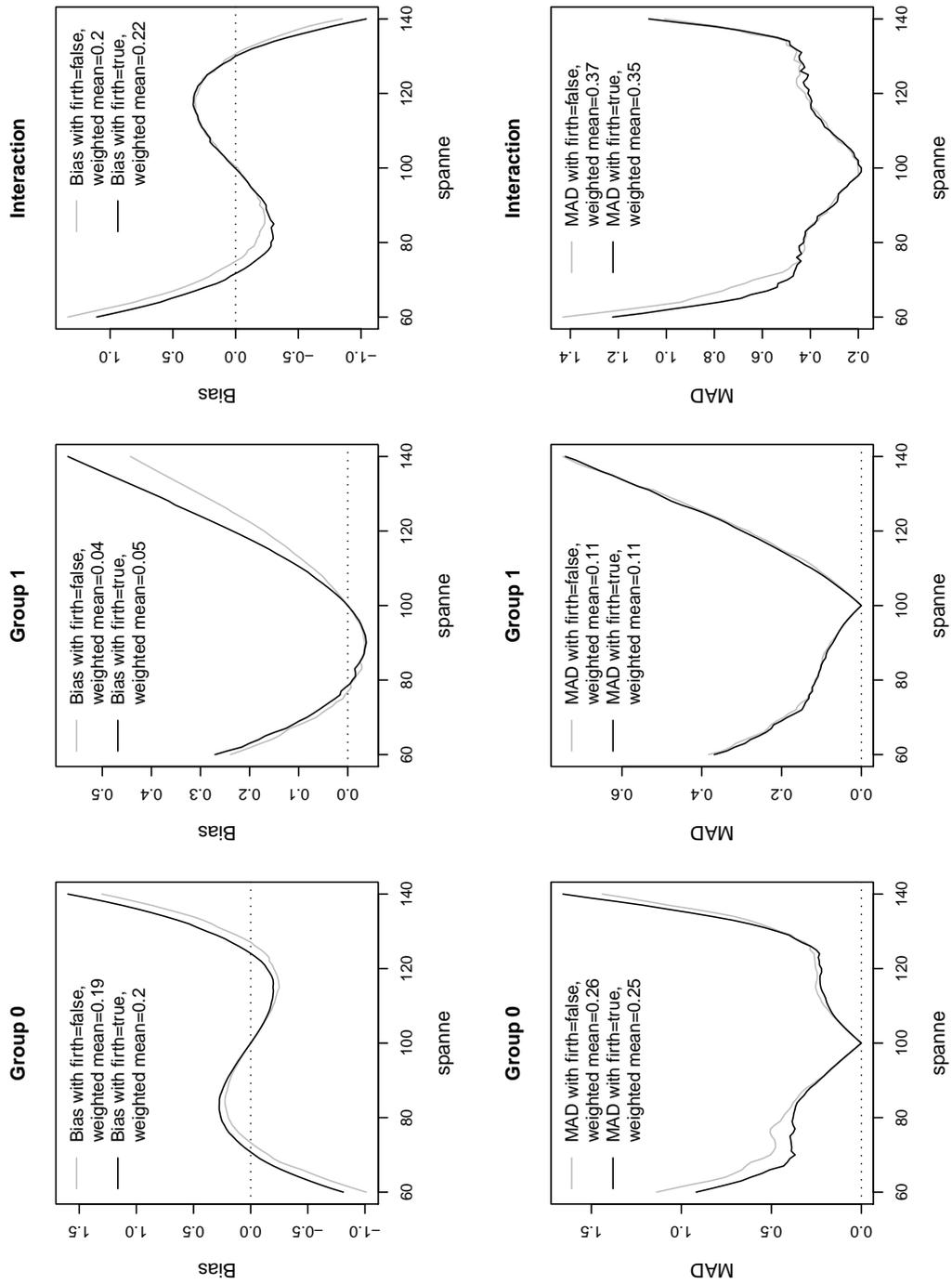


Abbildung 2.71: Abweichungen der Schätzer der Datenzelle 401

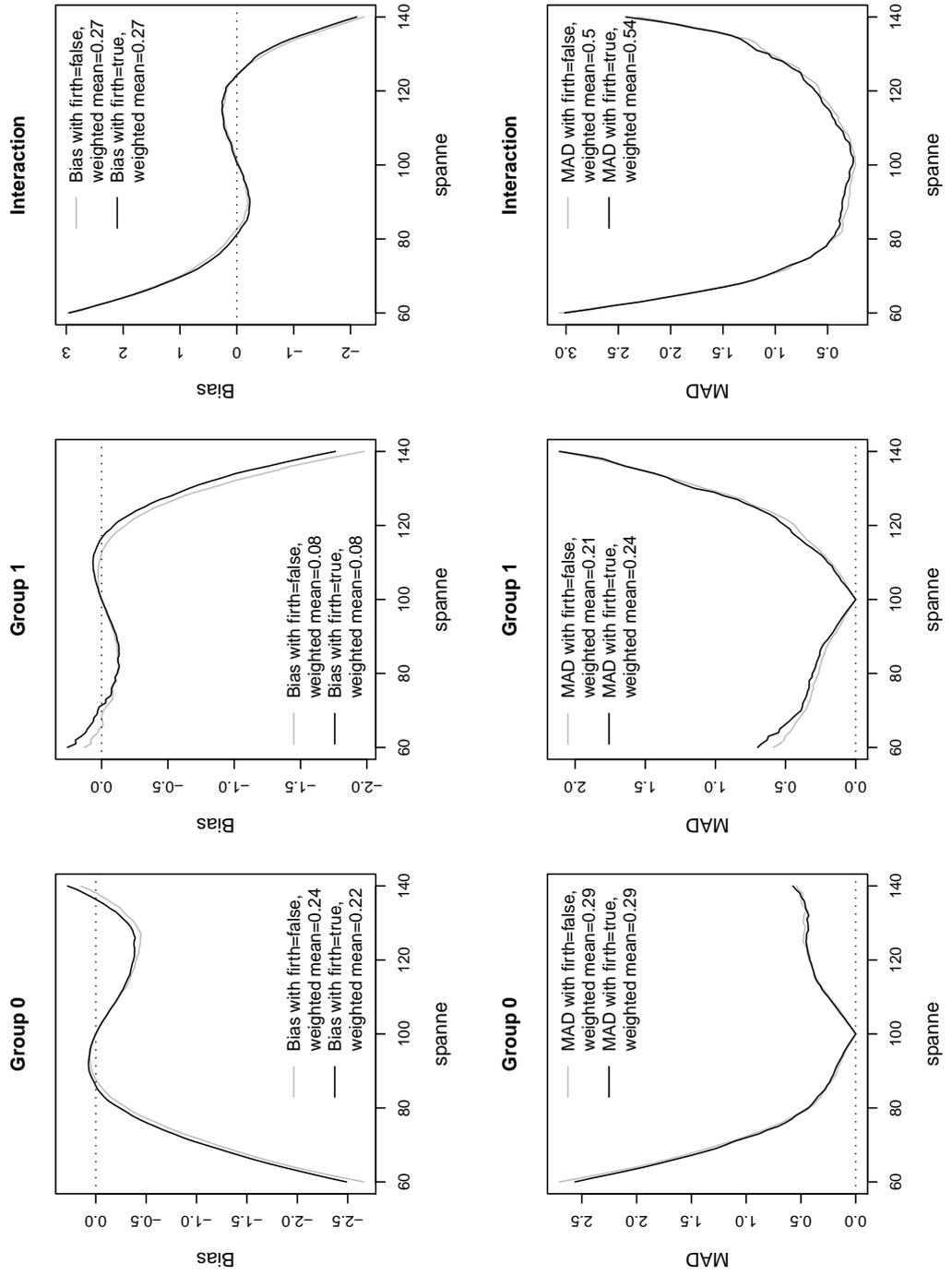


Abbildung 2.7.2: Abweichungen der Schätzer der Datenzelle 4331

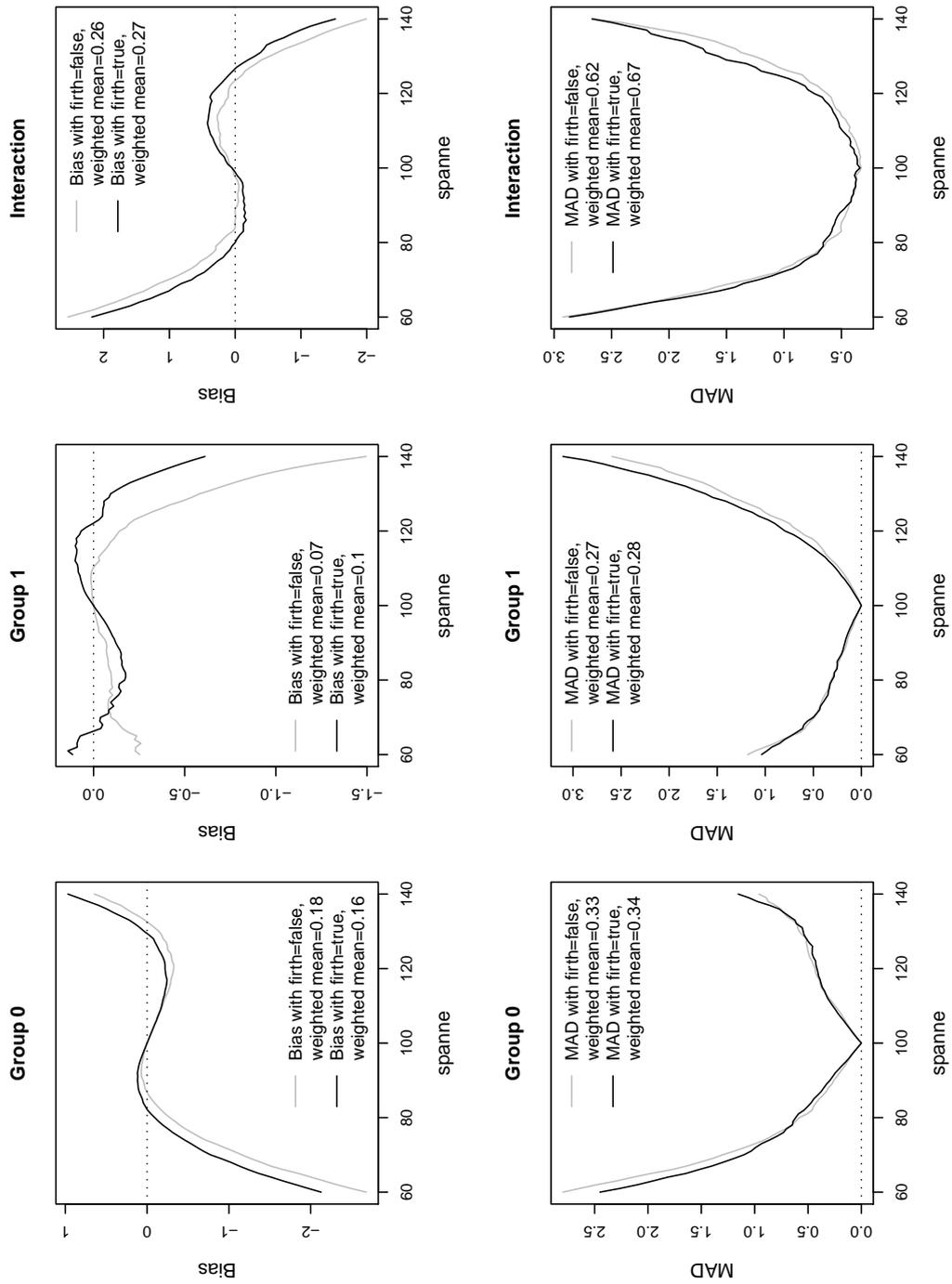


Abbildung 2.73: Abweichungen der Schätzer der Datenzelle 4671

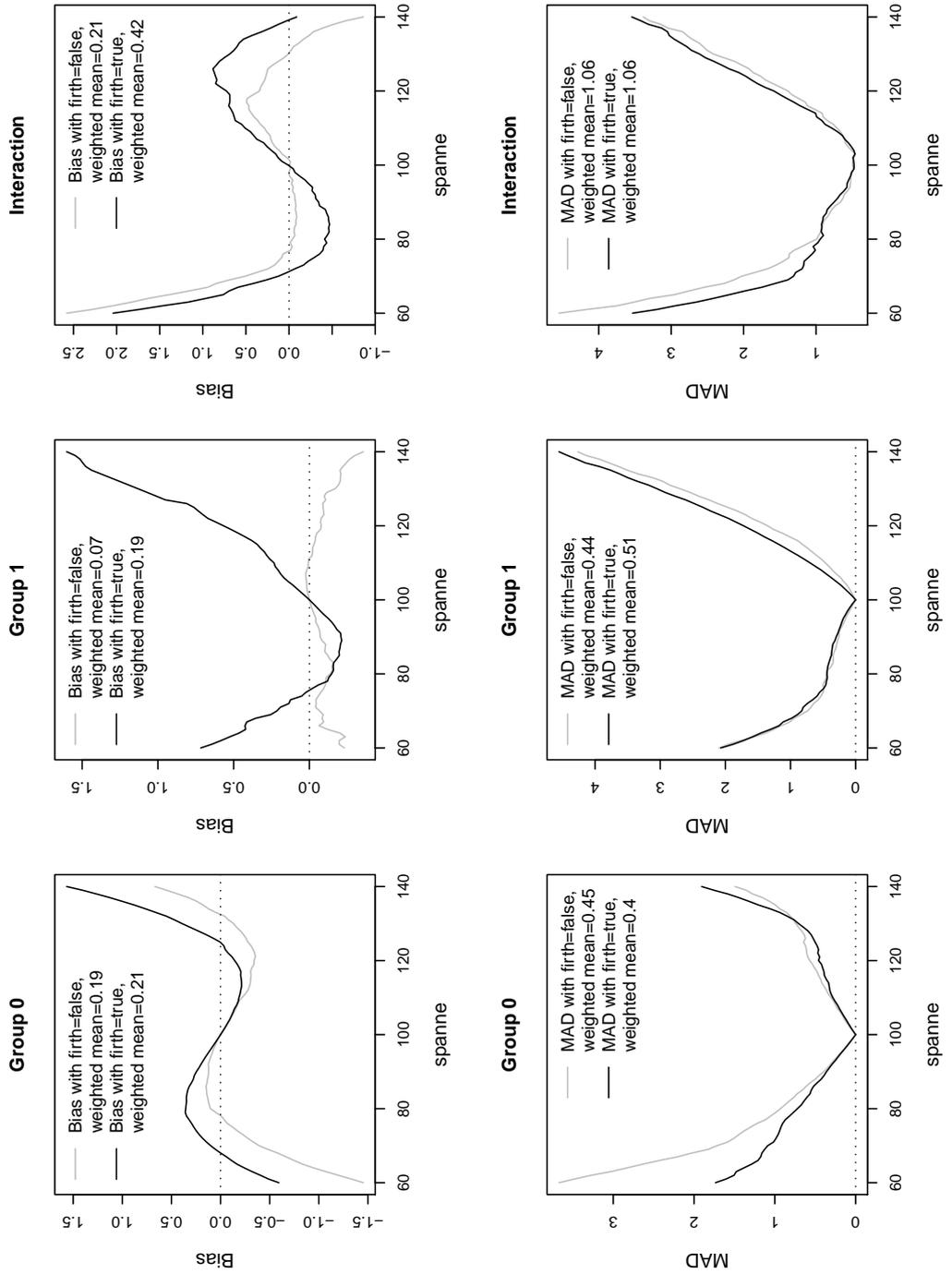


Abbildung 2.74: Abweichungen der Schätzer der Datenzelle 404

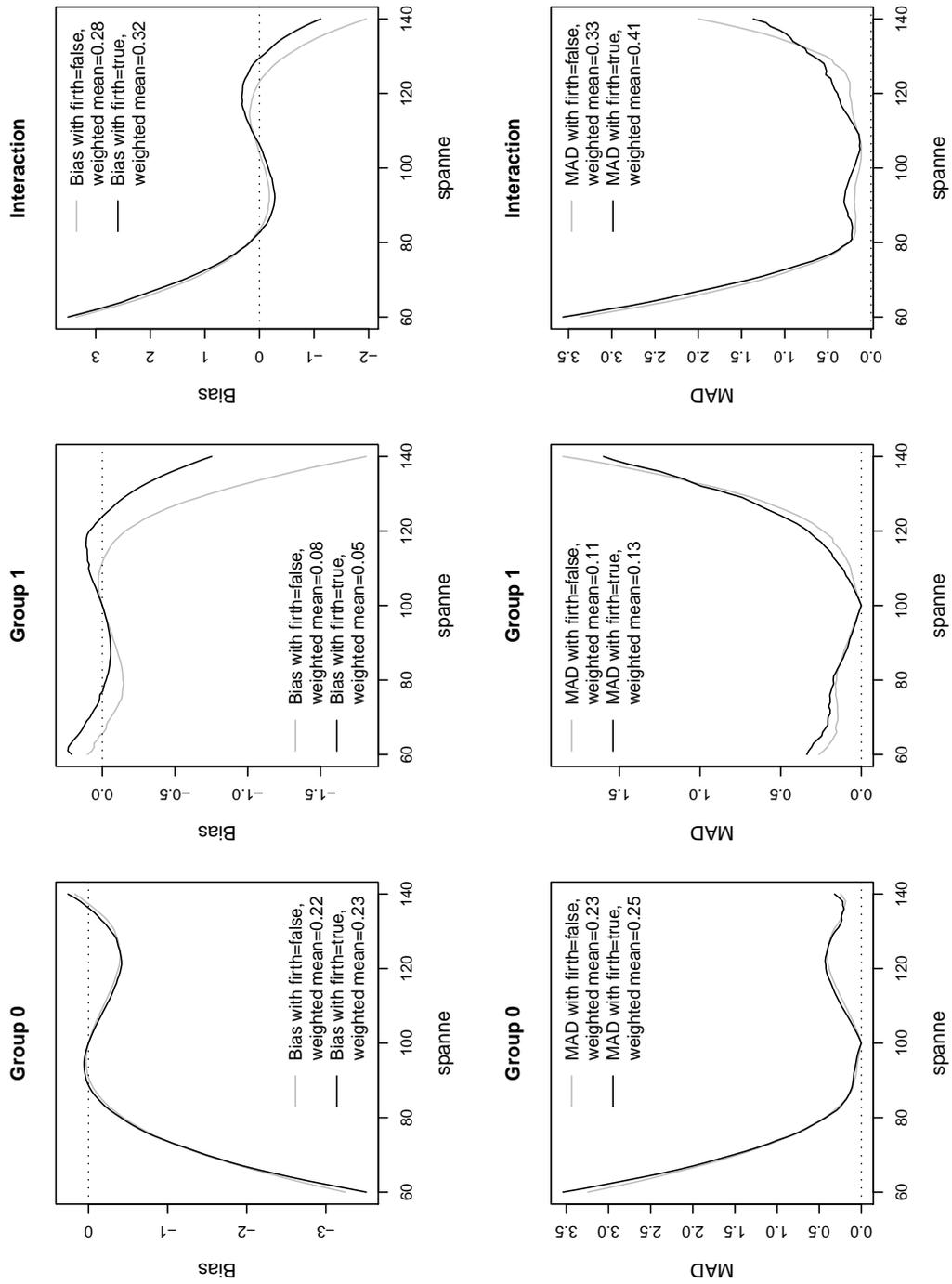


Abbildung 2.75: Abweichungen der Schätzer der Datenzelle 4334

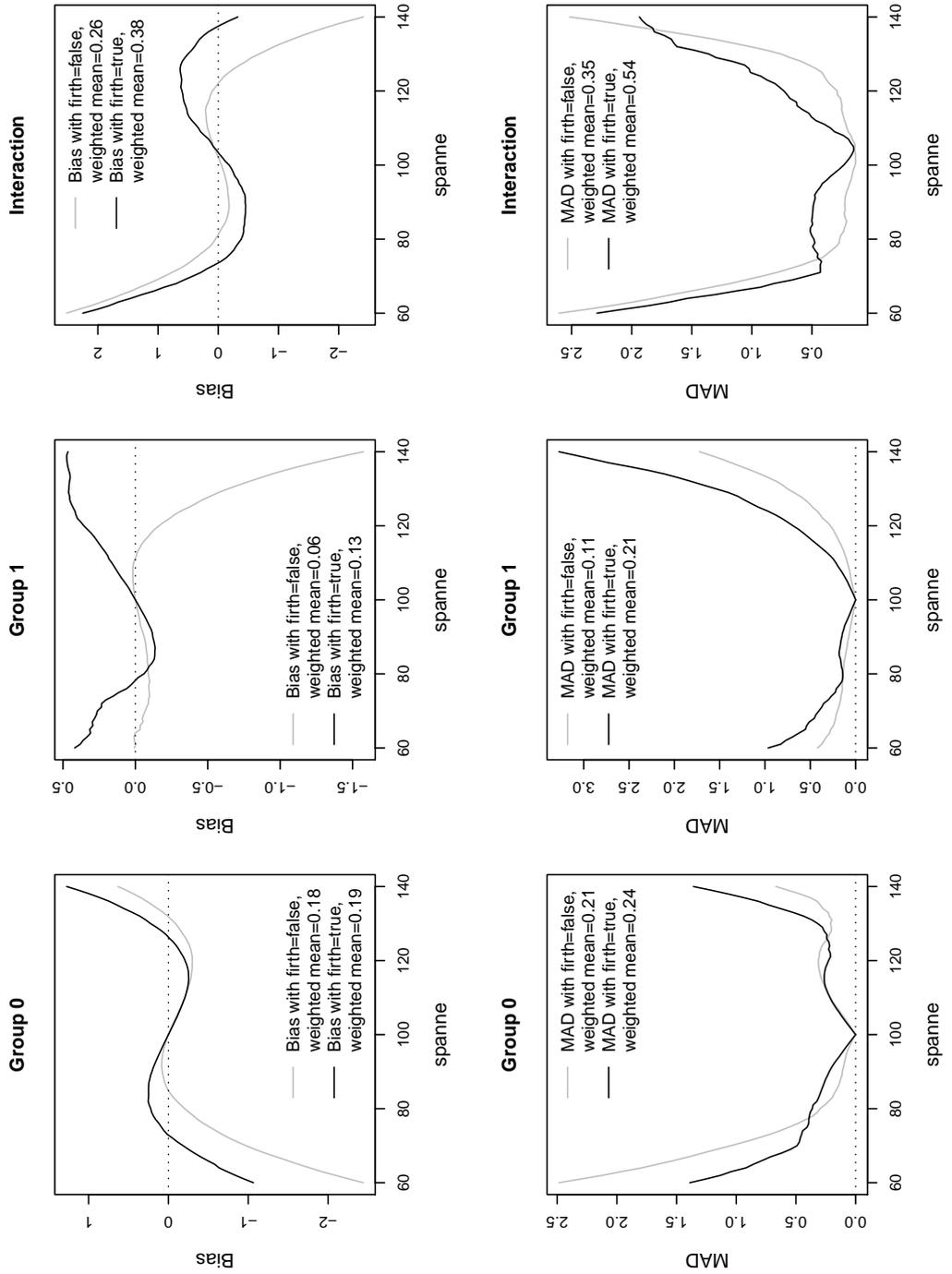


Abbildung 2.76: Abweichungen der Schätzer der Datenzelle 4674

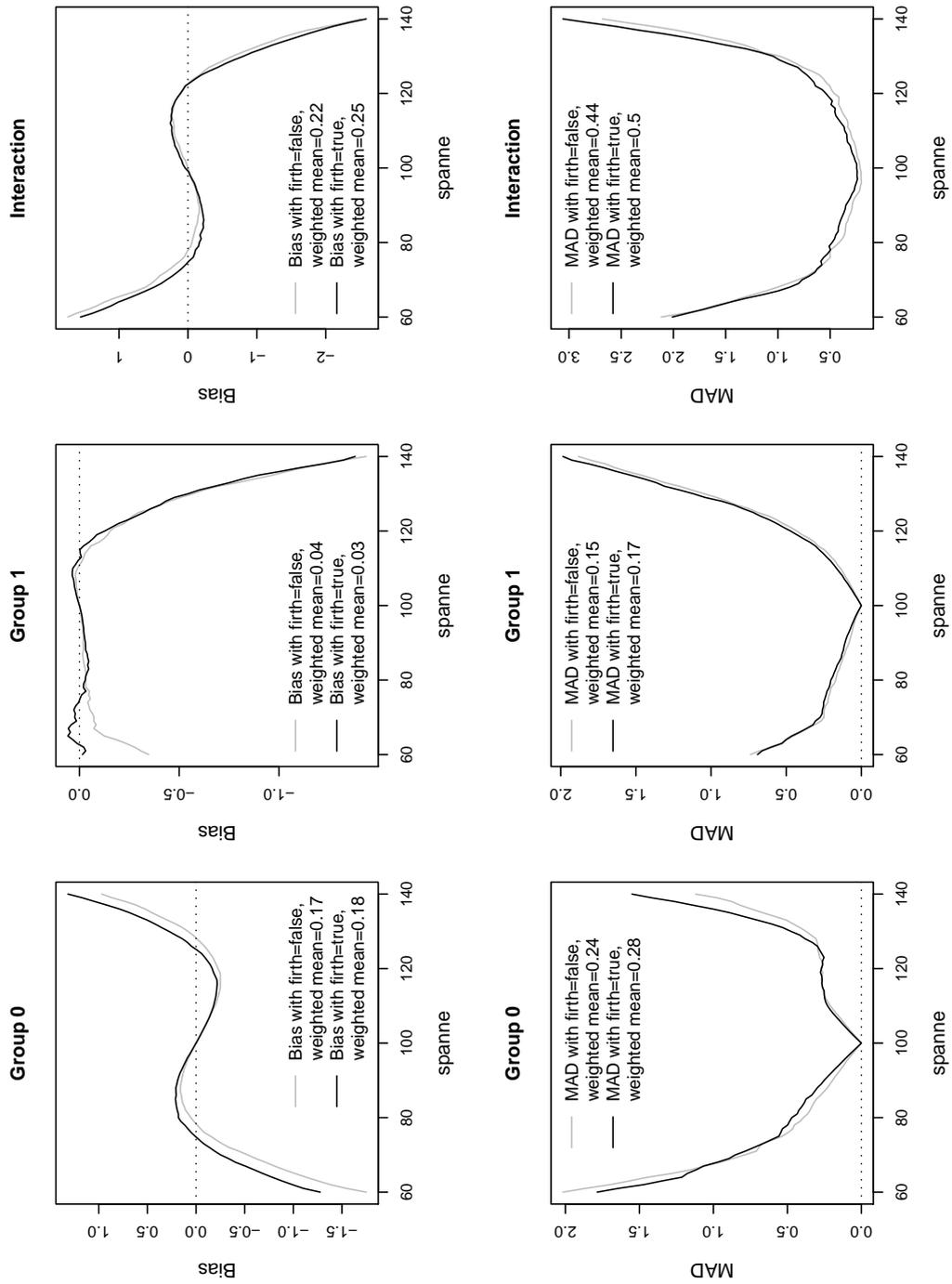


Abbildung 2.77: Abweichungen der Schätzer der Datenzelle 501

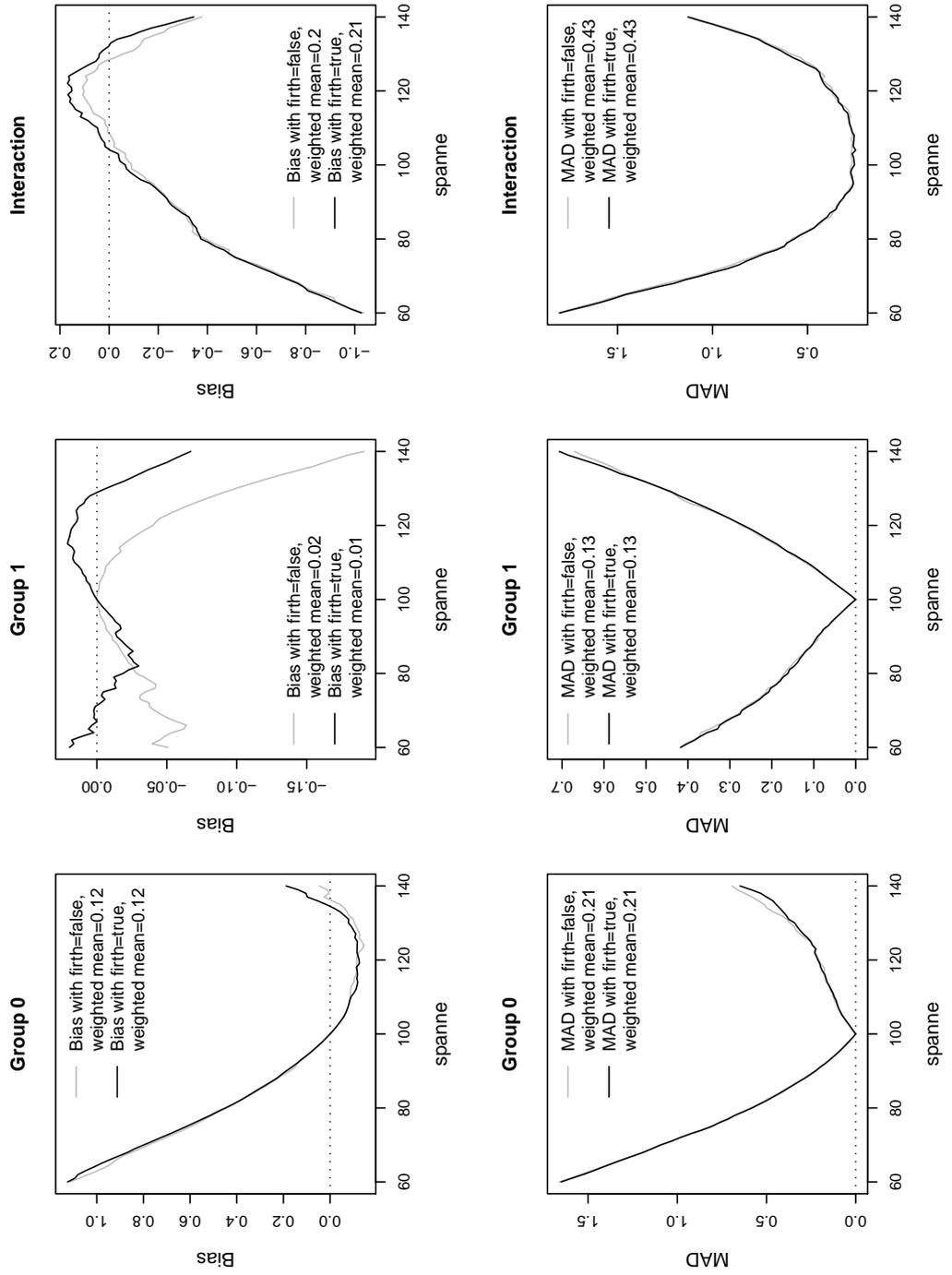


Abbildung 2.78: Abweichungen der Schätzer der Datenzelle 5331

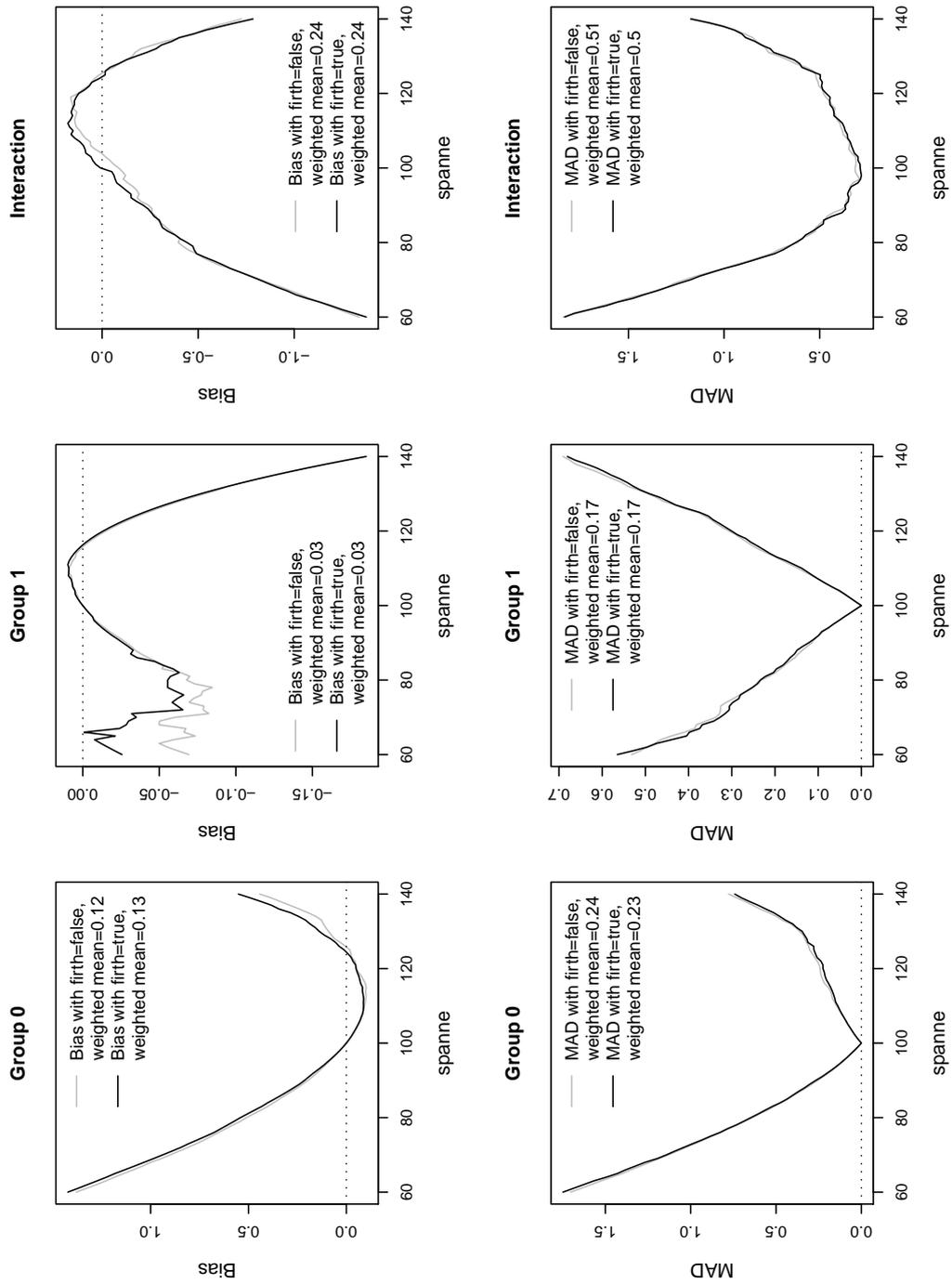


Abbildung 2.79: Abweichungen der Schätzer der Datenzelle 5671

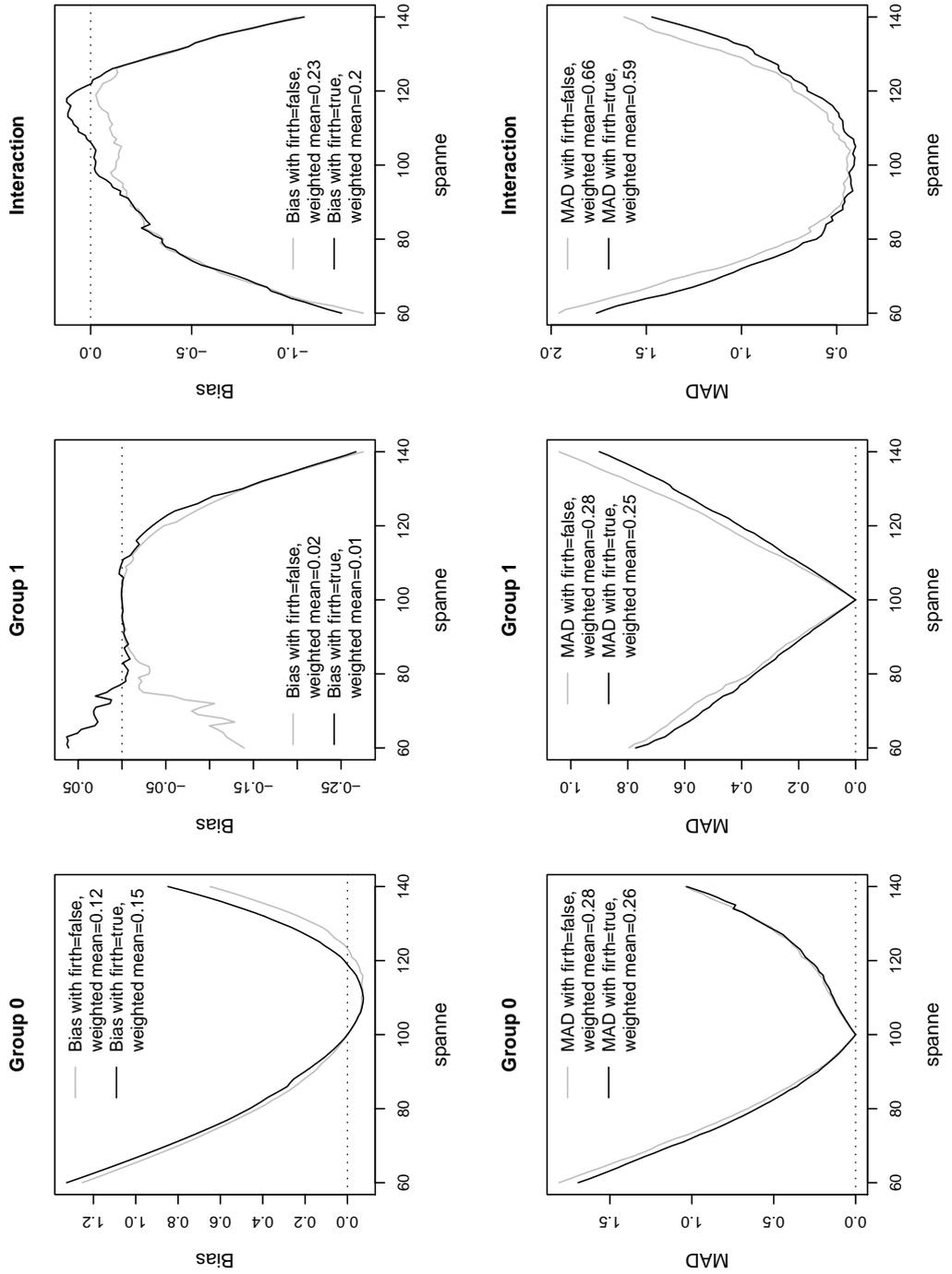


Abbildung 2.80: Abweichungen der Schätzer der Datenzelle 504

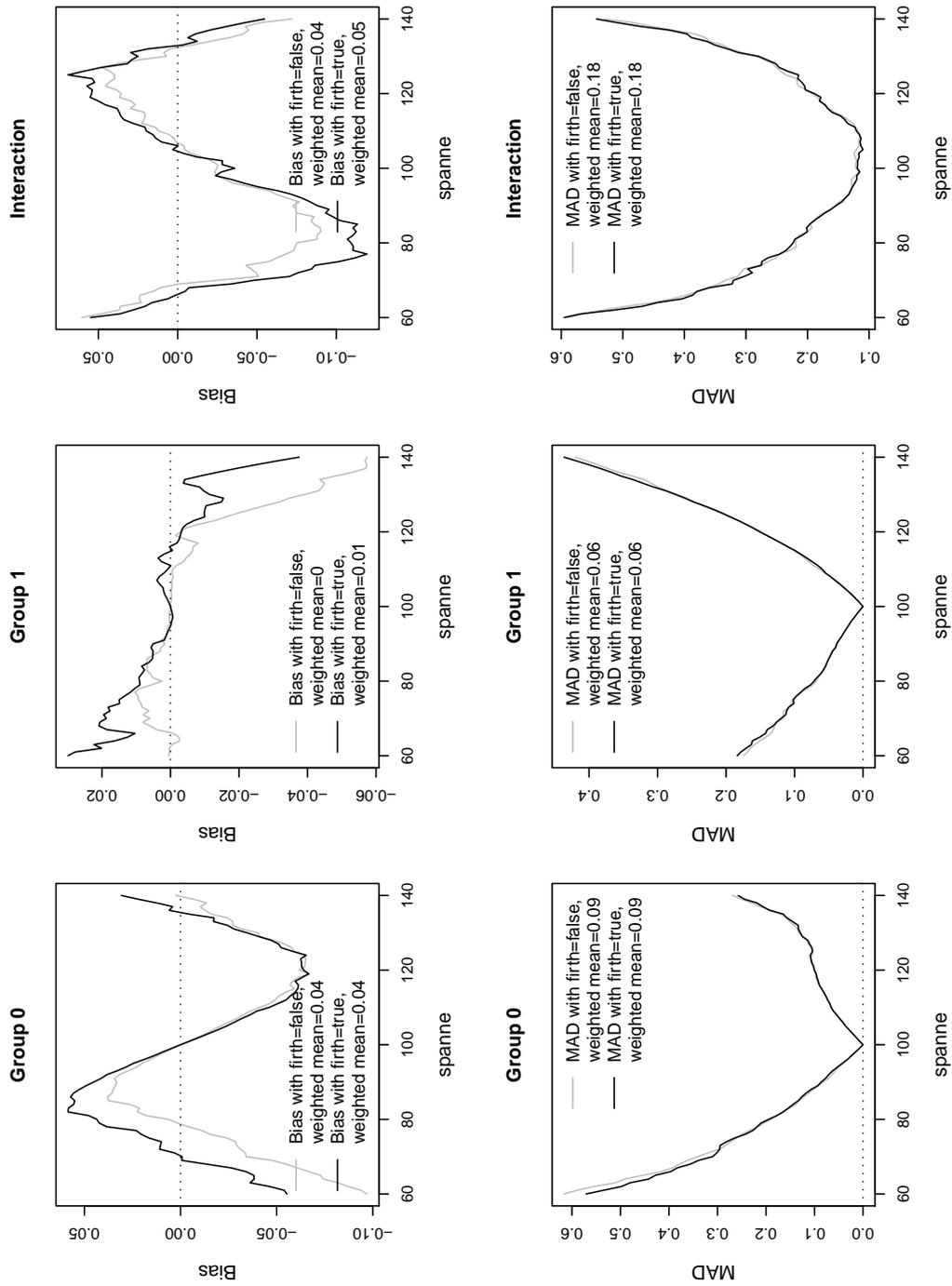


Abbildung 2.81: Abweichungen der Schätzer der Datenzelle 5334

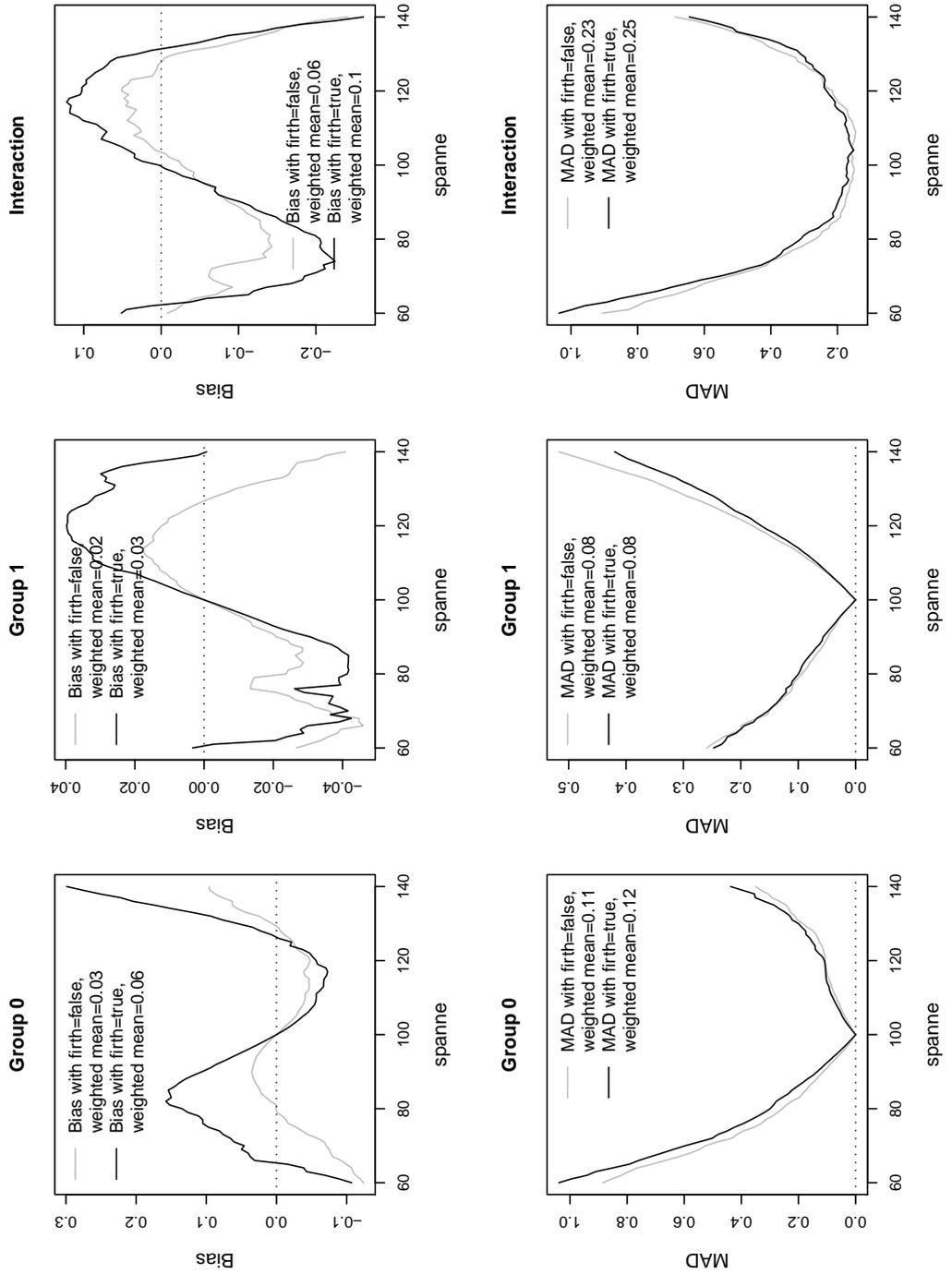


Abbildung 2.82: Abweichungen der Schätzer der Datenzelle 5674

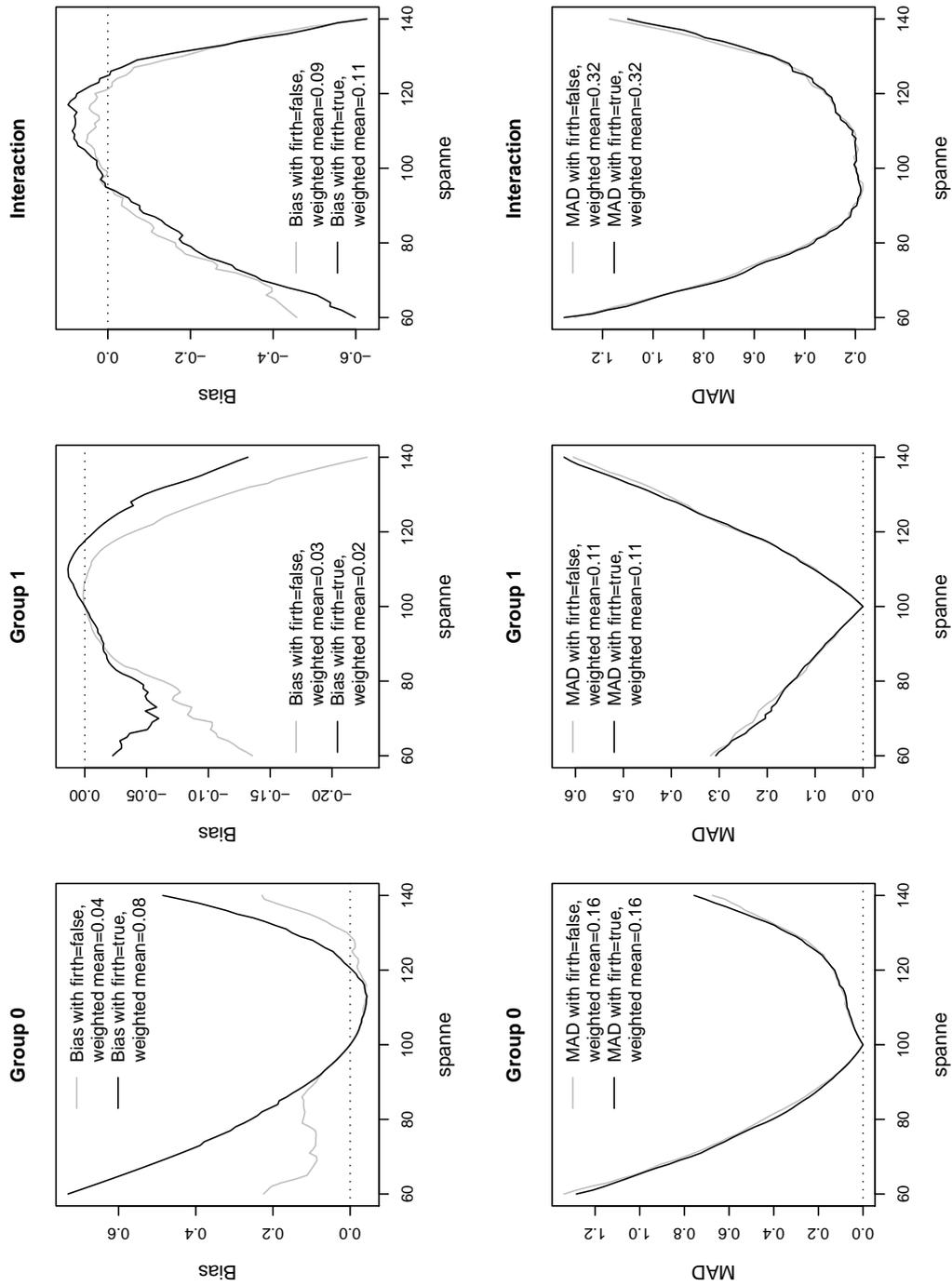


Abbildung 2.83: Abweichungen der Schätzer der Datenzelle 601

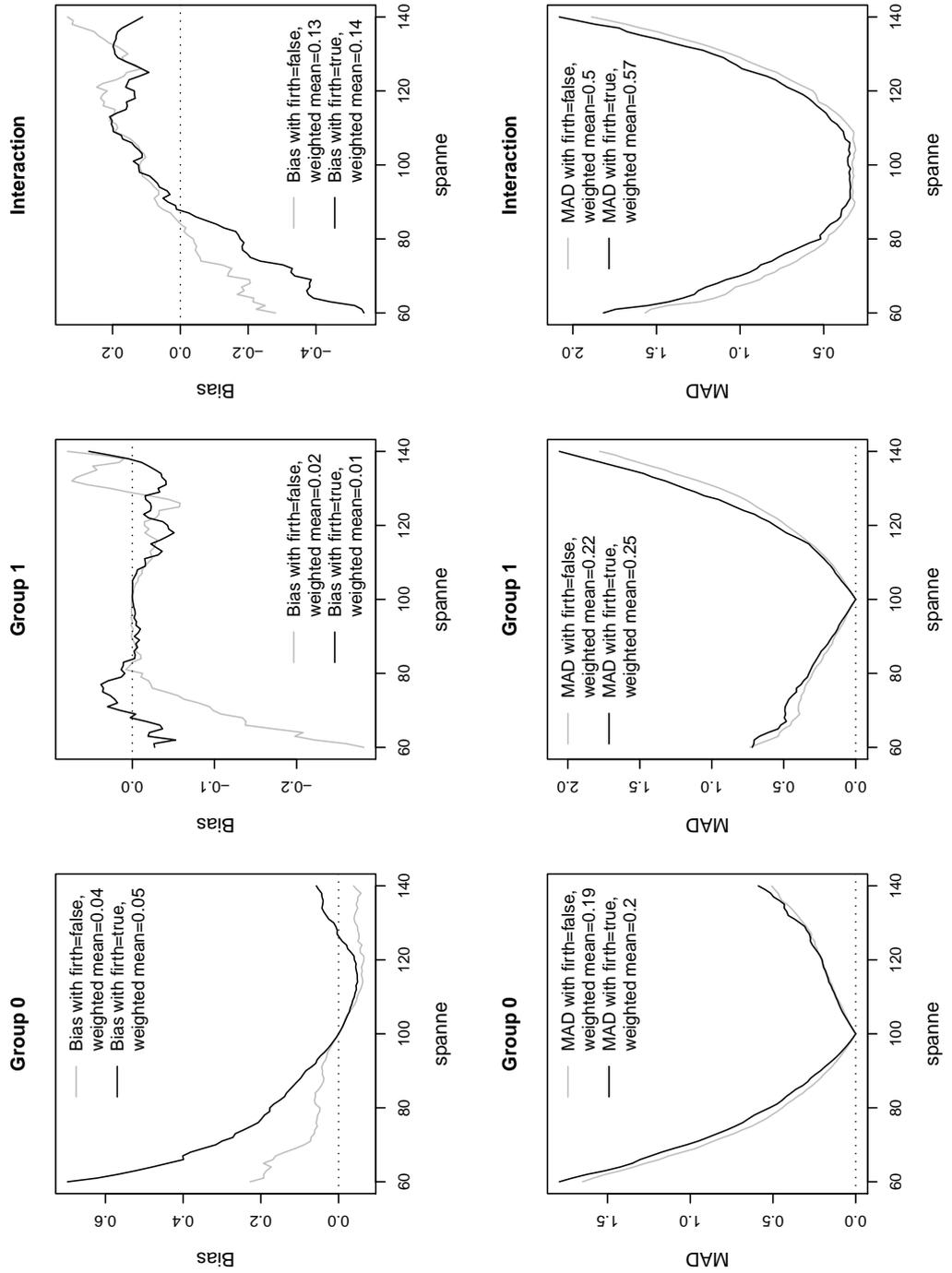


Abbildung 2.84: Abweichungen der Schätzer der Datenzelle 6331

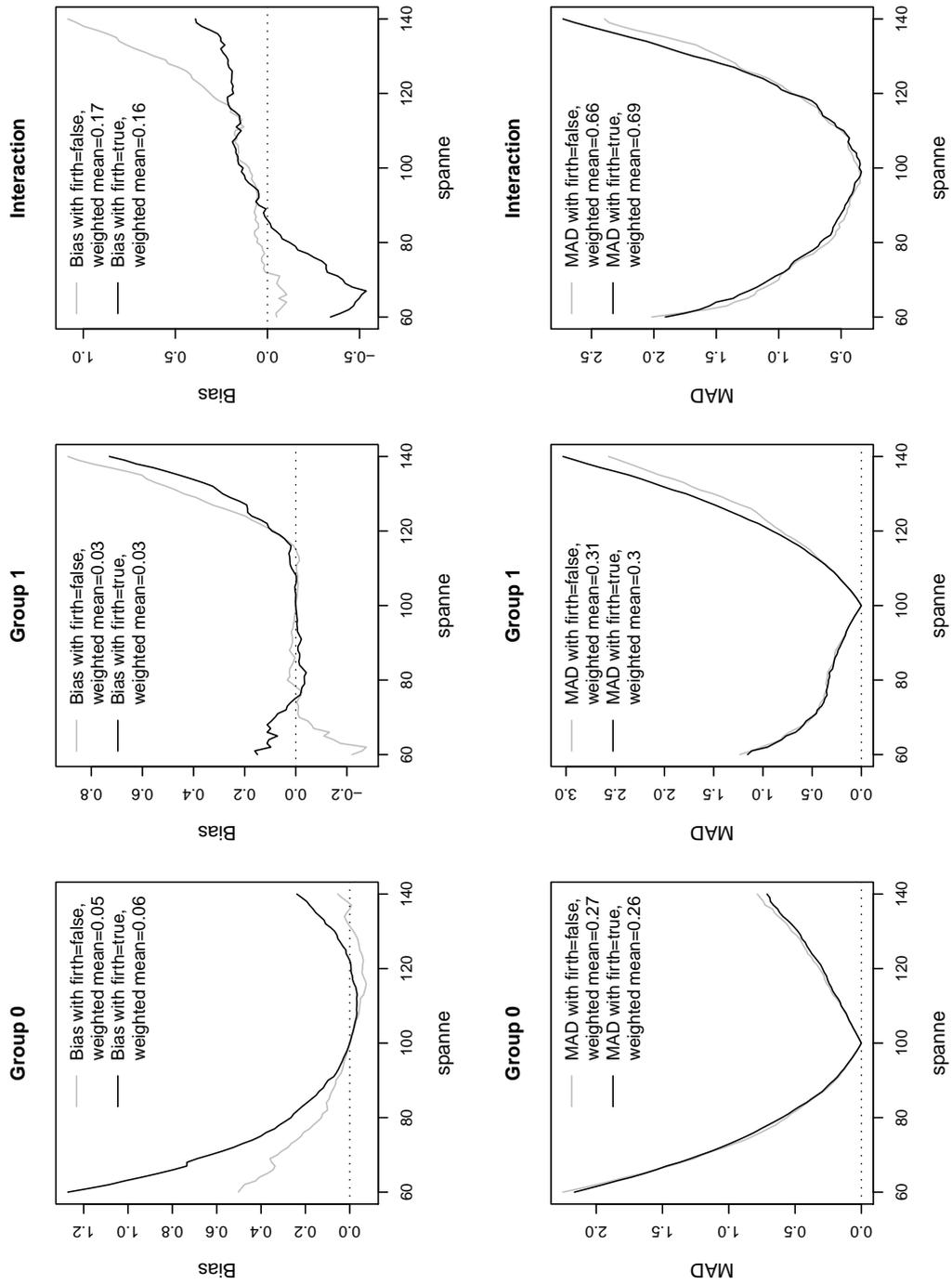


Abbildung 2.85: Abweichungen der Schätzer der Datenzelle 6671

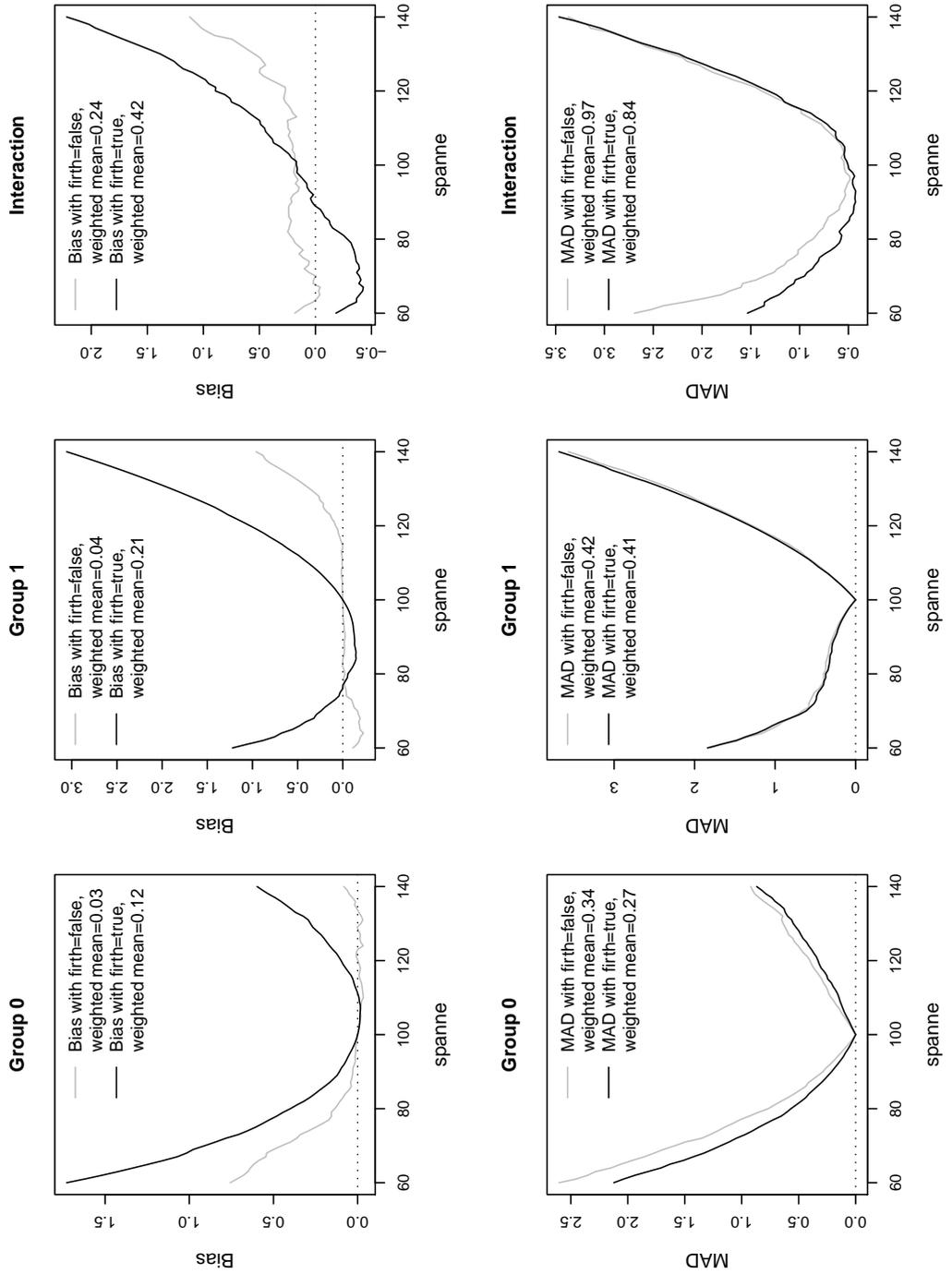


Abbildung 2.86: Abweichungen der Schätzer der Datenzelle 604

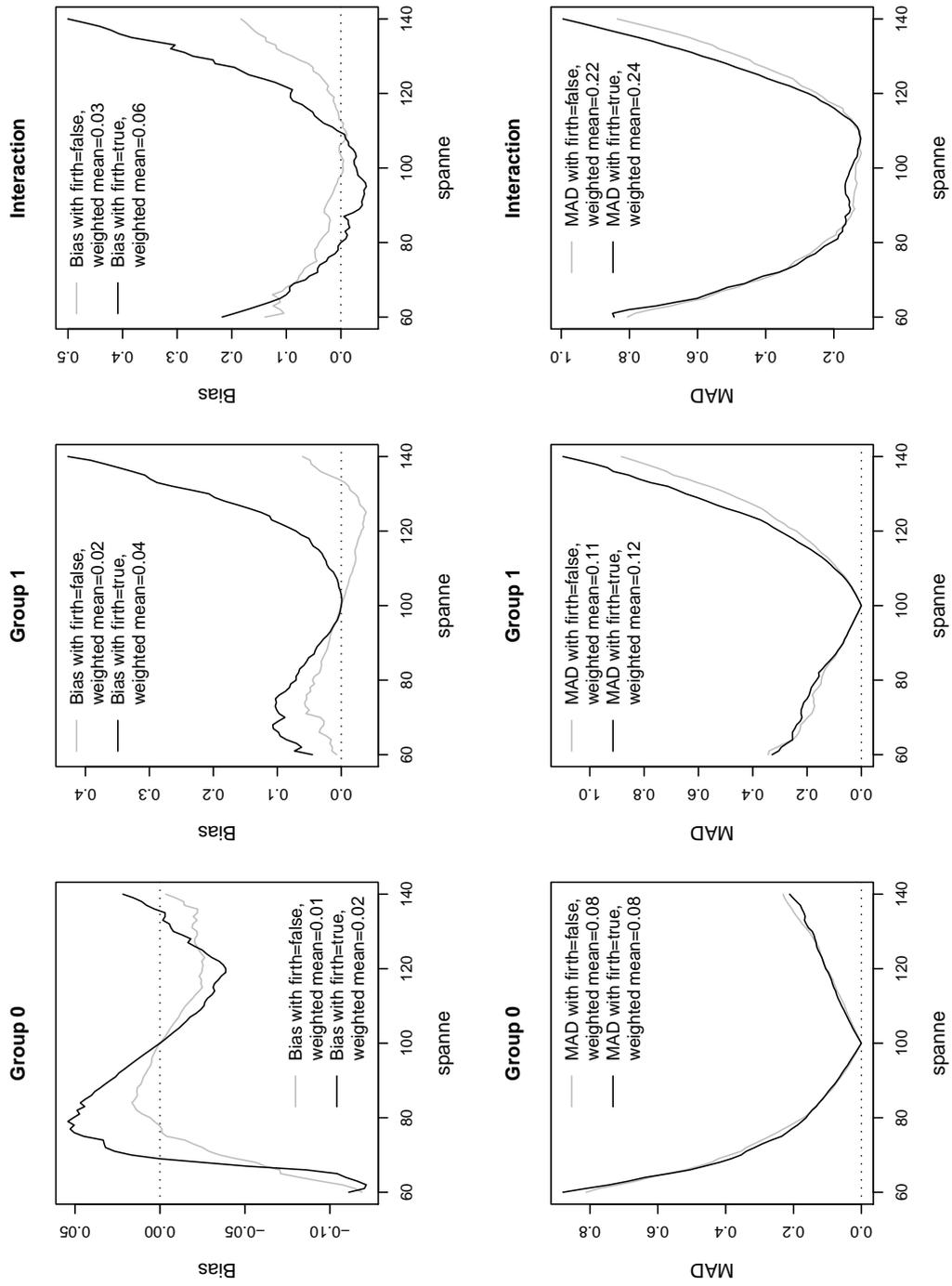


Abbildung 2.87: Abweichungen der Schätzer der Datenzelle 6334

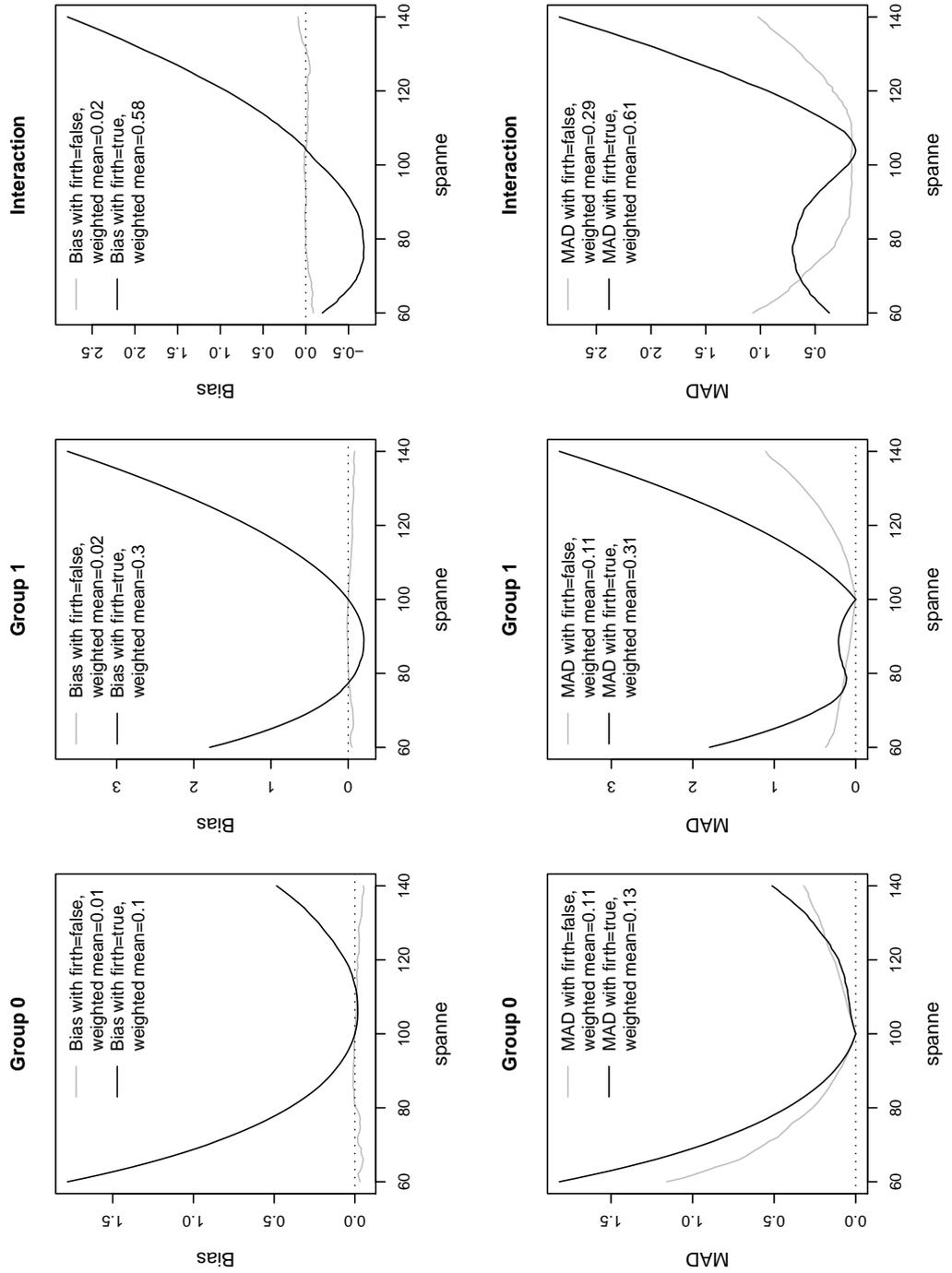


Abbildung 2.88: Abweichungen der Schätzer der Datenzelle 6674

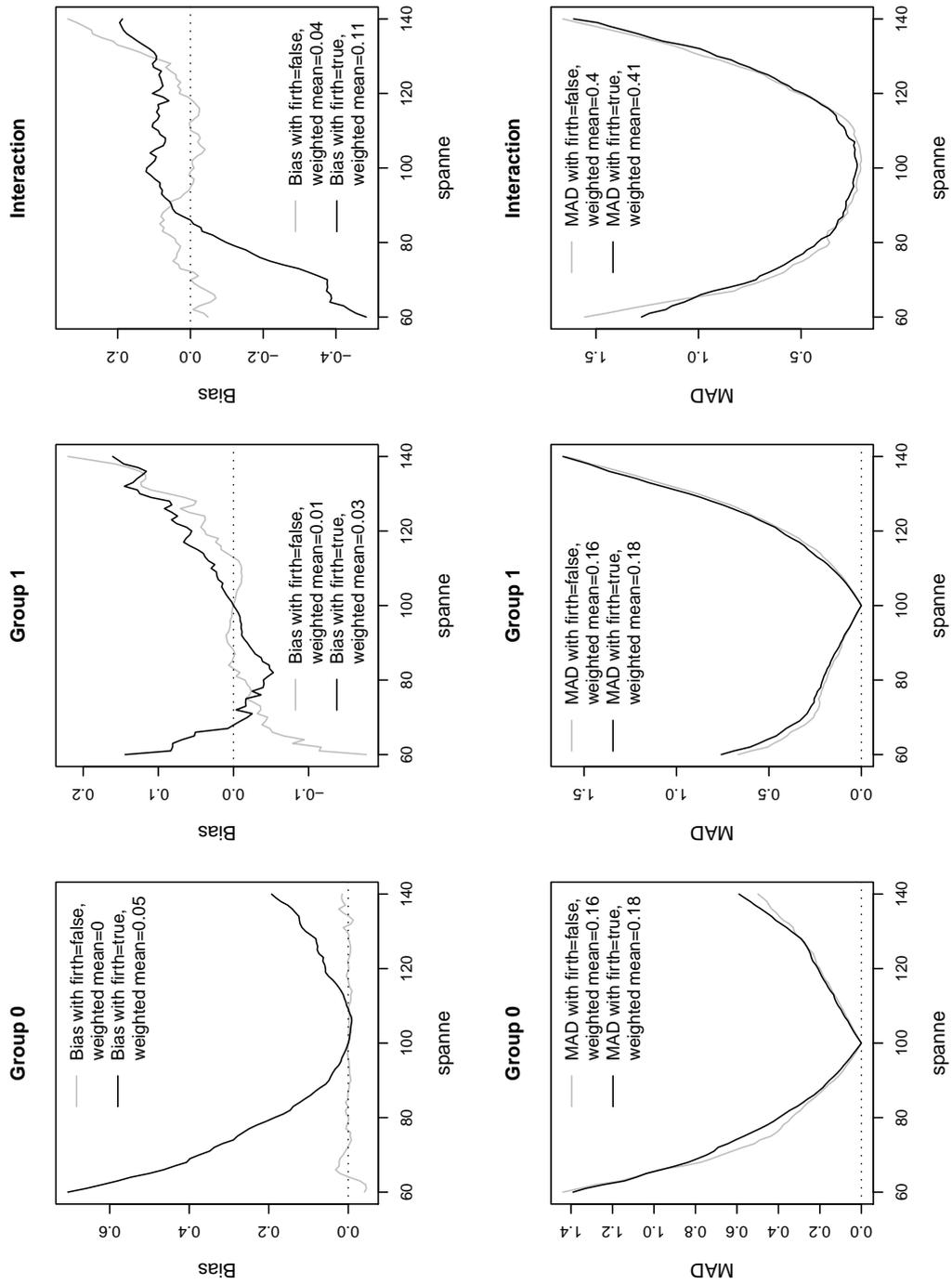


Abbildung 2.89: Abweichungen der Schätzer des fixierten Modells der Zelle 501

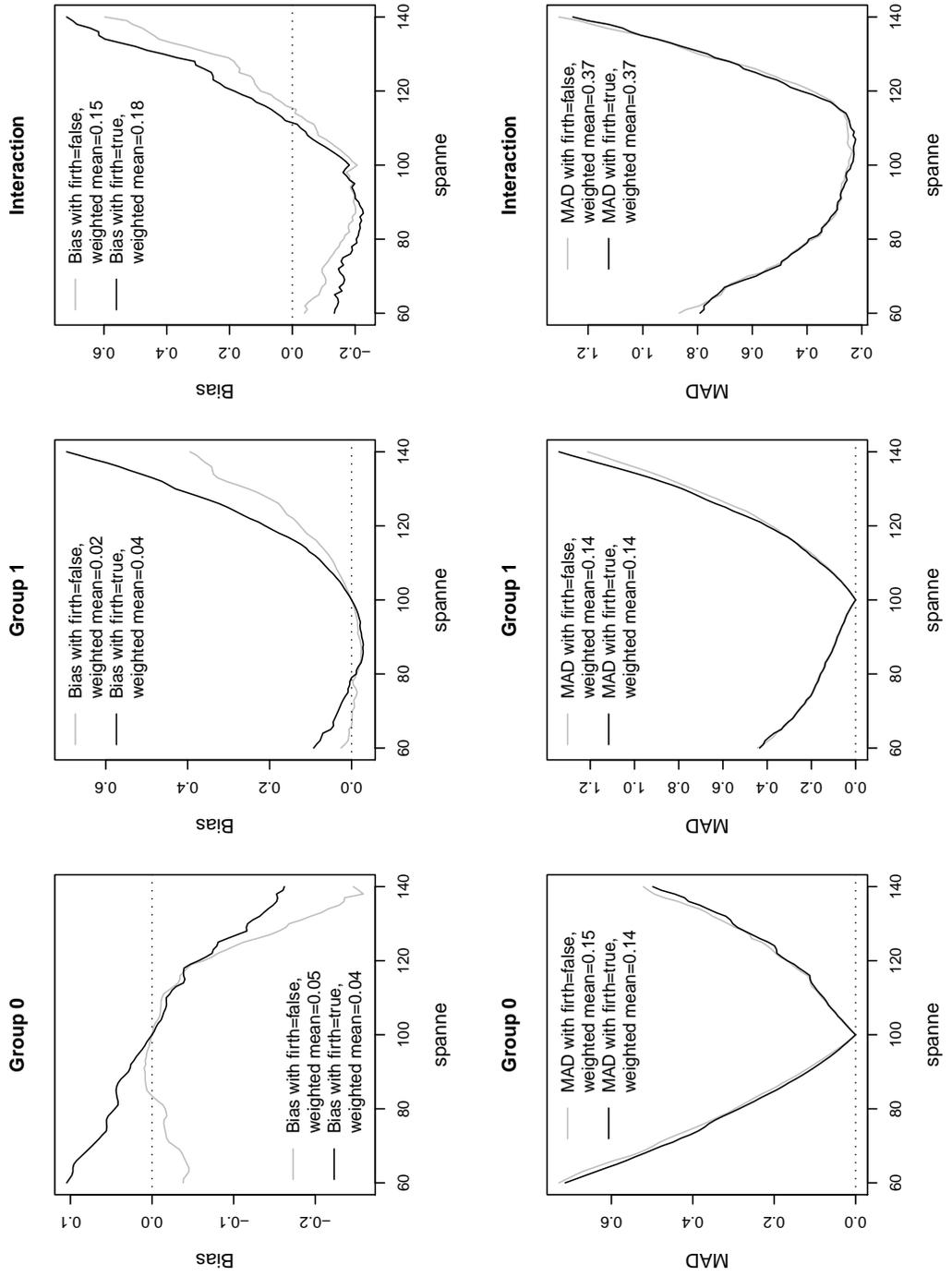


Abbildung 2.90: Abweichungen der Schätzer des fixierten Modells der Zelle 5331

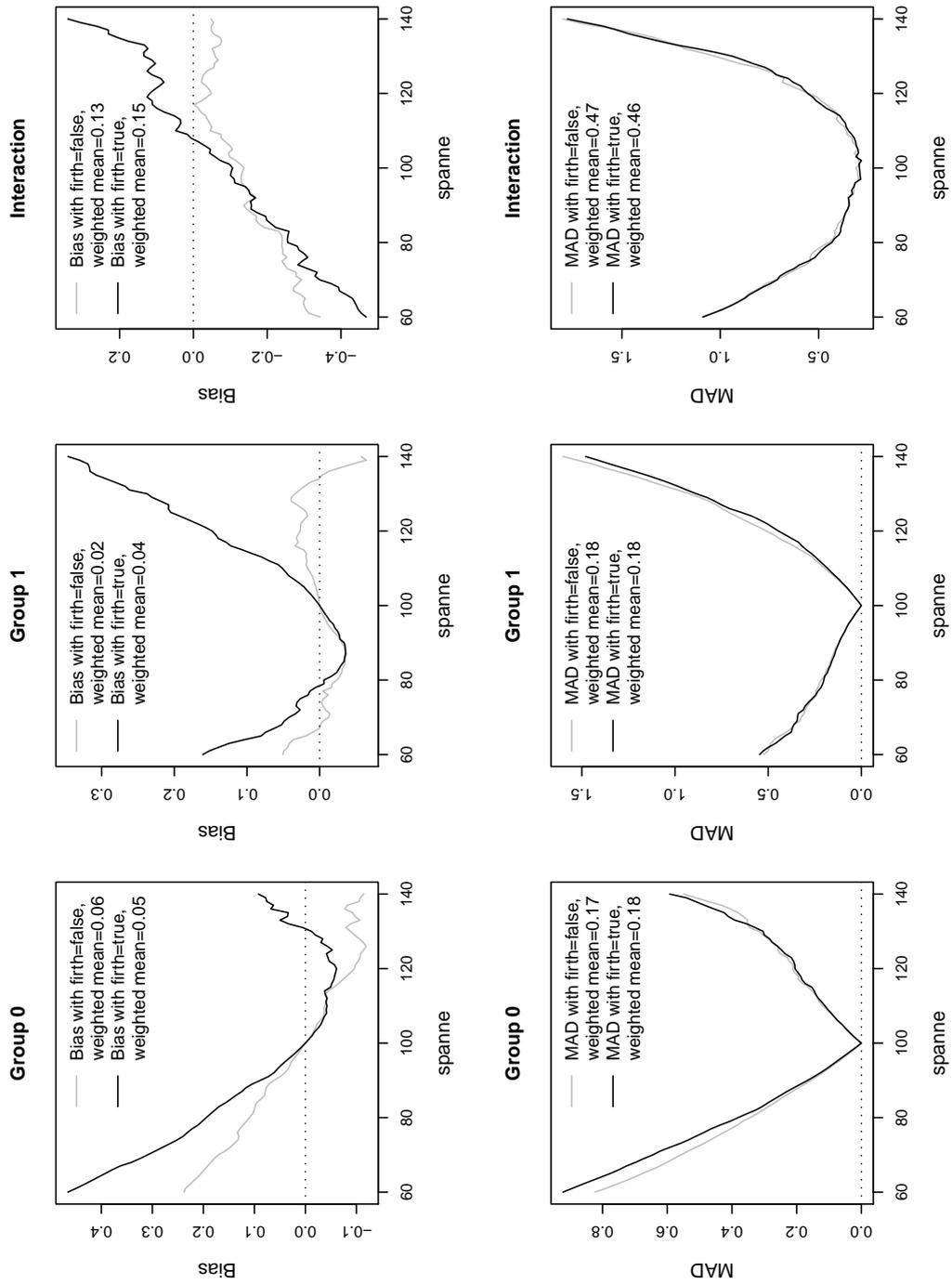


Abbildung 2.91: Abweichungen der Schätzer des fixierten Modells der Zelle 5671

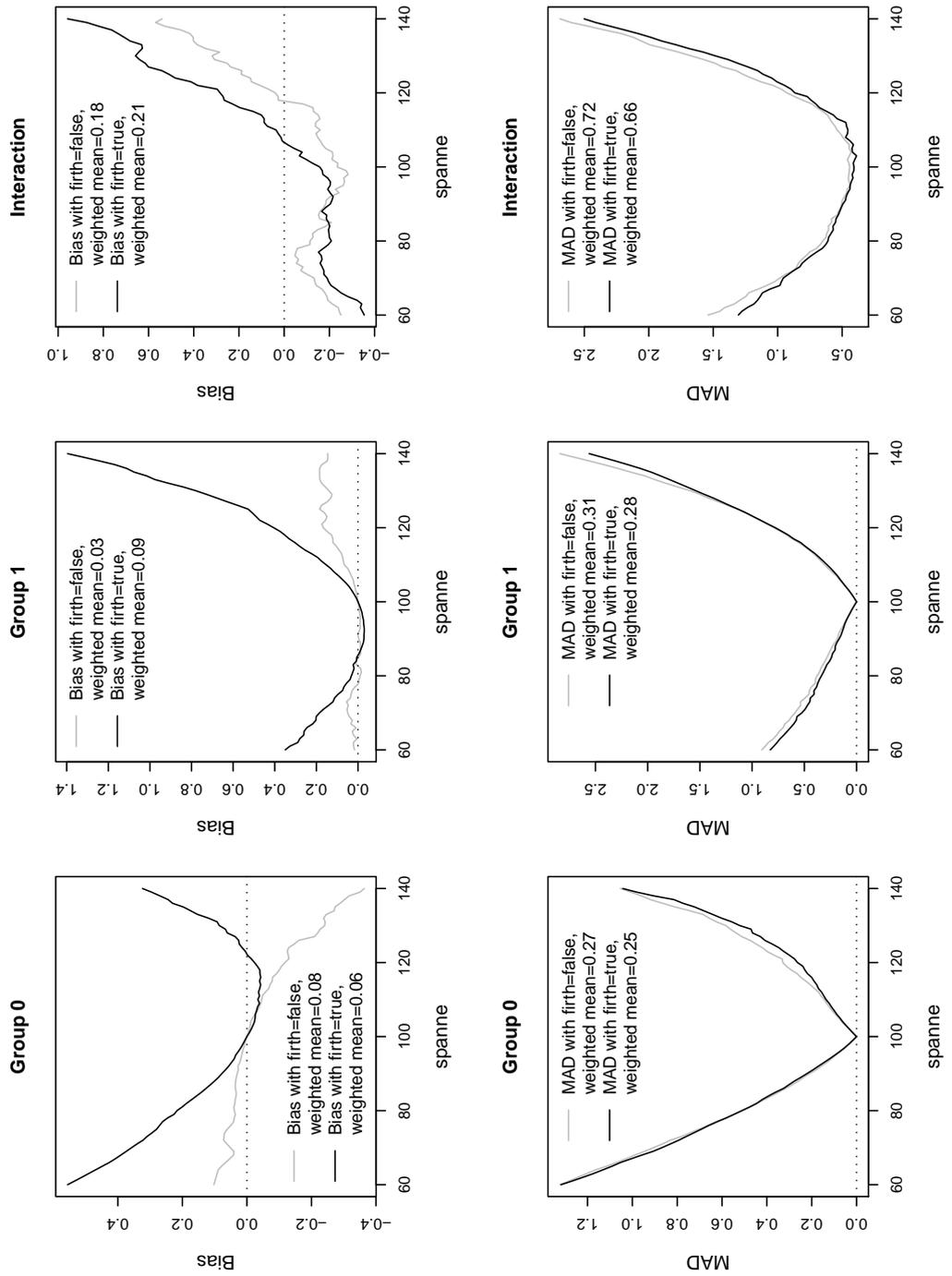


Abbildung 2.92: Abweichungen der Schätzer des fixierten Modells der Zelle 504

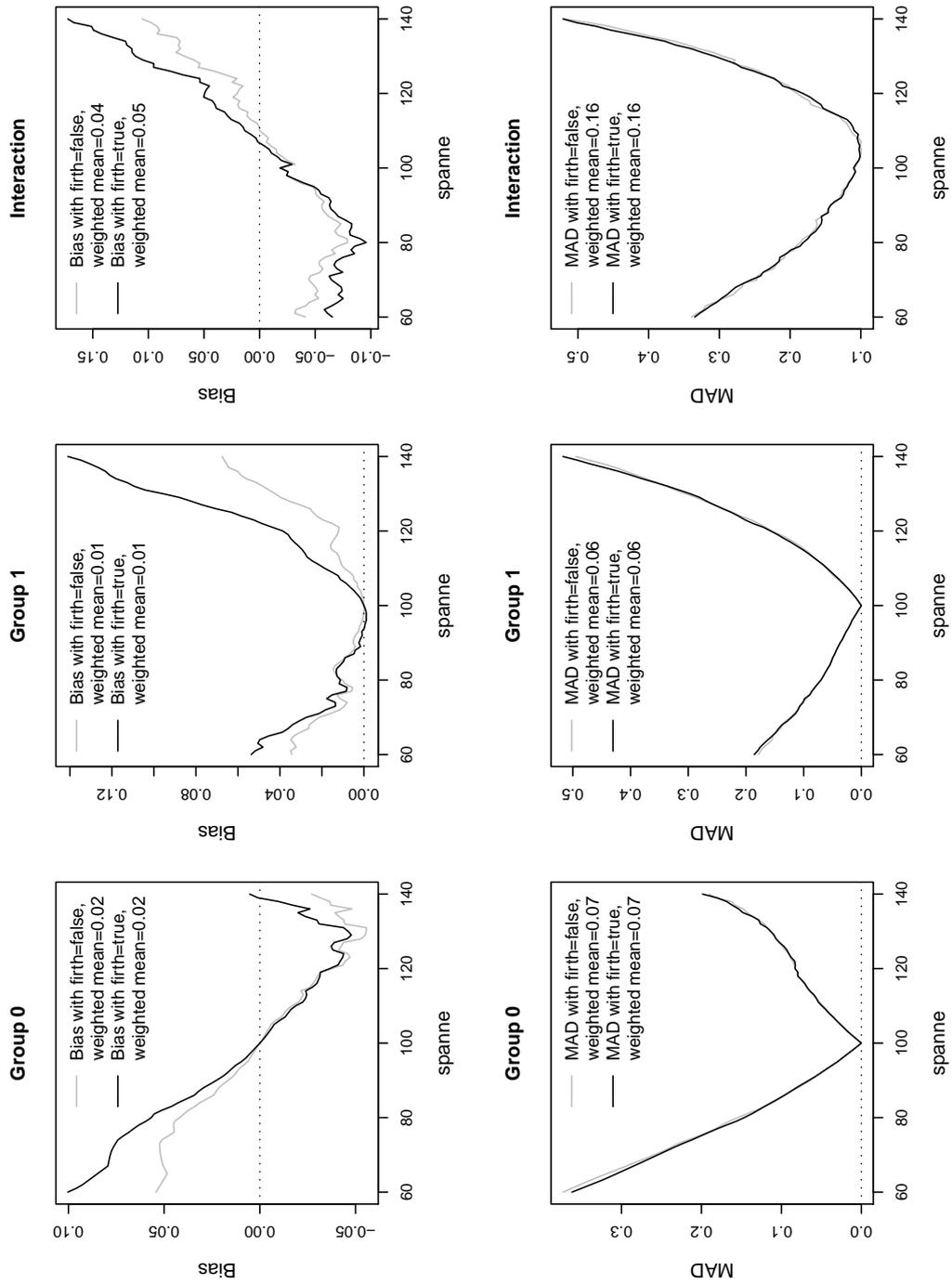


Abbildung 2.93: Abweichungen der Schätzer des fixierten Modells der Zelle 5334

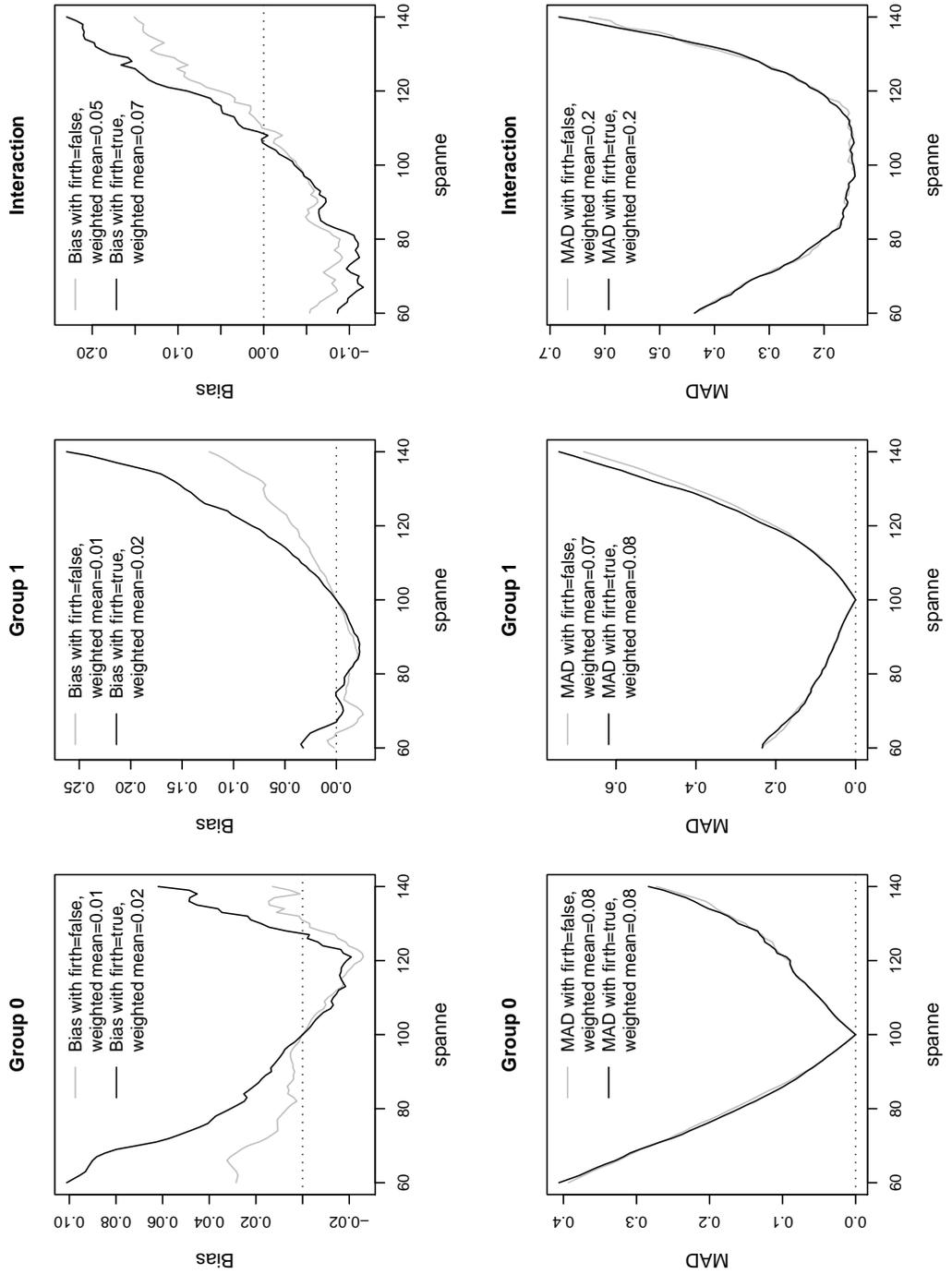


Abbildung 2.94: Abweichungen der Schätzer des fixierten Modells der Zelle 5674

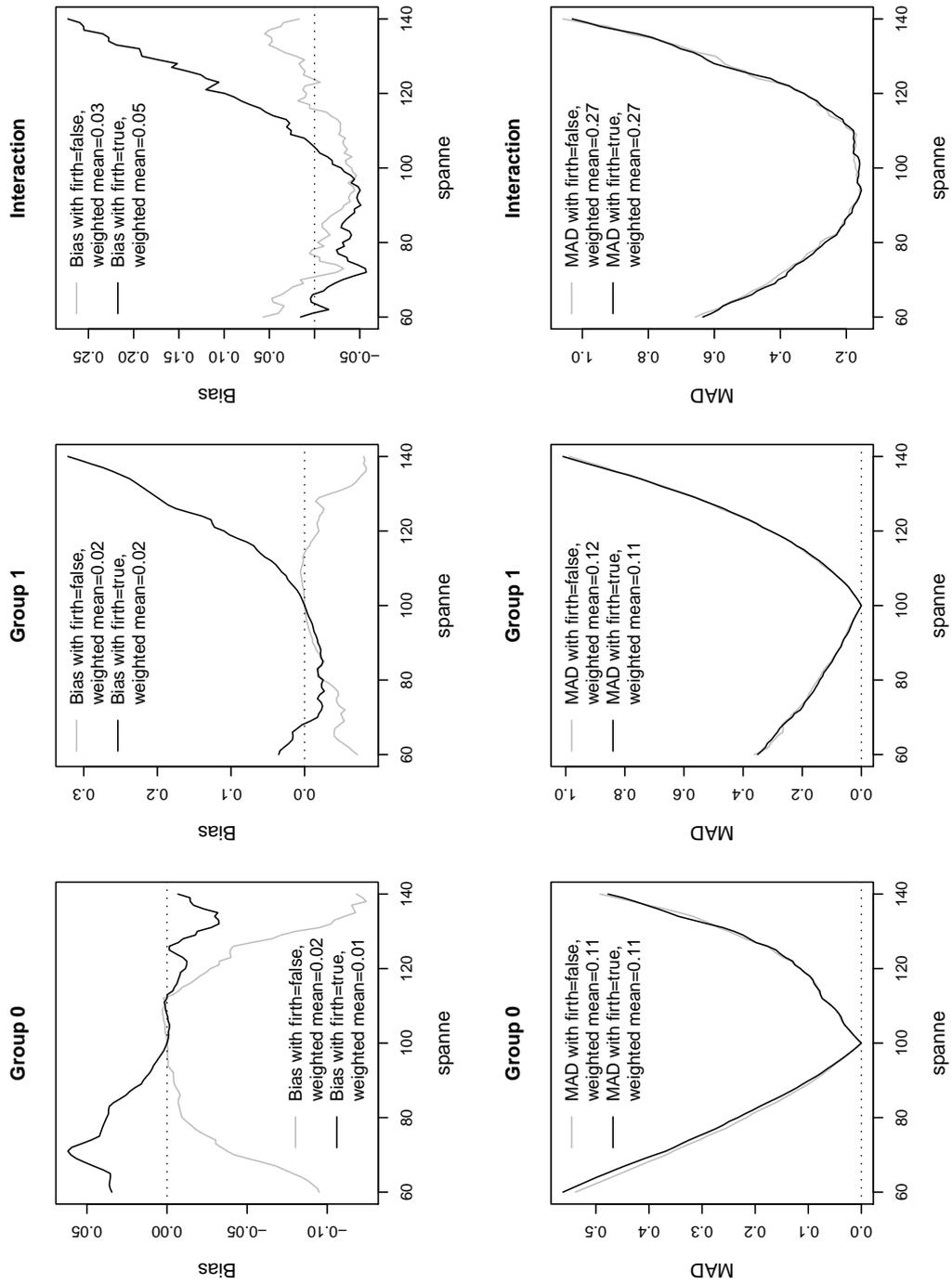


Abbildung 2.95: Abweichungen der Schätzer des fixierten Modells der Zelle 601

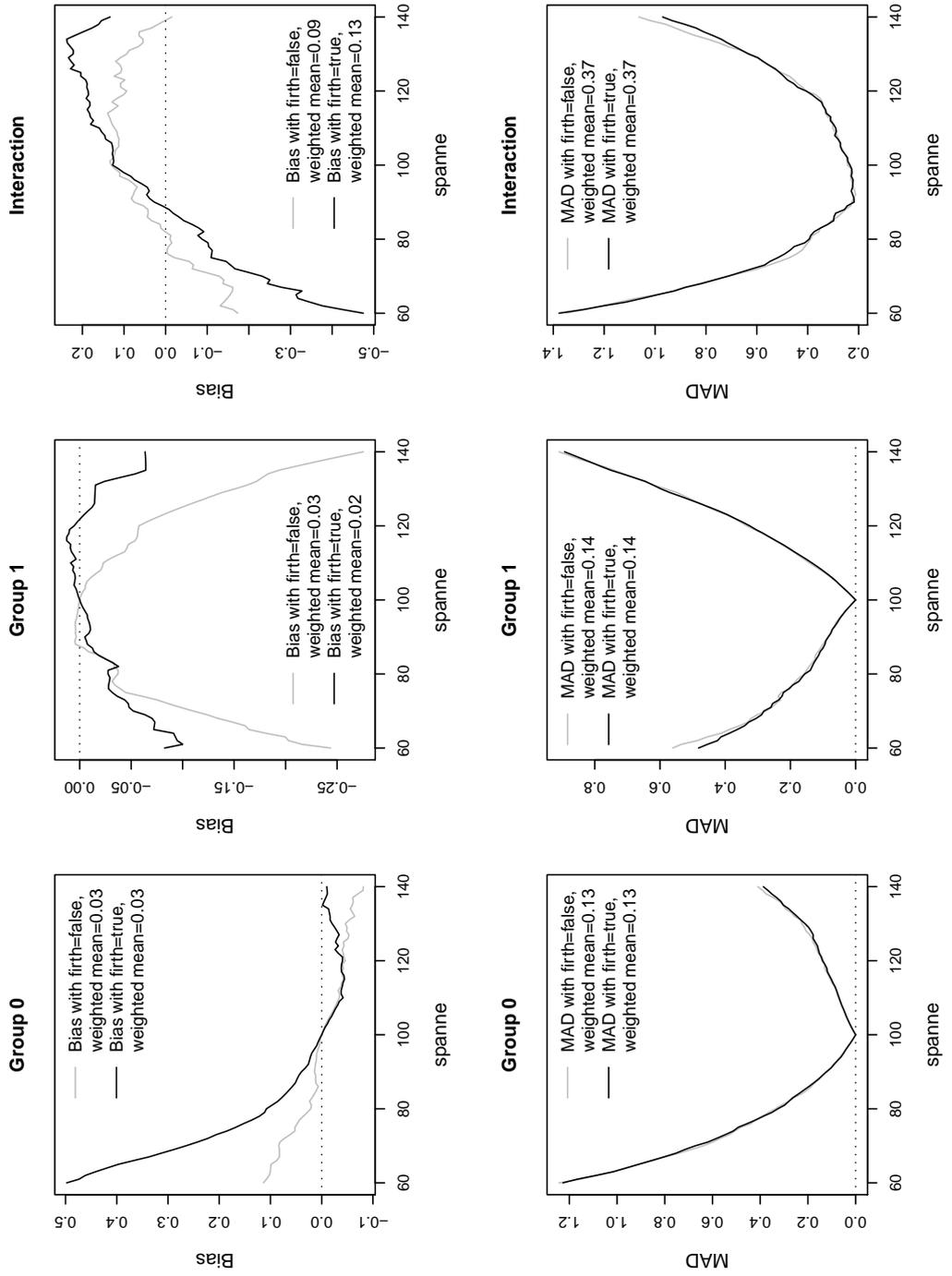


Abbildung 2.96: Abweichungen der Schätzer des fixierten Modells der Zelle 6331

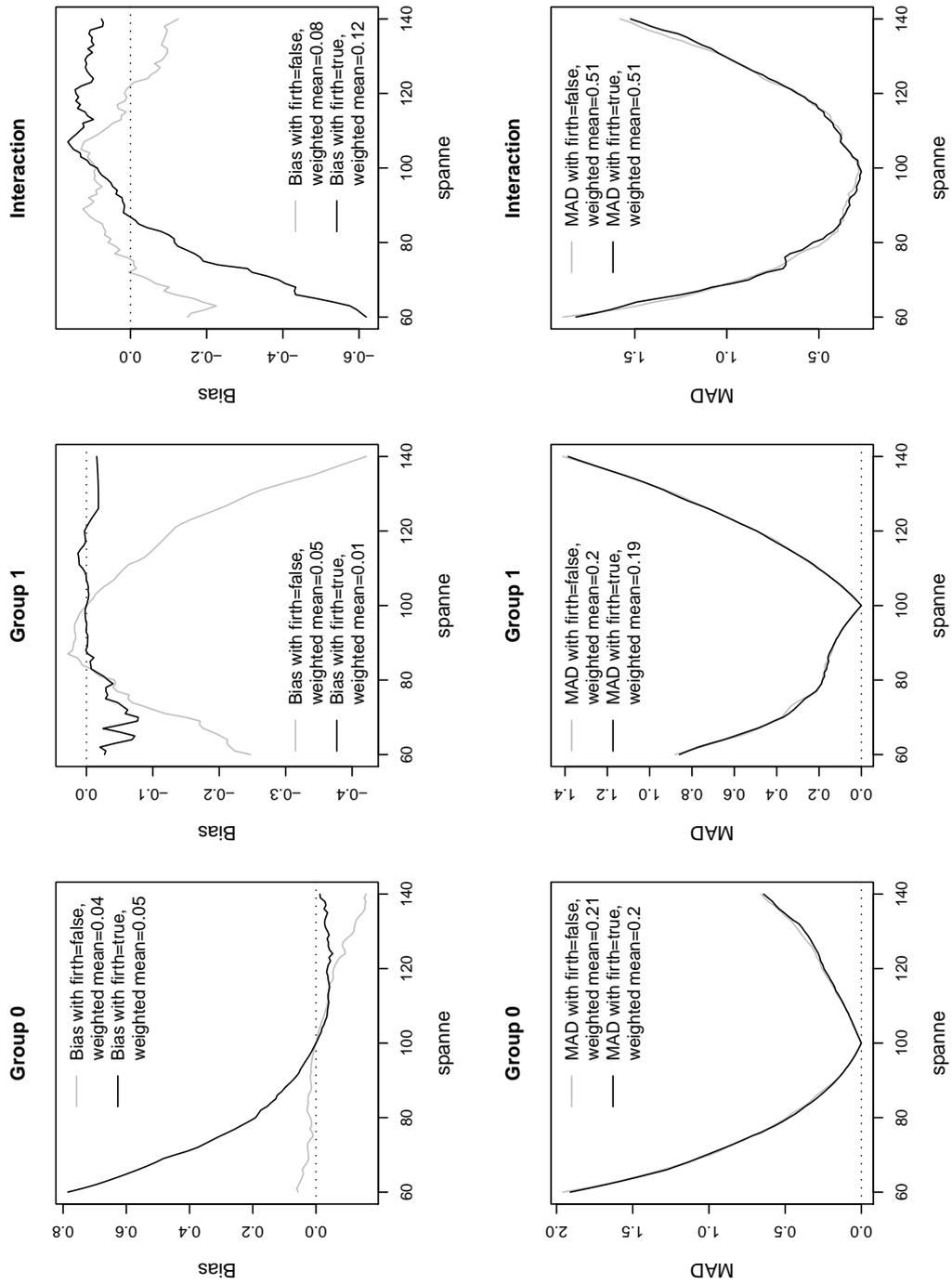


Abbildung 2.97: Abweichungen der Schätzer des fixierten Modells der Zelle 6671

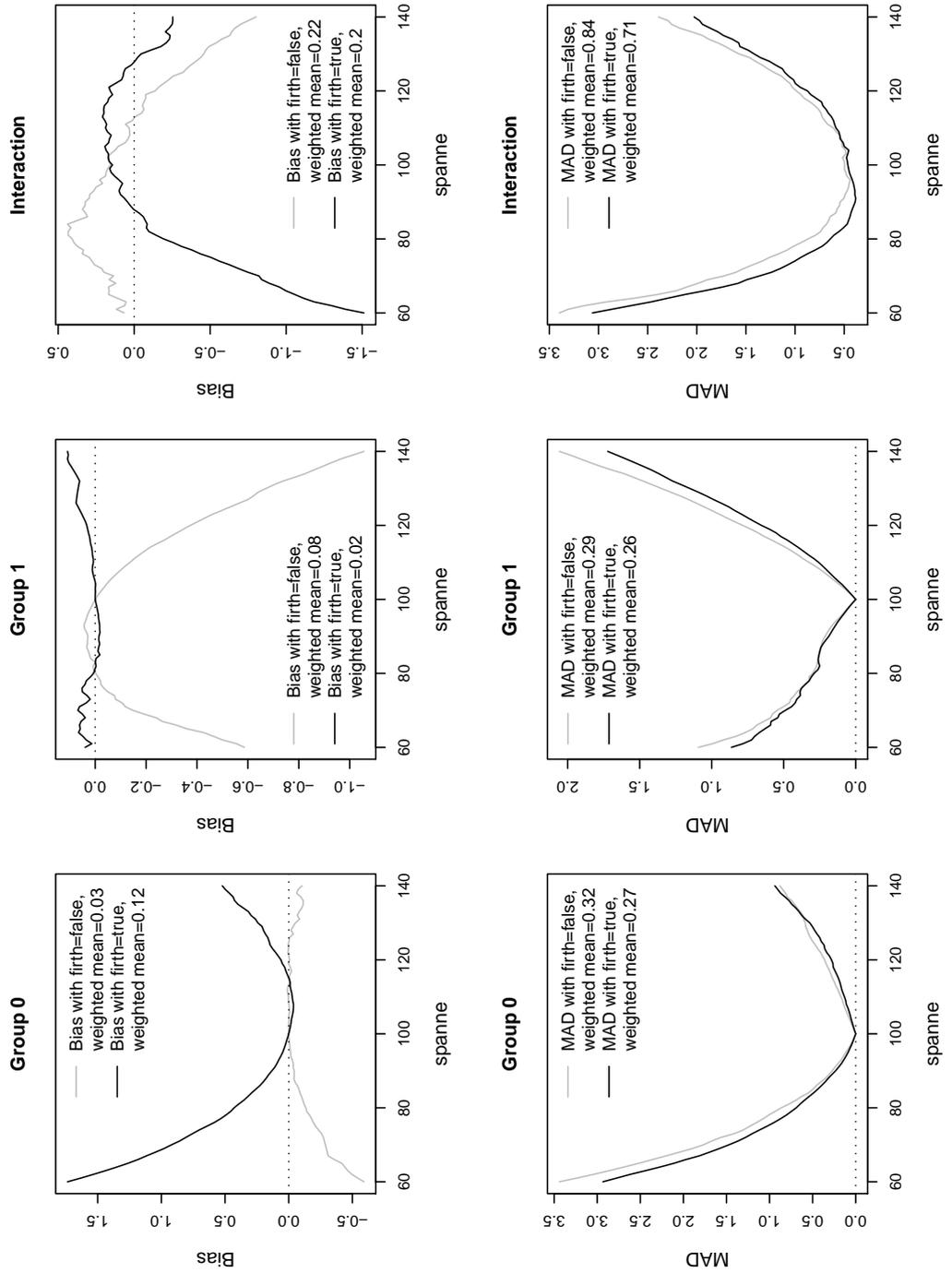


Abbildung 2.98: Abweichungen der Schätzer des fixierten Modells der Zelle 604

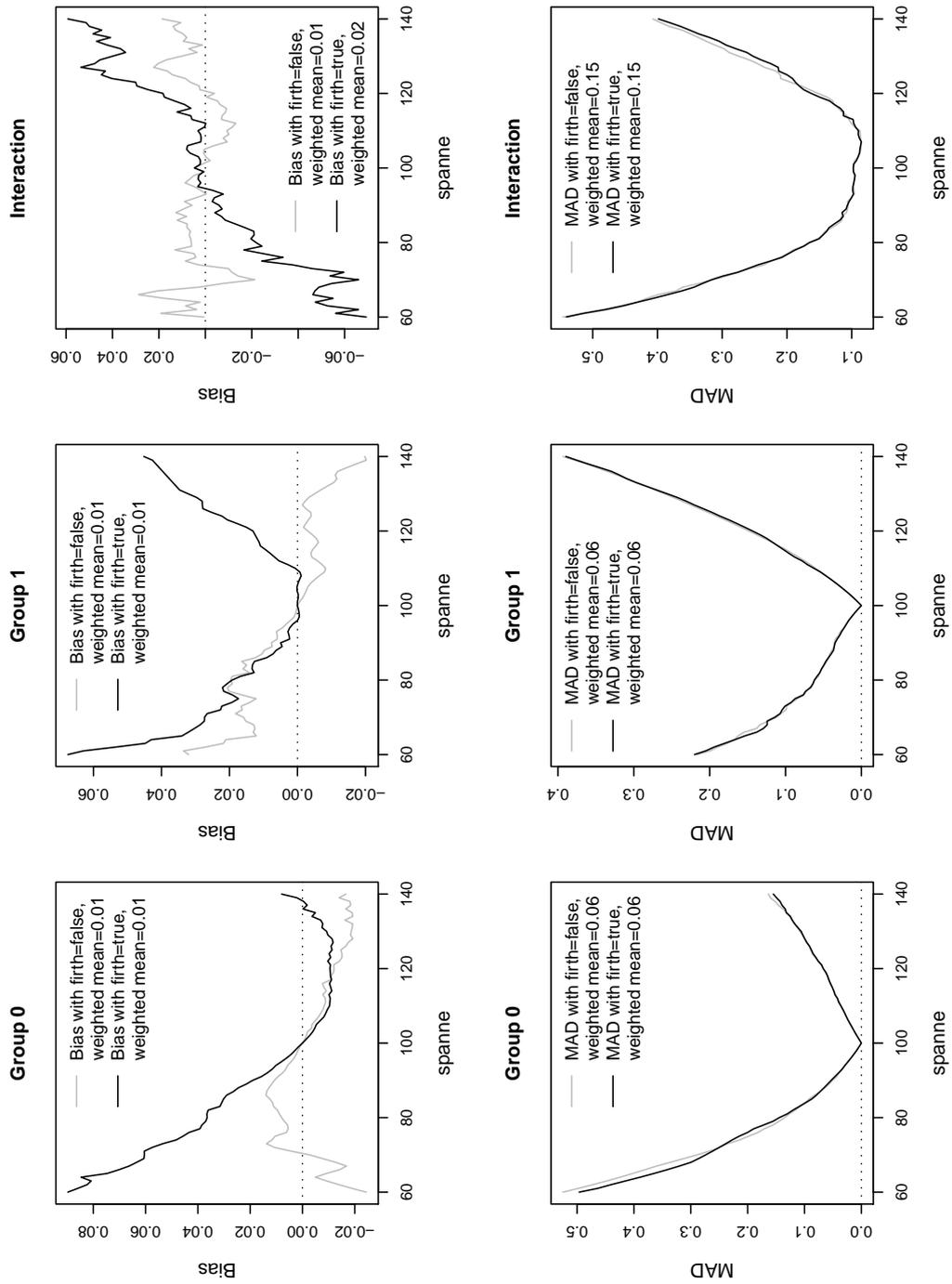


Abbildung 2.99: Abweichungen der Schätzer des fixierten Modells der Zelle 6334

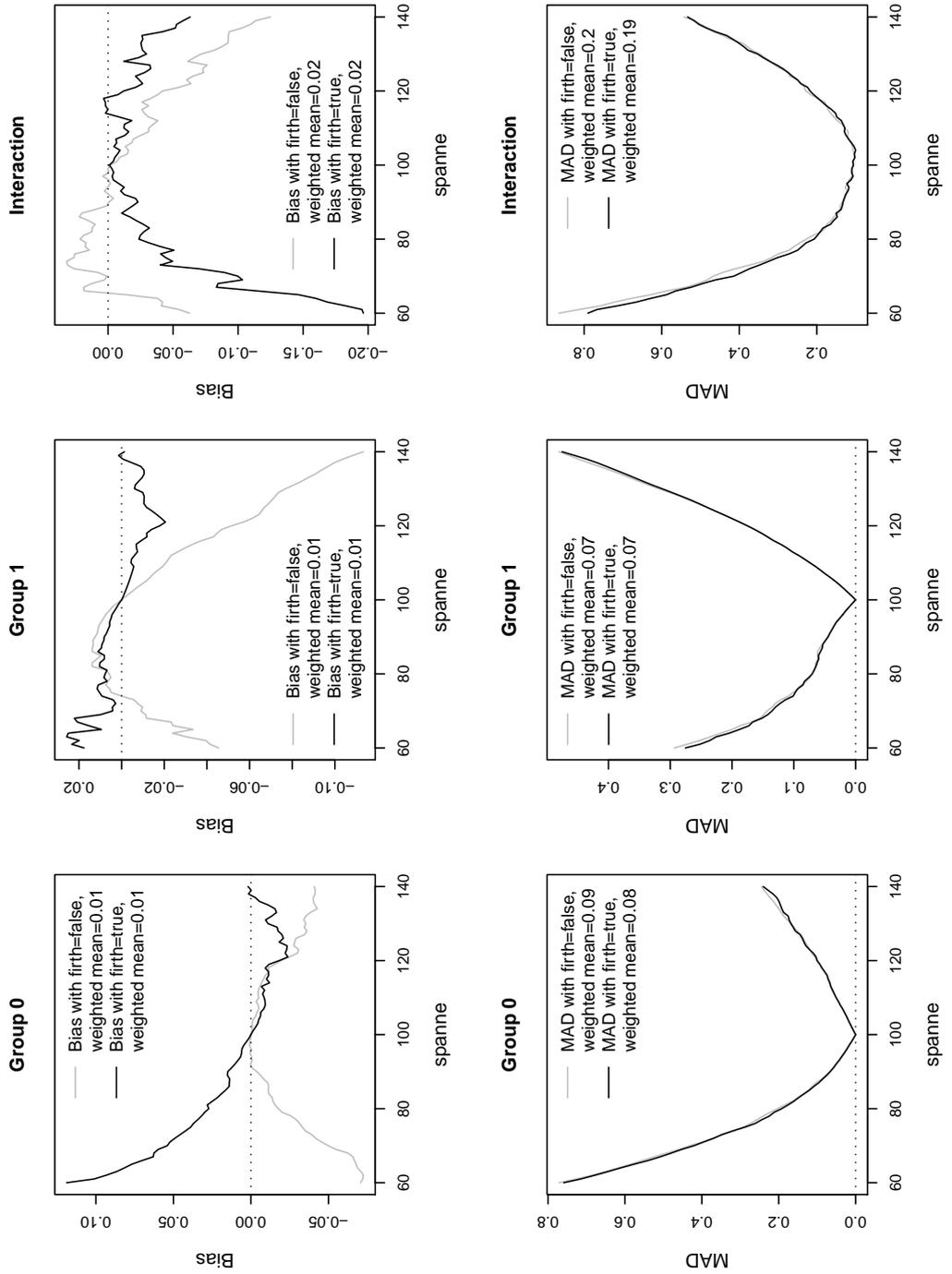
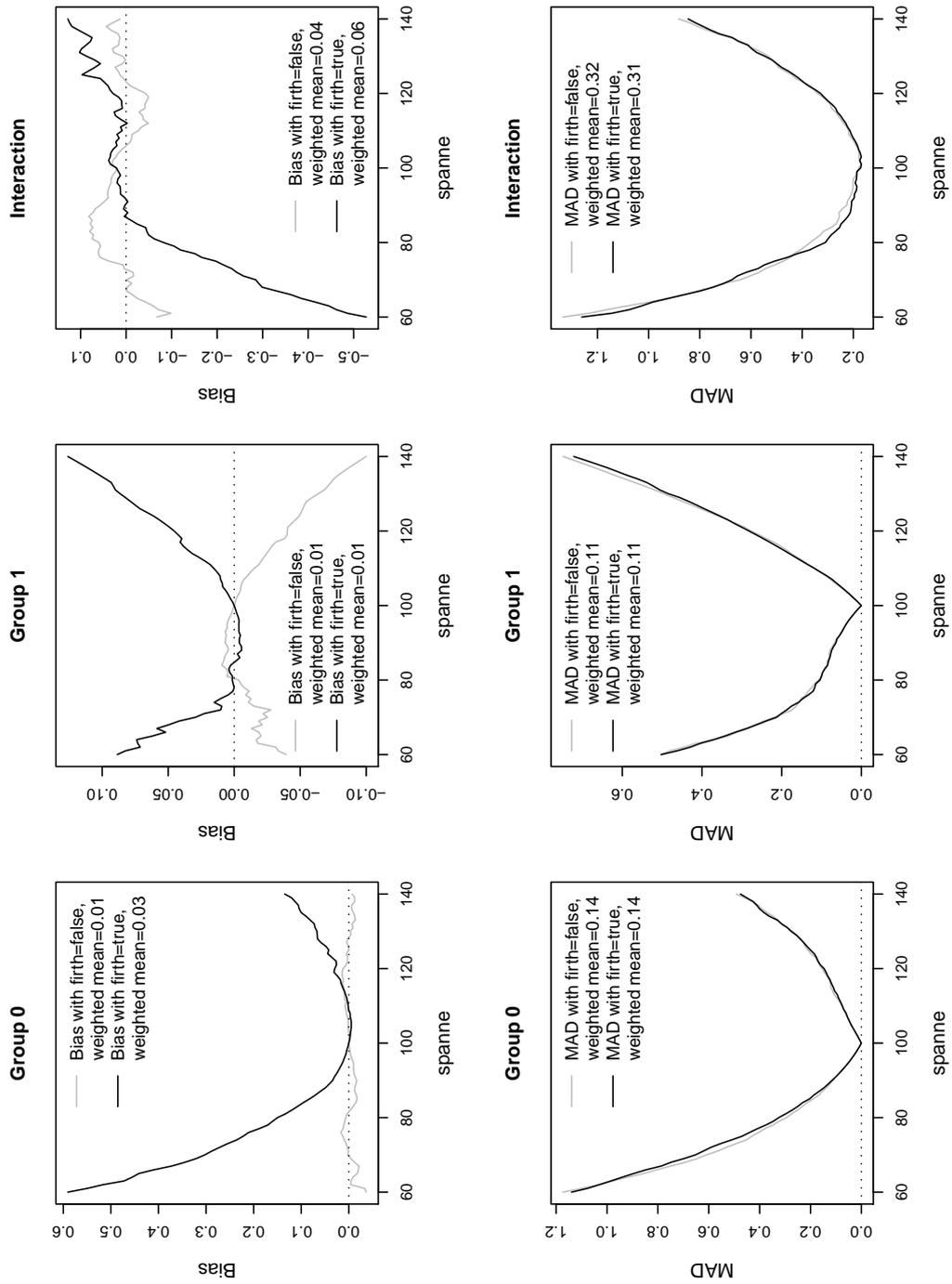


Abbildung 2.100: Abweichungen der Schätzer des fixierten Modells der Zelle 6674



In den Berechnungen mit $\text{Alpha}=\text{c}(1,1,1)$ sind die zuvor gewonnenen Erkenntnisse aus den Berechnungen mit $\text{Alpha}=\text{c}(1,0.157,0.157)$ teilweise nicht wiederzufinden, wie man in den Abbildungen (siehe elektronischen Anhang) sehen kann. Es ist aber immer noch zu erkennen, dass Datenzellen, welche Datensätze mit einer Stichprobengröße von 400 beinhalten, im Allgemeinen von der Maximum-Likelihood-Methode besser abgebildet werden können, während Datenzellen aus Datensätzen mit einer Stichprobengröße von 100 aufweisen eher kleinere Abweichungen bei der Firth-Methode liefern.

2.8 Interpretation

2.8.1 Die Punktschätzer

Zu den Punktschätzern lässt sich abschließend sagen, dass die Stichprobengröße der Datensätze natürlich einen wesentlichen Einfluss auf die Varianz der Schätzer hat, d.h. dass bei einer höheren Stichprobenanzahl die Varianz geringer ist. Die geringere Varianz hat aber keinerlei Einfluss auf die Güte der Anpassung, diese hängt wesentlich von der Komplexität der Funktionen der festgelegten Effekte ab. Grafisch ist deutlich ersichtlich, dass jene wahren Funktionen, die ausschließlich aus zwei Polynomtermen mit Potenzen aus dem Set S der Fraktionalen Polynome bestehen, besser abgebildet werden können als jene wahren Funktionen, welche mehr als zwei Terme beinhalten (wie die Szenarien 1 bis 4). Dies bedeutet, dass komplexere Funktionen schlechter von den FP angepasst werden können als einfache Funktionen. (Wobei unter *einfachen* Funktionen hier jene Funktionen verstanden werden, die aus zwei Termen mit Potenzen aus S bestehen.) In der Praxis weiß man jedoch nicht, wie die wahre zugrundeliegende Funktion aussieht, daher weiß man auch nicht, wie gut die Anpassung tatsächlich ist.

Der Vergleich der fixierten Modelle mit den Selektionsmodellen macht weiters deutlich, dass die fixierten Modelle nicht nur eine bessere Anpassung an die Originaleffekte liefern, sondern auch numerisch leichter lösbar sind, wie man an dem Konvergenzverhalten der Schätzer sehen kann. Daraus lässt sich schließen, dass die Suche nach dem bestmöglichen FP durch Parameteränderungen noch optimiert werden kann.

Ein Einfluss des Zensierungsanteils bei dieser Simulationsstudie wurde grafisch deutlich sichtbar in den Datenzellen 504, 5334, 5674, 604, 6334 und 6674 (Abbildung 2.32, 2.33, 2.34, 2.38, 2.39 und 2.40) der fixierten Modelle, während dieser Einfluss in allen anderen Datenzellen nur schwer bis kaum festgestellt werden konnte. Da gerade in den zuvor erwähnten Datenzellen die Anpassung perfekt ist, lässt sich sagen, dass der Effekt des Zensierungsanteils

in den anderen Datenzellen aufgrund der schlechteren Anpassung nicht so deutlich sichtbar ist. Dem Anwender in der Praxis muss bewusst sein, dass ein größerer Zensierungsanteil auch die Varianz seiner Schätzer erhöht.

2.8.2 Die Abweichungen der Schätzer

Es ist eine schwache Tendenz bezüglich der Stichprobengröße erkennbar, die darauf schließen lässt, dass die Firth-Methode für eine Stichprobengröße von 400 die Schätzer zu stark bestraft, d.h. gegen Null drängt. Eine Überlegung wäre, den Bestrafungsterm zu ändern, indem man die Penalisierung abschwächt.

Da die Firth-Methode besser beim MAD abschneidet und dieser die Variabilität der absoluten Abweichungen bewertet, kann man daraus folgern, dass die Firth-Methode die effizientere Methode ist. Im Gegensatz dazu liefert die ML-Methode die besseren Ergebnisse für den medianen Bias, wobei dieser nur die systematische Verzerrung berücksichtigt.

Der eindeutige Trend des Zensierungsanteils für den MAD kommt daher, dass die Streuung der Punktschätzer mit steigendem Zensierungsanteil zunimmt. Wegen der zugrundeliegenden Formel des MAD, die die absoluten Abweichungen beinhaltet, kommt es bei einem Anstieg von zensierten Beobachtungen zu größeren mittleren Abweichungen (MAD). Beim medianen Bias kann dieser Trend nicht gefunden werden, da sich die Abweichungen, wenn sie halbwegs symmetrisch sind, ausgleichen. Werden die Abweichungen größer, und zwar in beide Richtungen, so wirkt sich dies folglich nicht auf den medianen Bias aus. Da jedoch die besprochenen Unterschiede zwischen der Firth-Methode und der Maximum-Likelihood-Methode so minimal sind, macht es für die Anwendung in der Praxis bezüglich der Abweichungen der Schätzer keinen gravierenden Unterschied, für welche Methode man sich letztlich entscheidet.

Kapitel 3

Fallstudie zur Venenthrombose

3.1 Einleitung

Bei der Venenthrombose handelt es sich um eine chronische Gefäßerkrankung, bei der sich Blutgerinnsel in den Venen bilden. In 5% bis 9% der Fälle enden wiederkehrende Venenthrombosen tödlich (vgl. Eischer et al., 2011). Daher wurde eine große prospektive multizentrische Studie namens „Austrian Study on Recurrent Venous Thromboembolism“ (AUREC) durchgeführt. Aus dieser permanent laufenden Studie wird in Folge ein Datensatz betrachtet, dessen zugehörige Patienten in der Zeit von Juli 1992 bis August 2008 in vier Wiener Krankenhäusern rekrutiert wurden.

Das Ziel dieser Fallstudie war es herauszufinden, welchen Einfluß der Hämatokritwert eines Patienten auf die Wiederkehr einer Venenthrombose hat, und ob dieser Einfluß den Unterschied der Wiederkehr von Venenthrombosen bei Männern und Frauen erklärt. Unter Hämatokrit (Abkürzung: HCT) versteht man in der Medizin den Anteil am Blutvolumen, welcher aus zellulären Bestandteilen besteht. Es ist bereits bekannt, dass Männer ein höheres Risiko für eine wiederkehrende Venenthrombose haben als Frauen und dass sich die Geschlechter bezüglich des Hämatokritlevels unterscheiden.

In die Studie wurden ausschließlich Personen eingeschlossen, die älter als 18 Jahre waren, eine symptomatische Venenthrombose hatten und Antikoagulantien in einem Zeitraum von 3 bis 18 Monaten erhielten. Der Eintrittszeitpunkt für die Studie war mit der Einstellung der Antikoagulantien festgelegt. Der Studienendpunkt war mit dem wiederholten Auftreten einer tiefen Venenthrombose oder mit dem Auftreten einer Lungenembolie definiert. (vgl. Eischer et al., 2011)

3.2 Der HCT Datensatz

Der *HCT-Datensatz* (benannt nach dem hier interessierenden Hämatokrit) beinhaltet die Daten von 653 Patienten, wobei aufgrund fehlender Werte 34 Patienten aus der Datenanalyse ausgeschlossen werden. Die restlichen 619 Personen weisen einen Zensierungsanteil von 76,6% und eine mediane Nachbeobachtungszeit (vgl. Schemper u. Smith, 1996) von 46,43 Monaten auf. Bei den 5 Kovariablen (Tabelle 3.1) Geschlecht („*Sex*“), Raucherstatus („*Smoking*“), Hämatokrit („*HCT*“), Alter („*Age*“) und Body Mass Index („*BMI*“) handelt es sich ausschließlich um dichotome und kontinuierliche Variablen.

Tabelle 3.1: Kodierung der Variablen

Variable	Variablenlabel	Erläuterung
Sex	Geschlecht	0 männlich 1 weiblich
Smoking	Raucherstatus	0 Nichtraucher 1 Raucher
HCT	Hämatokritwert	
Age	Alter	zum Zeitpunkt der Venenthrombose
BMI	Body Mass Index	
censur survival	Zensierungsindikator Beobachtungszeitraum	0 zensiert 1 Rezidiv in Monaten

3.2.1 Univariate Betrachtung der HCT Daten

Um einen Überblick über den Datensatz zu bekommen, werden dessen Variablen zu Beginn der Analyse univariat betrachtet.

Histogramme

Für die Betrachtung der Verteilungen der einzelnen Variablen werden Histogramme herangezogen. Wie man in Abbildung 3.1 sehen kann, sind die Variablen *Age*, *HCT* und *BMI* approximativ normalverteilt. Die Variable *BMI* könnte durch Logarithmieren symmetrischer gemacht werden, jedoch würde sich dadurch die Interpretation des *BMI* erschweren, daher wird dies unterlassen.

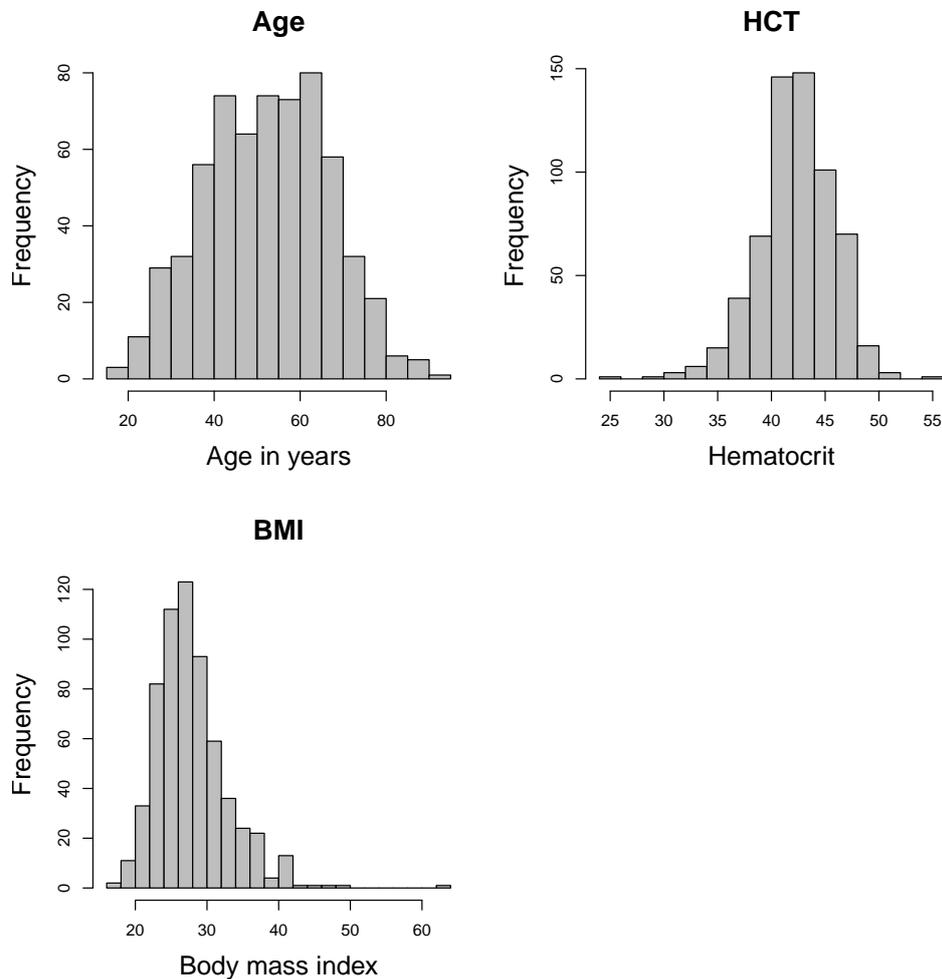


Abbildung 3.1: Verteilung der Variablen

Statistische Kennzahlen

Die Tabelle 3.2 enthält wichtige statistische Kennzahlen für alle Variablen, ob kontinuierlich oder dichotom.

Soll herausgefunden werden, ob der Raucherstatus bei Männern und Frauen unterschiedlich ist, so berechnet man einen χ^2 -Test. Der Test liefert ein nicht signifikantes Ergebnis mit einem p-Wert von 0.5593 (siehe Tabelle 3.2), d.h. es kann kein Unterschied im Rauchverhalten von Männern und Frauen angenommen werden. Um zu sehen, ob sich die Variablen *HCT*, *Age* und *BMI* für Männer und Frauen unterschiedlich verhalten, verwenden wir den Wilcoxon-Test. Dieser Test ist ein nichtparametrischer Test, welcher überprüft,

Tabelle 3.2: Die Kennzahlen der Variablen.

Variable	Label	Männer(0)	Frauen(1)	p-Wert
N (%)		406 (65.6)	213 (34.4)	
Smoking	0	281 (69.21)	153 (71.83)	0.5593*
	1	125 (30.79)	60 (28.17)	
HCT		43.7 (42.0,45.7)	40.0 (38.0,41.6)	< 0.0001†
Age		52.2 (41.2,61.4)	54.3 (44.0,66.4)	0.0161†
BMI		27.2 (24.6,29.8)	27.3 (24.2,31.3)	0.6998†

Kategorielle Variablen sind als abs. Häufigkeiten (Anteil in %) angegeben. Kontinuierliche Variablen sind mit Median (25%-Quantile, 75%-Quantile) angegeben. P-Werte basieren auf * χ^2 -Test, † Wilcoxon Rangsummentest.

ob zwei unabhängige Verteilungen (hier die für Männer und Frauen) aus derselben Grundgesamtheit stammen. Durch Betrachtung der p-Werte in der Tabelle 3.2 kann man erkennen, dass sich die Variable *BMI* per Geschlecht nicht unterscheidet, im Gegensatz zu den Variablen *HCT* und *Age*.

3.2.2 Kaplan-Meier-Kurven

Für die grafische Darstellung der Überlebenszeit wird in der Überlebensanalyse der Kaplan-Meier-Plot (vgl. Kaplan u. Meier, 1958) herangezogen. Dabei wird auf der horizontalen Achse die Zeit angegeben und auf der vertikalen Achse die Überlebenswahrscheinlichkeit. Bei der Kaplan-Meier-Kurve selbst handelt es sich im eigentlichen Sinne um eine Stufenfunktion, bei der die Stufe solange auf einem Level verweilt, bis das zu beobachtende Ereignis bei einem Individuum auftritt. In unserer Studie ist dieses Ereignis das Auftreten eines Rezidivs. Folglich erhalten wir unterschiedliche Stufenlängen (Zeitintervalle), d.h. die Ereignisse definieren die Beobachtungsintervalle. Die Stufenhöhe stellt das Rezidivrisiko zu einem bestimmten Zeitpunkt dar. Somit wird die Information, die den zensierten Beobachtungen zugrundeliegt, optimal genutzt. In der linken Grafik der Abbildung 3.2 sehen wir die Kaplan-Meier-Schätzer für unsere 619 Patienten, wobei bei 145 (23,4%) dieser Personen ein Rezidiv auftrat. Aus dem Kaplan-Meier-Plot lassen sich sowohl die kumulierten Rezidivraten als auch Quantile der Rezidivzeitverteilung ablesen, zum Beispiel der Median: dies ist jener Zeitpunkt, an dem die geschätzte kumulative Rezidivrate 50% beträgt. In unserem Plot existiert kein Median, da mehr als 50% der Patienten nach der Beobachtungsdauer noch rezidivfrei sind. Man kann jedoch auch gleichermaßen andere Perzentile wie das 75. und das 25. Perzentil der Rezidivzeitverteilung angeben. In diesem Fall beträgt im

54. Monat die geschätzte kumulierte Rezidivrate 25%. Kaplan-Meier-Kurven

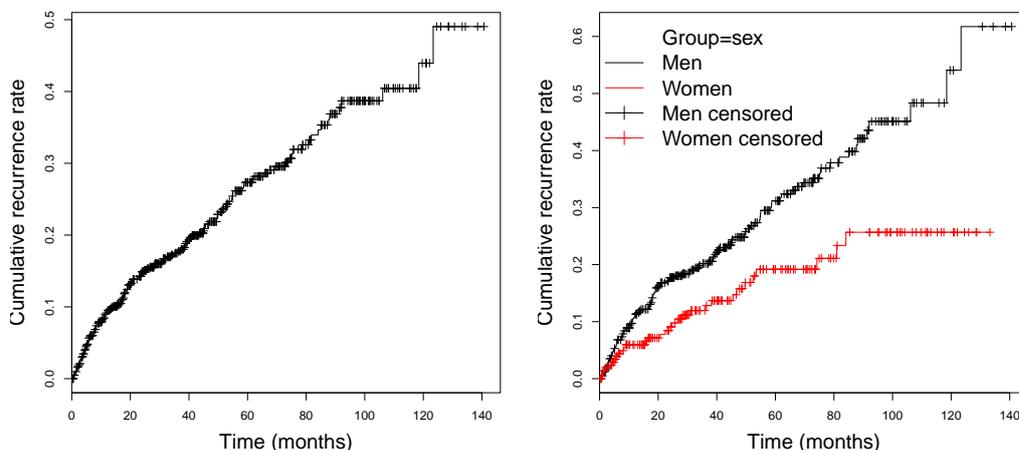


Abbildung 3.2: Kaplan-Meier-Schätzer für kumulierte Rezidivraten

können in weiterer Folge dafür verwendet werden, Gruppen zu vergleichen, wie dies auch in der rechten Grafik der Abbildung 3.2 geschieht. Aus der Grafik können wir ablesen, dass Frauen zu jedem Zeitpunkt eine niedrigere kumulierte Rezidivrate als Männer haben.

3.2.3 Proportionalitätsannahme

Die Proportionalitätsannahme (engl.: proportional hazard assumption) ist die zentrale Annahme im Cox Modell (vgl. Cox, 1972). Damit ist gemeint, dass der Quotient des momentanen Ausfallsrisikos (Hazard) zweier Patientengruppen über die Zeit konstant ist. Für die grafische Überprüfung dieser Annahme wird ein Plot der sogenannten Schönfeldresiduen gezeichnet.

Schönfeldresiduen

Diese Residuen wurden nach ihrem Urheber David Schönfeld (1982) benannt. Ihr großer Vorteil ist, dass sie sowohl für univariate als auch für multivariable Modelle verwendet werden können. Da wir für unsere Analyse ein multivariables Modell verwenden, wurden die Schönfeldresiduen jeder Variable nach den anderen Variablen des Modells adjustiert. Für jede Kovariable und jeden Zeitpunkt t wird ein Residuum berechnet, dann werden alle Residuen einer Kovariable gegen die Zeit geplottet. Um die Interpretation zu vereinfachen,

kann eine geglättete Linie der Residuen mit einem 95% Konfidenzintervall gezeichnet werden. Diese Linien sind bei Erfüllung der Proportionalitätsannahme konstant. Die Abbildungen 3.3 und 3.4 zeigen ausschließlich die Glättungslinien der Schönfeldresiduen.

Man kann in den Plots der Abbildung 3.3, welche die klinischen Variablen der Studie enthält, beobachten, dass die Proportionalitätsannahme nicht eklatant verletzt ist. Die Transformationen der Variable Hämatokrit sind in der Abbildung 3.4 enthalten, sie weisen ebenso keine Nichtproportionalitäten auf.

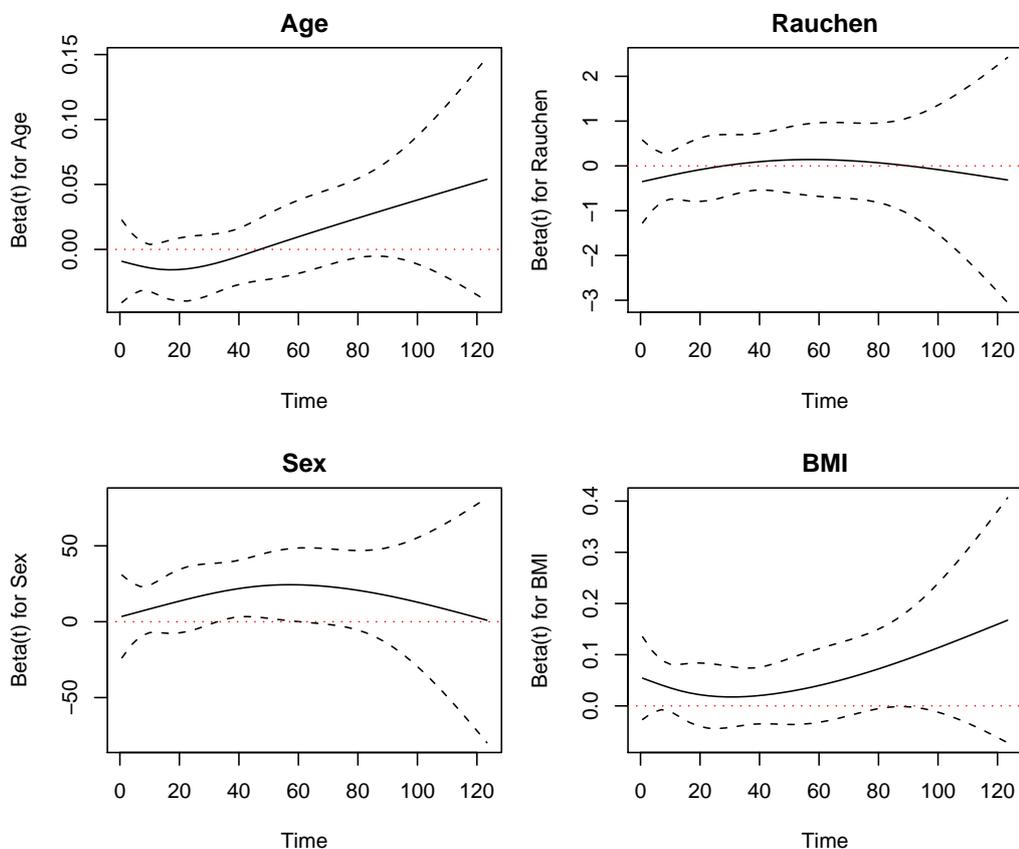


Abbildung 3.3: Glättungslinien der Schönfeldresiduen

3.2.4 Cox-Analyse am HCT Datensatz

Um herauszufinden, ob der Hämatokritwert einen Einfluß auf das Überleben der Patienten hat, benötigen wir Methoden für die Analyse von Überlebens-

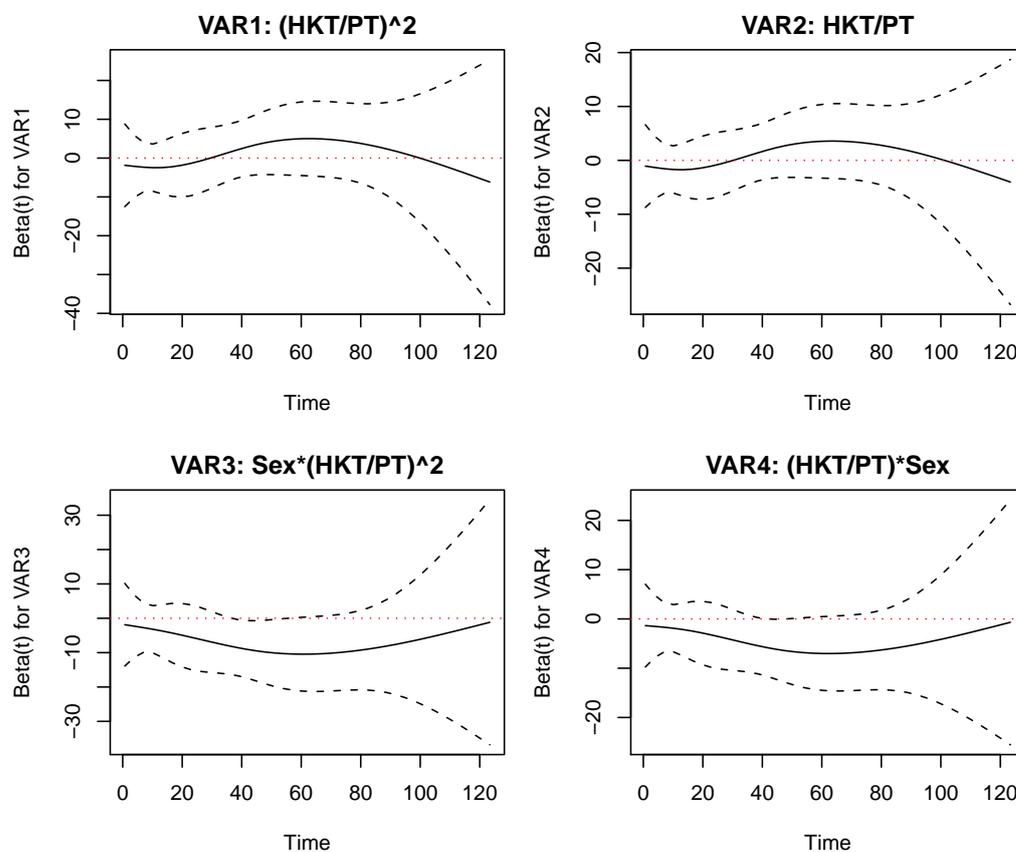


Abbildung 3.4: Glättungslinien der Schönfeldresiduen der Transformationen der Variable HB (mit $PT=100$)

daten, wie das Cox Proportional Hazards Modell. Da für das Modell noch die Annahmen gemacht wurden, dass es einen geschlechterspezifischen Effekt von Hämatokrit gibt, und dass dieser Effekt nicht linear ist, muss das Analyseverfahren nichtlineare Funktionen zulassen können. Um dies zu garantieren, wurden die Berechnungen sowohl mit dem Package `coxphf.beta` in R, welches nichtlineare Effekte mit Hilfe von Fraktionalen Polynomen modelliert, als auch mit dem SAS Makro `showinteractionv2`, welches RCS für die Modellierung der Effekte verwendet, durchgeführt.

Zuerst wurde das Modell ohne die Interaktion von HCT und Sex berechnet. Wobei dieses Modell zusätzlich nach den Variablen Age , $Smoking$, BMI und Sex adjustiert wurde, um mögliche Confounder zu berücksichtigen. Die Ergebnisse dieser Analyse sind in Tabelle 3.3 ersichtlich. Die Schätzer des Cox-Modells wurden mit Hilfe der Firth-Korrektur berechnet.

Tabelle 3.3: Cox Regression mit Firth-Korrektur ohne Interaktion

Variablen	Koeffizienten	Standardfehler	p-Wert
Age	-0.0028	0.0067	0.6719
Smoking	-0.1185	0.1969	0.5472
BMI	0.0420	0.0171	0.0143
powM2(HCT/100)	-2.9539	0.7771	0.0001
RpowM2(HCT/100)	-1.9911	0.5044	0.0001
Sex	-0.4736	0.2426	0.0509

Der *RA2-Algorithmus* entscheidet zugunsten eines nichtlinearen Effekts mit einem Polynom zweiten Grades. Für die beiden Potenzen p_1 und p_2 wählt er -2, daher sieht das Modell wie in Gleichung (3.1) aus.

$$h(t, x) = h_0(t) \exp\{\beta_1 Age + \beta_2 Smoking + \beta_3 BMI + \beta_4 \left(\frac{HCT}{100}\right)^{-2} + \beta_5 \left(\frac{HCT}{100}\right)^{-2} \log\left(\frac{HCT}{100}\right) + \beta_6 Sex\} \quad (3.1)$$

Ebenso ist aus der Tabelle 3.3 ersichtlich, dass der Body Mass Index (BMI) einen signifikanten Einfluß auf das Wiederauftreten einer Venenthrombose hat ($p=0.0143$). Ein geschlechterspezifischer Unterschied kann mit einem p-Wert von 0.0509 gerade nicht mehr angenommen werden.

Um in weiterer Folge herauszufinden, ob der Effekt des Hämatokrites auch geschlechterspezifisch unterschiedlich ist, bauen wir in das Modell einen Interaktionsterm des Hämatokritlevels mit dem Geschlecht ein. Dieses Modell wird in der Gleichung (3.2) dargestellt.

$$h(t, x) = h_0(t) \exp\{\beta_1 Age + \beta_2 Smoking + \beta_3 BMI + \beta_4 \left(\frac{HCT}{100}\right)^{-2} + \beta_5 \left(\frac{HCT}{100}\right)^{-2} \log\left(\frac{HCT}{100}\right) + \beta_6 Sex + \beta_7 \left(\frac{HCT}{100}\right)^{-2} Sex + \beta_8 \left(\frac{HCT}{100}\right)^{-2} \log\left(\frac{HCT}{100}\right) Sex\} \quad (3.2)$$

Die Resultate sind in der Tabelle 3.4 zu sehen.

Die Testung der hinzugefügten Interaktionsterme hängt von der Fragestellung des Anwenders ab. In Folge werden drei Möglichkeiten beschrieben, um eine Interaktion zu testen.

Tabelle 3.4: Cox-Regression mit Firth-Korrektur

Variablen	Koeffizienten	Standardfehler	p-Wert
Age	-0.0045	0.0068	0.5010
Smoking	-0.0879	0.1971	0.6554
BMI	0.0401	0.0172	0.0201
powM2(HCT/100)	-1.0690	2.2736	0.6382
RpowM2(HCT/100)	-0.7517	1.6374	0.6462
Sex	12.0381	5.8160	0.0385
powM2(HCT/100)*Sex	-4.7525	2.5688	0.0643
RpowM2(HCT/100)*Sex	-2.9558	1.7913	0.0989

Wird davon ausgegangen, dass man an der Darstellung von nichtlinearen Interaktionen interessiert ist, dann berechnet man mit `coxphf.beta` das Modell mit und ohne Interaktion und testet im Anschluss die Differenz der PLR-Statistik der beiden Modelle mit einem χ^2 -Test mit 2 Freiheitsgraden. Dabei können die beiden Modelle unterschiedliche Potenzen oder unterschiedliche Grade von Fraktionalen Polynomen aufweisen. In der Fallstudie sind die Potenzen beider Modelle gleich, wobei $p_1 = -2$ und $p_2 = -2$. Das Ergebnis ist mit einem p-Wert von 0.012 signifikant, d.h. es kann eine Wechselwirkung zwischen dem Hämatokritlevel und dem Geschlecht der Patienten angenommen werden.

Zuvor wurde mit einem komplexeren Modell gestartet. Möchte man aber bei der Evaluation mit einem möglichst einfachen Modell starten, so lässt man den *RA2-Algorithmus* das beste Modell ohne Interaktion berechnen. Dieses ist zufällig wieder das FP2 Modell mit $p_1 = -2$ und $p_2 = -2$. Zu diesem Modell fügt man nun die Interaktionsterme (hier $(HCT/100)^{-2} \cdot Sex$ und $(HCT/100)^{-2} \log(HCT/100) \cdot Sex$) hinzu und berechnet die Schätzer erneut. Das Modell mit der Interaktion wird gegen das genestete Modell getestet, wieder mit einem χ^2 -Test und 2 df. Da in der obigen Vorgehensweise der *RA2-Algorithmus* für das Modell ohne Interaktionen zufällig das genestete Modell des Interaktionsmodells wählte, erhält man auch bei dieser Vorgehensweise einen p-Wert von 0.012 (da die selben Modelle gegeneinander getestet werden). Diese Herangehensweise wird bei Royston und Sauerbrei (2008, S.155-156) im *MFPI-Algorithmus* praktiziert.

Zuletzt besteht noch die Möglichkeit das Interaktionsmodell, das vom *RA2-Algorithmus* ausgewählt wurde (mit $p_1 = -2$ und $p_2 = -2$), gegen das genestete Modell ohne Interaktionsterme zu testen (Umkehrung der vorangegangenen Prozedur). Dafür kann in R der Befehl `coxphftest.beta` verwendet werden, der

für die Testung der beiden Interaktionsterme den *penalized Likelihood-Ratio-Test* verwendet und zusätzlich als Option die Firth-Korrektur anbietet. Der daraus resultierende p-Wert von 0.0009 kommt zu denselben Schlussfolgerungen wie oben.

Jede der drei Evaluierungsarten hat ihre Berechtigung, solange der Anwender die Motivation seiner Modellierung begründen kann.

Aber wie ist der nichtlineare geschlechterspezifische Effekt von Hämatokrit zu interpretieren?

Bei der Interpretation der Resultate von Männern und Frauen wird ein Referenzwert verwendet. Mit diesem Wert werden alle anderen Punkte in Bezug gesetzt. In unserer Analyse wurde ein Referenzwert von 40 gewählt, d.h. alle relativen Hazardwerte der Hämatokritlevels von 30 bis 50 (relativ zum Baseline-Hazard) wurden mit dem relativen Hazardwert bei einem Hämatokritlevel von 40 verglichen. Bei den Männern (siehe Abbildung 3.5, linke obere Grafik) kommt man zu dem Ergebnis, dass das Risiko einer erneuten Venenthrombose nicht vom Hämatokritwert abhängt. Hingegen sieht man in der Abbildung der Frauen, dass bei einem Anstieg des Hämatokritwertes das Risiko eines Rezidivs ansteigt.

In der dritten Grafik der Abbildung 3.5 wird das Rezidivrisiko bei Männern und Frauen mit demselben Hämatokritwert verglichen. Dabei wird erkennbar, dass Männer bei einem Hämatokritwert zwischen 38 und 40 ein signifikant höheres Rezidivrisiko haben als Frauen. Hingegen haben Frauen ein höheres Rezidivrisiko bei höheren Hämatokritwerten (> 45). Bei der Interpretation sollte nicht außer Acht gelassen werden, dass es kaum Männer mit einem Hämatokritwert unter 37 gibt, daher sind die Konfidenzgrenzen unter diesem Wert sehr breit.

Die Cox Analyse wurde ein zweites Mal durchgeführt, diesmal wurden die Schätzer mit der Maximum-Likelihood-Methode berechnet. Der *RA2-Algorithmus* lieferte erneut ein nichtlineares Modell, wieder wählte er für p_1 und p_2 den Wert -2 . Betrachtet man die Koeffizienten dieser Analyse, so sieht man in Tabelle 3.5, dass diese nicht mit den Koeffizienten der Firth-Korrektur übereinstimmen. Die Ergebnisse, berechnet mit der Firth-Methode, liefern bei allen Variablen bis auf die Variable *BMI* Koeffizienten näher Null und besitzen kleinere Standardfehler mit einer Ausnahme, der Variablen *Age*.

Auftreten von Monotoner Likelihood

In den 1000 Bootstrapsamples trat bei keiner der Maximum-Likelihood-Schätzungen Monotone Likelihood auf. Der Grund dafür dürfte die Stichprobengröße von 619 sein.

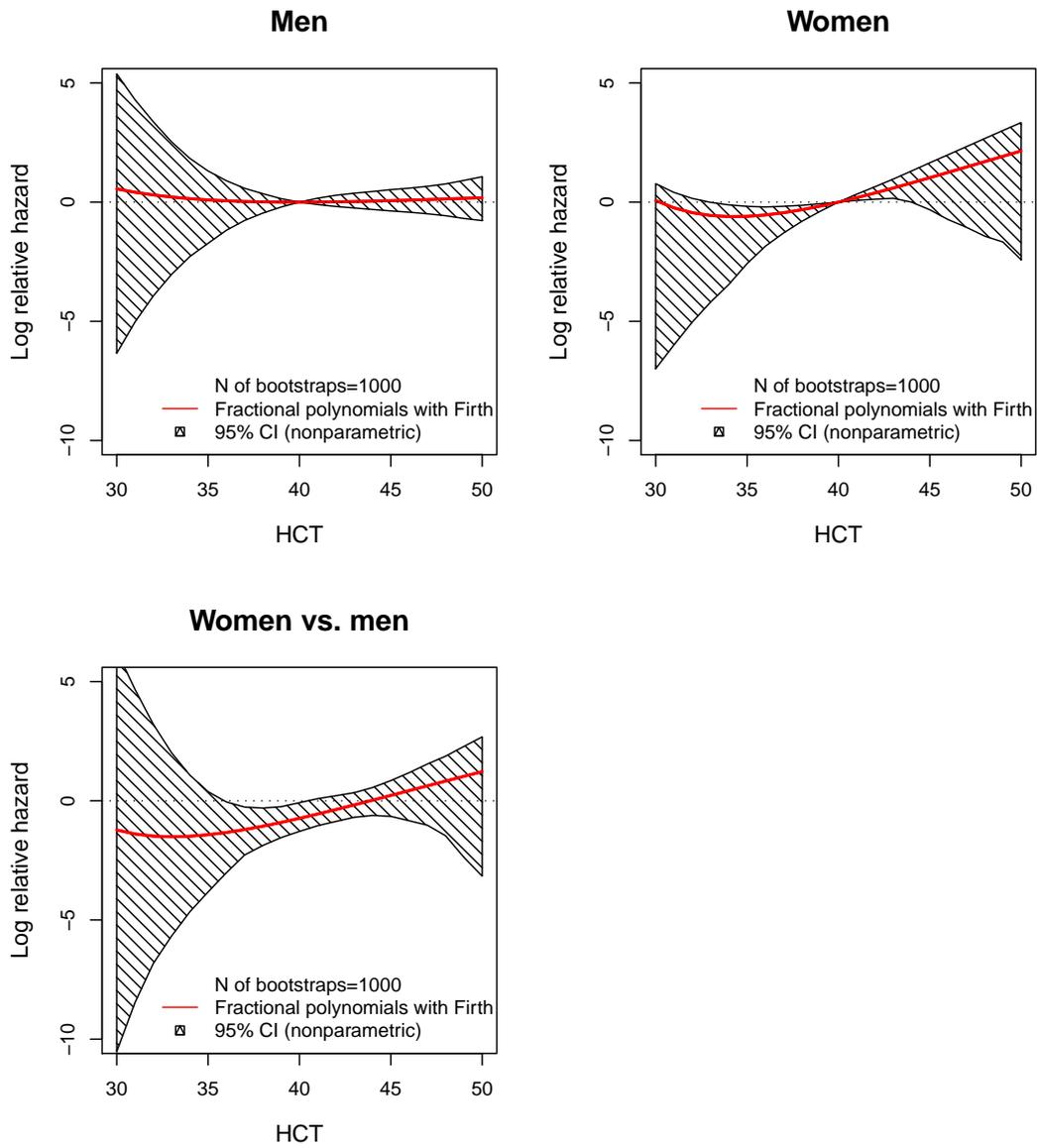


Abbildung 3.5: Geschlechterspezifischer Effekt von Hämatokrit

Tabelle 3.5: Cox Regression ohne Firth-Korrektur

Variablen	Koeffizienten	Standardfehler	p-Wert
Age	-0.0046	0.0067	0.5000
Smoking	-0.0954	0.1973	0.6286
BMI	0.0399	0.0173	0.0210
powM2(HCT/100)	0.2668	2.8251	0.9248
RpowM2(HCT/100)	0.2455	2.0626	0.9053
Sex	14.2372	6.7439	0.0348
powM2(HCT/100)*Sex	-5.8969	3.0844	0.0559
RpowM2(HCT/100)*Sex	-3.8202	2.1974	0.0821

3.2.5 Analyse der Konfidenzintervalle

In den Abbildungen dieses Kapitels werden pro Zeile drei Plots gezeigt. Der erste Plot zeigt die Log Relative Hazards der Hämatokritwerte für Männer, der mittlere Plot tut selbiges für die Frauen und der rechte Plot zeigt Männer und Frauen im Vergleich.

In den ersten sechs Plotzeilen werden für FP bzw. für RCS mit 3 Knoten oder RCS mit 5 Knoten jeweils die nichtparametrischen, parametrischen und modellbasierten (3 df) Konfidenzintervalle verglichen. Dies geschieht sowohl für die Maximum-Likelihood-Methode als auch für deren Korrektur mit der Firth-Methode. Um welche Schätzmethode, nichtlineare Modellierung oder Konfidenzintervalle es sich dabei handelt, kann in jeder Grafik durch die sich darin befindende Legende abgelesen werden. Bei den ersten beiden Abbildungen (3.6 und 3.7) betrachten wir die Log Relativen Hazardwerte der Fraktionalen Polynome, dabei sollte darauf geachtet werden, dass die y-Achsen der Plots nicht die gleiche Spanne aufweisen. Der Vergleich der Schätzmethode zeigt, dass alle drei Konfidenzarten jeweils engere Intervalle bei der Firth-Methode aufweisen als bei der Maximum-Likelihood-Methode. Dabei ist ebenso ersichtlich, dass die modellbasierten Konfidenzintervalle eindeutig enger sind als die anderen beiden Konfidenzintervalle. Die parametrischen KI sind breiter als die nichtparametrischen KI, deren eingeschlossene Fläche größtenteils im negativen Wertebereich liegt.

In den darauf folgenden Grafiken (Abbildung 3.8 und 3.9) lassen sich die drei Konfidenzintervallsarten für RCS mit drei Knoten vergleichen. Dabei sind die modellbasierten KI sowohl bei ML als auch bei Firth am schmalsten. Interessant ist, dass die parametrischen KI jeweils die größte obere Konfidenzintervallsgrenze haben, die untere kleinste Konfidenzintervallsgrenze

aber immer von einem nichtparametrischen KI stammt. Daraus lässt sich schließen, dass die Schätzer wohl nicht symmetrisch verteilt sind und daher nichtparametrische KI den parametrischen KI vorzuziehen sind. Zwischen den Abbildungen mit der Maximum-Likelihood-Methode und der Firth-Methode ist hier kaum ein Unterschied erkennbar.

Im Gegensatz dazu ist bei den RCS mit 5 Knoten (Abbildung 3.10 und 3.11) ein klarer Unterschied zwischen ML und Firth bei den parametrischen und nichtparametrischen Konfidenzintervallen erkennbar, da die KI bei der Firth-Korrektur eindeutig schmaler sind. Nur die modellbasierten KI sind bei beiden Schätzverfahren nahezu ident, gleichzeitig liefern sie in beiden Methoden die engsten Konfidenzintervalle.

In Folge werden jene Abbildungen (3.12 bis 3.17) betrachtet, welche die drei nichtlinearen Modellierungen (FP, RCS mit 3 Knoten, RCS mit 5 Knoten) vergleichen. Wenn man ausschließlich die drei geschätzten Log Relative Hazard Linien betrachtet, sieht man, dass die Schätzer der FP und die der RCS mit 3 Knoten sehr ähnlich sind, hingegen die Linie der RCS mit 5 Knoten eine sehr komplexe Funktion darstellt. Daraus lässt sich schließen, dass bei diesem Datensatz RCS mit 5 Knoten eine Überanpassung an die Daten liefern. Weiters gehen die dazugehörigen Konfidenzintervalle jeweils an den Enden sehr stark auseinander, bei den Männern speziell bei den niedrigen Hämatokritwerten, bei Frauen bei den hohen Hämatokritwerten. Der Grund dafür liegt wohl darin, dass in den Daten nur wenige Männer einen niedrigen Hämatokrit haben und nur wenige Frauen einen hohen Hämatokritwert aufweisen. Es ist bereits bekannt, dass die zugrundeliegende Verteilung des Hämatokrits bei Frauen und Männer unterschiedlich ist. In der Medizin wird davon ausgegangen, dass der Normalbereich des Hämatokrits bei Frauen von 38% bis 44% reicht, hingegen bei Männern 42% bis 48% als normal erachtet werden.

Je nach Gruppe liefern die modellbasierten KI (Abbildung 3.12 und 3.13) unterschiedliche Ergebnisse. So sind in der Gruppe der Männer die KI der RCS mit 3 Knoten deutlich enger als die anderen KI, während in der Gruppe der Frauen die KI der FP schmaler sind.

Bei den parametrischen KI (Abbildung 3.14 und 3.15) liefern die FP und die RCS mit 3 Knoten ähnliche Ergebnisse, die KI der RCS mit 5 Knoten sind größtenteils um ein vielfaches breiter.

Betrachtet man nur die nichtparametrischen Konfidenzintervalle (Abbildung 3.16 und 3.17), so ist klar, dass wieder einmal die RCS mit 5 Knoten die breitesten Konfidenzintervalle liefern. Die RCS mit 3 Knoten liefern deutlich engere nichtparametrische KI, trotzdem sind die der FP noch enger. Dies gilt sowohl für die Schätzer der Maximum-Likelihood-Methode als auch für die

der Firth-Methode.

In Folge werden die KI der Maximum-Likelihood-Methode und der Firth-Korrektur miteinander verglichen, dies geschieht anhand der Abbildungen 3.18 bis 3.26. Die ersten dieser Abbildungen (3.18 bis 3.20) zeigen den Vergleich der Schätzmethode bei den Fraktionalen Polynomen. Beide Methoden liefern fast idente Log Relative Hazard Linien, nur im unteren Hämatokritbereich weichen sie ein wenig voneinander ab. Bei den parametrischen Konfidenzintervallen (Abbildung 3.19) ist es nicht ersichtlich, welche Methode schmalere Konfidenzintervalle liefert, da diese sich je nach Geschlecht und Hämatokritbereich unterscheiden. Bei den modellbasierten KI (Abbildung 3.18) sind eindeutig jene, die der Firth-Methode zugeordnet werden, enger, gleiches gilt für die nichtparametrischen Konfidenzintervalle (Abbildung 3.20).

Der Vergleich der Schätzmethode bei den RCS mit 3 Knoten ist sehr eindeutig (Abbildung 3.21 bis 3.23). Die Funktionen der Log Relative Hazards sind fast vollkommen ident und alle drei Konfidenzintervalle sind bei der Firth-Methode minimal schmaler.

Die geschätzten Log Relative Hazards für Firth und ML der RCS mit 5 Knoten (Abbildung 3.24 bis 3.26) weichen ausschließlich im unteren und oberen Hämatokritbereich geringfügig voneinander ab. Die modellbasierten KI sind bei beiden Schätzmethode gleich (Abbildung 3.24), eigenartigerweise gibt es gravierende Unterschiede bei den parametrischen und nichtparametrischen KI (Abbildung 3.25 und 3.26). Diese sind bei der Maximum-Likelihood-Methode enorm breit, besonders im unteren Hämatokritbereich.

Tabelle 3.6: Abbildungsverzeichnis für Analyse der Konfidenzintervallen

Abbildung	Vergleich	Modellierung	Schätzung	Konfidenzintervall
3.6	KI	FP	ML	alle
3.7	KI	FP	Firth	alle
3.8	KI	RCS (3)	ML	alle
3.9	KI	RCS (3)	Firth	alle
3.10	KI	RCS (5)	ML	alle
3.11	KI	RCS (5)	Firth	alle
3.12	Modellierung	alle	ML	modellbasiert
3.13	Modellierung	alle	Firth	modellbasiert
3.14	Modellierung	alle	ML	parametrisch
3.15	Modellierung	alle	Firth	parametrisch
3.16	Modellierung	alle	ML	nichtparametrisch
3.17	Modellierung	alle	Firth	nichtparametrisch
3.18	ML vs. Firth	FP	beide	modellbasiert
3.19	ML vs. Firth	FP	beide	parametrisch
3.20	ML vs. Firth	FP	beide	nichtparametrisch
3.21	ML vs. Firth	RCS (3)	beide	modellbasiert
3.22	ML vs. Firth	RCS (3)	beide	parametrisch
3.23	ML vs. Firth	RCS (3)	beide	nichtparametrisch
3.24	ML vs. Firth	RCS (5)	beide	modellbasiert
3.25	ML vs. Firth	RCS (5)	beide	parametrisch
3.26	ML vs. Firth	RCS (5)	beide	nichtparametrisch

Abbildung 3.6: Fraktionale Polynome ohne Firth-Korrektur

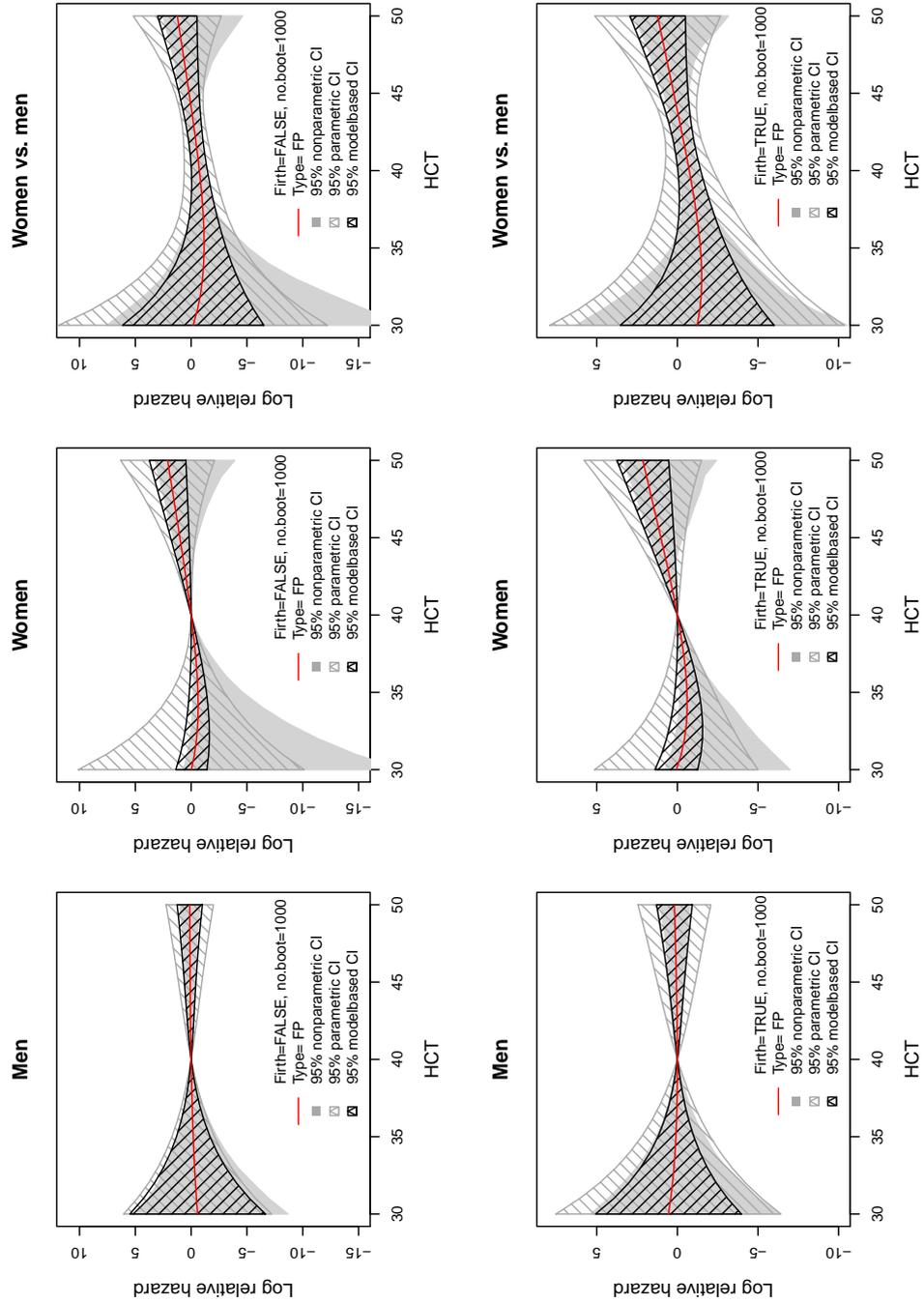


Abbildung 3.7: Fraktionale Polynome mit Firth-Korrektur

Abbildung 3.8: RCS (3 Knoten) ohne Firth-Korrektur

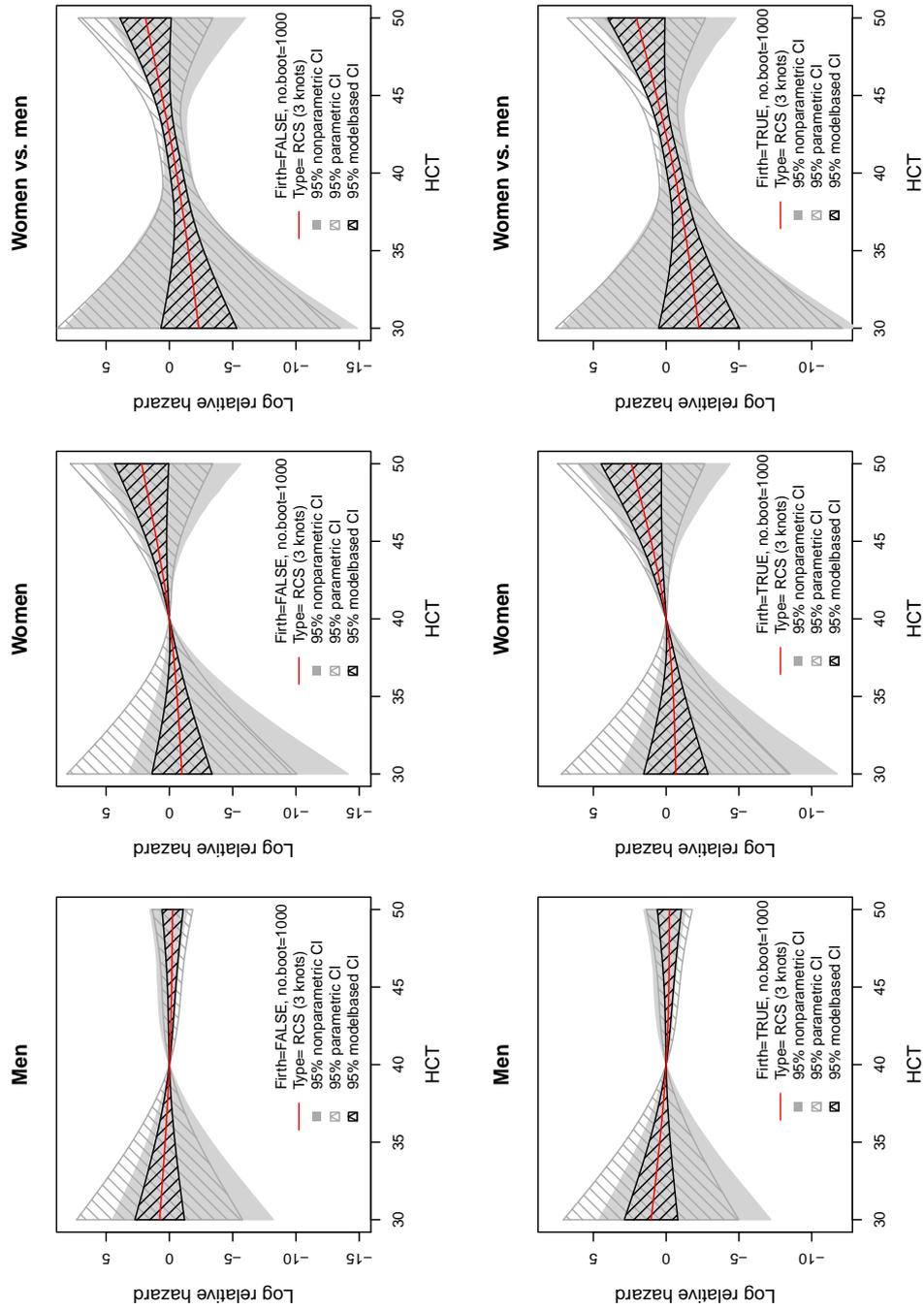


Abbildung 3.9: RCS (3 Knoten) mit Firth-Korrektur

Abbildung 3.10: RCS (5 Knoten) ohne Firth-Korrektur

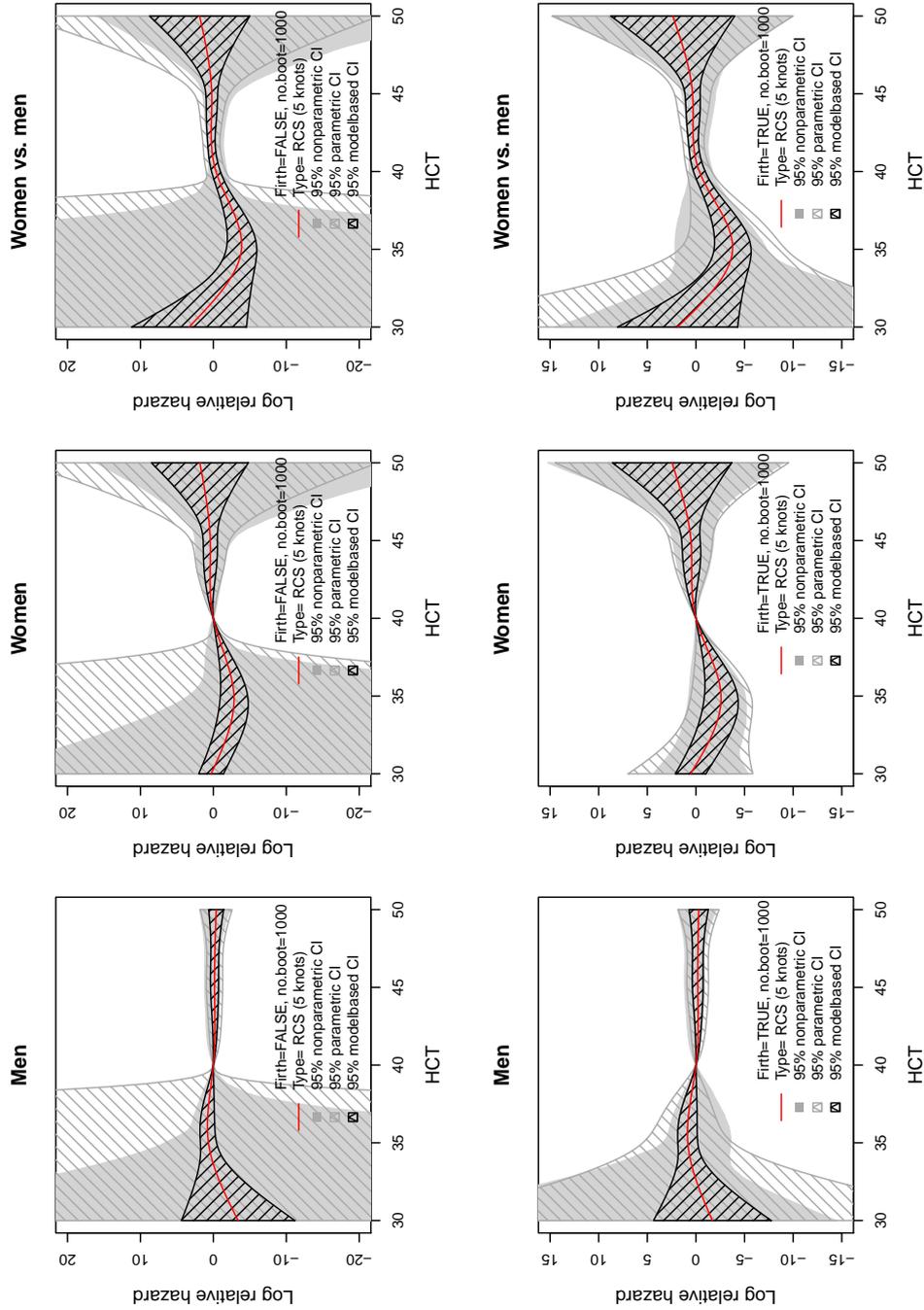


Abbildung 3.11: RCS (5 Knoten) mit Firth-Korrektur

Abbildung 3.12: Modellbasierte KI von Fraktionale Polynome und RCS ohne Firth-Korrektur

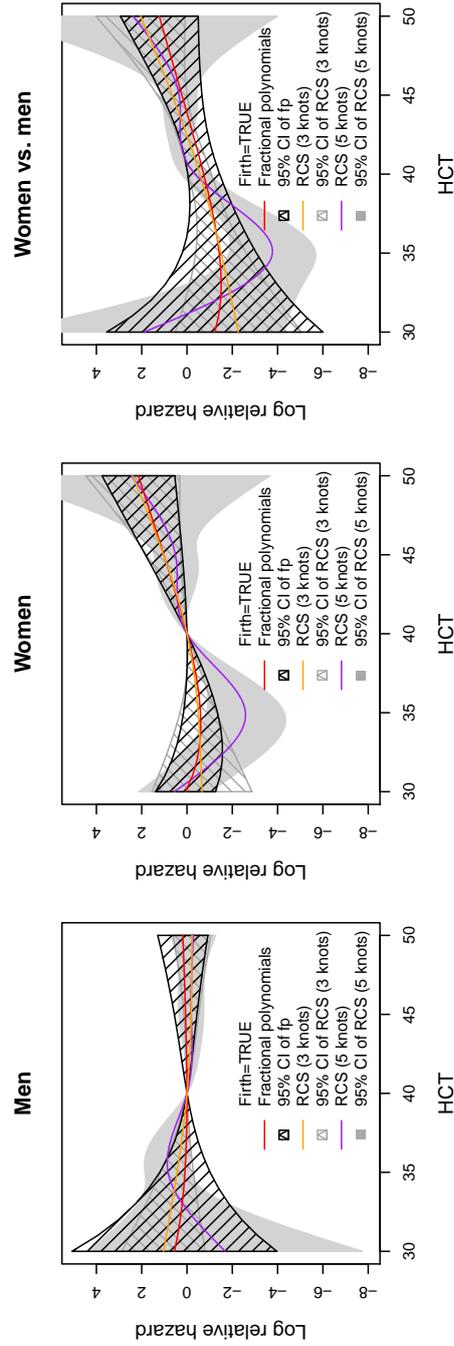
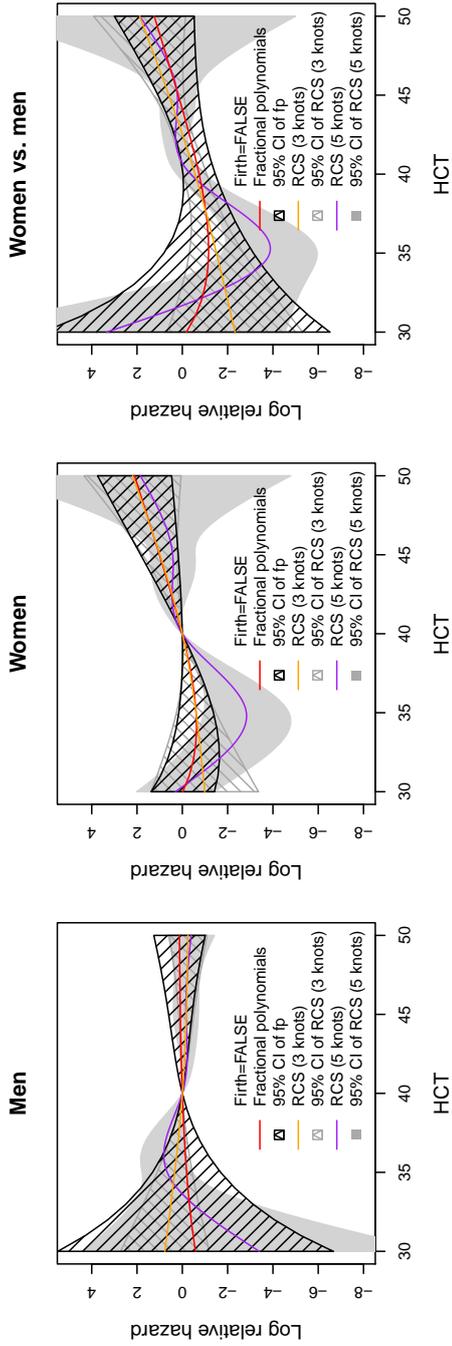


Abbildung 3.13: Modellbasierte KI von Fraktionale Polynome und RCS mit Firth-Korrektur

Abbildung 3.14: Parametrische KI von Fraktionale Polynome und RCS ohne Firth-Korrektur

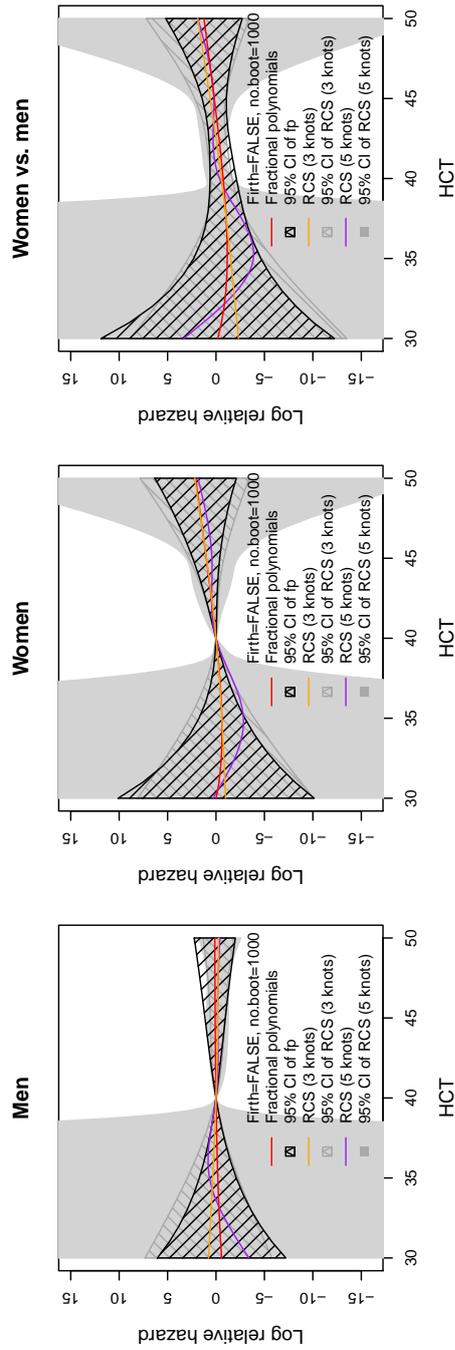


Abbildung 3.15: Parametrische KI von Fraktionale Polynome und RCS mit Firth-Korrektur

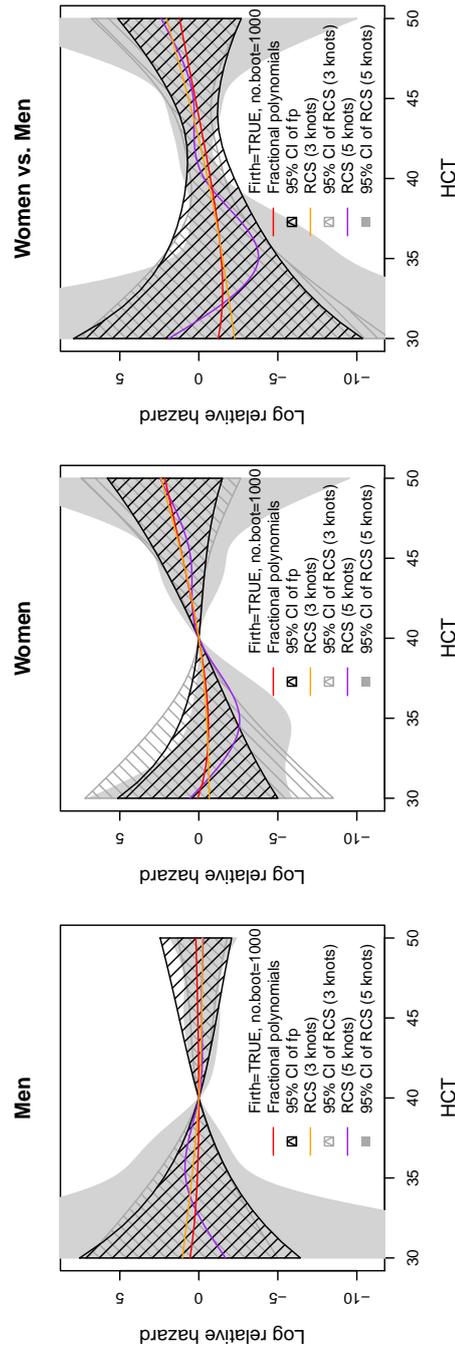


Abbildung 3.16: Nichtparametrische KI von Fraktionale Polynome und RCS ohne Firth-Korrektur

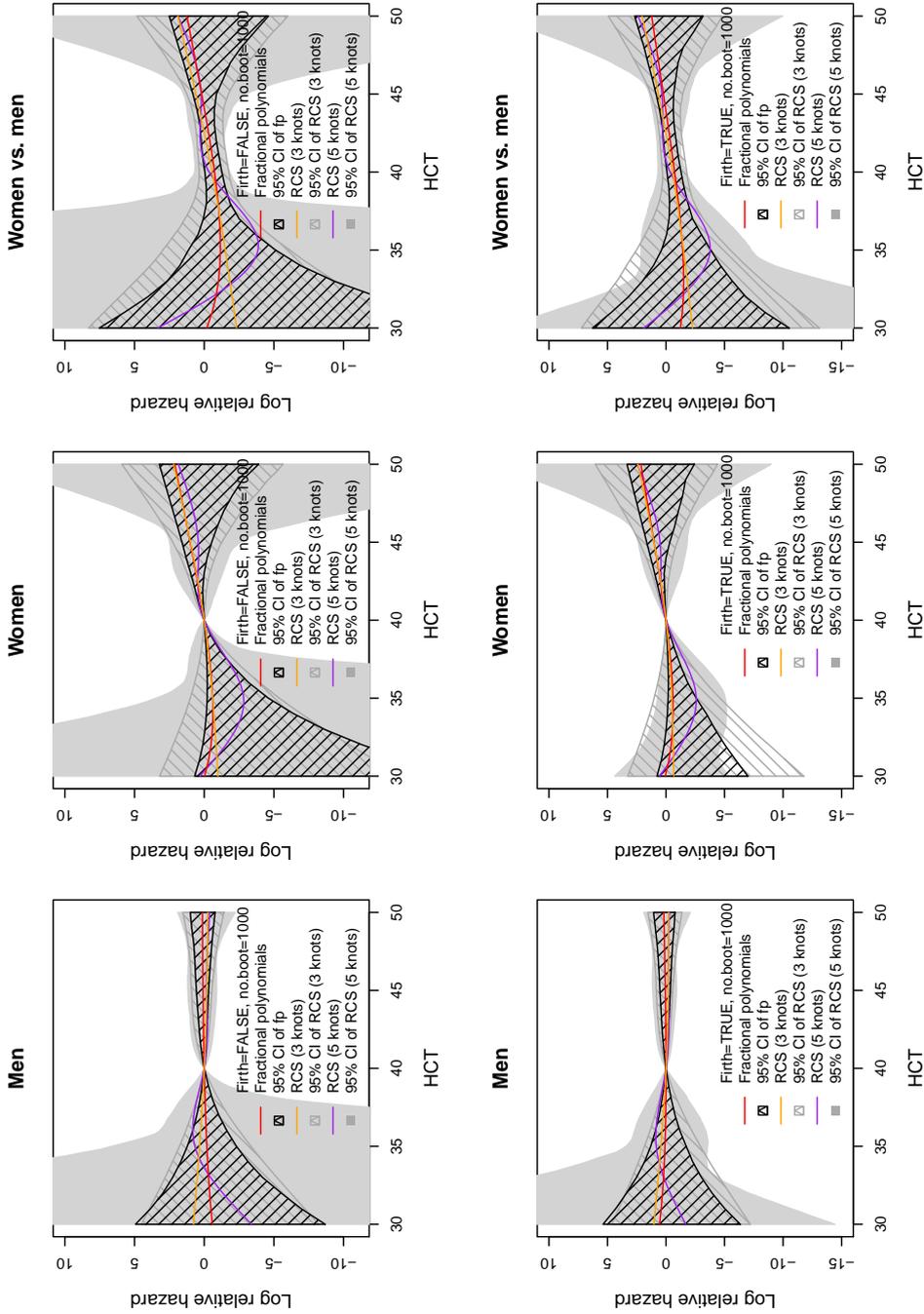


Abbildung 3.17: Nichtparametrische KI von Fraktionale Polynome und RCS mit Firth-Korrektur

Abbildung 3.18: Modellbasierte KI von Fraktionalen Polynomen mit und ohne Firth-Korrektur

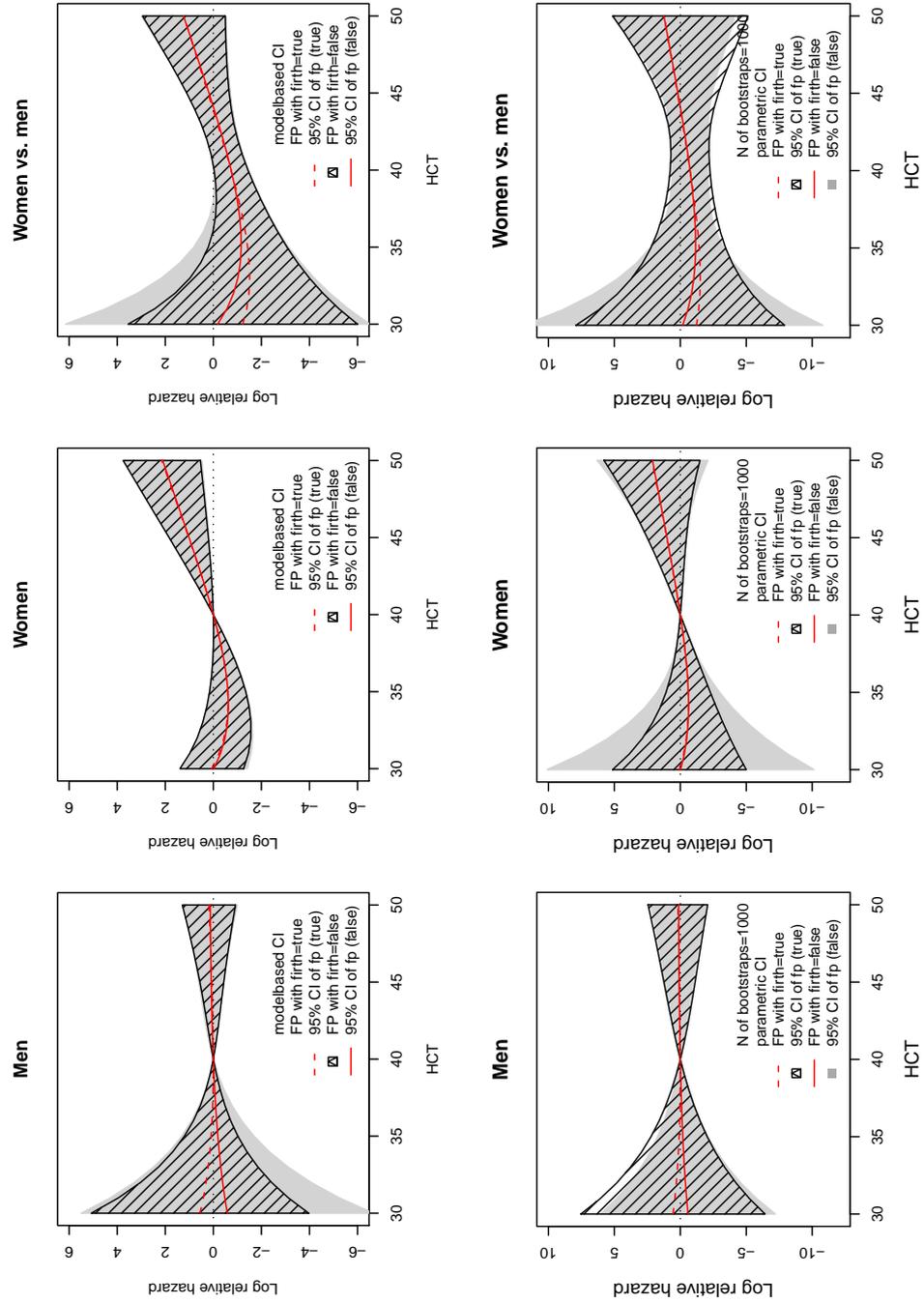


Abbildung 3.19: Parametrische KI von Fraktionalen Polynomen mit und ohne Firth-Korrektur

Abbildung 3.20: Nichtparametrische KI von Fraktionalen Polynomen mit und ohne Firth-Korrektur

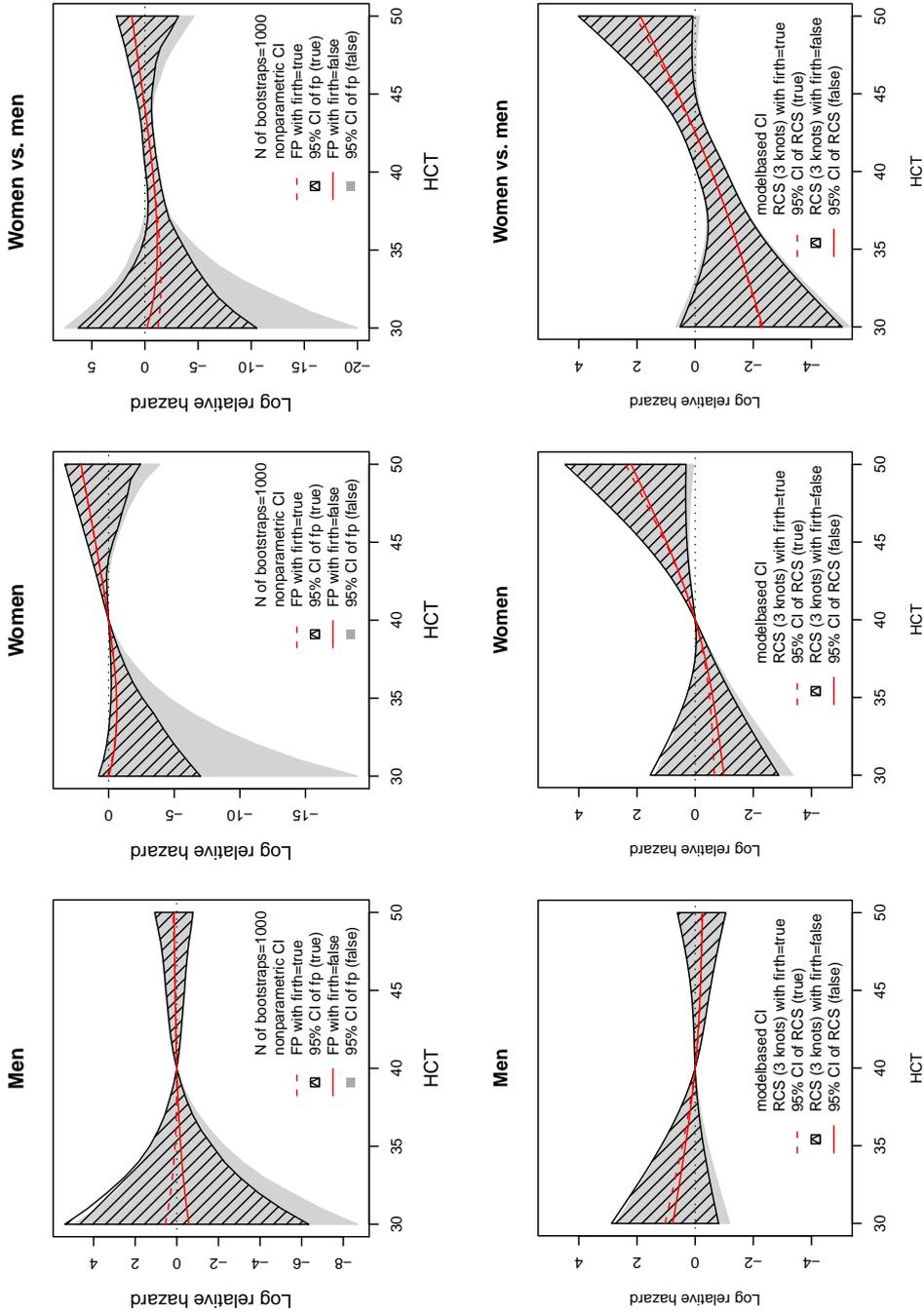


Abbildung 3.21: Modellbasierte KI von RCS (3 Knoten) mit und ohne Firth-Korrektur

Abbildung 3.22: Parametrische KI von RCS (3 Knoten) mit und ohne Firth-Korrektur

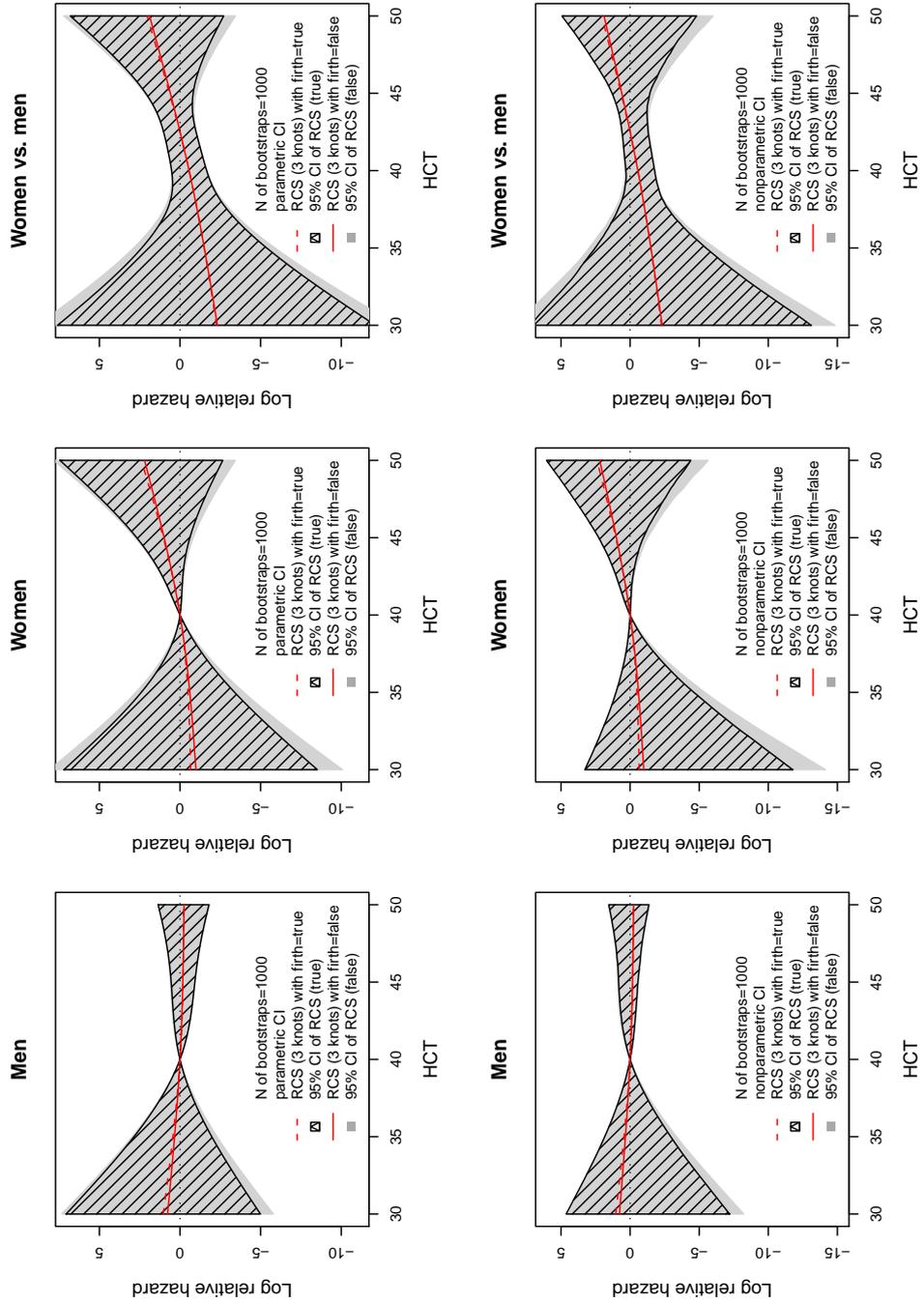


Abbildung 3.23: Nichtparametrische KI von RCS (3 Knoten) mit und ohne Firth-Korrektur

Abbildung 3.24: Modellbasierte KI von RCS (5 Knoten) mit und ohne Firth-Korrektur

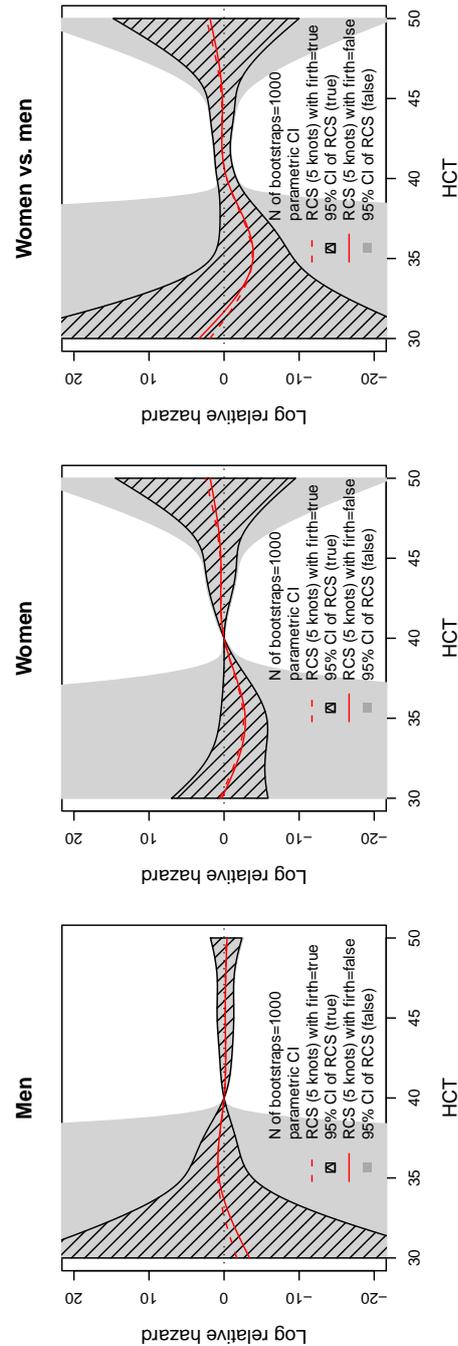
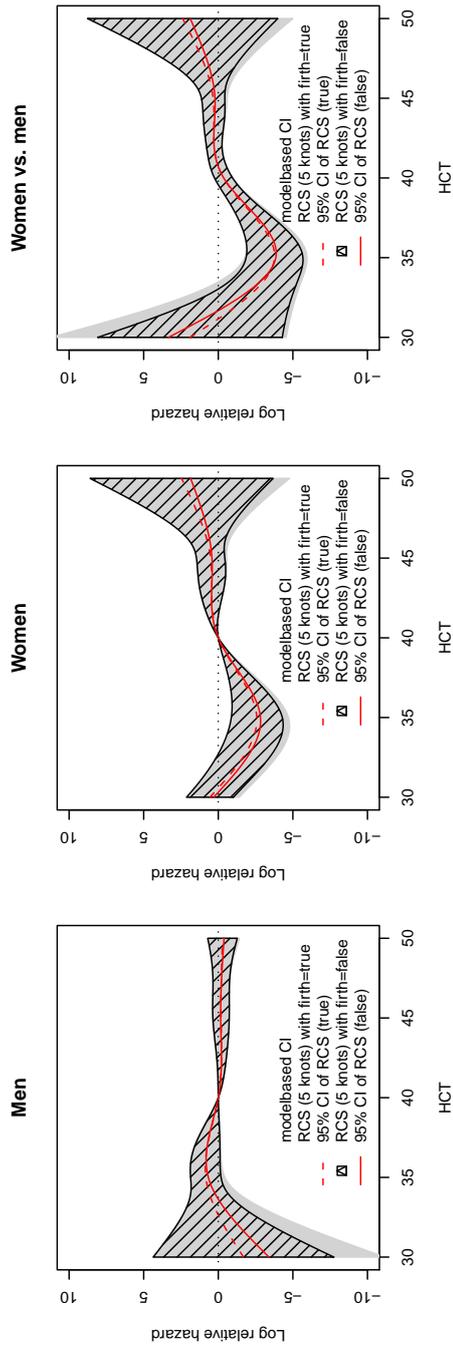


Abbildung 3.25: Parametrische KI von RCS (5 Knoten) mit und ohne Firth-Korrektur

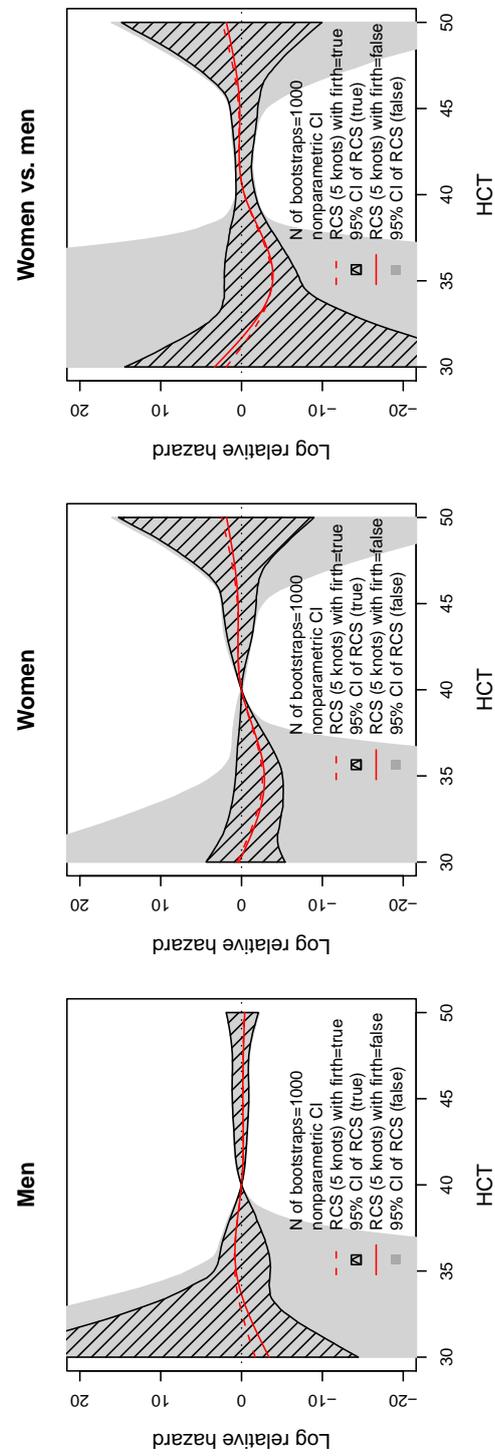


Abbildung 3.26: Nichtparametrische KI von RCS (5 Knoten) mit und ohne Firth-Korrektur

3.2.6 Fazit

Zu dieser Studie lässt sich abschließend sagen, dass ein signifikanter nichtlinearer Effekt von Hämatokrit vorliegt und dieser Effekt geschlechterspezifisch ist. Durch die Firth-Korrektur erhalten wir Schätzer näher bei Null mit kleineren Standardabweichungen, außerdem sind die Konfidenzintervalle der Schätzer enger. Daher ist es empfehlenswert, die Firth-Methode zu verwenden. Die parametrischen Konfidenzintervalle scheinen für diesen Datensatz, aufgrund der Nichtsymmetrie der Schätzer, ungeeignet. Ebenso wenig empfehlenswert sind die modellbasierten Konfidenzintervalle, da ihre geringe Breite im Vergleich zu den parametrischen und nichtparametrischen KI, für deren Berechnungen 1000 Bootstrapsamples verwendet wurden, zu optimistisch erscheinen. Für die Modellierung des nichtlinearen Effekts empfiehlt es sich, entweder Restricted Cubic Splines mit 3 Knoten oder Fraktionale Polynome zu verwenden, RCS mit 5 Knoten liefern eine Überanpassung an die Daten und sind daher nicht geeignet. Es darf jedoch nicht außer Acht gelassen werden, dass FP im Gegensatz zu den RCS einen linearen Zusammenhang widerspiegeln können und sogar eine Entscheidung zu Gunsten des Nullmodells treffen könnten. Sollte daher ein linearer Effekt nicht ausschließbar sein oder der Effekt der Variable überhaupt fragwürdig erscheinen, so sollten für die Modellierung FP herangezogen werden.

Diskussion

In der Simulationsstudie ist keine der beiden Schätzmethoden eindeutig besser, daher ist es empfehlenswert die Wahl des Schätzverfahrens von den vorliegenden Daten abhängig zu machen. Daraus folgt, dass bei einem Datensatz mit einer kleinen Stichprobe ($n = 100$) die Firth-Methode das geeignetere Schätzverfahren wäre; einerseits, da Monotone Likelihood in kleineren Stichproben deutlich häufiger vorkommt, wie es auch in der Simulationsstudie (siehe Tabelle 2.8 und 2.9) ersichtlich ist, andererseits, da eine schwache Tendenz erkennbar ist, dass die Firth-Methode bei kleineren Datensätzen auch kleinere Abweichungen der Schätzer aufweist. Die Maximum-Likelihood-Methode scheint hingegen ein geeignetes Schätzverfahren für große Datensätze ($n \geq 400$) zu sein, da sie im Gegensatz zur Firth-Methode kleinere Abweichungen bei großen Datensätzen liefert. Zusätzlich sei noch einmal erwähnt, dass die Firth-Methode im Vergleich mit der Maximum-Likelihood-Methode kleinere gewichtete Mittel beim MAD aufweist, aber größere gewichtete Mittel beim medianen Bias hat.

Der Vergleich der Punktschätzer der Selektionsmodelle und der fixierten Modelle zeigt, dass die Schätzung der fixierten Modelle numerisch einfacher ist und dass die Varianz ihrer Schätzer wesentlich kleiner ist.

In der Simulationsstudie ist deutlich ersichtlich, dass das Auftreten von Monotoner Likelihood von der Stichprobengröße, aber auch vom Zensierungsanteil abhängig ist.

Rückschließend kann daher gesagt werden, dass der Zensierungsanteil und die Stichprobengröße ausschlaggebend für Monotone Likelihood, die Abweichungen der Schätzer und die Varianz derselben sind.

Da FP und RCS anhand einer Fallstudie verglichen wurden, können keine allgemeinen Aussagen davon abgeleitet werden. Es macht auch hier Sinn, die Wahl der Modellierungsart von den Eigenschaften des Datensatzes abhängig zu machen, zum Beispiel ob es Ausreißer gibt. Weiters sollte bedacht werden, ob der Anwender ausschließlich von einem nichtlinearen Modell ausgehen möchte oder aber auch ein lineares Modell oder das Nullmodell zulassen will. Die Firth-Methode ist bereits in einigen statistischen Softwareprogrammen implementiert, wie zum Beispiel in SAS (`proc phreg`) oder in einigen Packages des R Systems (`coxphf`), daher für Anwender einfach verwendbar. Die Analyse von Daten mit Fraktionalen Polynomen ist ebenso einfach umsetzbar, da bereits einige Packages in R (`mfp`) und Stata (`fracpoly`) vorhanden sind, in denen unterschiedliche Regressionsmodelle bei der Modellierung auf FP zurückgreifen können.

Aufgrund der Ergebnisse dieser Masterarbeit wäre es von Interesse, in weiterführenden Simulationsstudien die verschiedenen Konfidenzintervalle und deren Überdeckungswahrscheinlichkeiten zu betrachten, damit die Aussagen bezüglich der Güte nicht nur auf den Abweichungen der Schätzer basieren. Zusätzlich sollten alle Analysen der Simulationen auch mit den Restricted Cubic Splines durchgeführt werden, wodurch sich allgemein gültige Aussagen über die RCS und den Vergleich mit FP ableiten ließen.

Überlegenswert könnte dabei sein, die RCS ebenso wie die FP mit einem Algorithmus zu analysieren, d.h. statt den Potenzen aus dem Set S die Anzahl und Position der Knoten variieren zu lassen, wie auch im *MVRS-Algorithmus* von Royston und Sauerbrei (2008, S.201 ff). Dadurch würde der Vergleich von RCS und FP auf denselben Voraussetzungen basieren.

In Zukunft könnte man auch den Strafterm a_r (siehe Gleichung 1.5) durch Vergrößerung und Verkleinerung ändern und die Unterschiede anhand einer Simulationsstudie vergleichen. Dadurch könnte untersucht werden, welcher Strafterm für welche Daten am besten geeignet ist oder ein Bestrafungsterm gefunden werden, der grundsätzlich für unterschiedliche Voraussetzungen gute Ergebnisse liefert.

Wegen der Verzerrung, die durch die optimale Selektion der Potenzen durch den *RA2-Algorithmus* entsteht, wäre es sinnvoll die Parameter zu *shrinken*. Sauerbrei (1999) schätzte den sogenannten *parameterwise shrinkage factor*, dabei wird für jeden Parameter des finalen Modells ein eigener Shrinkagefaktor berechnet. Die Multiplikation des Shrinkagefaktors mit dem Parameter soll somit die Verzerrung minimieren (siehe auch Royston und Sauerbrei, 2008, S.40 ff).

Die Ergebnisse der Masterarbeit zeigen eines deutlich: Die Entscheidung bezüglich der Methode und der Modellierung sollte immer auf den Eigenschaften der zugrundeliegenden Daten basieren.

Literaturverzeichnis

- AMBLER, G. & ROYSTON, P. (2001). Fractional polynomial model selection procedures: investigation of type i error rate. *Journal of Statistical Computation and Simulation* **69**, 89–108.
- BRYSON, M. C. & JOHNSON, M. E. (1981). The incidence of monotone likelihood in the cox model. *Technometrics* **23**, 381–383.
- CORDEIRO, G. M. & CRIBARI-NETO, F. (1998). On bias reduction in exponential and non-exponential family regression models. *Communications in Statistics - Simulation and Computation* **27**, 485–500.
- COX, D. R. (1972). Regression models and life-tables. *Journal of the Royal Statistical Society* **34**, 187–220.
- COX, D. R. (1975). Partial likelihood. *Biometrika* **62**, 269–276.
- DURRLEMAN, S. & SIMON, R. (1989). Flexible regression models with cubic splines. *Statistics in Medicine* **8**, 551–561.
- EFRON, B. & TIBSHIRANI, R. J. (1993). *An Introduction to the Bootstrap*. Chapman & Hall, New York.
- EISCHER, L., TSCHOLL, V., HEINZE, G., TRABY, L., KYRLE, P. A., & EICHINGER, S. (submitted manuscript, 2011). Hematocrit and the risk of recurrent venous thromboembolism: does it explain the difference in recurrence between men and women?
- FIRTH, D. (1993). Bias reduction of maximum likelihood estimates. *Biometrika* **80**, 27–38.
- HARRELL, F. J. (2001). *Regression modeling strategies. With applications to linear models, logistic regression, and survival analysis*. Springer Verlag, New York, Berlin, Heidelberg.

- HEINZE, G. (2006). A comparative investigation of methods for logistic regression with separated or nearly separated data. *Statistics in Medicine* **25**, 4216–4226.
- HEINZE, G. (unpublished, 2011). showinteractionv2 - sas macro for estimation of nonlinear interactions in cox regression using restricted cubic splines.
- HEINZE, G. & DUNKLER, D. (2008). Avoiding infinite estimates of time-dependent effects in small-sample survival studies. *Statistics in Medicine* **27**, 6455–6469.
- HEINZE, G., KAINZ, A., HÖRL, W. H., & OBERBAUER, R. (2009). Mortality in renal transplant recipients given erythropoietins to increase haemoglobin concentration: cohort study. *British Medical Journal* **339**, b4018.
- HEINZE, G. & SCHEMPER, M. (2001). A solution to the problem of monotone likelihood in cox regression. *Biometrics* **57**, 114–119.
- HEINZE, G. & SCHEMPER, M. (2002). A solution to the problem of separation in logistic regression. *Statistics in Medicine* **21**, 2409–2419.
- HEINZL, H. & KAIDER, A. (1997). Gaining more flexibility in cox proportional hazards regression models with cubic spline functions. *Computer Methods and Programs in Biomedicine* **54**, 201–208.
- HEINZL, H. & KAIDER, A. (2006). Manual for the sas-macro rcs (version 2.0). Technical report, Institut für Medizinische Computerwissenschaften, Abteilung für Klinische Biometrie.
- HEINZL, H., KAIDER, A., & ZLABINGER, G. (1996). Assessing interactions of binary time-dependent covariates with time in cox proportional hazards regression models using cubic spline functions. *Statistics in Medicine* **15**, 2589–2601.
- HESS, K. R. (1994). Assessing time-by-covariate interactions in proportional hazards regression models using cubic spline functions. *Statistics in Medicine* **13**, 1045–1062.
- KAPLAN, E. L. & MEIER, P. (1958). Nonparametric estimation from incomplete observations. *Journal of the American Statistical Association* **53**, 457–481.
- LOUGHIN, T. M. (1995). A residual bootstrap for regression parameters in proportional hazards models. *Journal of Statistical Computation and Simulation* **52**, 367–384.

- R DEVELOPMENT CORE TEAM (2009). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0.
- ROYSTON, P. & ALTMAN, D. G. (1994). Regression using fractional polynomials of continuous covariates: Parsimonious parametric modelling. *Journal of the Royal Statistical Society* **43**, 429–467.
- ROYSTON, P., ALTMAN, D. G., & SAUERBREI, W. (2006). Dichotomizing continuous predictors in multiple regression: a bad idea. *Statistics in Medicine* **25**, 127–141.
- ROYSTON, P., AMBLER, G., & SAUERBREI, W. (1999). The use of fractional polynomials to model continuous risk variables in epidemiology. *International Journal of Epidemiology* **28**, 964–974.
- ROYSTON, P. & SAUERBREI, W. (2004). A new approach to modelling interactions between treatment and continuous covariates in clinical trials by using fractional polynomials. *Statistics in Medicine* **23**, 2509–2525.
- ROYSTON, P. & SAUERBREI, W. (2008). *Multivariate Model-building - A pragmatic approach to regression analysis based on fractional polynomials for modelling continuous variables*. John Wiley & Sons Ltd, England.
- SAS INSTITUTE (2009). *SAS/STAT 9.2 User's Guide*. SAS Institute Inc., Cary, NC, USA.
- SAS INSTITUTE (2010). *SAS 9.2 Language Reference*. SAS Institute Inc., Cary, NC, USA.
- SAUERBREI, W. (1999). The use of resampling methods to simplify regression models in medical statistics. *Applied Statistics* **48**, 313–329.
- SCHEMPER, M. & SMITH, T. L. (1996). A note on quantifying follow-up in studies of failure time. *Controlled Clinical Trials* **17**, 343–346.
- SCHOENFELD, D. (1982). Partial residuals for the proportional hazards regression model. *Biometrika* **69**, 239–241.

Zusammenfassung

Für die Analyse von Überlebensdaten ist das proportionale Hazard-Modell (PH) nach Cox (1972) das am Häufigsten verwendete Regressionsmodell. Im PH-Modell wird für einen kontinuierlichen Faktor standardmäßig eine lineare Beziehung mit dem Log Hazard angenommen. Nichtlineare Effekte von kontinuierlichen Faktoren können zum Beispiel mit Hilfe von Fraktionalen Polynomen (FP) oder Restricted Cubic Splines (RCS) berechnet werden. Bei der Analyse wird aus einem vordefinierten Set von sogenannten 'Powers' jene Kombination von Transformationen (mit diesen 'Powers') der originalen Variable ausgewählt, welche den besten Fit liefert. Konfidenzintervalle für den Log Relative Hazard werden mit der Bootstrap-Methode geschätzt, wobei der Selektionsprozess bei jedem Bootstrapdatensatz wiederholt werden muss. In dieser Masterarbeit sollen im Speziellen Interaktionen von kontinuierlichen Faktoren, wie Hämatokrit-Level, mit binären Faktoren, wie Geschlecht, untersucht werden. Dabei wird davon ausgegangen, dass der kontinuierliche Faktor unterschiedliche nichtlineare Effekte für die Ausprägungen des binären Faktors annehmen kann.

Für die Schätzung der Modelle wird neben der Maximum-Likelihood-Methode (ML) die ML-Methode mit der Firth-Korrektur (FC; Firth, 1993) verwendet. Bei der FC wird von der Likelihood ein Strafterm abgezogen. Die Bestrafung der Likelihood nach Firth ermöglicht, dass keine unendlichen Parameterschätzer im Form von Monotoner Likelihood auftreten und eine geringere Verzerrung der Schätzer in kleinen und mittleren Datensätzen, in denen keine Monotone Likelihood auftritt (Heinze u. Schemper, 2001).

Für die Analyse wurde ein neues R Package, das FPs und die Firth-Korrektur kombiniert, verwendet. In einer Simulationsstudie wurden sechs verschiedene Szenarien, die sich durch den Zensierungsanteil, die Stichprobengröße und die Form des nichtlinearen Effekts unterscheiden, generiert. Von diesen Szenarien wurde das Auftreten von Monotoner Likelihood, das Schätzverhalten der Punktschätzer und die Abweichungen der Schätzer mittels medianen Bias und dem Median der absoluten Abweichungen (MAD) untersucht. Dabei war ersichtlich, dass die ML-Methode bessere Ergebnisse beim medianen Bias lieferte, während bei der zusätzlichen Verwendung der Firth-Korrektur bessere Ergebnisse beim MAD erzielt werden konnten. Bei steigendem Zensierungsanteil und kleiner werdenden Stichproben nahm die Häufigkeit von Monotoner Likelihood zu, während die Abweichungen der Schätzer größer wurden. Zusätzlich wurde der nichtlineare geschlechterspezifische Effekt von Hämatokrit in einer Fallstudie über Venenthrombosen aufgedeckt. Dabei wurde besonderes Augenmerk auf den Vergleich der Konfidenzintervalle von FP und

RCS gelegt. Dabei war erkennbar, dass die Schätzung der Konfidenzintervalle von nichtlinearen Interaktionen durch die Verwendung der Firth-Korrektur verbessert werden kann.

Abstract

Cox's proportional hazards (PH) model is frequently used in medical biostatistics to analyze survival outcomes (Cox, 1972). For continuous prognostic factors, the standard definition of the PH model assumes a linear relationship with the log hazard. Nonlinear effects of continuous prognostic factors could be accommodated, e. g., by the use of fractional polynomials (FP) or restricted cubic splines (RCS). The FP method selects so-called 'powers' to create nonlinear transformations of the original variable, which lead to the best fit in a subsequent Cox regression. Due to the optimal selection, confidence intervals of log hazard ratios have to be estimated by the bootstrap, repeating the selection process in resampled data sets. Here, we consider estimating the interaction of a binary prognostic factor, such as type of treatment or sex, with a continuous prognostic factor, such as haematocrit level. We further assume, as is often seen in practice, that the continuous factor could exhibit different types of nonlinear effects on survival time in both levels of the binary factor.

In this context, we explored Firth's bias corrective (FC) approach for estimating the parameters of the Cox model (Firth, 1993). The likelihood penalization imposed by FC not only prevents the occurrence of infinite parameter estimates in case of monotone likelihood, but it also provides more accurate estimates in small- or moderately-sized samples without monotone likelihood (Heinze and Schemper, 2001).

Using a newly developed R package for PH regression combining the FC and FP approaches, we performed a comparative analysis of a medical study inferring the gender-specific effect of haematocrit levels on the risk of recurrence of deep vein thrombosis. We were mainly interested in the comparison of confidence intervals from FPs and RCS. We conclude that confidence interval estimation for nonlinear interaction effects can be improved by using FC. We also report on a simulation study investigating properties of FC/FP point estimates of nonlinear interaction effects. The error in the point estimates were expressed by the median bias and the median absolute deviation (MAD). In the simulation study as well as in the case study we monitored the occurrence of monotone likelihood. It was distinctly and visibly that the proportion of censoring and the sample size have an influence on the error and on monotone likelihood. We also could see that the FC leads to better results by the MAD whereas the maximum likelihood method leads to better results by the median bias.

Lebenslauf

Persönliche Daten

Maria Kohl, Bakk. rer. soc. oec.

geboren am 10. Juni 1984 in Tulln, Niederösterreich

Ausbildung

09/90–07/94	Volksschule Michelhausen, Niederösterreich
09/94–07/98	Hauptschule Heiligeneich, Niederösterreich
09/98–07/03	Bundes-Bildungsanstalt für Sozialpädagogik, St.Pölten
10/04–00/08	Bakkalaureatsstudium aus Statistik, Universität Wien
00/08–03/11	Magisterstudium aus Statistik, Universität Wien

Arbeitserfahrung

08/03–07/04	Sozialpädagogin bei „Kinder in Wien“
10/08–06/10	Tutorin für Statistik an der Fakultät für Ernährungswissenschaften an der Universität Wien
07/09–10/09	Praktikum in der Abteilung Global Drug Discovery Statistics bei der Bayer Schering Pharma AG in Berlin
seit 04/10	Biostatistikerin an der Medizinischen Universität Wien im Institut für Klinische Biometrie

Elektronischer Anhang