
Statistische Analyse
von Kaufverhaltensmodellen

Inaugural-Dissertation

zur Erlangung des Doktorgrades
an den Naturwissenschaftlichen Fachbereichen
(Mathematik)
der Justus-Liebig-Universität Gießen

vorgelegt von
Kai Kopperschmidt

Frankfurt / Main, April 2005

Dekan: Prof. Dr. Volker Metag
Gutachter: Prof. Dr. Winfried Stute (Gießen)
Prof. Dr. Erich Häusler (Gießen)
Datum der Disputation: 27.05.2005

Ich möchte mich bei Prof. Dr. Winfried Stute herzlich für die hervorragende Betreuung dieser Arbeit bedanken. Ebenso gilt mein Dank Jürgen Christmann und Dr. Hans-Georg Lilge, ACNielsen Deutschland, für die fortwährende Unterstützung.

Dr. Axel Tenbusch hat die Kooperation zwischen ACNielsen Deutschland, und der Justus-Liebig-Universität, Gießen, zu diesem Projekt initiiert, wofür ich ihm sehr zu Dank verpflichtet bin.

Dr. Gerrit Eichner und Dr. Ulrich Pötter danke ich für viele intensive Diskussionen und hilfreiche Hinweise zum praktischen Teil der Arbeit.

Inhaltsverzeichnis

1	Die Messung von TV-Werbewirkung	1
1.1	Einführung	1
1.2	Theorien zur TV-Werbewirkung	2
1.3	Verfügbare Stichproben	4
1.4	Werbewirkungsmodelle auf Makrodaten	5
1.5	Schlußfolgerungen	6
2	Kaufverhaltensmodelle auf Mikrodaten	7
2.1	Einführung	7
2.2	Unbeobachtete Heterogenität	8
2.3	Modelle für Kaufzeitpunkte	11
2.3.1	Klassifikation	11
2.3.2	Modellierung mit Poissonprozessen	12
2.3.3	Modellierung mit Markovketten	13
2.3.4	Modellierung mit Erneuerungsprozessen	15
2.3.5	Modellierung mit allgemeinen Punktprozessen	19
3	Kleinst-Quadrate-Schätzer für Zählprozesse	21
3.1	Einführung	21
3.2	Voraussetzungen für die starke Konsistenz des KQS	23
3.2.1	Der Zählprozeß	23
3.2.2	Das parametrische Modell	23
3.2.3	Das Zählmartingal	24
3.2.4	Die i. i. d. Situation	24
3.2.5	Der Kleinst-Quadrate-Schätzer	25
3.2.6	Identifizierbarkeit	25
3.3	Die starke Konsistenz des Kleinst-Quadrate-Schätzers	26
3.4	Voraussetzungen für die asympt. Normalität des KQS	27
3.4.1	Lokale Lipschitz-Stetigkeit des Modells	27
3.4.2	Momentenbedingungen an das Modell	27
3.4.3	Die standardisierende Matrix	28
3.4.4	Die asymptotische Kovarianzmatrix	28
3.5	Die asympt. Normalität des Kleinst-Quadrate-Schätzers	29
4	Die Herleitung des Kaufverhaltensmodells	30
4.1	Einführung	30
4.2	Weitere Bezeichnungen	30
4.2.1	Die beobachteten Größen	30
4.2.2	Die Sprünge der Punktprozesse	30
4.3	Die Struktur des Datensatzes	31
4.4	Die Kaufgeschichte	35
4.5	Die Jahreszeiten	35
4.6	Die TV-Werbekontakte	37
4.7	Die Soziodemographie	38
4.8	Die Modellintensität	40
4.9	Die Wahl der Startparameter	41

5	Eine Simulationsrechnung	43
5.1	Einführung	43
5.2	Die Simulation der Prozesse	43
5.3	Der Verlauf der differentiellen Intensität	43
5.4	Die Kleinst-Quadrate-Schätzung	45
5.5	Die Wahl der Startparameter	47
5.6	Die Modellanpassung	48
5.7	Die Verteilung des KQS	49
5.8	Schlußfolgerungen	53
6	Die Anpassung des Kaufverhaltensmodells	54
6.1	Einführung	54
6.2	Die Modellanpassung	54
6.3	Die Parameterschätzung	64
6.4	Interpretation der Ergebnisse	66
7	Beweis der starken Konsistenz des KQS	67
7.1	Einführung	67
7.2	Technische Eigenschaften	68
7.3	Beweisschritte	72
8	Beweis der asymptotischen Normalität des KQS	80
8.1	Einführung	80
8.2	Technische Eigenschaften	81
8.2.1	Hilfsaussagen	81
8.2.2	Straffheit des α -Prozesses	85
8.2.3	Straffheit des β -Prozesses	95
8.2.4	Straffheit des γ -Prozesses	100
8.2.5	Straffheit der Matrix Φ_n	108
8.3	Beweisschritte	109
9	Regressionsmodelle für allgemeine Punktprozesse	119
9.1	Übersicht über die Literatur	119
9.2	Die Ergebnisse von [6]	120
9.3	Die Ergebnisse von [51, pp. 58-68]	123
10	Zusammenfassung und Ausblick	127
10.1	Zusammenfassung	127
10.2	Ausblick	128
A	Technische Lemmata	129
A.1	Konvergenz in Wahrscheinlichkeit	129
A.2	Pfadweise Integrale	130
A.3	L_2 -Techniken	139
A.4	Ein starkes Gesetz für Funktionale stochastischer Prozesse	143
A.5	Straffheit bei stetigen Funktionen mit mehreren Argumenten	151

B	Implementation des KQS	158
B.1	Einführung	158
B.2	Darstellungen relevanter Größen	158
B.2.1	Der Kleinst-Quadrate-Schätzer	158
B.2.2	Der mittlere Pfad	158
B.2.3	Die mittlere Intensität	158
B.2.4	Die standardisierende Matrix	159
B.2.5	Die i. i. d. Darstellung des KQS	160
B.3	Ein R-Algorithmus zur KQS-Bestimmung	162
B.3.1	Datenaufbereitung	162
B.3.2	Bestimmung des Kleinst-Quadrate-Schätzers	168
B.3.3	Bestimmung der standardisierenden Matrix	170
B.3.4	Bestimmung der i. i. d. Terme	172
B.3.5	Bestimmung der asymptotischen Kovarianzmatrix	174

1 Die Messung von TV-Werbewirkung

1.1 Einführung

Im Jahr 2001 hat die deutsche Wirtschaft etwa 66 Milliarden DM für TV-Werbekampagnen ausgegeben, [66, p. 77]. Entsprechend drängend ist das Verlangen, diesen Ausgaben eine erzielte "Werbewirkung" gegenüberzustellen. Üblicherweise wird erwartet, daß diese Effekte als Mehrumsatz oder -absatz im Anschluß an die Werbeausstrahlung ausgewiesen werden können.

Wir beziehen uns in der vorliegenden Arbeit ausschließlich auf Konsumgüter des täglichen Bedarfs. In diesem Bereich erheben kommerzielle Marktforschungsinstitute heute Einkaufsdaten, aus denen die genauen Einkäufe und Fernsehgewohnheiten einer großen Zahl Stichprobenhaushalte hervorgeht. Die Fachliteratur zum Thema Werbewirkung ist bereits sehr umfangreich und wächst ständig weiter. Die adäquate Abschätzung eines werbebedingten Mehrumsatzes oder -absatzes stellt jedoch immer noch eine Herausforderung dar.

Als grundlegendes Problem können wir ansehen, daß Werbung darauf abzielt, das Kaufverhalten der Konsumenten zu ändern. Dieses ist jedoch erstens schwer zu messen und unterliegt zweitens wie jede menschliche Verhaltensweise einer Vielzahl von Einflüssen. Insbesondere bilden sich bei einzelnen Haushalten unterschiedliche Gewohnheiten heraus, die von der Werbung durchbrochen werden müssen, um den erwarteten Mehrkonsum zu generieren.

Weiter erschwert wird die Untersuchung dadurch, daß die ständige Konzentration auf Hersteller- und Händlerseite zur Bildung von Konzernen mit enormen Finanzmitteln geführt hat. Diesen Unternehmen gelingt es ohne weiteres, ihre TV-Werbung jederzeit und überall auszustrahlen. Dadurch gibt es praktisch keine Gruppe von Konsumenten mehr, deren Kaufverhalten von Werbung völlig unbeeinflusst wäre. Mit anderen Worten, es fehlt die "Kontrollgruppe" und die Quantifizierung eines Werbeeffekts beruht letztlich immer zu einem gewissen Grad auf prinzipiell nicht überprüfaren Hypothesen. Es scheint zum Beispiel vermessen anzunehmen, wir könnten auf der Grundlage selbst beliebig umfangreicher Daten Aussagen darüber treffen, wie sich die Marke Coca-Cola entwickelt hätte, wenn die letzten Jahrzehnte nicht von Marketingaktivitäten eines Ausmaßes geprägt gewesen wären, die dieses Getränk und sein Logo zu einem weltweit bekannten Symbol gemacht haben.

Allerdings gibt es auch eine große Zahl praktisch relevanter Fälle, in denen eine angemessene Methode durchaus praktisch verwertbare Erkenntnisse über die genaue Werbewirkung für einzelne Produkte erwarten läßt. In der vorliegenden Arbeit werden wir uns von folgendem Beispiel leiten lassen. Die großen Markenhersteller in der Warengruppe Speiseeis begleiten seit Jahren den Beginn des Sommers mit groß angelegten Werbekampagnen für ihre Produkte. Damit stellt die jährliche Verkaufsspitze von Haushaltspackungen Speiseeis in den Sommermonaten zwar sicher größtenteils einen saisonalen Effekt dar. Allerdings wird sich darin auch immer ein gewisser Anteil "Werbewirkung" verbergen. Leider ist es aber wohl ziemlich unmöglich, in Deutschland einen Konsumenten zu finden, dessen sommerlicher Eisverzehr völlig unbeeinflusst von Werbung ist.

Außerdem ist die jahreszeitliche Wirkung auf den Konsum von Speiseeis selbst keineswegs offensichtlich. So können wir uns z. B. sicher mit Recht fragen, ob eigentlich der erhöhte Eiskonsum auch während längerer Hitzeperioden unverändert anhält. Womöglich wenden sich größere Teile der potentiellen Konsumenten bald weniger süßen Erfrischungsprodukten zu.

Schließlich finden wir in diesem Beispiel auch den erwähnten Gewohnheitsaspekt wieder, den wir vielleicht etwas allgemeingültiger als Einfluß der eigenen Kaufgeschichte auf das aktuelle Kaufverhalten verstehen können. Ein naheliegender Aspekt dieses Einflusses beruht darauf, daß Haushaltspackungen Speiseeis Produktmengen enthalten, die für einige Portionen ausreichen. Dies macht erneute Käufe innerhalb sehr kurzer Zeit, z. B. weniger Tage, recht unwahrscheinlich.

Ziel der vorliegenden Arbeit ist es, eine stochastische Modellklasse für Regressionen zu diskutieren, die geeignet erscheint, einige Einblicke in die umrissene ökonomische Situation zu bringen. In diesem Kapitel werden wir uns jedoch zunächst um die Grundlagen kümmern. In Abschnitt 1.2 stellen wir einige grundlegende Ideen zur Werbewirkung von TV-Kampagnen vor. Vor diesem Hintergrund betrachten wir in Abschnitt 1.3 die verschiedenen verfügbaren Datenbasen, die in der Literatur als Makro- und Mikrodaten bezeichnet werden. In Abschnitt 1.4 werfen wir einen kurzen Blick auf die vorhandenen Modelle auf Makrodaten. Wir vertiefen diese Betrachtung nicht, da uns Mikrodaten als der geeignetere Ausgangspunkt für unsere Arbeit erscheinen. Diesen Standpunkt diskutieren wir im Abschnitt 1.5.

1.2 Theorien zur TV-Werbewirkung

Viele Theorien zur Werbewirkung beruhen substantiell darauf, daß die Adressaten die Werbung bewußt oder unbewußt verarbeiten. Betrachten wir als Beispiel die beiden verbreiteten Konzepte "AIDA" und "ATR", [28, p. 167]:

Awareness → **I**nterest → **D**esire (oder Conviction) → **A**ction

und

Awareness → **T**rial → **R**einforcement

Beide enthalten lediglich *einen* Schritt, der von einem äußeren Beobachter zweifelsfrei festgestellt werden kann, nämlich "action" bzw. "trial". Deshalb erscheinen Modelle, die keinerlei Rücksicht auf Verhaltensdetails der Konsumenten nehmen, zur Untersuchung dieser Theorien schlecht geeignet. Demgegenüber hat die explizite Berücksichtigung menschlichen Verhaltens im Modell weitere Vorteile:

[54, p. 42]

The benefits of more behavioral detail include the following. One is that we force ourselves to be explicit about at least some of the stages in the behavioral process. [...] more detail provides opportunities for debate and reflection. Another advantage is that if a

model does not perform well, we want to know how we can improve it. With more detail there is greater opportunity to identify the model's weaknesses.

Da mentale Prozesse eine so entscheidende Rolle in der Werbewirkungstheorie spielen, ist die psychologische und kognitive Literatur zum Thema besonders umfangreich. Wir veweisen auf [23, pp. 29] für eine Übersicht der Feldstudien und experimentellen Arbeiten.

Andererseits wird auch immer wieder prinzipieller Zweifel an den Möglichkeiten der Werbung laut, so z. B. von dem angesehenen Forscher Ehrenberg im Journal of Advertising Research (Der Beitrag scheint eine Wiederauflage eines Artikels von 1974 zu sein):

[29, pp. 40]

The usual reason why people buy things is that they want them. Anyone who has washed dishes knows that the demand for nonstick frying pans or dishwashers did not have to be created. [...] There is no need to suppose that the role of advertising here is fundamental. It is a peculiar form of snobism to suppose that if other people want to smoke cigarettes, to smell nice, to have bathrooms, or to drive in motor cars, it is only because they have been manipulated by advertising. [...] The glossy images of affluence shown in advertisements and in the media generally reflect a real demand. Eliminating advertising would not eliminate the demanding consumer.

Wir wollen diese Meinung im folgenden nicht diskutieren. Es geht uns lediglich darum zu zeigen, daß bisher keine Theorie zur Werbewirkung einen unangefochtenen Stand erreicht hat. Dies gilt in vielleicht noch stärkerem Ausmaß für die praktischen Aspekte der Werbewirkungsforschung. Bei einer anderen Gelegenheit notiert Ehrenberg: *"The idea of giving a seminar on advertising at a leading US advertiser in Cincinnati was recently ruled out on humanitarian grounds: 'Do you want to get lynched?' I had to talk about promotions instead."* ([28, p. 167])

Wir strukturieren die konkurrierenden Theorien zur Werbewirkung nach [49] (zitiert nach [28, p. 167], wie auch die folgenden Textstellen). Dort wird nach "starken" und "schwachen" Werbewirkungstheorien unterschieden. Für die Einordnung ist entscheidend, welchen Einfluß die Theorie einer Werbung zugesteht, die nicht von vorheriger Erfahrung des Konsumenten mit dem Produkt begleitet wird. Das oben erwähnte AIDA Modell ist ein Beispiel für die *"traditional 'strong' theory of advertising"*, da sie unterstellt, die Werbung allein könne zum Kauf führen (*"turning a virgin non-user into a user"*). Demgegenüber repräsentiert das ATR Modell eine schwache Werbewirkungstheorie: *"Advertising can first arouse awareness and interest, nudge some customers towards a doubting first trial purchase (with the emphasis on trial, as in 'maybe I'll try it'), and then provide some reassurance and reinforcement after that first purchase."*

Sobald für einen Haushalt mehrere Werbekontakte und Einkaufsakte verzeichnet werden können, kommen viele weitere Aspekte möglicher Interaktionen hinzu: Denken wir etwa an Werbung für Bier in den Spielpausen von Fußballüber-

tragungen. Hier sorgt die Orientierung der Sendezeit an der Zielgruppe dafür, daß die erreichten Konsumenten vermutlich ohnehin einen höheren Verbrauch an den beworbenen Produkten haben als der Bevölkerungsschnitt. In diesem Sinne könnten wir die Auffassung vertreten, daß die vorangegangenen Einkäufe in ihrer Gesamtheit die Platzierung der Werbung im TV-Programm und damit teilweise die folgenden Werbekontakte determiniert haben. Auch in dem von uns besonders beachteten Beispiel des Speiseeises finden wir mit der jahreszeitlichen Orientierung eine bewußte Entscheidung der Hersteller, ihre Werbung hauptsächlich im Einklang mit anderen Faktoren zu schalten, die die Kaufbereitschaft der Empfänger erhöhen.

Schließlich drängt sich die Frage nach der gemeinsamen Wirkung mehrerer Werbekontakte auf, dem Effekt des "Werbedrucks". In der Praxis verspricht man sich oft etwas von dem Zusammenwirken zweier Kontakte. Dies ist wohl ein Grund dafür, daß häufiger innerhalb eines Werbeblocks dasselbe Produkt zwei Mal beworben wird. Dabei nimmt der zweite Spot oft die Form eines kurzen "Reminders" an. Entsprechend liegt es allerdings auch nahe, einen Sättigungseffekt der Werbewirkung zu vermuten, der bei immer höherer Werbekontaktzahl die Gesamtwirkung einem asymptotischen Wert zustreben läßt.

Die Untersuchung der Werbewirkung wird noch durch viele weitere Aspekte erschwert, auf die wir hier nicht im Detail eingehen. So scheint z. B. auch fragwürdig, ob Kaufintention und tatsächliches Verhalten immer übereinstimmen, vgl. [80]. [23] klassifiziert Effekte von Werbung auf Markenwahl und Kaufmenge, die wir jedoch zur Vereinfachung hier ebenfalls ausklammern.

1.3 Verfügbare Stichproben

Die unterschiedlichen Datenquellen, die für die panelbasierte Marktforschung genutzt werden können, werden in [38] ausführlich beschrieben. Der Begriff "Panel" bezeichnet dabei Stichproben, für die eine möglichst mehrjährige, kontinuierliche und gleichartige Beobachtung angestrebt wird. Diese Art der Erhebung wird manchmal auch als "Längsschnitt" bezeichnet. Einmalige Beobachtungen einer Stichprobe, wie sie sich zum Beispiel häufig bei Befragungen ergeben, werden demgegenüber als "Querschnitt" bezeichnet. Unsere Darstellung in Abschnitt 1.2 verdeutlicht die Vorzüge panelbasierter Modelle für die Werbewirkung.

Die zwei verbreitetsten Arten von Paneln sind gemäß [38, pp. 13 und pp. 25] und [54, p. 310] (Einzel-)handelspanels einerseits und Verbraucherpanel andererseits. Ein Handelspanel besteht aus Stichprobengeschäften, deren Abverkäufe beobachtet werden. Ein Verbraucherpanel erfaßt statt der Verkäufe einzelner Geschäfte die Einkäufe einzelner Haushalte. Daraus resultieren typische Unterschiede in der Art der verfügbaren Paneldaten, die von [44, p. 16 und pp. 35] aggregierte (Makro-)daten bzw. disaggregierte (Mikro-)daten genannt werden. In dieser Bezeichnung kommt zum Ausdruck, ob sich die Daten auf mehrere Konsumenten gleichzeitig oder auf einzelne Haushalte beziehen. Beispiele für aggregierte Daten, die gut aus Handelspaneln gewonnen werden können, sind Umsatz, Absatz oder Marktanteil eines Produkts. Interessieren wir uns hingegen für mittlere gekaufte Mengen je Haushalt und ähnliche Größen, so betrachten

wir Mikrodaten, die sich direkt aus Verbraucherpaneln bestimmen lassen.

Zusätzlich zu Informationen über das Kaufverhalten benötigen wir offenbar Daten zur Werbesituation. Derartige Größen werden als Mikrodaten in Fernsehzuschauerpaneln erhoben, vgl. [38, pp. 78]. Sie sind mitunter von eigenständigem Interesse, so z. B. bei der Ermittlung von Einschaltquoten. Außerdem existieren sie jedoch auch in Verbindung mit Verbraucherpaneln und ermöglichen so die Erhebung von Werbekontakten und Einkäufen "aus einer Quelle". Ein Beispiel ist das "Single Source Panel" von ACNielsen, Deutschland, auf dem unser Anwendungsbeispiel in Kapitel 4 beruht. Wir gehen auf diese Stichprobe in Abschnitt 4.3 genauer ein. Als Makrodaten zur Beschreibung der Werbesituation werden häufig "GRPs" (gross rating points) eingesetzt. In dieser Kenngröße werden Reichweite und Werbedruck innerhalb fixierter Zeiträume miteinander verrechnet.

1.4 Werbewirkungsmodelle auf Makrodaten

Bereits 1979 hat [55] (zitiert nach [44, pp. 196]) eine zusammenfassende Darstellung zu Werbewirkungsmodellen auf Makrodaten verfaßt. Dabei wurden theoretische Überlegungen und empirische Untersuchungen zugrundegelegt. Im Werbezeitraum selbst wird gefordert, daß der Absatz monoton wachsend vom Werbedruck abhängt, ab einem gewissen Wert sollte außerdem ein konkaver Funktionsverlauf angenommen werden. Dies trägt dem Sättigungseffekt Rechnung, den wir in Abschnitt 1.2 diskutiert haben. Mit zeitlichem Abstand zur Werbung sollte ihr Effekt nachlassen. Er braucht nicht vollständig zu verschwinden.

Für den genauen funktionalen Verlauf der Werbewirkung über der Zeit wurden viele Vorschläge gemacht und in verschiedenen Situationen untersucht, für eine Übersicht verweisen wir auf [44, pp. 32] und [54, pp. 87]. Zu den verbreitetsten zählt ein exponentieller Verlauf mit negativem Exponenten, der sich in diskreter Zeit als geometrischer Abfall darstellt. Diese Formalisierung ist in der Werbeforschung auch als "adstock" oder Koyckmodell bekannt, nach einem Autor, der sie in ökonomischem Kontext relativ früh eingesetzt hat, [50].

[13, p. 404] setzt eine geometrisch abfallende Werbewirkung in eine Markovkette mit zwei Zuständen ein. Jedem Zeitschritt entspricht eine Periode, innerhalb derer Werbung geschaltet werden und ein Marktanteil für das Fokusprodukt sinnvoll beobachtet werden kann, z. B. ein Jahr oder wohl zumindest mehrere Monate. Die vier Übergangswahrscheinlichkeiten der Markovkette geben an, wie sich der Marktanteil des Produkts zur Folgeperiode ändert: Der Marktanteil p wird dabei als $(p, 1 - p)$ kodiert und an die Übergangsmatrix multipliziert:

$$(p, 1 - p) \cdot \begin{pmatrix} p_{11} & 1 - p_{11} \\ 1 - p_{22} & p_{22} \end{pmatrix}$$

Der Marktanteil der Folgeperiode ergibt sich nun als $p_{11}p + (1 - p_{22})(1 - p)$. [13] schreibt die Übergangswahrscheinlichkeiten als parametrische Funktionen des Werbedrucks in den vergangenen Perioden und untersucht optimale Strategien zum Einsatz des Werbebudgets. Allerdings läßt der Autor als mögliche Strategien lediglich die gleichmäßige Verteilung sowie schnelle und langsame Pulsierung zu.

Das Modell wird nicht auf tatsächlichen Daten geschätzt, es findet nur eine formale Untersuchung unterschiedlicher Werbestrategien statt. In der Praxis dürfte es sich auch als schwierig erweisen, Daten entlang einer ausreichenden Zeitskala zu sammeln, da einzelne Zeitschritte ja in Monaten gemessen werden. Außerdem ist es fraglich, ob die Marktsituation über so lange Zeiträume tatsächlich ausreichend unverändert ist, um die Annahme eines konstanten Marktanteils bei fehlender Werbung zu rechtfertigen. Dann dürften sich über Jahre weder die Verbrauchsgewohnheiten der Konsumenten noch die Wettbewerbssituation ändern.

Der Einsatz einer Markovkette zur Beschreibung des Marktanteils (anstelle der Kaufentscheidung einzelner Haushalte) in längeren Perioden benötigt stark aggregierte Daten sowohl des Absatzes als auch des Werbeeinflusses. In diesem Rahmen lassen sich die verhaltensorientierten Werbetheorien aus Abschnitt 1.2 nur sehr bedingt abbilden.

Modelle dieser Art werden in der Literatur deshalb auch deutlich kritisiert. So finden wir z. B. folgende Aussage, die im ursprünglichen Zusammenhang sogar die Modellierung mit Mikrodaten einschließt:

[54, p. 239]

Researchers who applied the Markov model may not have been overly concerned with the restrictive nature of the underlying assumptions. They have used the model as a tool that captures some of the dynamics of markets. Generally speaking, the Markov model has often been applied because of its performance as a predictor of dynamic aggregate market behavior, not because it describes how individual consumers behave.

1.5 Schlußfolgerungen

In den meisten Werbewirkungstheorien spielen Kaufverhalten und Gewohnheit der Konsumenten eine grundlegende Rolle. In Makrodaten fehlt jedoch die Verbindung zwischen den einzelnen Einkaufsakte und Werbekontakten eines Haushalts. Deshalb sind den Möglichkeiten, Werbewirkungstheorien in Modellen auf aggregierten Daten umzusetzen, enge Grenzen gesetzt. Der häufig eingesetzten Kenngröße der GRPs wird in der Literatur eher der Stellenwert einer vorläufigen Hilfsgröße eingeräumt:

[44, p. 207]

Die Analyse von Single-Source Daten erlaubt die Ermittlung von durch unterschiedliche Medienbelegung bedingten Kaufwahrscheinlichkeiten bzw. Käufen und damit einer ökonomischen Werbewirkung, während Kontaktzahlen letztlich nur Ersatzkriterien für Absatzeffekte darstellen. Es ist zu erwarten, daß auf diese Weise gemessene ökonomische Effekte von Mediensaltungen in bestimmten Produktbereichen die heute in der Streuplanung verwendeten Kontaktzahlen [...] ablösen werden.

Wir betrachten derartige Modelle deshalb nicht weiter im Detail, sondern wenden uns im folgenden Kapitel der ökonometrischen Literatur zu, die Kaufverhalten mit Hilfe von Mikrodaten untersucht.

2 Kaufverhaltensmodelle auf Mikrodaten

2.1 Einführung

Die verstärkte Verfügbarkeit von Mikrodaten in den letzten Jahren wurde von der Literatur überwiegend geradezu euphorisch aufgenommen. [54] zählt einige Gründe dafür auf:

[54, p. 158]

The increasing availability of scanner-type data at the household level has led to a strong interest in models of individual buyer behavior and methods for aggregation. The unique opportunities for understanding consumer behavior and deriving implications for marketing actions include:

- *analyzing household brand switching and brand loyalty over time;*
- *monitoring new brand performance through measures such as trial and penetration rates;*
- *modeling household heterogeneity in purchase behavior and exploiting household differences for segmentation and targeting marketing actions, such as direct mail campaigns;*
- *understanding the impact of marketing variables on the timing of households purchases and stockpiling behavior for the planning of promotions by retailers;*
- *testing theories of consumer behavior.*

Zweifellos ermöglichen Mikrodaten einen vielversprechenden Blick ins Detail. Allerdings hat die Genauigkeit und Fülle dieser Daten auch die Diskussion über modelltechnische Notwendigkeiten zur adäquaten Abbildung an vielen Stellen neu entfacht. Einer der Punkte, die in fast allen Veröffentlichungen zum Kaufverhalten auftauchen, ist die Heterogenität der Konsumenten. Im Zitat oben wird sie als drittes angesprochen. Dort ist jedoch eigentlich primär von "beobachteter Heterogenität" die Rede. Mit diesem Begriff werden in der Literatur Unterschiede im Kaufverhalten der Haushalte bezeichnet, die mit Änderungen in beobachteten Merkmalen dieser Haushalte einhergehen. Je nach Warengruppe können z. B. die Haushaltsgröße oder das Einkommen die Einkäufe stark beeinflussen.

Die Vielzahl naheliegender und aufgedeckter Zusammenhänge zwischen Eigenschaften eines Haushaltes und seinem Verhalten hat die Ansicht bestärkt, daß in sozialwissenschaftlichen Anwendungen niemals ein vollständiger Satz von Einflußgrößen erhoben werden kann:

[52, p. 306]

[Dealing with] data that are not homogeneous [...] is a natural preoccupation of a social scientist. The diversity of people is surely greater than can be accounted for by the values of five or ten regressor variables.

Unterschiede im Verhalten einzelner Haushalte, die keiner vorliegenden Einflußgröße inhaltlich zugeordnet werden können, heißen "unbeobachtete Heterogenität". Wir diskutieren die Ansichten und Ansätze zu diesem Thema der Modellierung auf Mikrodaten in Abschnitt 2.2.

Eine weitere Folge der Datenfülle, die heute in Verbraucherpaneln zur Verfügung steht, ist die Zerlegung des Kaufverhaltens in einzelne Aspekte und deren separate Umsetzung in Modelle. Dabei hat sich spätestens seit [39] die Aufteilung Kaufzeitpunkt, Markenwahl und Mengenwahl in der Literatur weitgehend durchgesetzt. Die eingesetzten Modelle werden meist auch in dieser Reihenfolge aufeinander bedingt: So werden dem Markenwahlmodell lediglich Situationen vorgelegt, in denen bereits bekannt ist, daß der Haushalt ein Produkt der Warengruppe kauft. Das vorgeschaltete Modell zur Bestimmung des Kaufzeitpunktes wird dann auch genauer als "category purchase timing model" bezeichnet.

Um unsere Untersuchung möglichst übersichtlich zu halten, werden wir uns auf ein solches Modell für die Kaufzeitpunkte in einer Warengruppe beschränken.¹

Wir betrachten die Vielzahl unterschiedlicher Modelle, die für eine solche Verwendung in der Literatur bereits angegeben wurden, in Abschnitt 2.3. Nach einer einleitenden Klassifikation in Abschnitt 2.3.1 diskutieren wir in den Abschnitten 2.3.2 bis 2.3.4 Ansätze, die auf Poissonprozessen, Markovketten oder Erneuerungsprozessen beruhen.

Im Anschluß belegen wir in Abschnitt 2.3.5 unsere These, daß Regressionsmodelle für Punktprozesse mit stetigen stochastischen Kompensatoren eine flexible und angemessene Weiterentwicklung für alle früheren category purchase timing Modelle auf Mikrodaten darstellen.

2.2 Unbeobachtete Heterogenität

Die Möglichkeit zur Modellierung auf disaggregierten Daten hat in der Literatur die Frage aufgeworfen, ob es überhaupt angemessen ist, in einer Modellgleichung allen Haushalten dieselben Parameterwerte zuzuweisen. Im Rahmen der Verweildaueranalyse hat sich diese Unsicherheit zu der Sorge zugespitzt, daß Hazardraten einer Population zu gravierenden Fehlschlüssen Anlaß geben könnten:

[1, pp. 1]

Hazard rates play a fundamental role in survival analysis. Although useful, they may not be easy to understand. [...] Rates of divorce, for instance, as measured from time of marriage, first increase and

¹Die gesamte weitere Arbeit sollte aber direkt auf Modelle erweiterbar sein, die auch Marken- und Mengenwahl berücksichtigen, vgl. die Anmerkung in Abschnitt 10.2.

then decrease somewhat [...] This has resulted in speculation about a crisis in the typical marriage after a few years leading either to divorce or to consolidation of the marriage. Within the context of frailty theory a different explanation has been given. It has been pointed out that the hazard rate is not merely a measure of the development of risk in a single individual, but is also influenced by selections among individuals. The most frail individuals will tend to fail first, leaving the more robust ones. This will imply that a population hazard may decrease even though the individual rates are increasing.

Entsprechende Interpretationsprobleme lassen sich z. B. auch für die Markenwahl angeben: Betrachten wir eine Stichprobe, die zu einer Hälfte aus Exklusivverwendern der Marke A und zur anderen Hälfte aus Konsumenten besteht, die ausschließlich Marke B kaufen. Schätzen wir nun auf der Grundlage aller Einkaufsakte einen Markenwahlparameter, so werden wir den Quotienten aus der Anzahl Einkaufsakte der A-treuen und der B-treuen Kunden erhalten. Dieser Quotient gibt jedoch für keinen Haushalt des Samples die "individuelle" Markenwahlwahrscheinlichkeit an. Die ist ja nach Annahme entweder 0 oder 1.

Überlegungen dieser Art haben dazu geführt, daß die weit überwiegende Zahl von veröffentlichten Kaufverhaltensmodellen auf Mikrodaten eine der folgenden drei Methoden zur Berücksichtigung unbeobachteter Heterogenität verwendet:

- "Fixed Effects": Bei diesem Ansatz wird jedem Stichprobenhaushalt ein eigener Parameter zugeordnet, der zusätzlich zu den bereits vorhandenen Modellparametern geschätzt werden muß. Dies erhöht die Anzahl Parameter enorm. [34] verwenden dieses Verfahren als einzige unter den betrachteten Autoren. Ihr Vorgehen beruht darauf, daß sie einen Maximum-Likelihood-Schätzer verwenden. Die Bedingung an die erste Ableitung der Likelihood ergibt einen globalen, analytischen Zusammenhang zwischen den Maximum-Likelihood-Schätzungen der eigentlichen Modellparameter und denen der Haushaltsparameter. Letztere können daraufhin vor der Schätzung aus der Likelihood eliminiert werden. Dieses Verfahren wird als "concentrated likelihood" bezeichnet.
- "Random Effects": Bei diesem Verfahren werden die fixed effects durch haushaltsspezifische Realisationen einer Zufallsvariablen ersetzt. Lediglich deren Verteilung bleibt zu bestimmen. Die wird entweder parametrisch oder, nach einer Idee von [41], nichtparametrisch geschätzt. Dem nichtparametrischen Verfahren wird dabei die Anzahl Stützpunkte vorgegeben. Deren Lage und Gewichte bestimmt das Modell während der Schätzung. Um die Anzahl Stützpunkte zu bestimmen, werden Verfahren zum Modellvergleich herangezogen, z. B. Akaiikes Informationskriterium ([3]) oder Bayes' Informationskriterium. [76] verwenden diesen Ansatz.
- "Random Coefficients": Dieser Ansatz ist den "random effects" sehr ähnlich. Die Bezeichnung in der Literatur ist nicht einheitlich. Bei diesem Verfahren werden keine Haushaltsparameter eingeführt. Stattdessen werden die eigentlichen Modellparameter selbst als zufällig angenommen. Es

gibt wiederum parametrische und nichtparametrische Versionen dieses Ansatzes. Mit diesem Konzept ist es prinzipiell möglich, Kovarianzstrukturen zwischen Modellparametern zu schätzen. [47] untersuchen auf diese Weise den Zusammenhang zwischen Markenpräferenzen und der Reaktion auf Marketingaktivitäten. In der Mehrzahl der Anwendungen wird jedoch Unabhängigkeit der Verteilungen unterstellt. [21], [40], [17], [34], [36] verwenden Γ -verteilte random coefficients, [21], [4] setzen Normalverteilungen an, [21], [17], [47], [18], [20], [22] benutzen einen nichtparametrischen Ansatz. Weitere Verteilungen wurden bereits vorgeschlagen, vgl. [54, p. 224 und pp. 233].

Von den betrachteten Artikeln verzichten demgegenüber lediglich [2], [23], [46] und [75] auf die explizite Berücksichtigung unbeobachteter Heterogenität.

In vielen dieser Artikeln werden verschiedene Verfahren zur Berücksichtigung unbeobachteter Heterogenität miteinander verglichen, [21, p. 423] setzt dafür z. B. einen Spezifikationstest von [9] ein. Im Kontext der Verweildaueranalyse stellt [77] weitere verwandte Konzepte vor.

Wir gehen darauf hier nicht weiter ein, da einige generelle Überlegungen Zweifel an der Verwendung der vorgestellten Konzepte wecken: Zunächst einmal stellt ein "verbessertes" model fit durch Berücksichtigung der unbeobachteten Heterogenität unabhängig davon, wie er genau gemessen wird, keine Überraschung dar. Alle vorgestellten Methoden führen nämlich letztlich neue Parameter in das Modell ein, selbst die Positionen und Gewichte eines nichtparametrischen Ansatzes sind ja nichts anderes. Es ist intuitiv einsichtig, daß die gewonnene Flexibilität im Regelfall die Anpassung an die Daten verbessert.

Sofern die betrachtete Stichprobe der unterstellten Form der Heterogenität genügt, haben wir es hier sicher mit besonders interessanten Modellerweiterungen zu tun. Ob dies aber tatsächlich der Fall ist, kann häufig gar nicht festgestellt werden: Unbeobachtete Heterogenität ist meistens nicht identifizierbar. Denken wir z. B. an das einleitende Zitat über die Hazardraten für Ehescheidungen. Die erhobenen Daten liefern keinen Hinweis darauf, ob wir es mit einer homogenen Population zu tun haben, die das unterstellte Krisenjahr erlebt, oder ob sich zwei Arten von Partnerschaften herausbilden. Damit ist die Annahme unbeobachteter Heterogenität eine willkürliche Entscheidung des Modellierers, die jedenfalls nicht anhand der daraufhin geschätzten Modelle überprüft werden kann. Ob man diese Vermutung für so tragfähig halten sollte wie [34], die im Anschluß an die Schätzung der fixed effects für ihre Haushalte beginnen, nach Merkmalen zu suchen, die wiederum die unbeobachtete Heterogenität erklären, sei dahingestellt. Die Gefahr, ein Modell ungeeignet zu spezifizieren, wird durch die Verwendung eines Verfahrens zur Berücksichtigung unbeobachteter Heterogenität nicht kleiner. In der betrachteten Literatur zur Kaufverhaltensforschung weisen sowohl [76, p. 34] als auch [21, p. 418] auf die Gefahr inkonsistenter Parameterschätzungen als Folge fehlspezifizierter unbeobachteter Heterogenität hin.

Kommen wir nochmal auf das Beispiel der Ehescheidungen zurück. Hier wurde argumentiert, daß die geschätzte Hazardrate nicht ausschließlich ein Maß für das Scheidungsrisiko individueller Paare sei. Dies trifft jedoch auf sehr viele stochastische Konzepte zu, ohne daß ihre Bedeutung zur Beschreibung von Da-

tensätzen deshalb bestritten würde: Die erwartete Augenzahl von 3.5 läßt sich sicher als Eigenschaft des einzelnen Würfelwurfs verstehen, da es das erste Moment seiner Verteilung ist. Die Wichtigkeit dieses Wertes zum Verständnis des Würfelspiels wird gerade nicht dadurch geschmälert, daß er sich als Augenzahl eines einzelnen Wurfs gar nicht ergeben kann. Es ist natürlich möglich, daß die Gefahr von Fehlinterpretationen bei der Beurteilung zeitlicher Verläufe besonders hoch ist; diese Feststellung bedeutet aber keine Abwertung der beschreibenden statistischen Größen.

Schließlich sollten wir noch festhalten, daß bei Verwendung der Modellergebnisse auf Populationsniveau die unbeobachtete Heterogenität ohnehin wieder "herausintegriert" wird, was ihre praktische Relevanz weiter verringert.

2.3 Modelle für Kaufzeitpunkte

2.3.1 Klassifikation

In der Literatur werden sehr viele verschiedene Modelle zur Beschreibung des Kaufverhaltens eingesetzt. Die publizierten Ansätze zur Bestimmung des Kaufzeitpunktes sind dabei besonders zahlreich. Bereits 1970 hat [58] (zitiert nach [54, p. 222]) anhand der Bedeutung der modellierten Zielgröße eine Einteilung vorgenommen, die wir hier wiedergeben:

- "purchase decision models": Hier nimmt die abhängige Variable die Werte 0 oder 1 an und markiert so die Situationen, zu denen gekauft wurde. Bezugsgröße dabei können entweder feste Perioden oder Einkaufsgänge sein. [17] modelliert so die Wahrscheinlichkeit, daß bei einem gegebenen Einkaufsgang in der betrachteten Warengruppe eingekauft wird.
- "purchase incidence models": In diesen Modellen zählt die abhängige Variable die Anzahl Einkaufsakte innerhalb einer festen Periode. [63] verwenden einen derartigen Ansatz.
- "purchase timing models": Die abhängige Variable gibt hier die Länge des Kaufintervalles, also des zeitlichen Abstandes zwischen zwei aufeinanderfolgenden Einkaufsaktens desselben Haushaltes an. Die Zeitachse kann dabei diskret oder stetig sein.

Diese Aufteilung ist nicht ganz eindeutig, wie in [54, p. 226] angemerkt wird. Wenn wir ein purchase incidence Modell für ein Intervall $I \subset \mathbb{R}$ derart spezifizieren, daß wir für jedes Teilintervall $[a, b] \subset I$ die Verteilung der Anzahl Einkaufsakte angeben können, so haben wir damit auch gerade ein purchase timing Modell definiert. Ein einfaches Beispiel zweier solcher Modelle sind eine Poisson-verteilte Anzahl Einkaufsakte und exponentialverteilte Kaufintervalle, siehe z. B. [52, pp. 85]. Dabei müssen wir allerdings auch noch annehmen, daß die Kaufintervalle i. i. d. sind.

Für praktische Belange ist diese Mehrdeutigkeit nebensächlich, da purchase incidence Modelle meistens nur für disjunkte Intervalle des Untersuchungszeitraumes I formuliert werden, z. B. für jede Kalenderwoche.

Außerdem werden wir am Beispiel des NBD Modells in Abschnitt 2.3.2 sehen, daß ein purchase incidence Modell von der Literatur gelobt werden kann, während seine Interpretation als purchase timing Modell heftig kritisiert wird. In solchen Fällen ist also die Zuordnung zu einer der drei Klassen ganz entscheidend für die Bewertung durch Modellierer.

Interessanter für Anwendungen sind Überlegungen wie in [73]. Der Autor zeigt den engen Zusammenhang zwischen diskreten Hazardraten und sequentiellen Wahlmodellen der Logit- und Probit-Gestalt auf. Er wurde von [61] auf Arbeitsmarktmodelle angewendet, jedoch meines Wissens bisher nicht auf Kaufverhaltensmodelle.

Abschließend betrachten wir nun noch kurz die Bedeutung der Bezugsgröße "Kaufgelegenheit" oder "Einkaufsgang" für purchase decision Modelle. [17, p. 189, (6)] verwendet ein solches Modell zusammen mit nachgeschalteten Marktwahl- und Mengenmodellen.² Der Autor selbst bemerkt, daß eine Untersuchung der Effekte von Marketingaktivitäten mit diesem Modell nur bedingt auf die gegebenen Kaufgelegenheiten durchgeführt werden kann, [17, p. 190]. Falls also z. B. Zeitungsanzeigen oder TV-Werbekontakte zusätzliche Einkaufsgänge generieren, wäre dies nicht erkennbar.

2.3.2 Modellierung mit Poissonprozessen

Ein Ausgangspunkt für die Entwicklung von Kaufzeitpunktmodellen war die Annahme, daß die Anzahl Einkaufsakte eines Haushaltes in einer Warengruppe während eines festen Zeitraumes einer Poissonverteilung folgt. Der Parameter dieser Verteilung kann auch als gemeinsame konstante Einkaufsrate der Haushalte interpretiert werden. In der Literatur wurde dieser Ansatz jedoch zunächst primär als purchase incidence Modell verstanden. Um der unbeobachteten Heterogenität der Haushalte Rechnung zu tragen, hat [26] (zitiert nach [54, p. 222]) in diesem Zusammenhang die Verwendung eines über die Haushalte Γ -verteilten Parameters der Poisson-Verteilung vorgeschlagen. Damit ist die Anzahl Einkaufsakte negativ binomialverteilt.

Weitere Beiträge zur Kaufverhaltensforschung mit diesem Ansatz, der als NBD Modell in die ökonomische Literatur eingegangen ist, stammen unter anderem von [15], [25], [27] und [60]. [68] und [78] haben die Γ -Verteilung durch andere parametrische random effects ersetzt.

Ein inhaltliches Problem ergibt sich daraus, daß sowohl die Poisson- als auch die negative Binomialverteilung bei wachsender Intervalllänge letztlich jedem Haushalt einen Einkaufsakt zuschreiben. Da dies für viele Warengruppen unrealistisch ist, wurden "zero-inflated" Versionen dieser Verteilungen eingeführt, siehe [54, p. 224]. Diese Verteilungen haben einen weiteren Parameter, der den Anteil der Population angibt, der niemals kauft. Die Einkaufsakte der restlichen Stichprobe sind wie bisher Poisson- bzw. negativ binomialverteilt.

Die Fähigkeit dieser recht einfachen Modelle, sich an die tatsächliche Anzahl Einkaufsakte unterschiedlichster Warengruppen anzupassen, wurde in der Lite-

²Allerdings bezeichnet der Autor es abweichend von unserer Konvention als "purchase incidence model".

ratur mehrfach deutlich hervorgehoben, z. B. von [40, p. 1] und [54, p. 224]. Ein Schwachpunkt ist jedoch, daß die Beeinflussung des Einkaufsverhaltens durch punktuelle Marketingaktivitäten innerhalb des betrachteten Zeitraumes schwer abzubilden ist. Zwar können die Parameter der verwendeten Verteilungen durch parametrische Funktionen von Einflußgrößen ersetzt werden, vgl. [54, p. 225]. Die Zeitabhängigkeit dieser Einflüsse widerspricht jedoch in gewisser Weise der Funktion der Verteilungsparameter, die ja gerade zeitlich homogene Poissonprozesse bzw. Mischungen davon charakterisieren.

Als purchase timing Modell war dieser Ansatz aus ganz ähnlichen Gründen heftiger Kritik ausgesetzt. So stellt [60] fest, daß die resultierende Exponentialverteilung für die Kaufintervalle die tatsächliche Tendenz zur Ausbildung von Mustern im Einkaufsverhalten realer Haushalte nicht nachvollziehen kann. Diese Unzulänglichkeit hat unter anderem die Verwendung verweildaueranalytischer Modelle motiviert, auf die wir in Abschnitt 2.3.4 zu sprechen kommen.

Eine etwas andere Verallgemeinerung wurde von [30] vorgeschlagen. Dabei wechseln die Haushalte ihre ansonsten konstante Rate nach jedem Einkaufsakt mit einer gewissen Wahrscheinlichkeit. [74] ordnet das NBD Modell in die Klasse der $M/G/\infty$ queues ein.

2.3.3 Modellierung mit Markovketten

Markovketten wurden wie Poissonprozesse bereits sehr früh in der Kaufverhaltensforschung mit Mikrodaten eingesetzt. Das mag daran liegen, daß diese Modellklasse bereits vorher für Makrodaten verwendet wurde und den Anwendern daher vertraut war:

[54, p. 158]

Some models have been developed which have about the same structure at the individual (micro) level and the aggregate (macro) level. For example, Markov models are often used to accommodate the idea that the last brand chosen (in period t) affects the current purchase (in period $t + 1$).

Klassische Markovketten scheinen jedoch besser als Markenwahlmodell geeignet denn als Kaufzeitpunktmodell. Sie unterstellen nicht nur Stationarität (entlang der Zeitachse) und Homogenität (über die betrachteten Haushalte) in der Kaufentscheidung. Dazu kommt noch die unrealistische Annahme, daß jeder Haushalt in den festen Zeiträumen genau einmal kauft, vgl. [54, pp. 237]. Die sehr kritische Meinung von [54, p. 239] zur Verwendung von Markovketten haben wir bereits am Ende des Abschnitts 1.4 im Zusammenhang mit Modellen auf Makrodaten zitiert.

In der Literatur werden sowohl die Stationarität als auch die Homogenität der ursprünglichen Modelle stark aufgeweicht: [48], [53, p. 183] und [65] haben parametrische und nichtparametrische Verfahren zur Berücksichtigung unbeobachteter Heterogenität angewendet. Für Möglichkeiten, die Übergangswahrscheinlichkeiten nichtstationär zu modellieren, verweisen wir auf [54, p. 238] und die dort angegebene weitere Literatur.

Wir beschränken uns hier auf die Darstellung des weitreichendsten Ansatzes, der "Semi-Markov-Modelle". Diese Modelle unterliegen keiner der drei Einschränkungen, die wir für klassische Markovketten genannt haben. Es werden Kaufintervalle zufälliger Länge zwischen aufeinanderfolgenden Einkaufsakte zugelassen. Die Verteilung der Intervalllänge darf außerdem von der zu Beginn gekauften Marke abhängen und für die am Ende gekauften Marken variieren. Modelle dieser Art werden dem Bereich der "competing risks" zugeordnet. Sie berücksichtigen gleichzeitig Kaufzeitpunkt und Markenwahl.

Die Markoveigenschaft kommt zum Tragen, da weiterhin angenommen wird, daß Kaufintervalle desselben Haushaltes untereinander unabhängig sind. Ein großer verbleibender Nachteil ist somit, daß Vielkäufer durch ihre hohe Anzahl Kaufintervalle im Untersuchungszeitraum deutlich mehr Beobachtungen zum Datensatz beitragen als Haushalte, die selten kaufen. Da das Modell Unabhängigkeit zwischen diesen Intervallen unterstellt, läuft es Gefahr, kurze Abstände zwischen Einkaufsakte tendenziell überzubewerten. Man kann sich von Verfahren zur Berücksichtigung unbeobachteter Heterogenität Abhilfe erhoffen, entsprechend viele Veröffentlichungen legen auf diesen Aspekt ihrer Modellierung besonderen Wert. Allerdings kann im Rahmen der standardisierten Modelle häufig nicht ohne weiteres erreicht werden, daß der Haushaltsbezug der Kaufintervalle im Verfahren berücksichtigt wird. Damit sinkt die Hoffnung, die hohe Anzahl kurzer Intervalle durch die unbeobachtete Heterogenität aufzufangen.

[76] gehört zu den ersten Veröffentlichungen dieser Methode in der Kaufverhaltensforschung. Die Autoren verwenden eine nichtparametrische Technik zur Berücksichtigung unbeobachteter Heterogenität. Der Modellansatz findet sich ähnlich bei [18], [34] und [79].

Als baseline Hazardrate verwenden [76] "a restricted version of the Box-Cox formulation" (zitiert nach [22, p. 3]). [34] vergleichen eine quadratische Gompertzfunktion mit einer stückweise exponentiellen Formulierung. [79] verwenden ebenfalls die zuletzt genannte Funktion, [18] präferiert eine log-logistische Verteilung, da sie unter anderem die Box-Cox und die inverse Normalverteilung "on the three measures of model fit, predictive ability, and computational time" ([18, p. 255]) übertreffe.

Die meisten genannten Artikel verwenden eine proportional hazards Formulierung, um Einflußgrößen einzubinden. [18] unterstellt accelerated failure times.

Bei der Kodierung der Marketingaktivitäten als zentraler Einflußgrößen ähneln sich die betrachteten Veröffentlichungen sehr stark. Lediglich die Aktivitäten, die bei einem tatsächlich erfolgten Kauf beobachtet wurden, gehen in das Modell ein. Sie werden rückwirkend für das ganze vorangegangene Kaufintervall als konstant angenommen. Damit werden Aktivitäten von Marken, die nicht gekauft wurden, vernachlässigt. Dasselbe gilt für zwischenzeitliche verkaufsfördernde Maßnahmen.

In einem anderen Zusammenhang hat [40, p. 2] beispielhaft verdeutlicht, wie hier mit "verpaßten Kaufgelegenheiten" verfahren wird: "For example, if a consumer buys a product after k weeks, this approach uses the covariate values of the k^{th} week only and ignores the covariates of $(k - 1)$ weeks. In other words, the information about whether or not there was a big promotion in the previous

(k - 1) weeks is completely ignored.”

Mit Bezug auf [76] und andere Studien notiert [17, p. 185]: *”[...] the studies [...] use data that pertain to only those store visits in which purchases are made in the product category of interest. A pertinent question that arises is: what is the implication of such conditional-on-purchase analyses that ignore store visits in which households are exposed to the marketing activities of brands, but fail to make a category purchase?”* Der Autor bringt dann ebenfalls ein Beispiel und kommt zu dem Schluß: *”In order to determine the effects of marketing variables on purchase behavior, it would [...] be necessary to explicitly include in the analysis, store visits in which the consumer is exposed to the marketing variables [...] but does not make a category purchase [...].”*

2.3.4 Modellierung mit Erneuerungsprozessen

[22, p. 2]

The study of the purchase-timing behavior of households has seen renewed and growing activity in recent years in marketing, due largely to the application of hazard-function analysis to the investigation of this aspect of household purchase behavior.

Das Bestreben, Poisson-verteilte purchase incidence Modelle als purchase timing Modelle zu verwenden und die Einbeziehung von Kaufintervallen in Markovketten hat die Aufmerksamkeit der Anwender folgerichtig auf ”Hazardmodelle” gelenkt.

Da in diesem Zusammenhang die Begriffsbildung nicht eindeutig ist, müssen wir zunächst klären, welche Modellklasse wir nun genau betrachten wollen. Die Kaufverhaltensforschung verwendet Ergebnisse der ”Hazard-function analysis”, ”survival analysis” oder ”Verweildaueranalyse” zur Untersuchung von Sequenzen von Einkaufsakte, die bei einzelnen Haushalten während eines Untersuchungszeitraumes beobachtet wurden. Ursprünglich wurden diese Resultate jedoch für Anwendungen entwickelt, in denen jede beobachtete Einheit nur ein einziges Ereignis erfährt. Um sie trotzdem einsetzen zu können, wird nun jedes Kaufintervall einzeln als i. i. d. Realisation einer positiven Zufallsvariablen verstanden, die womöglich einer Rechtszensur am Ende des Beobachtungszeitraumes sowie zeitveränderlichen Einflußgrößen ausgesetzt ist.

Damit können wir diese Ansätze auch als markenunspezifische Version der Semi-Markov-Modelle des letzten Abschnitts begreifen. Als Teil mehrstufiger Modelle, wie wir sie im einleitenden Abschnitt 2.1 kurz vorgestellt haben, ist der Einsatz eines category purchase timing Modells nicht unbedingt als Rückschritt aufzufassen. Die zentrale Kritik bzgl. der Überbewertung von Vielkäufern, die wir für Semi-Markov-Modelle bereits formuliert haben, bleibt jedoch bestehen. Mit einer Begriffsbildung der Punktprozeßtheorie können wir die kritisierte Eigenschaft dieser Modelle genau fassen. Sie unterstellen, daß es sich bei der Folge der Einkäufe eines Haushaltes im Beobachtungszeitraum um die Realisation eines ”Erneuerungsprozesses” handelt. Genau für diese Punktprozesse trifft die Annahme von i. i. d. Intervallen zwischen ihren Ereignissen zu, vgl. z. B. [71, p. 316].

Durch die Verwendung speziell konstruierter Einflußgrößen wird in der ökonomischen Literatur versucht, die unrealistische i. i. d. Annahme für Kaufintervalle aufzuweichen. So verwendet [40, p. 7] einen wöchentlichen Lagerbestand der einzelnen Haushalte. Diese Größe wird einerseits um die Einkäufe in der jeweiligen Woche erhöht. Aus einer Vorperiode wird der wöchentliche Konsum des Haushaltes geschätzt. Er wird wöchentlich vom Lagerbestand abgezogen.

Einflußgrößen dieser Art stellen mehr oder weniger enge Bezüge zwischen aufeinanderfolgenden Kaufintervallen desselben Haushaltes her und sind deshalb prinzipiell geeignet, der Kritik an der i. i. d. Verteilungsannahme für diese Intervalle zu begegnen. Sie bleiben aber eine relativ grobe Technik. So müssen die Bestände des Haushaltes zu Untersuchungsbeginn mehr oder weniger willkürlich gesetzt werden, der Verbrauch wird oft mangels Alternative stationär formuliert. Diese Einschränkungen führen z. B. dazu, daß häufig negative Bestände konstruiert werden, die vor Modellschätzung "hart" auf Null gesetzt werden müssen.

Zumindest der Kritik an Poissonprozessen wegen der Unregelmäßigkeit ihrer Intervalllängen haben Erneuerungsprozesse aber etwas entgegenzusetzen. [16] und [40] schlagen als Alternative zur Exponentialverteilung, die von homogenen Poissonprozessen impliziert wird, die Erlang-2-Verteilung zur Beschreibung der Kaufintervalllänge vor. Viele andere Verteilungen wurden mittlerweile angewendet: [69] nutzt inverse Normalverteilung, Lognormalverteilung und Weibull (zitiert nach [54, p. 224]). [75] verwendet verallgemeinerte Log-Gamma-Verteilungen.

In der Abbildung von Marketingaktivitäten und ihren Auswirkungen auf das Kaufverhalten kommen die Einschränkungen einer auf gängigen Methoden der Verweildaueranalyse basierenden Modellbildung jedoch wieder zum Ausdruck. Wir betrachten zwei weit verbreitete Formulierungen.

[40] verwendet im Rahmen eines proportional hazards Modells die Normalpreise, Preisreduktionen und Indikatoren für spezielle verkaufsfördernde Aktivitäten in den Geschäften. Diese geschäfts- und produktgenauen Werte werden aus dem Datensatz wochengenau ermittelt. Bevor sie in das Modell einfließen, werden sie für jeden Haushalt zu je einer Einflußgröße für den Normalpreiseffekt, den Preisreduktionseffekt und den Effekt verkaufsfördernder Maßnahmen aggregiert. Dabei werden die marken- und geschäftsgenauen Ausprägungen gewichtet aufaddiert: *"The weights used are individual-household-level brand and store shares. [...] Different weighting schemes were used, such as equal weights, market share weights, and weight of 1 to favorite brand and store. All weighting schemes gave approximately the same results."* ([40, p. 7])

Für ein category purchase timing Modell ist eine derartige Aggregation eine ansprechende und technisch gut umsetzbare Idee. Ein echter Schwachpunkt tritt erst bei der Berücksichtigung im proportional hazards Modell auf. Dort gehen die verschiedenen Einflüsse x_k üblicherweise als parametergewichtete Linearkombination und exponentiell ein, um für beliebige Parameterschätzungen β_k eine positive Hazardrate $\lambda(t)$ sicherzustellen: $\lambda(t) := \lambda_o(t) \cdot \exp \sum \beta_k x_k$. Diese funktionale Festlegung ist aber sehr restriktiv, es sei denn die Einflußgröße x_k ist ein Indikator. In diesem Fall besteht der Effekt von x_k offenbar darin, die Grundrate $\lambda_o(t)$ mit dem positiven Faktor $\exp \beta_k$ zu multiplizieren. Sobald x_k aber drei oder mehr Werte annehmen kann, ist die exponentielle Abhängig-

keit zwischen $\lambda(t)$ und x_k häufig inhaltlich nicht mehr gerechtfertigt. Durch die von [40] vorgenommene Gewichtung werden nun aber selbst die Indikatoren verkaufsfördernder Aktivitäten zunächst zu reellwertigen Einflußgrößen umkodiert.

Das folgende Konzept zur Berücksichtigung von Marketingaktivitäten hat einen engen Bezug zur Nutzentheorie. Diese Theorie ist zentraler Bestandteil vieler Forschungen in der Ökonomie. Entsprechend umfangreich ist ihre Ausarbeitung und Darstellung in der Literatur. Wir verweisen auf die Lehrbücher [57], [44, pp. 36 und pp. 42], [54, pp. 159, pp. 241 und pp. 291] und [33, pp. 53 und pp. 59] für eine ausführliche Darstellung und beschränken uns hier auf das Notwendige zum Verständnis der Anwendung in Kaufzeitpunktmodellen, z. B. bei [22]. Die Nutzentheorie geht davon aus, daß die Haushalte jedem Produkt einen stochastischen Nutzen zuordnen, der sich als reelle Zahl ausdrücken läßt. Der Haushalt wird sich bei der Markenwahl an den Artikel halten, der ihm den höchsten Nutzen in diesem formalen Sinne bietet. Der stochastische Nutzen zerfällt additiv in einen deterministischen Teil und eine Zufallsvariable. Der deterministische Teil wird in Anwendungen häufig als parametrisierte Linearkombination objektiver Produkteigenschaften formuliert, der stochastische Teil hingegen als extremwertverteilte Zufallsvariable. Diese Annahmen führen dazu, daß der maximale Nutzen über alle Produkte ebenfalls eine Extremwertverteilung hat. Außerdem ergibt sich sein erstes Moment, also der erwartete maximale Nutzen der Warengruppe für den Haushalt, explizit aus den deterministischen Einzelnutzen der Artikel.

Wenn wir nun also die Marketingaktivitäten als Produkteigenschaften auffassen, die den Nutzen einzelner Artikel für die Haushalte wie gerade skizziert beeinflussen, so läßt sich die Verwendung des maximalen erwarteten Nutzens als Einflußgröße im category purchase timing Modell mit nutzentheoretischen Argumenten rechtfertigen. Die praktische Verwendung dieser Größe unterliegt weiterhin den technischen Restriktionen, die wir bereits für die gewichteten Marketingaktivitäten bei [40] formuliert haben.

Da die Nutzentheorie eine zentrale Position in der Ökonomie einnimmt, besprechen wir nun noch ihre Verwendung im umrissenen Rahmen für Kaufzeitpunktmodelle. Ein großer praktischer Vorteil ist sicher, daß sie aus den vielen möglichen Gewichtungen der markenspezifischen Aktivitäten zu einer Warengruppengröße eine auszeichnet.

Die zugrundeliegenden Annahmen dieser Theorie scheinen aber für eine Anwendung auf das Kaufverhalten von Konsumgütern nicht uneingeschränkt geeignet. Sie unterstellen ja insbesondere, daß sich der einkaufende Haushalt aller bestehenden Kaufoptionen bewußt ist. Darüberhinaus wird angenommen, daß er alle diese Möglichkeiten gegeneinander abwägen kann. Hier spiegelt sich das Bild des "rationalen Akteurs" der Nutzentheorie im Konsumenten. Die Wirklichkeit der Handelslandschaft vermittelt jedoch einen anderen Eindruck. Die Information des Konsumenten über bestehende Wahlmöglichkeiten beruht im allgemeinen wohl eher auf einer Mischung aus Erfahrung mit den einzelnen Handelsketten, ihrer Erreichbarkeit und verfügbaren Wurfpostsendungen, sofern sie zur Kenntnis genommen werden. Die Atmosphäre und Ausstattung eines typischen Geschäftes zeigt sich ebenfalls kaum im Einklang mit der rationalen, auf vollständiger Information beruhenden Vorstellung der Nutzentheorie. Teu-

re Marken sind in Augenhöhe im Regal platziert, billigere Alternativen in Bodennähe verbannt. Impulskäufe in der Kassenschlange werden durch die Sonderplatzierung einzelner Produkte in diesem Bereich gefördert. Viele Handelsketten greifen zu untermalender Musik und anderen Mitteln, um den Einkaufsgang zu einem möglichst angenehmen Erlebnis zu machen. Auch die typische Verpackung eines Konsumguts spricht wohl primär das Unterbewußte, Emotionale an und weniger den kalkulierenden Verstand. Frohe Farben, Formen und Grafiken bestimmen das Bild. Wie die Abbildung bekannter Comicfiguren auf der Verpackung mit dem "Nutzen" des Verpackten für den Verbraucher in Verbindung zu bringen ist, erscheint jedenfalls nicht unmittelbar einsichtig. "Technisch bedingte Füllhöhen" verleiten den Konsumenten dazu, eine große Pappverpackung zu erstehen, die aber zu einem gewissen Teil leer ist. Einige Hersteller verfolgen anscheinend weiterhin die Strategie, ihre Produkte möglichst in einzigartigen Packungsgrößen auf den Markt zu bringen; offenbar, um den direkten Preisvergleich zu erschweren. Immerhin an dieser Stelle bieten nun die meisten Händler dem Verbraucher kleine Hilfestellungen, wenn sie kleingedruckte Kilogrammpreise am Regal notieren.

Mit dieser Beschreibung wollen wir der Nutzentheorie keineswegs ihren zentralen Platz in der modernen Ökonomie streitig machen. Wir wollen lediglich unsere Position verdeutlichen, daß ein Kaufverhaltensmodell nicht ausschließlich danach zu bewerten ist, ob es eine Interpretation im Sinne eines formalen Verbrauchernutzens gestattet. Mit dieser relativierenden Einschätzung stehen wir nicht allein. So notiert [17] zwar noch 1993:

[17, pp. 186]

In this paper, we investigate the impact of marketing variables on the category purchase, brand choice and purchase quantity decisions of households for frequently purchased packaged goods. We formulate a model based on microeconomic foundations of household utility maximizing behavior. From this single utility maximizing problem for the household on a store visit, we obtain: (1) the probability that no purchase is made in the category, (2) the probabilities of the household purchasing the different brands in the category and (3) the purchase quantity if a brand is bought. As the three choice decisions are the consequences of a single utility maximizing problem, the model ensures that the decisions, in combination, provide the highest utility to the household.

Fünf Jahre später ist jedoch die Einbettung in die Nutzentheorie für denselben Erstautor ein Kriterium zur Modellbewertung unter vielen:

[22, p. 3]

Each individual component of the dynamic McFadden model has been derived based on utility-maximizing behavior [...]. The composite model is ad hoc in its formulation, however. Nevertheless, we feel that the loss in conceptual appeal is compensated for by the model's ability to capture two important aspects of household purchase behavior, purchase timing and brand choice.

2.3.5 Modellierung mit allgemeinen Punktprozessen

Unter einem "allgemeinen Punktprozeß" wollen wir in dieser Arbeit einen Punktprozeß auf einem kompakten Intervall $I := [t_{\min}, t_{\max}] \subset \mathbb{R}$ verstehen, dessen stochastischer Kompensator eine absolutstetige Intensität mit rechtsstetiger Ableitung ist.

Derartige Prozesse sind zwar bei weitem noch nicht die allgemeinsten, die in der stochastischen Literatur untersucht werden, vgl. [12]. Trotzdem erscheint diese Bezeichnung in unserem Zusammenhang angemessen. Erstens werden wir gleich sehen, daß alle wesentlichen Modellklassen der Kaufverhaltensforschung als Spezialfälle dieses Typs von Punktprozessen angesehen werden können.

Zweitens haben wir praktische Belange der Modellbildung zu berücksichtigen. So bietet es sich an, Regressionsmodelle in Anwendungen für die differentielle Intensität zu formulieren. Dabei sind wir darauf angewiesen, Hypothesen aus der angewandten Forschung formal in die Modellgleichungen zu überführen. Genauso sollten wir die Parameterschätzungen eines Regressionsmodells auch wieder für die praktische Situation interpretieren können. Vor diesem Hintergrund erweist sich die Klasse der rechtsstetigen, stochastischen (differentiellen) Intensitäten als ausgesprochen mächtig.

Wir illustrieren nun unsere Behauptung, mit der angesprochenen Klasse von Punktprozessen eine Verallgemeinerung der auf Poissonprozessen, Markovketten und Erneuerungsprozessen basierenden category purchase timing Modelle angegeben zu haben. Sei dazu $N(\cdot) : I \rightarrow \mathbb{N}$ der Zählprozeß der Einkaufsakte eines Haushaltes für die betrachtete Warengruppe im Beobachtungszeitraum $I := [t_{\min}, t_{\max}]$. Mit $t_{\min} < Y_1 < Y_2 < \dots < Y_K < t_{\max}$ bezeichnen wir die aufsteigende Folge von Zeitpunkten, zu denen der Haushalt in der Warengruppe gekauft hat. $\Lambda(\cdot) : I \rightarrow \mathbb{R}_+$ sei die kumulative Intensität von N , ihre Ableitung nach der Zeit nennen wir $\lambda(\cdot)$.

Wir lassen zunächst die Einflußgrößen und die unbeobachtete Heterogenität außer acht. Dann erlauben alle auf Poissonprozessen basierenden Kaufzeitpunktmodelle eine Darstellung mit deterministischer Intensität $\lambda(t)$.

Die Semi-Markov-Modelle scheinen zunächst eine Sonderrolle zu spielen, da sie Markenwahlaspekte berücksichtigen. Diese Einschätzung stimmt jedoch nur bedingt. Zwar werden wir uns in dieser Arbeit auf unmarkierte Punktprozesse beschränken, um insbesondere auch die Beweise in den Kapiteln 7 und 8 nicht technisch zu überladen. Es sollte jedoch direkt möglich sein, unsere theoretischen Ergebnisse, die Theoreme 3.3.1 und 3.5.1, auf den Fall markierter Punktprozesse mit Bildraum \mathbb{N}^q , $q \in \mathbb{N}$, zu übertragen. Damit wäre die Markenwahl des Konsumenten in sehr allgemeiner Weise berücksichtigt. Selbst der Mengenwahlaspekt sollte, bei weiterhin unveränderter Beweisstruktur, mit dem Bildraum \mathbb{R}_+^q erfaßbar sein.

Die markenunspezifische Formulierung der Semi-Markov-Modelle entspricht, wie bereits dargestellt, gerade den Verweildauermodellen. Deren Pendant aus Punktprozeßsicht, der Erneuerungsprozeß, hat eine stochastische Intensität, die nur vom aktuellen Abstand zum letzten Ereignis abhängt: $\lambda(t) \equiv f(t - Y_{N(t)})$, vgl. [71, p. 316, Korollar zu Theorem 6.3.4].

Untersuchen wir nun, wie sich Marketingaktivitäten als Einflußgrößen im Rahmen allgemeiner Punktprozesse darstellen lassen. Sowohl die Werbewirkungstheorien aus Abschnitt 1.2 als auch die technischen Bedingungen für Werbewirkung in Makromodellen aus Abschnitt 1.4 legen dabei folgendes Vorgehen nahe: Die Marketingaktivitäten, denen ein Haushalt ausgesetzt ist, werden ihrerseits als Prozesse auf I betrachtet.³ Diese Prozesse erzeugen eine eigene Filtration über dem Beobachtungsintervall. Lassen wir nun zu, daß die modellierte Intensität meßbar bzgl. der gemeinsamen Filtration des Einkaufszählprozesses und der jeweiligen Marketingprozesse ist. Dann stehen uns sehr viele Möglichkeiten offen, die zeitveränderlichen Auswirkungen der verkaufsfördernden Aktivitäten auf den weiteren Einkaufsprozess über seine Intensität abzubilden. Diese Form der Modellierung ist aus anderen Anwendungsbereichen als "Filtering" bekannt, wir verweisen auf die umfangreiche Beispielsammlung "gefilterter Poissonprozesse" in [71, pp. 366, Abschnitt 7.4].

Nun wollen wir untersuchen, inwieweit sich die Vorgehensweisen zur Abbildung von Marketingaktivitäten aus der Literatur zur Kaufverhaltensforschung in diesem Rahmen darstellen lassen. Am Ende des Abschnitts 2.3.3 hatten wir die Modellierungstechnik angesprochen, jeweils die verkaufsfördernden Maßnahmen, die beim Einkauf vorlagen, rückwirkend für das gesamte vorherige Kaufintervall anzunehmen. Wir hatten auch gesehen, daß dieses Vorgehen bereits aus angewandter Sicht recht stark kritisiert wurde. Nun müssen wir feststellen, daß es sich zudem einer direkten Übertragung in den schlüssigen Rahmen gefilterter Prozesse widersetzt. Den Wert eines beeinflussenden Prozesses ab einem Einkaufsakt Y_j können wir nämlich erst festlegen, wenn wir die tatsächlichen Gegebenheiten zum Zeitpunkt des nächsten Einkaufs Y_{j+1} kennen. Anstelle eines vorgegebenen Einflußprozesses, der zu jedem Zeitpunkt eine Funktion der vergangenen Marketingaktivitäten ist, müssen wir hier einen künstlichen Prozeß konstruieren, dessen Wert von zukünftigen Ereignissen abhängt. Zu diesen Ereignissen zählt zu allem Überfluß auch noch der nächste Einkaufsakt, also eine Eigenschaft des beeinflussten Prozesses. Die Verwendung eines derartigen Kunstprozesses ist natürlich technisch möglich, da uns die Daten des gesamten Beobachtungszeitraumes vorliegen. Sie ist jedoch hochgradig unbefriedigend. Wir merken als weiteren Schwachpunkt noch an, daß der Einfluß dieses Prozesses auf den Einkaufsprozess als konstant während eines Kaufintervalles angenommen wird.

Die beiden Verfahren zur Beeinflussung des Kaufverhaltens durch Marketingaktivitäten aus Abschnitt 2.3.4 erlauben hingegen eine schlüssige Darstellung als Einflußprozesse. Die einzige Einschränkung besteht in den Möglichkeiten, zeitveränderliche Kovariable in standardisierten Verweildauermodellen zu berücksichtigen. In den betrachteten Artikeln zur Kaufverhaltensforschung führt dies zur Verwendung reellwertiger Stufenprozesse. Bei [40] sind diese Prozesse zusätzlich haushaltsspezifisch, wenn z. B. die Aktivitäten präferierter Produkte höher

³Es bleibt das erhebungstechnische Problem zu lösen, wie wir von verkaufsfördernden Aktivitäten erfahren, die dem Haushalt während eines Kaufganges vorlagen, von ihm aber nicht wahrgenommen wurden. Im Rahmen der von uns primär betrachteten TV-Werbewirkung ist dies kein Problem, da die TV-Kontakte unabhängig von Einkäufen gemessen werden. Preisreduktionen und ähnliche Aktivitäten lassen sich oft aus den Einkaufsinformationen der gesamten Stichprobe und der Kenntnis der Marketingstrategie einzelner Handelsorganisationen angemessen approximieren.

gewichtet werden.

Nun betrachten wir die Verfahren zur Berücksichtigung unbeobachteter Heterogenität. Lügen uns sämtliche haushaltsspezifischen Werte der Parameter vor, so könnten wir direkt ein allen Haushalten gemeinsames Modell für die stochastische Intensität angeben. Tatsächlich sind uns diese Größen aber nicht bekannt. Unsere Unkenntnis läßt sich formalisieren, wenn wir die stochastische Intensität mit einem geeigneten, gröberen Filter bedingen. Das bedeutet aber, daß wir bei Berücksichtigung unbeobachteter Heterogenität nicht mehr den Kompensator der beobachteten Zählprozesse modellieren, sondern eine bedingte Fassung dieser stochastischen Funktion. Direkt im Anschluß an die Formulierung unserer theoretischen Resultate in den Abschnitten 3.3 und 3.5 werden wir diese Situation behandeln. Es wird sich zeigen, daß sich die beiden Theoreme 3.3.1 und 3.5.1 entsprechend verallgemeinern lassen. Ihre Anwendbarkeit in praktischen Problemen hängt davon ab, ob sich die bedingten Intensitäten der verwendeten Modelle ausrechnen lassen. Der einfachste Fall sind die random effects, wenn sie unabhängig von allen anderen stochastischen Größen des Modells formuliert werden. Für die praktische Behandlung dieses Falls geben wir am Ende des Abschnitts 3.3 Hinweise.

Die Übersicht über die stochastische Literatur zu Regressionsmodellen mit Punktprozessen verschieben wir auf Kapitel 9. Erst dann können wir die in der Literatur verwendeten Beweistechniken im Detail mit unseren Methoden aus Kapitel 7 und 8 vergleichen.

3 Kleinst-Quadrate-Schätzer für Zählprozesse

3.1 Einführung

In diesem Kapitel untersuchen wir das Verhalten von Kleinst-Quadrate-Schätzern bei wachsendem Stichprobenumfang in folgender Situation:

Der Pfad eines Zählprozesses $N(\cdot)$ wird über einen endlichen Zeitraum I hinweg beobachtet. Für die Intensität $\Lambda_\circ(\cdot)$ dieses Prozesses ist ein parametrisches Modell $\mathbf{\Lambda} := \{\Lambda_\vartheta(\cdot), \vartheta \in \Theta \subset \mathbb{R}^d\}$ gegeben, d. h. es gelte $\Lambda_\circ(\cdot) = \Lambda_{\vartheta_\circ}(\cdot)$ für ein unbekanntes $\vartheta_\circ \in \Theta$.

Zu jedem Zeitpunkt t innerhalb des Beobachtungsintervalles I hängt die Intensität $\Lambda_{\vartheta_\circ}(t)$ parametrisch von dem bisherigen Verlauf des Punktprozesses $N(s)$, $s \leq t$, selbst ab. Zusätzlich können weitere Größen auf die Intensität Einfluß nehmen. Darunter befinden sich u. U. Punktprozesse über I , deren gesamter vergangener Verlauf bis t Auswirkungen auf $\Lambda_{\vartheta_\circ}(t)$ haben kann.

Unsere Stichprobe besteht aus i. i. d. Punktprozessen $N_k(\cdot)$, $k = 1, \dots, n$. Zu jedem dieser Punktprozesse gehört ein eigener Satz an Einflußgrößen. Deshalb werden sich die Intensitäten $\Lambda_{\vartheta_\circ, k}(\cdot)$ der Punktprozesse in der Regel voneinander unterscheiden.

Unser Interesse gilt einem Schätzer ϑ_n^* des Parameters $\vartheta_\circ \in \Theta$ auf der beschriebenen Datengrundlage. ϑ_n^* minimiere dabei den quadratischen Abstand

zwischen den gemittelten beobachteten Pfaden der Zählprozesse $N_k(\cdot)$, $k = 1, \dots, n$, und dem entsprechend gemittelten Modell:

$$\vartheta_n^* := \arg \inf_{\vartheta \in \Theta} \int_I \left(\frac{1}{n} \sum_{k=1}^n \Lambda_{\vartheta, k}(t) - \frac{1}{n} \sum_{k=1}^n N_k(t) \right)^2 \frac{1}{n} \sum_{k=1}^n N_k(dt)$$

In Theorem 3.3.1 werden wir zeigen, daß ϑ_n^* unter geeigneten Glattheitsannahmen stark konsistent ist, daß also

$$\vartheta_n^* \xrightarrow{\text{f. s.}} \vartheta_\circ \quad \text{für } n \rightarrow \infty$$

gilt.

Mit weiteren Voraussetzungen folgt außerdem die asymptotische Normalität von $n^{1/2}(\vartheta_n^* - \vartheta_\circ)$. Genauer gilt

$$n^{1/2} \Phi_n(\vartheta_\circ)(\vartheta_n^* - \vartheta_\circ) \xrightarrow{\mathcal{D}} N(0, C(\vartheta_\circ)) \quad \text{für } n \rightarrow \infty,$$

wie wir in Theorem 3.5.1 sehen werden.

Bei der standardisierenden Matrix $\Phi_n(\vartheta_\circ)$ handelt es sich um eine stochastische Größe mit Werten in $\mathbb{R}^{d \times d}$; die asymptotische Kovarianzmatrix $C(\vartheta_\circ) \in \mathbb{R}^{d \times d}$ mit derselben Dimension ist deterministisch. Wir werden in den Abschnitten 3.4.3 bzw. 3.4.4 explizite Darstellungen für beide Größen angeben. Wie die Schreibweise bereits andeutet, muß der wahre Parameterwert ϑ_\circ bekannt sein, um die standardisierende Matrix und die asymptotische Kovarianzmatrix in Anwendungen auszurechnen. ϑ_\circ wird in diesem Fall approximativ durch ϑ_n^* ersetzt. Während $\Phi_n(\vartheta_\circ)$ mit dieser Einschränkung aus den Daten geschätzt werden kann, werden in der Darstellung von $C(\vartheta_\circ)$ außerdem die theoretischen Größen $E \Lambda_{\vartheta_\circ}(\cdot)$ und $E \frac{\partial}{\partial \vartheta} \Lambda_{\vartheta_\circ}(\cdot)$ benötigt. Ihre Werte können durch

$$\frac{1}{n} \sum_{k=1}^n \Lambda_{\vartheta_n^*, k}(\cdot)$$

bzw.

$$\frac{1}{n} \sum_{k=1}^n \frac{\partial}{\partial \vartheta} \Lambda_{\vartheta_n^*, k}(\cdot)$$

angenähert werden.

Zum Abschluß dieser Einführung geben wir nun noch einen Überblick über die Struktur von Kapitel 3.

In Abschnitt 3.2 führen wir zunächst die bereits informell verwendeten Bezeichnungen sowie die Voraussetzungen formal ein, die der starken Konsistenz zugrundeliegen. Es handelt sich überwiegend um Stetigkeits- und Differenzierbarkeitsbedingungen. Die dabei definierten Zufallsvariablen sind auf einem gemeinsamen W -Maßraum (Ω, \mathcal{A}, P) definiert. Entsprechend sind σ -Algebren immer Unter- σ -Algebren von \mathcal{A} , und "f. s." bezieht sich auf P . Diese Zusammenhänge werden wir in der Notation unterdrücken. \mathbb{N} und \mathbb{R} bezeichnen wie üblich die natürlichen bzw. die reellen Zahlen.

Die starke Konsistenz des Kleinst-Quadrate-Schätzers formulieren wir in Abschnitt 3.3.

Die zusätzlichen Bezeichnungen und Voraussetzungen für die asymptotische Normalität von $n^{1/2}(\vartheta_n^* - \vartheta_o)$ folgen in Abschnitt 3.4, die Aussage selbst stellen wir in Abschnitt 3.5 vor.

3.2 Voraussetzungen für die starke Konsistenz des Kleinst-Quadrate-Schätzers

3.2.1 Der Zählprozeß

Bezeichne $I := [t_{\min}, t_{\max}]$ ein kompaktes Intervall und $\mathcal{A}(\cdot)$ einen Filter auf I . Sei $N(\cdot) : I \rightarrow \mathbb{N}$ ein Zählprozeß mit $\sigma(N(s), s \leq t) \subset \mathcal{A}(t)$ für alle $t \in I$.

Die dritten Momente seien endlich. Wegen der Monotonie von N genügt es, dies am rechten Intervallende zu verlangen:

$$\mathbb{E} N^3(t_{\max}) < \infty \quad (3.2.1)$$

Die u. U. stochastische kumulative Intensität von $N(\cdot)$ bezeichnen wir mit $\Lambda_o(\cdot)$. Wir nehmen an, daß sie absolutstetig ist:

$$\Lambda_o(\cdot) = \int_{t_{\min}}^{\cdot} \lambda_o(s) ds \quad (3.2.2)$$

Ihre Lebesgue-Dichte $\lambda_o(\cdot)$ sei dabei rechtsstetig. Gemäß [12, S. 288, D32 und T33] können wir Versionen von $\Lambda_o(\cdot)$ und $\lambda_o(\cdot)$ wählen, die progressiv meßbar bzgl. $\mathcal{A}(\cdot)$ sind.

3.2.2 Das parametrische Modell

Sei $\Theta \subset \mathbb{R}^d$, $d \in \mathbb{N}$, ein Parameterraum. Zu jedem Parameterwert $\vartheta \in \Theta$ sei $\Lambda_\vartheta(\cdot)$ eine absolutstetige Intensität mit rechtsstetiger Lebesgue-Dichte $\lambda_\vartheta(\cdot)$. Beide Funktionen seien progressiv meßbar bzgl. $\mathcal{A}(\cdot)$ gewählt. Wir nehmen an, daß $\Lambda_\vartheta(t)$ bei festem $t \in I$ nach dem Parameter ϑ stetig differenzierbar ist:

$$\vartheta \mapsto \Lambda_\vartheta(t) \in C^1(\Theta), \quad t \in I \quad (3.2.3)$$

Weiter werden wir benötigen, daß

$$\Theta \subset \mathbb{R}^d \quad \text{offen und beschränkt} \quad (3.2.4)$$

und

$$(t, \vartheta) \in I \times \bar{\Theta} \mapsto \Lambda_\vartheta(t) \quad \text{f. s. stetig} \quad (3.2.5)$$

ist. Die Funktion $(t, \vartheta) \in I \times \Theta \mapsto \Lambda_\vartheta(t)$ muß also auf den Rand von $I \times \Theta$ f. s. stetig fortsetzbar sein. Ohne (3.2.5) ließe sich die Beschränktheit in (3.2.4) durch eine geeignete Umparametrisierung immer erreichen: Ein Parameter mit

unbeschränktem Definitionsbereich kann etwa durch eine vorgeschaltete, diffeomorphe Transformation des offenen Einheitsintervalls $(0, 1)$ auf die reelle Achse \mathbb{R} in einen beschränkten Parameter überführt werden, ohne (3.2.3) zu verletzen.

Wie bereits für die Zählprozesse in (3.2.1) benötigen wir auch für die Intensitäten die dritten Momente. Wegen der Monotonie genügt es wiederum, diese Bedingung am rechten Intervallende zu stellen. Allerdings müssen wir eine im Parameter lokal gleichmäßige Integrierbarkeit des Modells fordern: Es bezeichne dazu

$$B_r(\vartheta) := \{\vartheta' \in \bar{\Theta} : |\vartheta' - \vartheta| < r\}, \quad r > 0, \vartheta \in \bar{\Theta}$$

den Schnitt der r -Kugel um $\vartheta \in \bar{\Theta}$ mit dem Abschluß $\bar{\Theta}$ des Parameterraumes. Dann existiere zu jedem Parameterwert $\vartheta \in \bar{\Theta}$ ein positiver Radius $r_\circ(\vartheta) \equiv r_\circ > 0$, so daß

$$\mathbb{E} \left(\sup_{\vartheta' \in B_{r_\circ}(\vartheta)} \Lambda_{\vartheta'}(t_{\max}) \right)^3 < \infty \quad (3.2.6)$$

gilt. (3.2.6) ist dann auch für jeden kleineren Radius $0 \leq r < r_\circ$, also insbesondere punktweise im Parameter $\vartheta \in \Theta$, richtig. Schließlich enthalte das Modell $\Lambda := \{\Lambda_\vartheta(\cdot), \vartheta \in \Theta\}$ die Intensität des Zählprozesses $N(\cdot)$. Es gebe also einen Parameterwert $\vartheta_\circ \in \Theta$ mit

$$\Lambda_{\vartheta_\circ}(\cdot) \equiv \Lambda_\circ(\cdot).$$

3.2.3 Das Zählmartingal

Die Intensität $\Lambda_{\vartheta_\circ}(\cdot)$ ist der Kompensator des Zählprozesses $N(\cdot)$. Dies macht ihre Differenz zu einem $\mathcal{A}(\cdot)$ -Martingal, vgl. [12, S. 27, T8 (α)]. Wir nennen es ZÄHLMARTINGAL und bezeichnen es mit $M(\cdot)$:

$$M(\cdot) := N(\cdot) - \Lambda_{\vartheta_\circ}(\cdot)$$

3.2.4 Die i. i. d. Situation

Wir führen nun Zählprozesse $N_k(\cdot) : I \rightarrow \mathbb{N}$, $k \in \mathbb{N}$, ein. Es handle sich dabei um i. i. d. "Kopien" von $N(\cdot)$. Die zugehörigen Filter auf I bezeichnen wir mit $\mathcal{A}_k(\cdot)$, die absolutstetigen, $\mathcal{A}_k(\cdot)$ -progressiv meßbaren kumulativen Intensitäten mit $\Lambda_{\vartheta_\circ, k}(\cdot)$, die $\mathcal{A}_k(\cdot)$ -Zählmartingale mit

$$M_k(\cdot) := N_k(\cdot) - \Lambda_{\vartheta_\circ, k}(\cdot).$$

Den gemeinsamen Filter der ersten n Zählprozesse nennen wir

$$\check{\mathcal{A}}_n(\cdot) := \bigvee_{k=1}^n \mathcal{A}_k(\cdot).$$

Die Prozesse $N_k(\cdot)$, $k = 1, \dots, n$, definieren einen gemittelten Prozeß

$$\bar{N}_n(\cdot) := \frac{1}{n} \sum_{k=1}^n N_k(\cdot).$$

Wir führen zusätzlich die entsprechend gemittelte Intensität

$$\bar{\Lambda}_{\vartheta,n}(\cdot) := \frac{1}{n} \sum_{k=1}^n \Lambda_{\vartheta,k}(\cdot), \quad \vartheta \in \Theta,$$

ein. Dann hat der Zählprozeß $n \cdot \bar{N}_n(\cdot)$ die Intensität $n \cdot \bar{\Lambda}_{\vartheta_o,n}(\cdot)$. Mit

$$\bar{M}_n(\cdot) := \frac{1}{n} \sum_{k=1}^n M_k(\cdot)$$

bezeichnen wir das zugehörige MITTELWERTMARTINGAL.

3.2.5 Der Kleinst-Quadrate-Schätzer

Wir verwenden die Prozesse $\bar{N}_n(\cdot)$ und $E \Lambda_{\vartheta_o}(\cdot)$ nun, um Halbnormen für geeignete stochastische Funktionen auf dem Intervall I einzuführen. Zunächst erhalten wir die Familie der stochastischen Halbnormen

$$\|f(\cdot)\|_{\bar{N}_n} := \left(\int_I f(t)^2 \bar{N}_n(dt) \right)^{1/2}, \quad n \in \mathbb{N},$$

für $\bar{N}_n(\cdot)$ -quadrat-integrable, u. U. stochastische Funktionen $f(\cdot) : I \rightarrow \mathbb{R}$, siehe [7, S. 90]. Dies erlaubt die Einführung der stochastischen Cramér-von Mises-Distanzen oder Pseudometriken, siehe [72, S. 234]:

$$\|f(\cdot) - g(\cdot)\|_{\bar{N}_n}$$

$\|\cdot\|_{\bar{N}_n}$ hat als Halbnorm alle Eigenschaften einer Norm, nur kann der Abstand zweier unterschiedlicher Funktionen auf I verschwinden.

Wir definieren den Kleinst-Quadrate-Schätzer ϑ_n^* durch

$$\vartheta_n^*(\omega) \equiv \vartheta_n^* := \arg \inf_{\vartheta \in \Theta} \|\bar{N}_n(\cdot) - \bar{\Lambda}_{\vartheta,n}(\cdot)\|_{\bar{N}_n}.$$

Im Zusammenhang mit der Konsistenz von ϑ_n^* in Theorem 3.3.1 wird außerdem die deterministische Halbnorm

$$\|f(\cdot)\|_{E \Lambda_o} := \left(\int_I f(t)^2 E \Lambda_{\vartheta_o}(dt) \right)^{1/2}$$

benötigt. Die Konvergenz von $\|\bar{N}_n(\cdot) - \bar{\Lambda}_{\vartheta,n}(\cdot)\|_{\bar{N}_n}$ gegen $\|E \Lambda_o(\cdot) - E \Lambda_{\vartheta}(\cdot)\|_{E \Lambda_o}$ für wachsendes $n \in \mathbb{N}$, die wir in den Lemmata 7.3.1 und 7.3.2 genauer untersuchen werden, ist entscheidend für den Nachweis der Konsistenz von ϑ_n^* in Theorem 3.3.1.

3.2.6 Identifizierbarkeit

Wir wollen mit Hilfe des Schätzers ϑ_n^* den Parameter $\vartheta_o \in \Theta$ der Intensität $\Lambda_{\vartheta_o}(\cdot)$ aus den gemittelten Beobachtungen $\bar{N}_n(\cdot)$ zurückgewinnen. Unser Schätzer

verwendet gemäß Abschnitt 3.2.5 die Größe $\|\bar{N}_n(\cdot) - \bar{\Lambda}_{\vartheta,n}(\cdot)\|_{\bar{N}_n}$, um die Abweichung der parametrisierten Intensität von den gemittelten Beobachtungen zu bestimmen. In den Lemmata 7.3.1 und 7.3.2 werden wir zeigen, daß sich $\|\bar{N}_n(\cdot) - \bar{\Lambda}_{\vartheta,n}(\cdot)\|_{\bar{N}_n}$ für großes $n \in \mathbb{N}$ in geeignetem Sinne der deterministischen Distanz $\|\mathbb{E}\Lambda_{\circ}(\cdot) - \mathbb{E}\Lambda_{\vartheta}(\cdot)\|_{\mathbb{E}\Lambda_{\circ}}$ nähert. Ohne dies hier zu präzisieren, ist damit folgende Überlegung einleuchtend: der Schätzer ϑ_n^* ist darauf angewiesen, daß sich unterschiedliche Parameterwerte $\vartheta \neq \vartheta'$ in $\|\mathbb{E}\Lambda_{\vartheta}(\cdot) - \mathbb{E}\Lambda_{\vartheta'}(\cdot)\|_{\mathbb{E}\Lambda_{\circ}}$ niederschlagen. In Anlehnung an [64, S. 50] treffen wir deshalb die Annahme:

$$\forall \varepsilon > 0 : \quad \inf_{\vartheta: |\vartheta - \vartheta_{\circ}| \geq \varepsilon} \|\mathbb{E}\Lambda_{\circ}(\cdot) - \mathbb{E}\Lambda_{\vartheta}(\cdot)\|_{\mathbb{E}\Lambda_{\circ}} > 0 \quad (3.2.7)$$

3.3 Die starke Konsistenz des Kleinst-Quadrate-Schätzers

Wir formulieren das erste zentrale Resultat dieses Kapitels:

3.3.1 Theorem *Mit den in Abschnitt 3.2 eingeführten Notationen und unter den Annahmen (3.2.1) bis (3.2.7) gilt*

$$\vartheta_n^* \xrightarrow{f. s.} \vartheta_{\circ} \quad \text{für } n \rightarrow \infty.$$

Wir werden Theorem 3.3.1 in Kapitel 7 beweisen. Im Beweis wird kein Gebrauch davon gemacht, daß $N - \Lambda_{\circ}$ ein Martingal auf I ist. Es ist lediglich von Bedeutung, daß dieser Prozeß zentriert ist.

Folglich könnten wir in der Definition des KQS in Abschnitt 3.2.5 die Intensität $\bar{\Lambda}_{\vartheta,n}$ durch einen Prozeß $\mathbb{E}(\bar{\Lambda}_{\vartheta,n}|\mathcal{C})$ ersetzen, der bei Verwendung einer beliebigen σ -Algebra $\mathcal{C} \subset \check{\mathcal{A}}_n$ aus $\bar{\Lambda}_{\vartheta,n}$ hervorgeht, z. B. die deterministische Funktion $\mathbb{E}\Lambda_{\vartheta}$. Jeder derart modifizierte Schätzer ist stark konsistent, der Beweis des Theorems 3.3.1 ist wörtlich übertragbar.

Diese Bemerkung ist einerseits von theoretischem Interesse, da bei der Untersuchung von Punktprozessen in der Literatur die oben erwähnte Martingalstruktur oft eine zentrale Rolle spielt, vgl. etwa [5] und [12].

Für Anwendungen ist diese Überlegung ebenfalls von Bedeutung, wenn wir zwar alle Punktprozesse $N_k(\cdot)$ vollständig beobachtet haben, die Verläufe einiger Einflußgrößen jedoch nicht vorliegen. Es kann dann evtl. trotzdem noch möglich sein, für jedes $N_k(\cdot)$ und jeden Parameterwert $\vartheta \in \Theta$ aus den Daten eine "bedingte Intensität" $\mathbb{E}(\Lambda_{\vartheta}|\mathcal{C}_k)$ zu berechnen, wobei die σ -Algebren \mathcal{C}_k angeben, welche Einflüsse in unseren Daten zur Verfügung stehen.

Eine derartige Situation liegt z. B. dann vor, wenn die unbeobachteten Einflüsse von den beobachteten unabhängig sind und ihre Verteilung bekannt ist, siehe [8, S. 128, Exercise 7]. In diesem Fall kann der "KQS bzgl. der bedingten Intensität" eingesetzt werden.

3.4 Voraussetzungen für die asymptotische Normalität des Kleinst-Quadrate-Schätzers

3.4.1 Lokale Lipschitz-Stetigkeit des Modells

Über (3.2.3) hinaus fordern wir, daß

$$\vartheta \mapsto \Lambda_\vartheta(t) \in C^2(\Theta), \quad t \in I \quad (3.4.1)$$

gilt. Insbesondere ist $\frac{\partial}{\partial \vartheta} \Lambda_\vartheta(\cdot)$ lokal Lipschitz-stetig. Zu jedem Parameterwert $\vartheta \in \Theta$ existieren demnach eine Umgebung $\Theta_\circ(\vartheta)$ und eine in der Regel stochastische Funktion $L'(\cdot)$, so daß gilt:

$$\forall \vartheta' \in \Theta_\circ(\vartheta) : \left| \frac{\partial}{\partial \vartheta} \Lambda_\vartheta(\cdot) - \frac{\partial}{\partial \vartheta'} \Lambda_{\vartheta'}(\cdot) \right| \leq L'(\cdot) \cdot |\vartheta - \vartheta'| \quad (3.4.2)$$

Die stochastische Funktion $L'(\cdot)$ erfülle folgende Integrierbarkeitsbedingungen:

$$\mathbb{E} \int_I L'^4(t) N(dt) < \infty \quad \forall \vartheta \in \Theta \quad (3.4.3)$$

$$\mathbb{E} \int_I L'^4(t) \Lambda_\vartheta(dt) < \infty \quad \forall \vartheta \in \Theta \quad (3.4.4)$$

$$\mathbb{E} \int_I L'^4(t) \mathbb{E} \Lambda_\circ(dt) < \infty \quad (3.4.5)$$

Die Funktion, welche die entsprechende Rolle für $\frac{\partial}{\partial \vartheta} \Lambda_{\vartheta,i}(\cdot)$, $i = 1, \dots, n$, spielt, bezeichnen wir mit $L'_i(\cdot)$.

3.4.2 Momentenbedingungen an das Modell

Um über die starke Konsistenz hinaus die asymptotische Normalität des KQS zu erhalten, benötigen wir weitere Momente des Zählprozesses und seiner Intensität. Einerseits verschärfen wir (3.2.1) zu

$$\mathbb{E} N^6(t_{\max}) < \infty \quad (3.4.6)$$

und die punktweise Version von (3.2.6) zu

$$\mathbb{E} \Lambda_\vartheta^6(t_{\max}) < \infty \quad \forall \vartheta \in \Theta. \quad (3.4.7)$$

Andererseits führen wir für die ersten beiden Ableitungen der Intensität Momentenannahmen ein, nämlich

$$\mathbb{E} \int_I \left| \frac{\partial}{\partial \vartheta} \Lambda_\vartheta(t) \right|^4 N(dt) < \infty, \quad \mathbb{E} \int_I \left| \frac{\partial^2}{\partial \vartheta^2} \Lambda_\vartheta(t) \right|^4 N(dt) < \infty, \quad \forall \vartheta \in \Theta, \quad (3.4.8)$$

$$\mathbb{E} \int_I \left| \frac{\partial}{\partial \vartheta} \Lambda_\vartheta(t) \right|^4 \Lambda_\circ(dt) < \infty, \quad \mathbb{E} \int_I \left| \frac{\partial^2}{\partial \vartheta^2} \Lambda_\vartheta(t) \right|^4 \Lambda_\circ(dt) < \infty, \quad \forall \vartheta \in \Theta, \quad (3.4.9)$$

und

$$\mathbb{E} \int_I \left| \frac{\partial}{\partial \vartheta} \Lambda_\vartheta(t) \right|^4 \mathbb{E} \Lambda_\circ(dt) < \infty, \quad \mathbb{E} \int_I \left| \frac{\partial^2}{\partial \vartheta^2} \Lambda_\vartheta(t) \right|^4 \mathbb{E} \Lambda_\circ(dt) < \infty, \quad \forall \vartheta \in \Theta. \quad (3.4.10)$$

Außerdem müssen wir fordern, daß der Parameterraum Θ über (3.2.4) hinaus auch einfach zusammenhängend ist.

Schließlich setzen wir voraus, daß

$$\mathbb{E} \Lambda_\circ(\cdot) \quad \text{streng monoton wachsend} \quad (3.4.11)$$

ist und zu jedem Vektor $x \in \mathbb{R}^d$ ein $t_x \in I$ existiert, welches

$$x^\top \mathbb{E} \frac{\partial}{\partial \vartheta} \Lambda_{\vartheta_\circ}(t_x) \neq 0 \quad (3.4.12)$$

erfüllt.

3.4.3 Die standardisierende Matrix

In Abschnitt 3.1 tauchte bereits informell die standardisierende Matrix $\Phi_n(\vartheta_\circ)$ auf. Allgemein setzen wir für beliebiges $\vartheta \in \Theta$:

$$\Phi_n(\vartheta) := \frac{\partial}{\partial \vartheta} \int_I (\bar{\Lambda}_{\vartheta,n}(t) - \bar{\Lambda}_{\vartheta_\circ,n}(t)) \frac{\partial}{\partial \vartheta} \bar{\Lambda}_{\vartheta,n}(t) \bar{\Lambda}_{\vartheta_\circ,n}(dt)$$

Die C^2 -Glattheitsvoraussetzung (3.4.1) wird hier bereits benötigt. Wir führen außerdem die deterministische Funktion

$$\Phi_\circ(\vartheta) = \frac{\partial}{\partial \vartheta} \int_I (\mathbb{E} \Lambda_\vartheta(t) - \mathbb{E} \Lambda_\circ(t)) \mathbb{E} \frac{\partial}{\partial \vartheta} \Lambda_\vartheta(t) \mathbb{E} \Lambda_\circ(dt)$$

auf Θ ein.

3.4.4 Die asymptotische Kovarianzmatrix

Wir definieren schließlich die asymptotische Kovarianzmatrix $C(\vartheta_\circ) \in \mathbb{R}^{d \times d}$, die in der Einführung 3.1 bereits informell erwähnt wurde. Für die Elemente $C_{ij}(\vartheta_\circ)$, $1 \leq i, j \leq d$, dieser Matrix setzen wir:

$$C_{ij}(\vartheta_\circ) := \int_I \left(\int_{[x, t_{\max}]} (\mathbb{E} \frac{\partial}{\partial \vartheta} \Lambda_{\vartheta_\circ}(t))_i \mathbb{E} \Lambda_\circ(dt) \right) \cdot \left(\int_{[x, t_{\max}]} (\mathbb{E} \frac{\partial}{\partial \vartheta} \Lambda_{\vartheta_\circ}(t))_j \mathbb{E} \Lambda_\circ(dt) \right) \mathbb{E} \Lambda_\circ(dx)$$

3.5 Die asymptotische Normalität des Kleinst-Quadrate-Schätzers

Wir formulieren nun das zweite zentrale Resultat dieses Kapitels:

3.5.1 Theorem *Mit den in den Abschnitten 3.2 und 3.4 eingeführten Notationen sowie unter den Annahmen (3.2.1) bis (3.2.7) und (3.4.1) bis (3.4.12) gilt*

$$n^{1/2}\Phi_n(\vartheta_\circ)(\vartheta_n^* - \vartheta_\circ) \xrightarrow{\mathcal{D}} N(0, C(\vartheta_\circ)) \quad \text{für } n \rightarrow \infty.$$

Theorem 3.5.1 wird in Kapitel 8 bewiesen. Die Bemerkung, die wir bereits bei der Formulierung der starken Konsistenz in Abschnitt 3.3 gemacht haben, bleibt auch für den Nachweis der asymptotischen Normalität weitgehend gültig: Im Hauptteil des Beweises werden wir nicht benötigen, daß es sich bei $N - \Lambda_\circ$ um ein Martingal handelt, nur die Zentriertheit des Prozesses wird eine Rolle spielen.

Wir werden also auch den Nachweis der asymptotischen Normalität führen, ohne auf zentrale Hilfsmittel zurückzugreifen, die die Literatur für Punktprozesse bereitstellt.

Für Anwendungen ist relevant, daß sich die asymptotische Kovarianzmatrix ändert, wenn der KQS bzgl. einer bedingten Intensität berechnet wird. Um diese Aussage präzise zu fassen, betrachten wir die in Abschnitt 3.3 eingeführten σ -Algebren \mathcal{C}_k , $k = 1, \dots, n$, und die bedingten Intensitäten

$$\Lambda_{\vartheta_\circ, k}^{\mathcal{C}_k}(\cdot) := E(\Lambda_{\vartheta_\circ, k}(\cdot) | \mathcal{C}_k). \quad (3.5.1)$$

Unter Verwendung dieser Größen definieren wir die Matrix $\Delta C(\vartheta_\circ) \in \mathbb{R}^{d \times d}$ mit den Elementen

$$\begin{aligned} \Delta C_{ij}(\vartheta_\circ) := E \left(\int_I \left(\int_{[x, t_{\max}]} (E \frac{\partial}{\partial \vartheta} \Lambda_{\vartheta_\circ}(t))_i E \Lambda_\circ(dt) \right) (\Lambda_\circ(dx) - \Lambda_{\vartheta_\circ}^{\mathcal{C}}(dx)) \right. \\ \left. \cdot \int_I \left(\int_{[x, t_{\max}]} (E \frac{\partial}{\partial \vartheta} \Lambda_{\vartheta_\circ}(t))_j E \Lambda_\circ(dt) \right) (\Lambda_\circ(dx) - \Lambda_{\vartheta_\circ}^{\mathcal{C}}(dx)) \right) \end{aligned}$$

Dann erhalten wir:

3.5.2 Korollar *Wenn der KQS in der Situation des Theorems 3.5.1 bzgl. der bedingten Intensitäten $\Lambda_{\vartheta_\circ, k}^{\mathcal{C}_k}(\cdot)$ definiert wird, gilt*

$$n^{1/2}\Phi_n(\vartheta_\circ)(\vartheta_n^* - \vartheta_\circ) \xrightarrow{\mathcal{D}} N(0, C(\vartheta_\circ) + \Delta C(\vartheta_\circ)) \quad \text{für } n \rightarrow \infty.$$

Wir beweisen Korollar 3.5.2 im Anschluß an Theorem 3.5.1 in Kapitel 8.

$\Delta C(\vartheta_\circ)$ ist die Kovarianzmatrix des d -dimensionalen Vektors mit den Elementen

$$\int_I \left(\int_{[x, t_{\max}]} (E \frac{\partial}{\partial \vartheta} \Lambda_{\vartheta_\circ}(t))_i E \Lambda_\circ(dt) \right) (\Lambda_\circ(dx) - \Lambda_{\vartheta_\circ}^{\mathcal{C}}(dx))$$

für $i = 1, \dots, d$. Folglich ist $\Delta C(\vartheta_\circ)$ positiv semidefinit. In diesem Sinne ist die asymptotische Varianz $C(\vartheta_\circ)$ des KQS also "am kleinsten", wenn ihm ein Modell der Intensität $\Lambda_{\vartheta_\circ}(\cdot)$ zugrundegelegt wird. Demgegenüber kann die Verwendung einer bedingten Intensität $\Lambda_{\vartheta_\circ}^C(\cdot)$ als "Informationsverlust" interpretiert werden, der die asymptotische Kovarianz des Schätzers um die positiv semidefinite Matrix $\Delta C(\vartheta_\circ)$ "erhöht".

Folglich ist es bei der Verwendung unserer Schätzer in Anwendungen ratsam, die σ -Algebren \mathcal{C}_k so groß wie möglich zu wählen.

4 Die Herleitung des Kaufverhaltensmodells

4.1 Einführung

In diesem Abschnitt stellen wir die Anwendung unserer Modellklasse auf eine ökonometrische Fragestellung der Kaufverhaltensforschung vor. In Abschnitt 4.2 präzisieren und vervollständigen wir dazu unsere Bezeichnungen aus Kapitel 3. Anschließend diskutieren wir in Abschnitt 4.3 kurz die verwendete Stichprobe und betrachten den Datensatz, der im weiteren Verwendung findet. In den Abschnitten 4.4 bis 4.7 entwickeln wir schrittweise das Modell. Abschnitt 4.8 stellt eine für die weitere Untersuchung günstige Zerlegung der Modellintensität bereit. In Abschnitt 4.9 beschäftigen wir uns mit der Wahl geeigneter Anfangswerte für die Parameterschätzung.

Das hergeleitete Modell werden wir in Kapitel 5 zunächst mit einer Simulationsstudie untersuchen. Die Modellergebnisse auf den realen Daten folgen in Kapitel 6.

4.2 Weitere Bezeichnungen

4.2.1 Die beobachteten Größen

Um Modellgleichungen in Anwendungen zu entwickeln, ist es notwendig, die Notationen des einleitenden Abschnitts 3.1 zu präzisieren.

Der Prozeß N_i , $i = 1, \dots, n$, werde von dem Punktprozeß $W_i : I = [t_{\min}, t_{\max}] \rightarrow \mathbb{N}$ beeinflusst.

Es gilt also

$$\mathcal{A}_i(t) = \sigma(N_i(s), W_i(s); s \leq t) \quad t \in I$$

4.2.2 Die Sprünge der Punktprozesse

Außerdem wird es sich als zweckmäßig erweisen, die Bezeichnungen der Abschnitte 3.2 und 3.4 zu ergänzen.

Es bezeichne

$$K_j := N_j(t_{\max}) - N_j(t_{\min}) \quad j = 1, \dots, n,$$

die Anzahl Sprünge des Prozesses N_j . Ihre Positionen nennen wir im folgenden:

$$t_{\min} < Y_{j1} < Y_{j2} < \cdots < Y_{jK_j} < t_{\max}$$

Die folgenden Betrachtungen werden sich vereinfachen, wenn wir auch Indizierungen außerhalb des Bereiches $[1, \dots, K_j]$ zulassen:

$$\begin{aligned} Y_{jk} &:= t_{\min} && \text{für } k \leq 0 \\ Y_{jk} &:= t_{\max} && \text{für } k > K_j \end{aligned}$$

Da die Sprünge im Datensatz f. s. nicht auf die Endpunkte des Beobachtungsintervalles fallen, sind die "künstlichen" Sprünge direkt an ihrem Wert zu erkennen.

Die Gesamtzahl der Sprünge in der Stichprobe bezeichnen wir mit

$$K := \sum_{j=1}^n K_j,$$

die maximale Sprungzahl in einem Pfad mit

$$S := \max_{j=1, \dots, n} K_j.$$

Ganz entsprechend beschreiben wir den beeinflussenden Punktprozeß W_j . Wir setzen dazu

$$J_j := W_j(t_{\max}) - W_j(t_{\min}) \quad j = 1, \dots, n,$$

und bezeichnen die Sprünge von W_j mit

$$\begin{aligned} t_{\min} &< X_{j1} < X_{j2} < \cdots < X_{jJ_j} < t_{\max}, \\ X_{jk} &:= t_{\min} && \text{für } k \leq 0, \\ X_{jk} &:= t_{\max} && \text{für } k > J_j. \end{aligned}$$

Wir benötigen außerdem die Größen

$$\begin{aligned} J &:= \sum_{j=1}^n J_j, \\ R &:= \max_{j=1, \dots, n} J_j. \end{aligned}$$

4.3 Die Struktur des Datensatzes

Der Datensatz, den wir in unserem Anwendungsbeispiel verwenden, beruht auf dem "Single Source Panel" von ACNielsen, Deutschland. Für eine ausführliche Darstellung der Marktforschung mit Panels verweisen wir auf [38, insb. pp. 74]. Im folgenden stellen wir lediglich die Aspekte des Panels vor, die für das Verständnis unserer Anwendung unerlässlich sind.

Die Haushalte des Single Source Panels werden für die Dauer ihrer Teilnahme von ACNielsen mit technischen Geräten ausgestattet, die die Erfassung ihres Einkaufs- und Fernsehverhaltens ermöglichen.

Sie liefern tagesgenaue Angaben über ihre gesamten Einkäufe von Konsumgütern des täglichen Bedarfs. Dies geschieht, indem sie den Barcode der eingekauften Artikel zuhause erneut einscannen. Der hierfür verwendete Scanner ermöglicht die Eingabe weiterer Details, wie etwa der Einkaufsstätte. Von diesen Informationen machen wir jedoch im folgenden keinen Gebrauch. Um unser Modell möglichst überschaubar zu halten, werden wir die getätigten Einkaufsakte in der relevanten Warengruppe Haushaltspackungen Speiseeis lediglich zählen; sowohl die gekaufte Menge als auch die Marke werden wir vernachlässigen.

Die Panelhaushalte erhalten ein zweites Gerät, welches sekundengenau die Zeit erfaßt, zu der der Fernseher des Haushaltes eingeschaltet war. Außerdem wird automatisch festgestellt, welches Programm empfangen wurde. Da die Sendezeiten und -kanäle der relevanten Werbespots unabhängig vom Panel bestimmt werden können, ist es möglich, so die "Werbekontakte" der Panelhaushalte zu ermitteln. Entsprechend zu unserem Vorgehen bei den Einkäufen beschränken wir uns auch bei den Werbekontakten darauf, solche für Produkte der relevanten Warengruppe zu zählen, ohne genauer auf das beworbene Produkt oder das Werbemotiv einzugehen.

Streng genommen kann natürlich nicht festgestellt werden, ob zum Zeitpunkt der Werbespotausstrahlung überhaupt ein Mitglied des Haushaltes das laufende Fernsehprogramm verfolgt hat. Weiterhin müssen wir vereinfachend voraussetzen, daß eine wahrgenommene Werbung auch der haushaltsführenden Person zur Kenntnis gelangt.

Schließlich geben die Panelteilnehmer in einem standardisierten jährlichen Fragebogen Auskunft über relevante Eigenschaften ihres Haushaltes, z. B. die Haushaltsgröße sowie das Alter, Geschlecht und die Ausbildung der Personen. Derartige Informationen werden auch als soziodemographische Eigenschaften bezeichnet. Sie spielen z. B. eine Rolle, wenn Daten der Stichprobe auf die deutsche Bevölkerung hochgerechnet werden sollen. Diesen Aspekt blenden wir jedoch im folgenden vollständig aus. Wir verwenden in unserer Anwendung die Information über die Anzahl im Haushalt wohnender Personen und das zur Verfügung stehende monatliche Einkommen, um die Haushalte in Untergruppen zu zerlegen, für die wir jeweils eigene Parameterschätzungen durchführen.

Nun müssen wir noch dafür sorgen, daß die tagesgenau erhobenen Einkaufsakte als Sprünge eines Punktprozesses in stetiger Zeit aufgefaßt werden können. Dazu numerieren wir die Tage des Jahres 2000 zunächst durch, ermitteln zu jedem erfaßten Einkauf den zugehörigen Tag und addieren eine auf $[0, 0.5]$ gleichverteilte Zufallsvariable. Diese Zufallsvariablen seien für unterschiedliche Einkaufsakte in der Stichprobe i. i. d.. Die resultierenden Werte im Intervall $I := [0, 365]$ entsprechen den Ereignissen Y_{jk} in der Notation des vorangegangenen Abschnitts. Dabei identifiziert j den Haushalt und k seine chronologisch durchnummerierten Einkaufsakte in der Warengruppe Haushaltspackungen Speiseeis.

Auf diese Weise haben wir die Bindungen im Einkaufsdatensatz gelöst. Allerdings vermitteln die sekundengenauen Werbekontaktdaten nun eine Scheinge-

n	2208
Anz. kaufende HH	1669
K	8604
Anz. erreichte HH	2151
J	89324
Beginn	3 January 2000
Ende	31 December 2000
Länge (d)	364

Tabelle 1: Informationen zum Datensatz

nauigkeit, da sie zufällig vor oder nach einem am selben Tag erfolgten Einkauf liegen. Deshalb verwenden wir die Werbekontaktdaten ebenfalls nur tagesgenau und subtrahieren weitere i. i. d. Variablen, die auf $[0, 0.5]$ gleichverteilt sind. Wir nehmen also insbesondere an, daß ein Werbekontakt, der am selben Tag wie ein Einkaufsakt stattfindet, diesem immer vorausgeht. Die derart entstandenen Zeitpunkte übernehmen die Rolle der X_{jk} .

Tabelle 1 liefert einige zusammenfassende Informationen über unseren Datensatz. In unser Modell gehen insgesamt $n = 2208$ Haushalte ein, von denen 1669 mindestens einmal in der von uns betrachteten Warengruppe eingekauft haben. Für 539 Haushalte gilt mithin $K_j = 0$. Insgesamt beobachten wir $K = 8604$ Einkäufe in der Stichprobe.

Fast alle Haushalte, nämlich 2151 von ihnen, hatten Werbekontakte mit Produkten der Warengruppe. Auch in der Summe beobachten wir deutlich mehr Werbekontakte als Einkaufsakte, insgesamt $J = 89324$ in unserer Stichprobe. Lediglich 57 Haushalte haben im Jahr 2000 keine einzige TV-Werbung für Haushaltspackungen Speiseeis gesehen.

Abschließend vermittelt die Abbildung 1 noch einen etwas genaueren Einblick in die Anzahl Einkaufsakte und Werbekontakte unserer Stichprobenhaushalte. In der oberen Grafik sehen wir für jede mögliche Gesamtzahl von Einkaufsaktiven k die Anzahl Stichprobenhaushalte, die genau so häufig im Jahr 2000 Produkte der Warengruppe gekauft haben. Neben den 539 Abstinenzlern identifizieren wir nun 354 Einmalkäufer, 284 Zweimalkäufer, usw.. Schlechter zu erkennen sind die insgesamt 12 Haushalte, die mehr als 30 Mal kaufen. Die höchste Zahl beobachteter Einkaufsakte für einen Haushalt beträgt 92, die zweithöchste 56.

Wie aufgrund der viel höheren Gesamtzahl bereits zu erwarten war, fällt die entsprechende Kurve für die Werbekontakte deutlich langsamer ab. Die untere Grafik der Abbildung 1 weist z. B. noch 5 Haushalte aus, die jeweils 100 Werbekontakte hatten. Die höchste beobachtete Anzahl liegt hier bei 210.

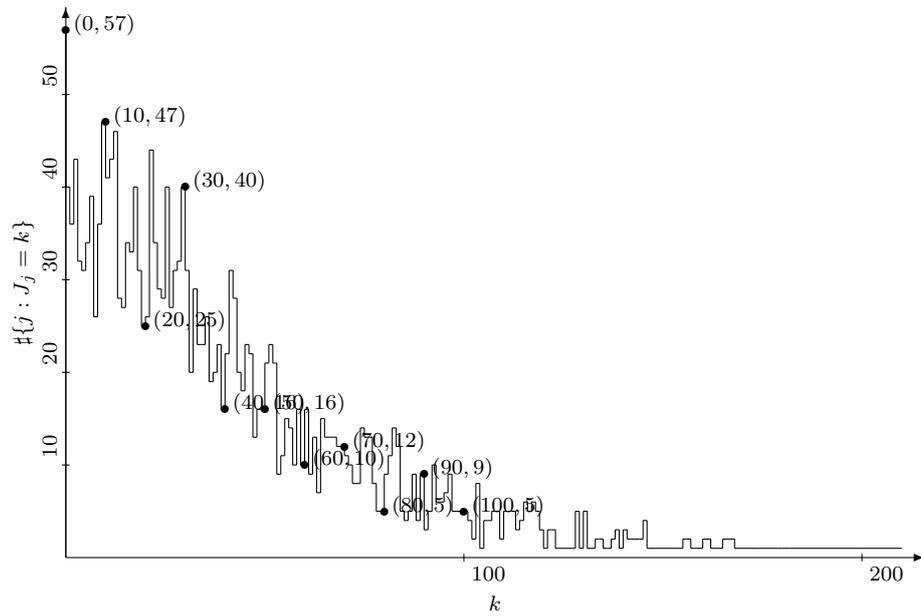
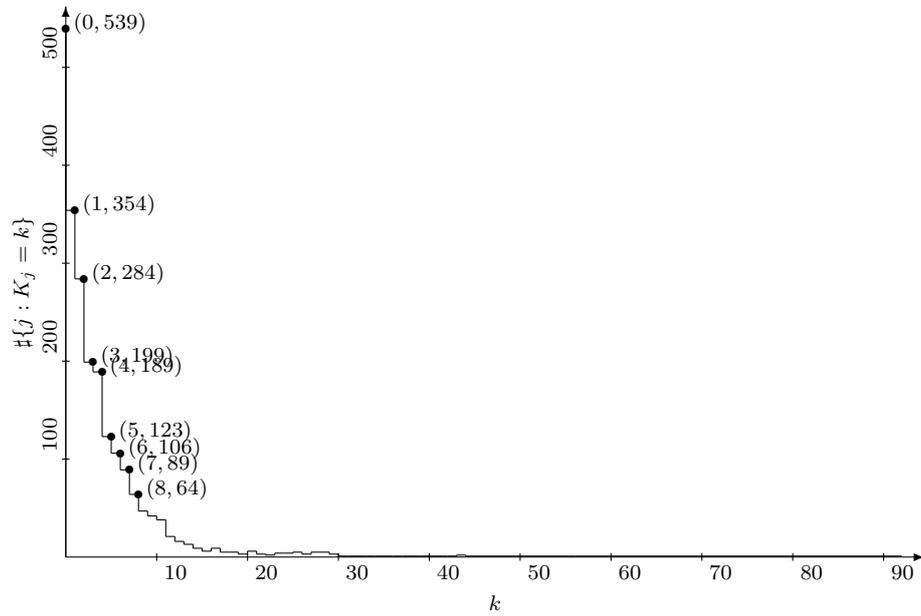


Abbildung 1: Zu jeder Anzahl Einkaufsakte und Werbekontakte k die Anzahl Stichprobenhaushalte $j = 1, \dots, n$, mit $K_j = k$ bzw. $J_j = k$

4.4 Die Kaufgeschichte

Wir beginnen nun damit, unser Modell vorzustellen. Zunächst werden wir eine Parametrisierung für die Auswirkung der eigenen Vergangenheit des Haushaltes angeben. Um die Modellformulierung möglichst einfach zu halten, werden wir uns auf einen Aspekt des Kaufverhaltens beschränken, der intuitiv recht naheliegt. Bei Haushaltspackungen Speiseeis handelt es sich um Süßspeisen, die bezogen auf ihre typische Portionierung in relativ großen Mengen verkauft werden. Deshalb erscheint es zweckmäßig, direkt nach einem Einkauf in der Warengruppe eine kurzzeitige "Sättigung" des Haushaltes anzunehmen, in der weitere Einkäufe nur noch mit geringer Intensität erfolgen. Diese dämpfende Wirkung eines Einkaufes auf weitere Einkaufsakte sollte dann jedoch ihrerseits abklingen, da der Haushalt seinen Vorrat an Speiseeis sukzessive verbrauchen wird.

Derartige Effekte treten auch in vielen anderen Anwendungsbereichen auf, so diskutiert [71, p. 319] z. B. die elektrische Aktivität im Hörnerv, die ebenfalls als Punktprozeß mit einer "selbstdämpfenden" Wirkung elektrischer Entladungen modelliert werden kann. [59, p. 659, (2)] untersucht Punktprozesse $j = 1, \dots, n$, deren Intensität sich in unserer Notation als

$$\lambda_{\vartheta,j}(t) := \vartheta_1 \cdot (1 - \vartheta_2 e^{\vartheta_3(t - Y_j N_j(t))}) \cdot \mathbf{1}\{N_j(t) > 0\} \quad (4.4.1)$$

mit dreidimensionalem Parametervektor schreiben läßt. Eine solche Formulierung genügt unserer Anforderung, wenn wir $\vartheta_2 := 1$ wählen und $\vartheta_3 < 0$ annehmen.

Von der konstanten "Grundrate" ϑ_1 werden wir uns im nächsten Abschnitt trennen. Es ist jedoch interessant, sich kurz zu vergegenwärtigen, wie sich typische Realisierungen eines Prozesses mit der Intensität (4.4.1) auf dem Intervall $I = [t_{\min}, t_{\max}]$ von der eines homogenen Poissonprozesses mit Rate

$$\vartheta'_1 := \frac{\mathbb{E} \Lambda_{\vartheta,j}(t_{\max}) - \mathbb{E} \Lambda_{\vartheta,j}(t_{\min})}{t_{\max} - t_{\min}}$$

unterscheiden. Nach Wahl von ϑ'_1 erwarten wir für beide Prozesse in I dieselbe Anzahl Ereignisse. Allerdings wird der Poissonprozeß eine höhere Anzahl sehr kurzer Intervalle zwischen zwei Ereignissen aufweisen, während der Prozeß mit Rate (4.4.1) tendentiell längere Intervalle bevorzugt. Da er dabei im Schnitt das Intervall I in dieselbe Anzahl Teilintervalle zerlegt, können wir sein Verhalten im Vergleich zum Poissonprozeß als "regelmäßiger" bezeichnen. Diese Neigung zur Musterbildung, die nicht von außen angeregt zu werden braucht, kommt unserer Vorstellung vom Kaufverhalten als einer gewohnheitsmäßigen Tätigkeit sehr entgegen.

4.5 Die Jahreszeiten

In der betrachteten Warengruppe Haushaltspackungen Speiseeis erscheint es angemessen, einen ausgeprägten jahreszeitlichen Effekt auf das Kaufverhalten anzunehmen. Wir nähern uns der Saisonalität, indem wir die Anzahl Einkaufsakte in der Stichprobe je Kalenderwoche aggregieren. Damit blenden wir die Wochentagseffekte aus, insbesondere die minimalen Einkäufe am Sonntag.

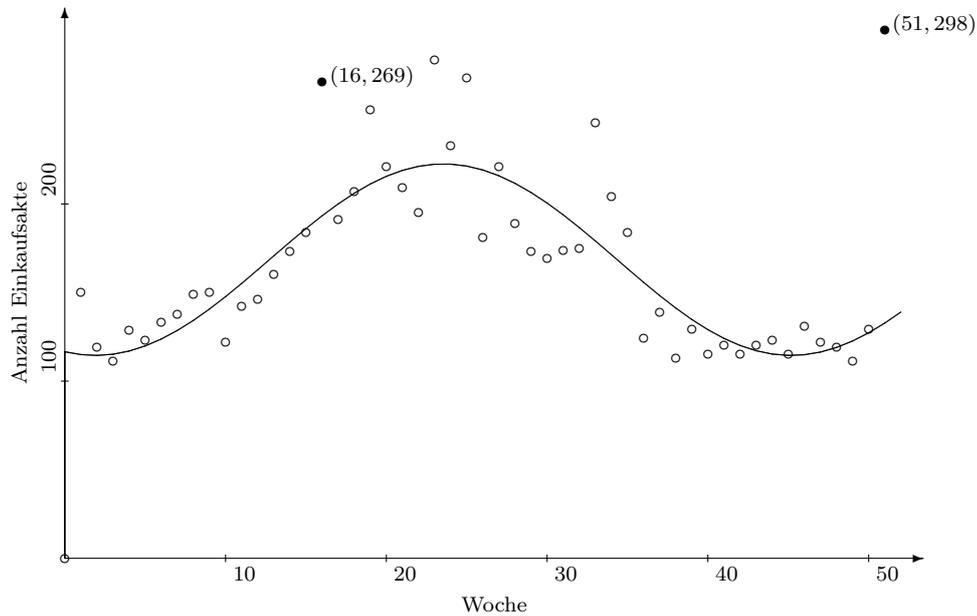


Abbildung 2: Die Anzahl Einkaufsakte der Stichprobe je Beobachtungswoche und die univariat geschätzte Sinuskurve. Oster- und Weihnachtswoche sind markiert, sie fließen nicht in die gezeigte Schätzung ein.

Abbildung 2 zeigt die derart gewonnenen Werte als Punkte. Die Oster- und Weihnachtswoche 16 bzw. 51 sind mit ihren Koordinaten gekennzeichnet.

Wir sehen in dieser Abbildung außerdem eine Kurve aus der parametrisierten Familie

$$f(t) = y_0 + y_1 \cdot \sin(x_1 \cdot (t + x_0)), \quad t = 1, \dots, 52, \quad (4.5.1)$$

die wir univariat mit einem KQS an die Wochendaten angepaßt haben. Die Oster- und Weihnachtswoche haben wir dabei nicht verwendet. Da sich die Anzahl Einkaufsakte für das gesamte Panel recht gut mit der Sinuskurve beschreiben läßt, werden wir eine entsprechende Parametrisierung der Saisonalität auch für die Intensität der i. i. d. Einkaufsprozesse einzelner Haushalte annehmen.

Die jahreszeitlichen Einflüsse lassen sich mit dem Effekt der jeweiligen Kaufgeschichte aus dem vorherigen Abschnitt technisch leicht verknüpfen, wenn wir annehmen, daß die differentielle Intensität in zwei Faktoren zerfällt. Dies haben wir in der Diskussion nach (4.4.1) bereits angedeutet. Wir ersetzen in unserer Anwendung den konstanten Faktor ϑ_1 durch eine deterministische Funktion der Zeit von der Gestalt (4.5.1). Um nicht in Konflikt mit den bisher verwendeten Parameterbezeichnungen zu kommen, verwenden wir von nun an die Notation $(\alpha, \beta, \gamma, \delta, \varepsilon)$ in der folgenden Form:

$$(\alpha \cdot \sin(\beta t + \gamma) + \delta) \cdot (1 - e^{-\varepsilon \cdot (t - Y_{iN_i(t-)})}) \cdot \mathbf{1}\{t > Y_{i1j}\} \quad (4.5.2)$$

Unsere bisherige Diskussion verdeutlicht auch, daß wir

$$\delta > \alpha > 0 \quad \beta \approx \frac{2\pi}{t_{\max} - t_{\min}} \quad \gamma \approx -\frac{\pi}{2} \quad \varepsilon > 0$$

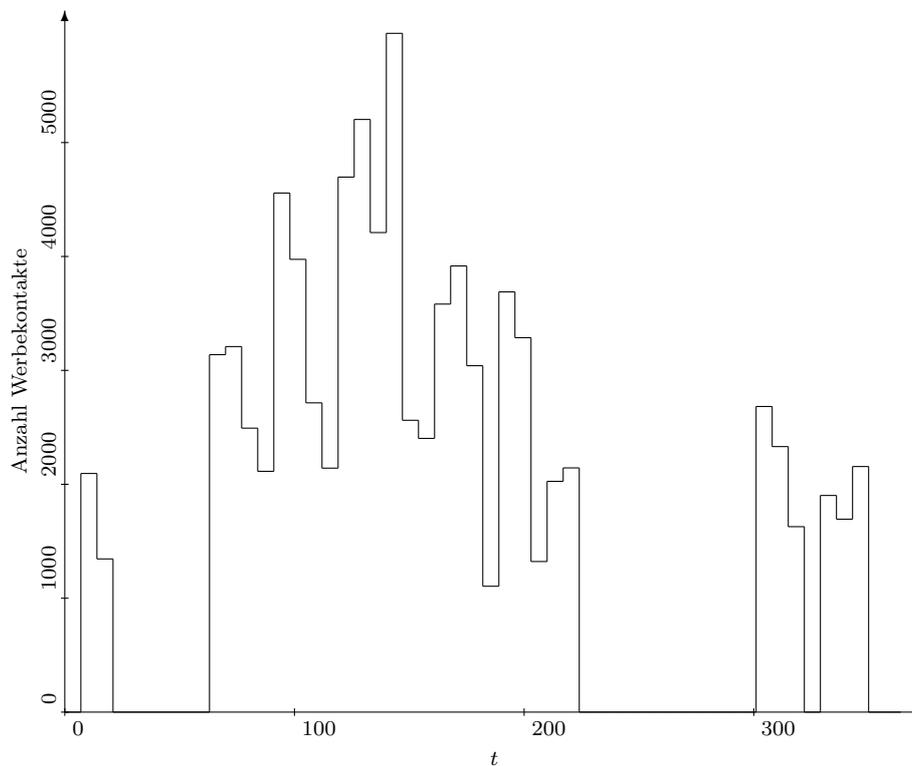


Abbildung 3: Zu jeder Kalenderwoche des Jahres 2000 die Anzahl beobachteter Werbekontakte in der Stichprobe

erwarten.

4.6 Die TV-Werbekontakte

Wir vervollständigen unser Modell (4.5.2) nun mit dem Einfluß der Werbekontakte X_{ik} , $k = 1, \dots, J_i$. Wie bereits in den beiden vorangegangenen Abschnitten konzentrieren wir uns auch hier wieder auf ein hervorstechendes Problem. Uns ist primär daran gelegen, nachzuweisen, daß mit der in Kapitel 3 vorgestellten Modellklasse typische Probleme der Kaufverhaltenstheorie prinzipiell behandelt werden können.

Hier interessiert uns vor allem die in Abbildung 3 gezeigte Tatsache, daß sich die Werbung für unsere Warengruppe selbst in einem jahreszeitlichen Rhythmus bewegt. Abgesehen von einer leichten Spitze zu Weihnachten findet sich der Hauptteil der Werbekontakte im Frühsommer. Um eine gerechte Bewertung der Werbewirkung vorzunehmen, muß unser Modell nun also die fast gleichzeitig auftretende saisonal bedingte Änderung des Kaufverhaltens identifizieren können. Natürlich sind die Werbekampagnen zweifellos gerade so geschaltet, um vom großen Sommergeschäft und der punktuellen Abverkaufsspitze zu Weihnachten bestmöglich zu profitieren. Ebenso offensichtlich sollten wir aber den überwiegenden Anteil der Absatzsteigerung auf die Jahreszeiten zurückführen,

da er zumindest auf Warenebene sicherlich auch ohne erhöhten Werbe-
druck aufgetreten wäre.

Der Schlüssel zur Auflösung dieser zeitlichen Korrelation zweier Einflüsse liegt letztlich in der Struktur der Single Source Daten. Sie ermöglichen es uns ja zumindest für die Werbekontakte, einen sehr differenzierten und haushaltsspezifischen Verlauf dieser Einflußgröße anzugeben, während wir den saisonalen Einfluß für alle Haushalte gleich annehmen. Allerdings muß das verwendete Modell natürlich technisch in der Lage sein, die funktionale Form $W_i : I \mapsto \mathbb{N}$ des Werbeverlaufs auch angemessen weiterzuverarbeiten. Und hier scheint unsere Modellklasse wegen der konsequenten Verwendung der parametrisierten stochastischen Rate $\lambda_{\vartheta,i}(\cdot)$ besonders geeignet. Wir werden zwar eine weit verbreitete und geradezu klassische Idee der Werbewirkung verwenden, den sogenannten "adstock", der uns bereits in Abschnitt 1.4 begegnet ist. Meines Wissens ist dies jedoch das erste Mal, daß dieses funktionale Konzept direkt in ein Kaufverhaltensmodell auf disaggregierten Daten einfließt. In allen mir bekannten Modellklassen muß die grundlegende Einflußfunktion zunächst relativ stark vereinfacht, z. B. aggregiert werden, bevor es technisch möglich ist, sie in die eigentliche Modellformulierung einfließen zu lassen.

Dem Konzept des adstocks liegen additive Wirkungen der einzelnen Werbekontakte zugrunde, die mit zeitlichem Abstand zum auslösenden Werbekontakt langsam abnehmen.⁴ Wir operationalisieren dies mit einer naheliegenden und technisch einfach zu verarbeitenden Funktion. Zu der Intensität (4.5.2) addieren wir folgende durch $(\zeta, \eta) \in \mathbb{R}^2$ parametrisierte Funktion auf I , die an $\sigma(W_i(t))$ adaptiert ist:

$$\zeta \sum_{h=1}^{W_i(t)} e^{\eta \cdot (t - X_{ih})} \quad (4.6.1)$$

Diese stochastische Funktion springt an jedem Werbekontakt X_{ih} des Haushaltes i um ζ und fällt im Anschluß von dem erreichten Wert exponentiell ab. Dabei haben wir

$$\zeta > 0 \qquad \eta < 0$$

vorausgesetzt. Die Auswirkungen der vergangenen Werbekontakte überlagern sich additiv, wie wir bereits angekündigt hatten.

4.7 Die Soziodemographie

Wir haben bisher das Modell

$$\begin{aligned} \lambda_{\vartheta,i}(t) := & (\alpha \cdot \sin(\beta t + \gamma) + \delta) \cdot (1 - e^{-\varepsilon \cdot (t - Y_{iN_i(t-)})}) \cdot \mathbf{1}\{t > Y_{i1}\} \\ & + \zeta \sum_{h=1}^{W_i(t)} e^{\eta \cdot (t - X_{ih})} \end{aligned} \quad (4.7.1)$$

für den Parametervektor

$$\vartheta = (\alpha, \beta, \gamma, \delta, \varepsilon, \zeta, \eta) \in \mathbb{R}^7$$

⁴"adstock" läßt sich frei als "Vorrat an Werbekontakten" übersetzen.

Haushaltsgröße					
1 Person		2 Personen		≥ 3 Personen	
Eink. (DM)	Anz. HH	Eink. (DM)	Anz. HH	Eink. (DM)	Anz. HH
< 2000	132	< 2500	100	< 3000	92
2000-2999	171	2500-2999	100	3000-3499	110
≥ 3000	96	3000-3499	135	3500-3999	129
		3500-3999	116	4000-4499	141
		4000-4999	184	4500-4999	119
		5000-5999	126	5000-5999	194
		≥ 6000	96	≥ 6000	167

Tabelle 2: Die Einteilung der Haushalte nach Größe und Einkommen

vorgesehen, vgl. (4.5.2) und (4.6.1). Für den Parameter erwarten wir dabei:

$$\begin{aligned} \delta > \alpha > 0 & & \beta \approx \frac{2\pi}{t_{\max} - t_{\min}} & & \gamma \approx -\frac{\pi}{2} \\ \varepsilon > 0 & & \zeta > 0 & & \eta < 0 \end{aligned}$$

Die letzten Einflüsse, die wir in unserem Modell noch berücksichtigen wollen, sind soziodemographische Variablen. Diese Einflußgrößen fristen in ökonometrischen Modellen häufig ein Schattendasein, da ihre adäquate Formalisierung offenbar problematisch ist:

[34, p. 1226]

In general, marketing researchers routinely exclude demographic variables from the analysis of brand choice and purchase behavior, because their effects are marginal or insignificant in the presence of market variables such as price and promotion ([14], [35], [40]).

Um die größtmögliche Flexibilität für diese Einflüsse in unserem Modell zu erreichen, ordnen wir die Haushalte entsprechend ihrer Soziodemographie in unterschiedliche Gruppen ein, für die wir jeweils ein eigenes Modell (4.7.1) schätzen. Damit erhalten wir mehrere Schätzungen von $\vartheta \in \mathbb{R}^7$, die jeweils die Auswirkungen der Kaufgeschichte, der Saison und der Werbekontakte für eine soziodemographische Gruppe angeben. Diese Schätzungen können wir auch untereinander vergleichen, womit wir einen weiteren Plausibilitätscheck für unsere Modellformulierung bekommen.

Die Gruppierung der Samplehaushalte nehmen wir wie in Tabelle 2 angegeben anhand der Haushaltsgröße und des monatlichen Haushaltsnettoeinkommens vor. Für die Haushaltsgröße wählen wir dabei die naheliegende Aufteilung in Einpersonen-, Zweipersonen- und größere Haushalte. Die Einkommensstufen fassen wir dann so zusammen, daß wir Gruppengrößen von etwa 100 bis 200 Haushalten bekommen.

4.8 Die Modellintensität

In diesem Abschnitt betrachten wir nochmals die Intensität unseres Modells (4.7.1). Die dabei hergeleiteten Darstellungen werden sowohl beim Verständnis der Simulationsergebnisse als auch bei der Implementation des Schätzers hilfreich sein. Die genaue Formulierung unseres Verfahrens als R-Programm findet sich in Anhang B.

Es erweist sich als rechentechnisch günstig, eine additive Zerlegung der Form

$$\begin{aligned}\lambda_{\vartheta,i}(t) &= (\alpha \cdot \sin(\beta t + \gamma) + \delta) \cdot (1 - e^{-\varepsilon \cdot (t - Y_{iN_i(t-)})}) \cdot \mathbf{1}\{t > Y_{i1}\} \\ &\quad + \zeta \sum_{h=1}^{W_i(t)} e^{\eta \cdot (t - X_{ih})} \\ &= \lambda_{\vartheta}^{\text{Poi}}(t) + \lambda_{\vartheta,i}^{\text{SE}}(t) + \lambda_{\vartheta,i}^{\text{Ext}}(t)\end{aligned}\quad (4.8.1)$$

vorzunehmen, wobei

$$\begin{aligned}\lambda_{\vartheta}^{\text{Poi}}(t) &:= \alpha \cdot \sin(\beta t + \gamma) + \delta \\ \lambda_{\vartheta,i}^{\text{SE}}(t) &:= -(\alpha \cdot \sin(\beta t + \gamma) + \delta) \cdot e^{-\varepsilon \cdot (t - Y_{iN_i(t-)})} \cdot \mathbf{1}\{t > Y_{i1}\} \\ \lambda_{\vartheta,i}^{\text{Ext}}(t) &:= \zeta \sum_{h=1}^{W_i(t)} e^{\eta \cdot (t - X_{ih})}\end{aligned}\quad (4.8.2)$$

gelte. $\lambda_{\vartheta}^{\text{Poi}}$ umfaßt dabei den gesamten, vom bisherigen Verlauf der Prozesse N_i und W_i unabhängigen Teil. Die Funktion ist für alle $i = 1, \dots, n$ gleich. Sie kann als Intensität eines Poissonprozesses aufgefaßt werden. $\lambda_{\vartheta,i}^{\text{SE}}$ beschreibt die Auswirkungen der Vergangenheit des Prozesses N_i , $\lambda_{\vartheta,i}^{\text{Ext}}$ quantifiziert den Einfluß des Prozesses W_i . Wir sind in unserem Kaufverhaltensmodell davon ausgegangen, daß kein Interaktionsterm zwischen N_i und W_i berücksichtigt zu werden braucht. Dies ermöglicht später einige technische Vereinfachungen in der algorithmischen Darstellung.

Für die kumulative Intensität erhalten wir aus (4.8.1):

$$\Lambda_{\vartheta,i}(t) = \Lambda_{\vartheta}^{\text{Poi}}(t) + \Lambda_{\vartheta,i}^{\text{SE}}(t) + \Lambda_{\vartheta,i}^{\text{Ext}}(t), \quad (4.8.3)$$

mit

$$\begin{aligned}\Lambda_{\vartheta}^{\text{Poi}}(t) &:= \frac{\alpha}{\beta} (\cos \gamma - \cos(\beta t + \gamma)) + \delta t, \\ \Lambda_{\vartheta,i}^{\text{SE}}(t) &:= \frac{\delta}{\varepsilon} \sum_{h=2}^{N_i(t-)} (e^{-\varepsilon \cdot (Y_{ih} - Y_{i,h-1})} - 1) + \frac{\delta}{\varepsilon} (e^{-\varepsilon \cdot (t - Y_{iN_i(t-)})} - 1) \cdot \mathbf{1}\{t > Y_{i1}\} \\ &\quad + \frac{\alpha}{\beta^2 + \varepsilon^2} \sum_{h=2}^{N_i(t-)} \left(e^{-\varepsilon \cdot (Y_{ih} - Y_{i,h-1})} (\varepsilon \sin(\beta Y_{ih} + \gamma) + \beta \cos(\beta Y_{ih} + \gamma)) \right. \\ &\quad \left. - (\varepsilon \sin(\beta Y_{i,h-1} + \gamma) + \beta \cos(\beta Y_{i,h-1} + \gamma)) \right) \\ &\quad + \frac{\alpha}{\beta^2 + \varepsilon^2} \left(e^{-\varepsilon \cdot (t - Y_{iN_i(t-)})} (\varepsilon \sin(\beta t + \gamma) + \beta \cos(\beta t + \gamma)) \right. \\ &\quad \left. - (\varepsilon \sin(\beta Y_{iN_i(t-)} + \gamma) + \beta \cos(\beta Y_{iN_i(t-)} + \gamma)) \right) \cdot \mathbf{1}\{t > Y_{i1}\}\end{aligned}$$

und

$$\Lambda_{\vartheta,i}^{\text{Ext}}(t) := \frac{\zeta}{\eta} \sum_{h=1}^{W_i(t)} (e^{\eta \cdot (t - X_{ih})} - 1).$$

Die Darstellung von $\Lambda_{\vartheta,i}^{\text{SE}}(t)$ ergibt sich dabei so: Für $t > Y_{i1}$ gilt

$$\begin{aligned} \int_{Y_{iN_i(t-)}}^t \lambda_{\vartheta,i}^{\text{SE}}(s) ds &= \frac{\delta}{\varepsilon} (e^{-\varepsilon \cdot (t - Y_{iN_i(t-)})} - 1) \\ &+ \frac{\alpha}{\beta^2 + \varepsilon^2} \left(e^{-\varepsilon \cdot (t - Y_{iN_i(t-)})} (\varepsilon \sin(\beta t + \gamma) + \beta \cos(\beta t + \gamma)) \right. \\ &\left. - (\varepsilon \sin(\beta Y_{iN_i(t-)} + \gamma) + \beta \cos(\beta Y_{iN_i(t-)} + \gamma)) \right). \end{aligned}$$

Dies setzen wir in

$$\begin{aligned} \int_{t_{\min}}^t \lambda_{\vartheta,i}^{\text{SE}}(s) ds &= \sum_{h=2}^{N_i(t-)} \int_{Y_{i,h-1}}^{Y_{ih}} \lambda_{\vartheta,i}^{\text{SE}}(s) ds \\ &+ \int_{Y_{iN_i(t-)}}^t \lambda_{\vartheta,i}^{\text{SE}}(s) ds \cdot \mathbf{1}\{t > Y_{i1}\} \end{aligned}$$

ein. Dabei können wir von der Beziehung $N_i(Y_{ih-}) = h - 1$ Gebrauch machen.

4.9 Die Wahl der Startparameter

Wir werden eine generische Optimierungsroutine verwenden, um die Parameter unseres Modells zu schätzen. Das Verfahren ist darauf angewiesen, daß wir geeignete Startwerte für die Parameter bereitstellen. In diesem Abschnitt diskutieren wir, wie sich plausible Werte für gegebene Daten wählen lassen.

Die Anfangswerte für β und γ definieren wir anhand unserer Erwartung an die Saison:

$$\begin{aligned} \beta &:= \frac{2\pi}{t_{\max} - t_{\min}} \\ \gamma &:= -\frac{\pi}{2} \end{aligned}$$

Den Parameter ε legen wir so fest, daß die dämpfende Wirkung eines Einkaufs-aktes auf die Kaufintensität nach einer Woche zu 99% abgeklungen ist:

$$\begin{aligned} 1 - \exp(-7\varepsilon) &= 0.99 \\ \varepsilon &:= \frac{\log 100}{7} \end{aligned}$$

Für die Wahl der weiteren Parameter fixieren wir zunächst einen plausiblen Anteil der Werbewirkung an der Kaufintensität für den Startparameter ϑ :

$$\frac{n^{-1} \sum_{i=1}^n \Lambda_{\vartheta,i}^{\text{Ext}}(t_{\max})}{\bar{\Lambda}_{\vartheta,n}(t_{\max})} := \frac{K}{J}$$

Die Idee hierbei ist, daß die zu erwartende Werbewirkung je Werbekontakt steigt, wenn die beobachtete Anzahl Kaufakte K wächst oder die beobachtete Anzahl Werbekontakte J sinkt.

Der Wert $\bar{\Lambda}_{\vartheta,n}(t_{\max})$ wird wesentlich von δ beeinflusst. Wir wollen mit den Startparametern $\bar{\Lambda}_{\vartheta,n}(t_{\max}) \approx \bar{N}_n(t_{\max})$ erzielen und dabei $\Lambda_{\vartheta}^{\text{Poi}}(t_{\max})$ den Anteil $1 - K/J$ einräumen. Deshalb wählen wir den Parameter δ so, daß die Gerade $\delta \cdot (t - t_{\min})$ durch die Punkte $(t_{\min}, 0)$ und $(t_{\max}, \bar{N}_n(t_{\max}) \cdot (1 - K/J))$ verläuft:

$$\delta := \frac{\bar{N}_n(t_{\max})}{t_{\max} - t_{\min}} \cdot \left(1 - \frac{K}{J}\right)$$

Den Anfangswert von α fixieren wir nun als den KQS für

$$\arg \min_{\alpha} \left((\Lambda_{\vartheta}^{\text{Poi}}(0.25t_{\max}) - (1 - K/J)\bar{N}_n(0.25t_{\max}))^2 + (\Lambda_{\vartheta}^{\text{Poi}}(0.75t_{\max}) - (1 - K/J)\bar{N}_n(0.75t_{\max}))^2 \right).$$

Dies entspricht der Wahl

$$\alpha := \frac{\beta}{2} \cdot \left(\bar{N}_n(0.75t_{\max}) - \frac{\delta \cdot t_{\max}}{2} - \bar{N}_n(0.25t_{\max}) \right) \cdot \left(1 - \frac{K}{J}\right).$$

Der Grund für diese Belegung liegt darin, daß wir im Frühling $0.25t_{\max}$ und im Herbst $0.75t_{\max}$ die größte Abweichung zwischen $\bar{N}_n(\cdot)$ und der Geraden $\delta \cdot (t - t_{\min})$ erwarten. α beschreibt die "Schwingung" von $\Lambda_{\vartheta}^{\text{Poi}}(\cdot)$ um diese Gerade. Der Parameter kann deshalb dazu verwendet werden, die Abweichung zum mittleren beobachteten Pfad an diesen Stellen gering zu halten.

Schließlich betrachten wir noch die beiden Parameter, die die Einflüsse der Werbekontakte beschreiben. Wir beginnen mit dem Parameter η , der die Nachwirkung eines Werbekontaktes quantifiziert. Um einen vernünftigen Startwert zu bekommen, betrachten wir die Abstände

$$\{Y_{ik} - X_{ih}; i = 1, \dots, n; k = 1, \dots, K_i; h = 1, \dots, J_i; Y_{ik} > X_{ih}\}$$

der Werbekontakte zu den späteren Einkaufsakte der Haushalte. Den Median dieser Menge nennen wir x . Dann setzen wir η derart, daß die Fläche unter der exponentiell abfallenden Intensität nach einem Werbekontakt in der Zeitspanne x gerade die Hälfte der Fläche $-1/\eta$ unter $e^{\eta t}$ für $t \geq 0$ beträgt:

$$\int_0^x e^{\eta t} dt = \frac{e^{\eta t}}{\eta} \Big|_0^x = \frac{e^{\eta x} - 1}{\eta} = -\frac{1}{2\eta}$$

$$\eta := \frac{\log 0.5}{x}$$

Die initiale Sprunghöhe ζ legen wir nun so fest, daß die kumulative Intensität, die durch die Werbekontakte generiert wird, gerade dem Anteil K/J an den beobachteten Einkaufsakte entspricht:

$$\frac{1}{n} \sum_{i=1}^n \Lambda_{\vartheta,i}^{\text{Ext}}(t_{\max}) \approx \frac{K}{J} \bar{N}_n(t_{\max})$$

$$\frac{J}{n} \zeta \int_0^{\infty} e^{\eta t} dt = \frac{K^2}{nJ}$$

Das Integral hat den Wert $-1/\eta$, die letzte Gleichung wird deshalb von

$$\zeta := -\eta \frac{K^2}{J^2}$$

gelöst.

5 Eine Simulationsrechnung

5.1 Einführung

Die Ergebnisse einer Simulation ergänzen unsere theoretischen Erkenntnisse über das Verhalten des KQS. Der Vollständigkeit halber geben wir kurze R-Befehle an, mit denen die präsentierten Darstellungen erzeugt werden können. Dabei gehen wir davon aus, daß der in Anhang B.3 angegebene Programmcode ausgeführt wurde und die erstellten Objekte im R-Workspace vorliegen.

5.2 Die Simulation der Prozesse

Wir verwenden das Modell (4.7.1), das unserem ökonomischen Beispiel zugrundeliegt. Den Parametervektor ϑ_\circ setzen wir auf

$$\vartheta_\circ := (0.006, 2\pi/t_{\max}, -\pi/2, 0.013, 0.1, 0.002, -0.5) \quad (5.2.1)$$

für $t_{\max} := 365$ und $t_{\min} := 0$. Diese Werte liegen in dem Bereich, der auch für unsere Anwendung realistisch ist. Wir simulieren $n = 150$ i. i. d. Pfade für dieses Modell, womit wir ebenfalls die typische Größe der Haushaltgruppen widerspiegeln. Als externe Einflußprozesse W_j , $j = 1, \dots, 150$, simulieren wir homogene i. i. d. Poissonprozesse auf $[50, 270]$ mit Rate 1 und schneiden nach spätestens 50 Ereignissen ab.

5.3 Der Verlauf der differentiellen Intensität

Wir verschaffen uns zunächst einen Eindruck vom Verlauf der differentiellen Intensität $\lambda_{\vartheta_\circ, i}(\cdot)$. Ihn erhalten wir so:

```
parameter <- c(0.006, 2*pi/tmax, -pi/2, 0.013, 0.1, 0.002, -0.5)
pathno <- 1
examplediffintens <-
  cbind(0:365, diffintensvec(0:365, parameter, pathno))
```

Die (366×2) -Matrix `examplediffintens` enthält nun in ihrer ersten Spalte eine aufsteigende Zerlegung des Intervalls $I = [0, 365]$. Die zweite Spalte liefert die zugehörigen Funktionswerte von $\lambda_{\vartheta, i}(\cdot)$ für $\vartheta = \text{parameter}$ und $i = \text{pathno}$. Abbildung 4 zeigt die Verläufe von $\lambda_{\vartheta_\circ, i}(\cdot)$ über I aus unserer Simulation für $i = 1, 2, 3$. Alle drei Intensitäten haben die Sinusschwingung $\lambda_\vartheta^{\text{Poi}}(\cdot)$ gemeinsam,

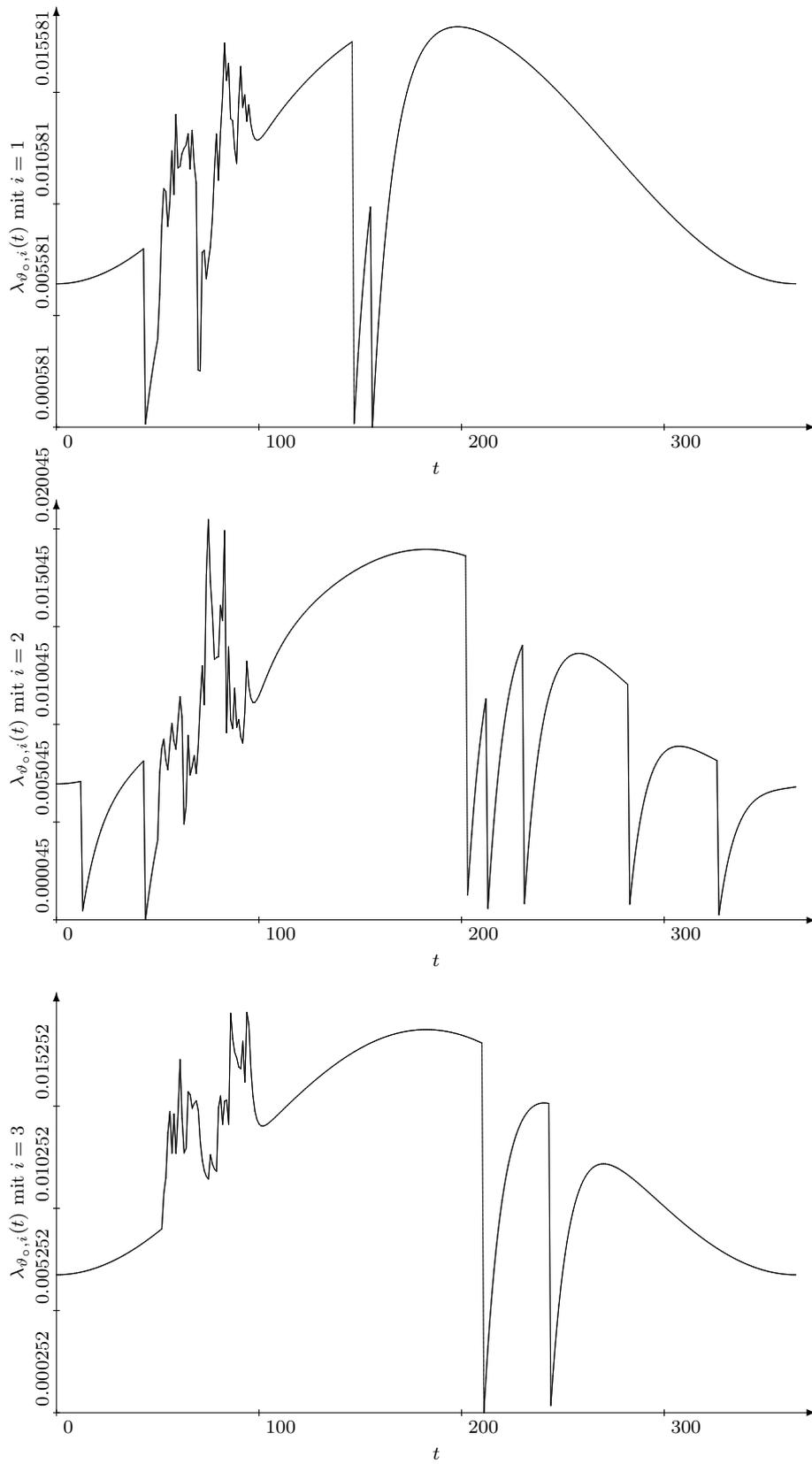


Abbildung 4: Der Verlauf von drei simulierten Intensitäten $\lambda_{\theta, i}(\cdot)$ über $I = [0, 365]$.

die für $t = 0$ und $t = 365$ zu Funktionswerten von etwa $\delta - \alpha = 0.007$ und in der Intervallmitte zu Werten von etwa $\delta + \alpha = 0.019$ führt.

Sehr deutlich treten jedoch auch die pfadspezifischen Auswirkungen $\lambda_{\vartheta,i}^{\text{SE}}(\cdot)$ eigener Sprünge hervor. Aus den Abbildungen ist leicht $K_1 = 4$, $K_2 = 7$ und $K_3 = 2$ ablesbar. An den Positionen, zu denen der zugehörige Punktprozeß selbst einen Sprung aufweist, fällt die Intensität auf Null ab⁵ und steigt im Anschluß gemäß $\lambda_{\vartheta,i}^{\text{SE}}(\cdot)$ exponentiell gegen $\lambda_{\vartheta}^{\text{Poi}}(\cdot)$ an.

Schließlich erkennen wir bei $t \in [50, 100]$ den Einfluß der externen Prozesse W_i gemäß $\lambda_{\vartheta,i}^{\text{Ext}}(\cdot)$. Die Intensität $\lambda_{\vartheta,i}(\cdot)$ reagiert auf Ereignisse X_{ik} durch einen Sprung um 0.002 nach oben, gefolgt von einem exponentiellen Abfall mit der Rate -0.5 gegen $\lambda_{\vartheta}^{\text{Poi}}(\cdot)$. Die Einflüsse für $k = 1, 2, 3, \dots$ überlagern sich additiv.

5.4 Die Kleinst-Quadrate-Schätzung

Als Nächstes betrachten wir den simulierten Verlauf von $\bar{N}_n(\cdot)$ und $\bar{\Lambda}_{\vartheta_o,n}(\cdot)$. Wieder erstellen wir dazu zweispaltige Matrizen mit Argumenten und Funktionswerten, diesmal verwenden wir jedoch die Sprungstellen der Stichprobe als Argumente:

```
examplemeanpath <-
  cbind(sort(Naug), sort(rowSums(NoJumpsNiBeforeYjk)/n))
exampletrueintens <-
  cbind(sort(Naug), sort(mittlintensitaet(parameter)))
```

Abbildung 5(a) zeigt beide Kurven über dem Intervall I . Die Cramér-von Mises-Distanz zwischen den beiden Funktionen beträgt ungefähr

$$\|\bar{N}_n(\cdot) - \bar{\Lambda}_{\vartheta_o,n}(\cdot)\|_{\bar{N}_n}^2 \approx 5.476.$$

Nun bestimmen wir den Kleinst-Quadrate-Schätzer ϑ_n^* auf dieser simulierten Datenbasis mit dem in den Abschnitten B.3.1 und B.3.2 beschriebenen R-Programm. Die Wahl der Startparameter in der Simulation ist Thema des nächsten Abschnittes. Wir erhalten

$$\begin{aligned} \vartheta_n^* &\approx (0.00668, 0.0156, -1.32, 0.012, 0.0948, 0.00215, -0.524), \\ \vartheta_o &\approx (0.006, 0.0172, -1.57, 0.013, 0.1, 0.002, -0.5), \end{aligned}$$

wobei wir den wahren Wert zum Vergleich mit derselben Genauigkeit wie die Schätzung angegeben haben. Für das Modell (4.8.1) liefert der KQS auf 150 Datensätzen also bereits eine akzeptable Näherung an ϑ_o .

Abbildung 5(b) zeigt den beobachteten mittleren Pfad $\bar{N}_n(\cdot)$ zusammen mit der geschätzten Intensität $\bar{\Lambda}_{\vartheta_n^*,n}(\cdot)$. Die Cramér-von Mises-Distanz beträgt in diesem Fall etwa

$$\|\bar{N}_n(\cdot) - \bar{\Lambda}_{\vartheta_n^*,n}(\cdot)\|_{\bar{N}_n}^2 \approx 0.2602.$$

⁵Da es sich bei den gezeigten Kurven um numerische Näherungen handelt, sieht dies in den Abbildungen fälschlich wie ein Abfall auf kleine, positive Werte aus.

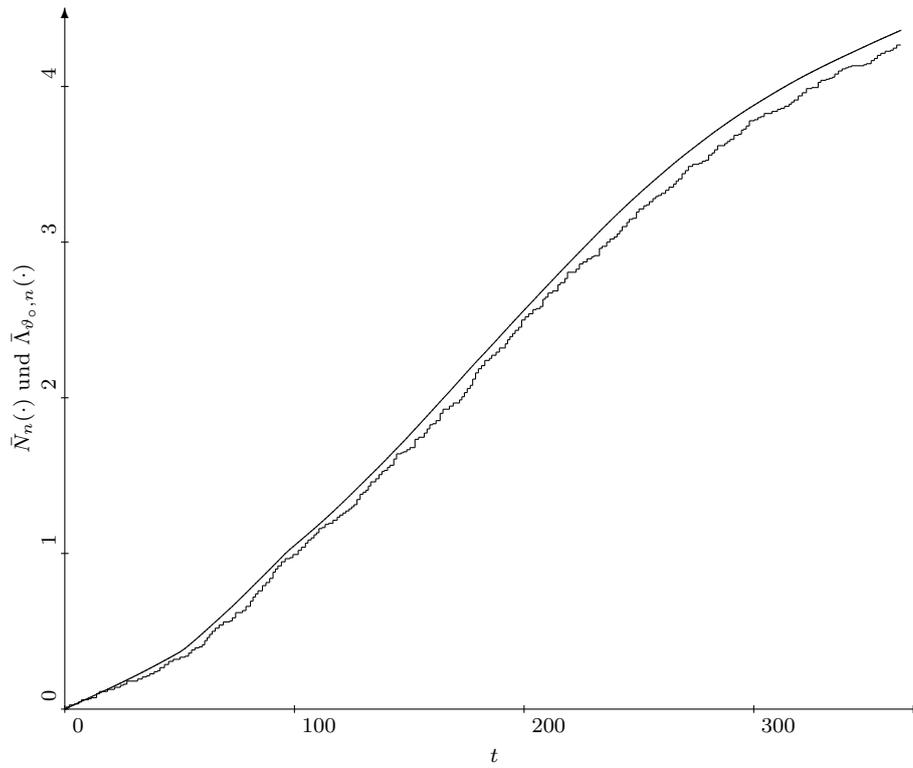
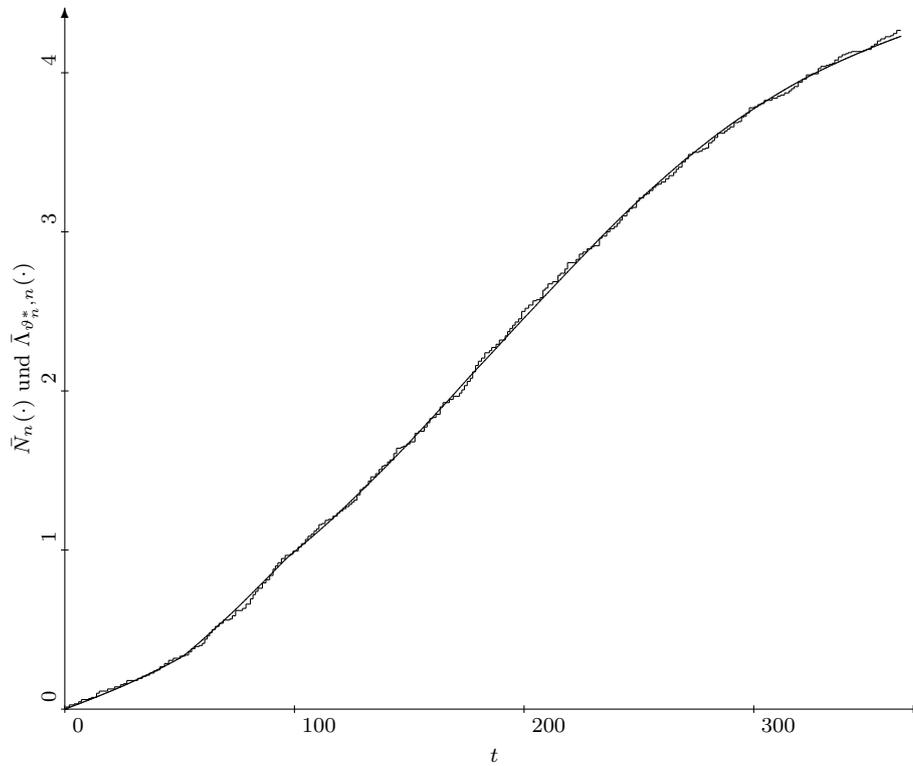
(a) Der mittlere simulierte Zählprozeß $\bar{N}_n(\cdot)$ und seine kumulative Intensität $\bar{\Lambda}_{\vartheta_0, n}(\cdot)$ (b) Der mittlere simulierte Zählprozeß $\bar{N}_n(\cdot)$ und die geschätzte Intensität $\bar{\Lambda}_{\vartheta_n^*, n}(\cdot)$

Abbildung 5: Die Kleinst-Quadrate-Schätzung in der Simulation

5.5 Die Wahl der Startparameter

Die generierten Daten unserer Simulation folgen nach Konstruktion für $i = 1, \dots, 150$ dem unterstellten Modell $\Lambda_{\vartheta,i}(\cdot)$ für $\vartheta = \vartheta_0$. Deshalb können wir Fragen bzgl. der geeigneten Modellwahl und Folgen möglicher Fehlspezifikation hier außer acht lassen. Trotzdem vermischen sich bereits in der Simulation die prinzipiellen Aussagen der Theoreme 3.3.1 und 3.5.1 mit spezifischen Eigenschaften des betrachteten Modells (4.8.3) und dem Verhalten des numerischen Verfahrens, das wir einsetzen, um den KQS zu bestimmen:

Die praktische Berechnung des KQS auf gegebenen Daten entspricht der Suche nach einem Minimum der Funktion

$$\vartheta \mapsto \|\bar{N}_n(\cdot) - \bar{\Lambda}_{\vartheta,n}(\cdot)\|_{\bar{N}_n} \quad (5.5.1)$$

auf $\Theta \subset \mathbb{R}^d$. Numerische Verfahren, die diese Aufgabe lösen, starten üblicherweise von einem vorzugebenden Startwert für ϑ und suchen in der Umgebung dieses Wertes nach Parametern, die einen kleineren Funktionswert liefern. Dieses Vorgehen wird iteriert, bis sich die Folge der erzeugten Parameterwerte bis auf eine geringe, der endlichen Rechengenauigkeit geschuldete Abweichung einem lokalen Minimum genähert hat.

Bei ungünstiger Vorgabe der Startparameter kann ein solches Verfahren offenbar anstelle des KQS ein anderes, lediglich lokales Minimum der Funktion (5.5.1) zurückgeben. Je nach gewählter Modellgleichung und vorheriger Kenntnisse über plausible Parameterwerte kann es also unterschiedlich schwierig sein, den tatsächlichen KQS in Anwendungen überhaupt zu bestimmen.

Ein weiteres Problem kann dadurch entstehen, daß sich einige der modellierten Einflußgrößen in ihrer Wirkung auf die Modellintensität zumindest teilweise aufheben.⁶ Dies wird in der Literatur als Identifikationsproblem bezeichnet. Wir erwarten diese Schwierigkeit in unserem Modell bei dem jahreszeitlichen Effekt und den TV-Werbewirkungen, die ebenfalls den erhöhten Eiskonsum im Sommer zum Teil erklären können. Aus Sicht des numerischen Verfahrens führt dies zu einem Bereich in Θ , innerhalb dessen sich die Zielfunktion (5.5.1) kaum ändert. Die endliche Präzision, mit der das Minimum dieser Funktion bestimmt wird, kann dann ebenfalls zu Problemen führen, den wahren KQS zu bestimmen.

Ein ganz ähnlich gelagerter Fall entsteht, wenn die Modellanpassung (5.5.1) auf kleine Änderungen der Komponenten des Parameters ϑ unterschiedlich stark reagiert. Die Literatur faßt dies unter den Begriff der Parametersensitivität. Falls das Modell nun auf eine Parameterkomponente viel insensitiver reagiert als auf die anderen, so kann dies zu einem Bereich in Θ entlang der Koordinatenachse der betroffenen Parameterkomponente führen, in dem sich der Zielfunktionswert kaum ändert. Dann bestimmt wiederum die numerische Präzision des Verfahrens zu einem großen Teil den Wert des geschätzten KQS.

Es ist somit klar, daß die Wahl der Startparameter einen entscheidenden Einfluß auf den berechneten Schätzer haben kann. Außerdem wird deutlich, daß ein geeignetes Verfahren zur Bestimmung von Startwerten sowohl die Funkti-

⁶Die Möglichkeit, daß sich Einflüsse vollständig aufheben, haben wir durch die Voraussetzung (3.2.7) ausgeschlossen.

	$\ N_n(\cdot) - \Lambda_{\vartheta,n}(\cdot)\ _{N_n}^2$	$\ N_n(\cdot) - \Lambda_{\vartheta_n^*,n}(\cdot)\ _{N_n}^2$
Min.	2.086	0.1835
1st Qu.	23.12	0.4101
Median	61.53	0.5824
Mean	95.11	0.6247
3rd Qu.	134.8	0.7374
Max.	591.5	1.935
Std. Dev.	96.41	0.30396
CV	1.0137	0.4866

Tabelle 3: Cramér-von Mises-Distanzen in der Simulation

on (5.5.1), und damit die Modellgleichung und die erhobenen Daten, als auch inhaltliches Verständnis der modellierten Situation berücksichtigen sollte.

Da die Funktion (5.5.1) für unser Kaufverhaltensmodell (4.8.3) kompliziert zu untersuchen ist, haben wir in Abschnitt 4.9 besonders auf die inhaltliche Plausibilität der Startwerte gesetzt.

Im Unterschied zu der Modellierung realer Daten bietet uns die Simulation die Möglichkeit, die Einflüsse der Startparameterwahl auf die Güte der Parameterschätzung direkt zu untersuchen. Um diese Möglichkeit auszunutzen, verwenden wir in der Simulation ein anderes Verfahren zur Bestimmung der Startwerte.

Dazu gehen wir von den wahren Parameterwerten aus und multiplizieren sie mit i. i. d. Zufallsvariablen, die auf $[0.75, 1.25]$ gleichverteilt sind. Mit anderen Worten, wir starten mit Werten, die um bis zu ein Viertel von den tatsächlichen Parametern abweichen. Wir hoffen, daß wir damit nahe genug am KQS starten, so daß die mögliche Existenz anderer lokaler Minima von (5.5.1) nur eine untergeordnete Rolle spielt. Der Vergleich mit den geschätzten Parametern sollte dann Aussagen zur Identifikation und Parametersensitivität ermöglichen. Wir werden diese Untersuchung in Abschnitt 5.7 durchführen, nachdem wir zuvor einen Blick auf die erzielten Modellanpassungen geworfen haben.

5.6 Die Modellanpassung

Wir wiederholen die in den Abschnitten 5.2 und 5.5 beschriebene Simulation von 150 i. i. d. Haushalten und die anschließende numerische Bestimmung des KQS 200 Mal. In Tabelle 3 fassen wir die Modellanpassung der Startparameter ϑ und KQS ϑ_n^* zusammen. In der ersten Spalte erkennen wir, daß die Startparameter im Mittel zu einer Abweichung zwischen Intensität und beobachtetem Pfad von 95.11 führen. Die Streuung dieser Werte beträgt mit 96.41 etwa genauso viel.

Nachdem die numerische Anpassung der Parameter erfolgt ist, ist die mittlere Cramér-von Mises-Distanz auf 0.6247 gesunken. Die Streuung beträgt nur noch 0.30396, knapp die Hälfte dieses Durchschnitts. Die kleinste erzielte Anpassung liegt bei 0.1835, die größte bei 1.935. Zum Vergleich verweisen wir nochmal auf die Abbildungen 5(a) und 5(b). Dort sind, wie in Abschnitt 5.4 besprochen, exemplarisch die Anpassungen 5.476 und 0.2602 zu sehen.

	α	β	γ	δ	ε	ζ	η
Min.	0.004241	0.01466	-2.051	0.01074	0.07538	-0.0002955	-0.6263
1st Qu.	0.00552	0.01647	-1.76	0.01244	0.08847	0.001426	-0.5612
Median	0.00597	0.01726	-1.608	0.0129	0.1023	0.002084	-0.5076
ϑ_0	0.006	0.01721	-1.571	0.013	0.1	0.002	-0.5
Mean	0.00603	0.01731	-1.597	0.013	0.1212	0.002168	-0.5039
3rd Qu.	0.006519	0.01819	-1.422	0.01359	0.1147	0.002926	-0.4358
Max.	0.008841	0.02063	-1.18	0.01546	4.036	0.005568	-0.3759
Std. Dev.	0.000804	0.001269	0.2196	0.000824	0.2772	0.0011	0.07421
CV	0.1333	0.07332	-0.1375	0.06339	2.286	0.5074	-0.1473

Tabelle 4: Die Verteilung der Parameterschätzungen für die Simulation

Wir können festhalten, daß die modellierte Intensität in allen 200 Simulationen sehr genau dem Verlauf des jeweils beobachteten mittleren Pfades folgt. Dies sagt allerdings aus den in Abschnitt 5.5 angeführten Gründen noch nichts über die Güte der Parameterschätzung aus, der wir uns im nächsten Abschnitt zuwenden.

5.7 Die Verteilung des KQS

Wir wollen nun die Güte der Parameterschätzungen in den 200 Simulationen diskutieren. Dazu geben wir in Tabelle 4 das arithmetische Mittel, einige Quantile und die Streuung der dabei bestimmten KQS zusammen mit dem wahren Parameterwert ϑ_0 an. Alle Werte sind mit vier signifikanten Stellen angegeben. Die starke Konsistenz $\vartheta_n^* \xrightarrow{\text{f. s.}} \vartheta_0$ ist mit den Daten der Tabelle nicht direkt zu illustrieren; wir können allerdings eng verwandte Eigenschaften wie $E(\vartheta_n^*) \rightarrow \vartheta_0$ aus dem Vergleich des durchschnittlichen geschätzten und des wahren Parameterwertes erkennen.

Der Variationskoeffizient der Startparameterverteilung hat den Betrag $48^{-0.5} \approx 0.1443$. Die Verteilung der KQS-Komponenten β und δ hat etwa halb so große Variationskoeffizienten, die von ε und ζ sind hingegen größer. Für ε scheint der hohe Variationskoeffizient allerdings auf wenige Schätzungen zurückzuführen zu sein, die zwischen dem 75%-Quantil 0.1147 und dem Maximum 4.036 liegen. Bei den restlichen drei Komponenten hat sich der Variationskoeffizient nur geringfügig geändert.

In den Abbildungen 6 bis 8 sehen wir die Verteilungen der Komponenten des Parameters, und zwar jeweils die Verteilung des Startparameters und des geschätzten Wertes im Vergleich. Für die Komponenten α , β , δ und ζ unterscheiden sich die Start- und Endverteilung deutlich voneinander. Außerdem sind diese Komponenten des KQS jeweils ungefähr normalverteilt, wie es Theorem 3.5.1 erwarten läßt.

Die drei anderen Parameter γ , ε und η hingegen scheinen von der numerischen Optimierung fast nicht angetastet worden zu sein. Start- und Endverteilung sind praktisch gleich. Dies deutet darauf hin, daß das Modell bzgl. der genauen Sai-

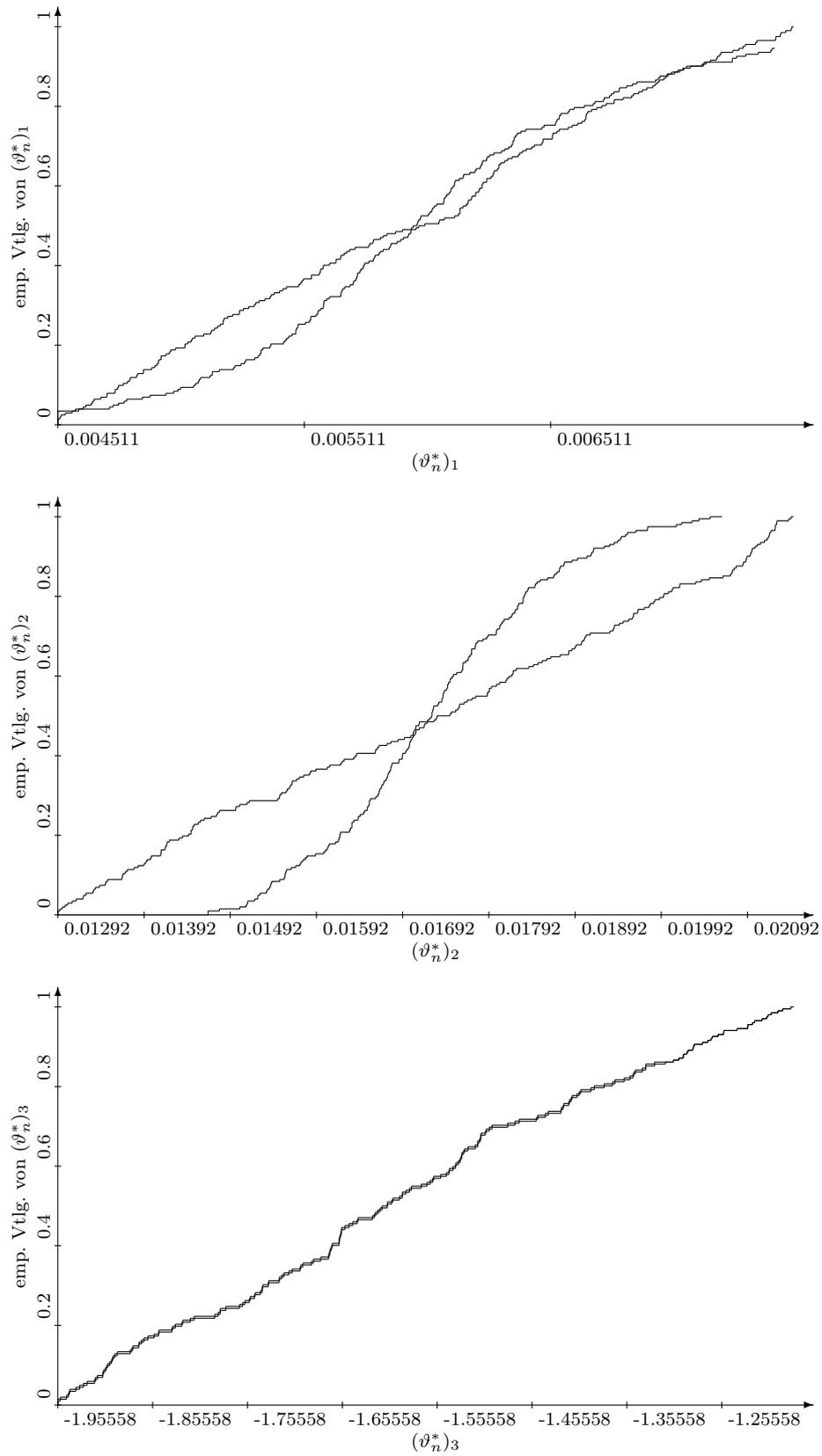


Abbildung 6: Die emp. Vtlg. des KQS und die Startvtlg.

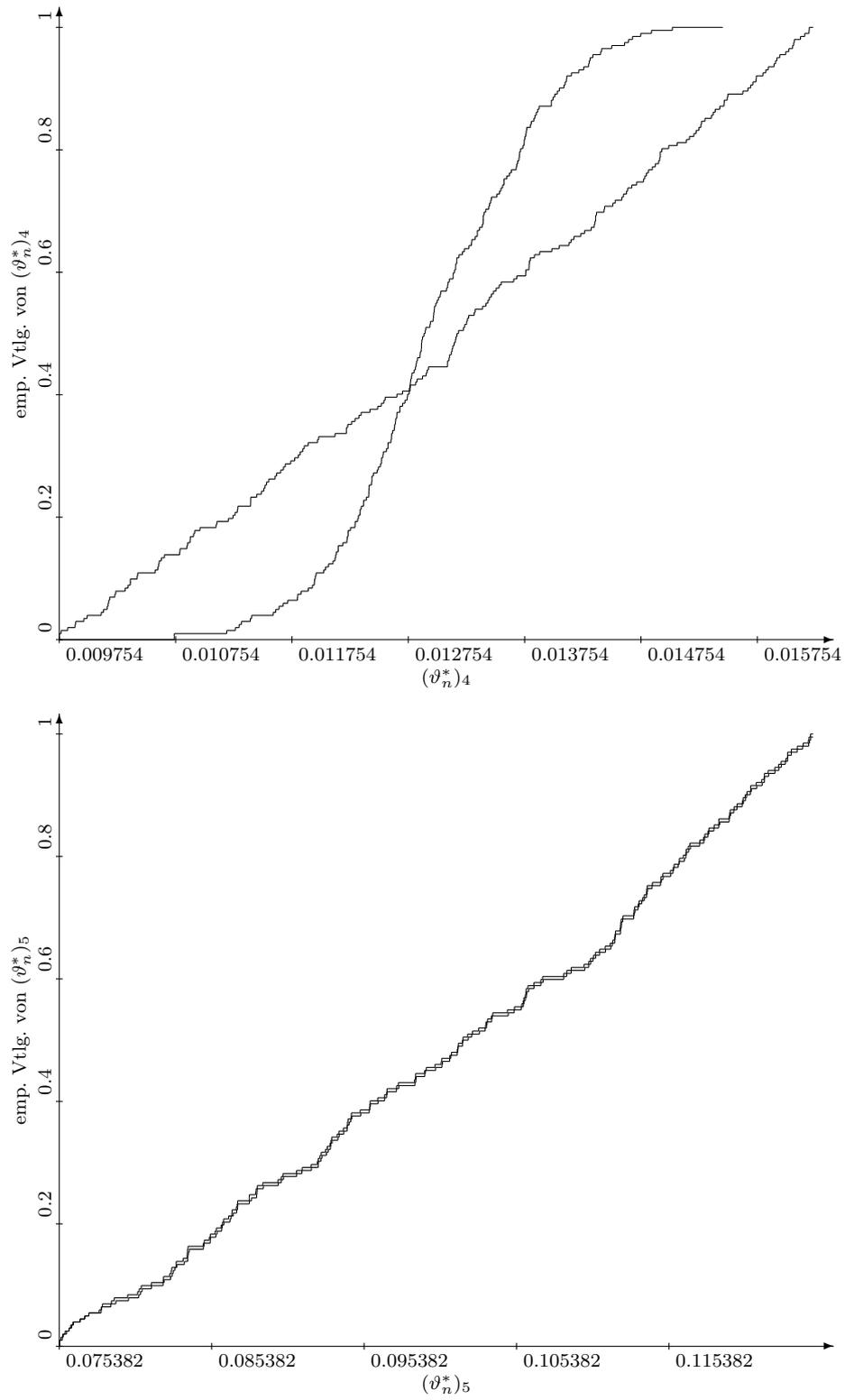


Abbildung 7: Die emp. Vtlg. des KQS und die Startvtlg.

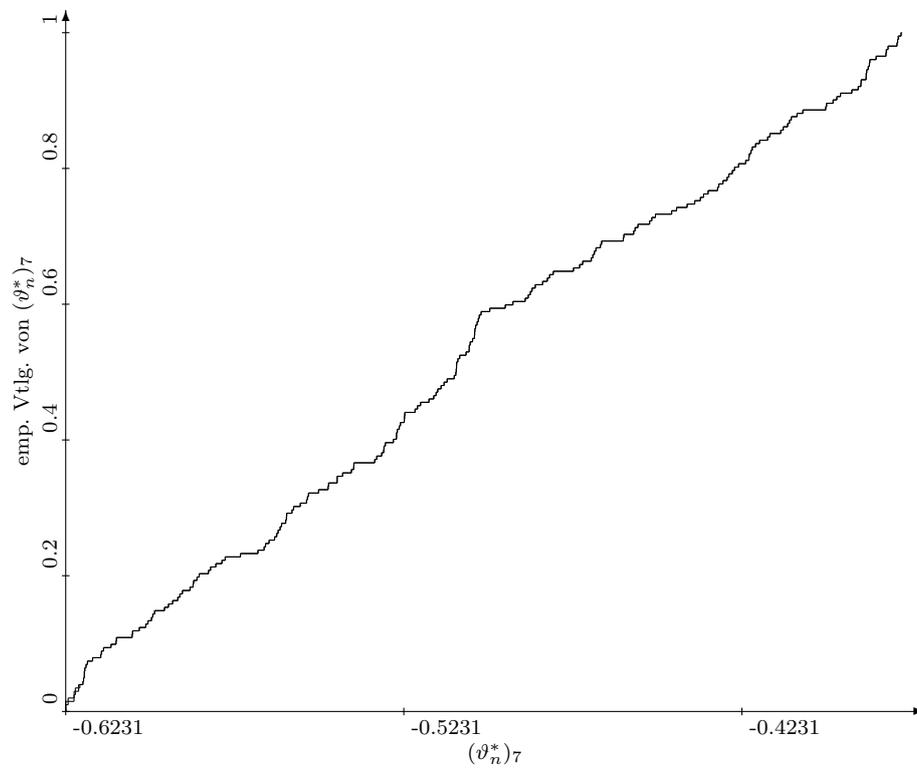
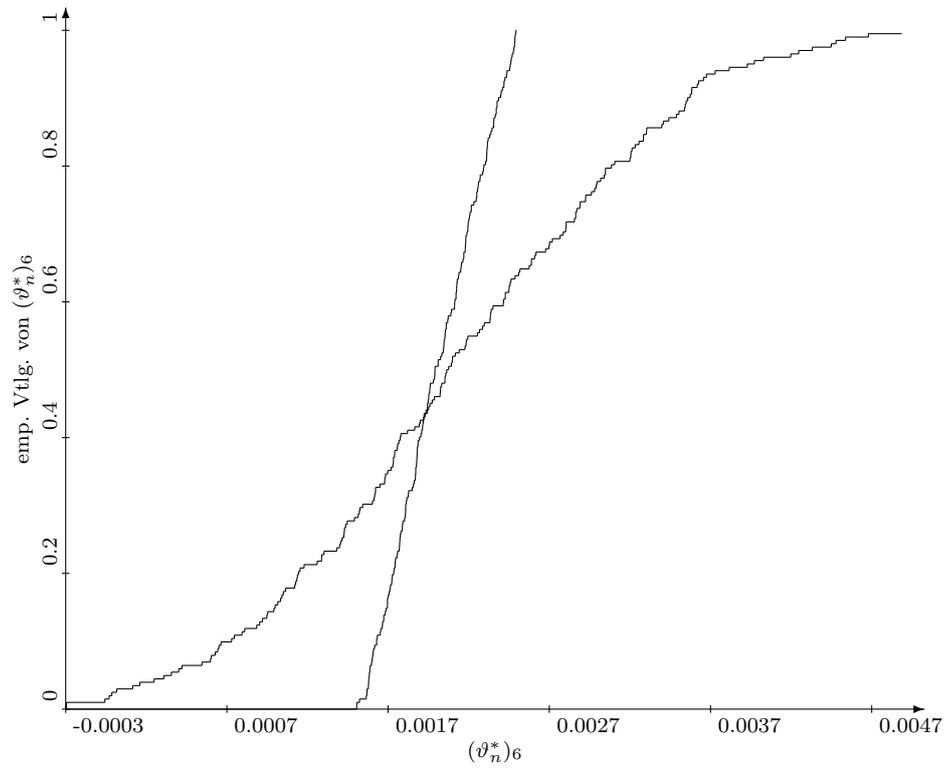


Abbildung 8: Die emp. Vtlg. des KQS und die Startvtlg.

sonspitze, der Erholungszeit nach einem Einkaufsakt und der Nachwirkung eines Werbekontaktes insensitiv ist. Hier haben also die endliche Rechengenauigkeit und die Tatsache, daß die Modellanpassung durch Veränderung der anderen vier Variablen viel stärker verbessert werden konnte, die theoretisch zu erwartende Normalität der Parameterschätzungen verhindert.

Übrigens läßt sich bei anderer Wahl der Startparameter die Normalverteilung aller Komponentenschätzungen in der Simulation erreichen: Wenn als Startparameter der wahre Wert verwendet wird, so sind auch die Schätzer der insensitiven Komponenten normalverteilt. Sie haben dann eine verschwindend geringe Streuung. Ihre Verteilung ist wohl weniger auf das Modell zurückzuführen als auf eine gewisse "Zufälligkeit", mit der die numerische Optimierung auch diese Parameterwerte versuchsweise variiert, um sich dann wegen minimaler Auswirkungen auf die Anpassungsgüte auf die anderen Parameter zu konzentrieren.

Wenden wir uns nun nochmal den normalverteilten Parameterschätzungen zu, also der Saisonamplitude α , der Saisonlänge β , der Grundrate δ und der initialen Werbewirkung ζ . Wir hatten bereits festgestellt, daß der Variationskoeffizient der KQS-Verteilungen von β und δ gegenüber der Startverteilung deutlich reduziert ist, während er für α ungefähr konstant geblieben und für ζ deutlich angestiegen ist. Es ist plausibel, die hohe Streuung der Saisonamplitude und der initialen Werbewirkung als Folge eines Identifikationsproblem anzu sehen, da beide Einflüsse die höheren Absätze von Speiseeis im Sommer erklären können. Unter dieser Situation scheint vor allem die Präzision des Werbewirkungsschätzers zu leiden. Er wird aufgrund seiner hohen Streuung manchmal sogar negativ geschätzt.

5.8 Schlußfolgerungen

Die Simulationsstudie hat ergeben, daß die praktische Schätzung unseres Modells (4.8.3) einige Probleme mit der Identifikation von Effekten und der Sensitivität einzelner Parameter hat. Diese Unzulänglichkeiten sind einerseits eine Folge der Datenlage, hier haben wir insbesondere die Saisonalität der geschalteten Werbung bemerkt, die mit einem direkten Effekt der Saison auf den Speiseeiskonsum konkurriert.

Andererseits haben einige Einflüsse, die sich größtenteils aus den ökonometrischen Thesen zu Kaufverhalten und Werbewirkung ergeben, nur eine geringe Auswirkung auf die Modellanpassung und werden deshalb von der numerischen Optimierung vernachlässigt.

Während das zuerst genannte Problem die Kovarianz des KQS lediglich vergrößert, stellt die zweite Schwierigkeit die theoretisch zu erwartende, asymptotische Normalität des Schätzers in Frage. Stattdessen hängt die Verteilung der geschätzten Parameter wesentlich von der Verteilung der Startwerte ab.

Es macht vor diesem Hintergrund wenig Sinn, die empirische Kovarianzmatrix aus den Simulationen zu untersuchen.

Auch die Approximation der Kovarianzmatrix über die i. i. d. Darstellung (B.2.5) aus den Daten eines einzigen Modellaufs ist wenig erfolgversprechend,

da wir an der Normalität der praktisch bestimmten Größen zweifeln müssen.

Hier tritt zudem noch ein weiteres schwerwiegendes numerisches Problem auf: Starke Unterschiede in der Parametersensitivität können dazu führen, daß die standardisierende Matrix $\Phi_n(\vartheta_n^*)$ eine dem Betrage nach sehr große Determinante hat. In der Simulation aus Abschnitt 5.4 erhalten wir etwa $1.55 \cdot 10^8$. Die inverse Matrix, die gemäß (B.2.6) bei der Approximation der Kovarianz des Schätzers eingesetzt werden muß, ist damit praktisch singulär. Ihre Verwendung ist aus numerischer Sicht ungünstig, die endliche Rechengenauigkeit kann zu großer Ungenauigkeit des Ergebnisses führen.

Wir werden deshalb bei der Untersuchung des Kaufverhaltensmodells auf realen Daten im nächsten Kapitel auf die Bestimmung der Kovarianzmatrizen verzichten und dies bei der Interpretation der Modellergebnisse berücksichtigen.

6 Die Anpassung des Kaufverhaltensmodells

6.1 Einführung

In diesem Kapitel stellen wir die Ergebnisse vor, die sich bei der Anpassung des in Kapitel 4 entwickelten Kaufverhaltensmodells auf die in Abschnitt 4.3 vorgestellten Daten ergeben. Dabei werden die in Kapitel 5 im Rahmen der Simulationsstudie gewonnenen Erkenntnisse eine wichtige Rolle spielen.

Wir beginnen in Abschnitt 6.2 mit einer kurzen Betrachtung zur erzielten Modellanpassung, in Abschnitt 6.3 untersuchen wir die zugehörigen Parameterschätzungen im einzelnen. Wir schließen in Abschnitt 6.4 mit einer inhaltlichen Interpretation der Modellergebnisse.

6.2 Die Modellanpassung

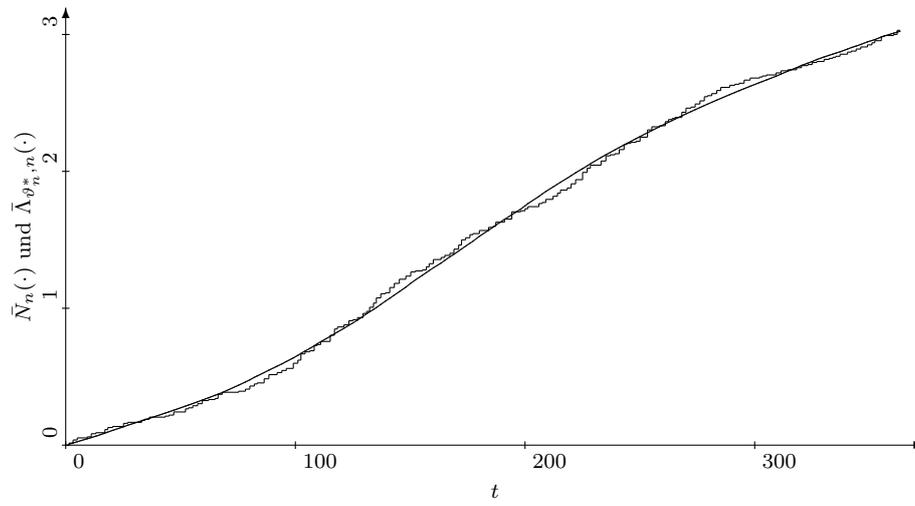
In den Abbildungen 9 bis 16 haben wir für alle siebzehn Haushaltsgruppen den mittleren Zählprozeß $\bar{N}_n(\cdot)$ und die geschätzte Intensität $\bar{\Lambda}_{\vartheta_n^*, n}(\cdot)$ aufgetragen. Diese Darstellungen vermitteln den Eindruck, daß das Modell alle Datensätze sehr gut nachbilden kann. Ein Vergleich der Cramér-von Mises-Distanzen aus Tabelle 5 mit der Simulation aus Abschnitt 5.6 bestätigt dies. Die Modellanpassung an die realen Daten ist mit der in der Simulation erzielten, sehr guten Anpassung durchaus vergleichbar: Der mittleren quadrierten Cramér-von Mises-Distanz 0.6247 in der Simulation steht hier eine quadrierte Distanz von 0.7203 gegenüber. Die maximale quadrierte Distanz beträgt unter den Modellanpassungen der Simulation 1.935; unter den Anpassungen an reale Daten liegt sie mit 2.17 nur geringfügig darüber.

Wir vergleichen außerdem die initialen Cramér-von Mises-Distanzen und damit die verschiedenen Verfahren zur Wahl der Startparameter. Es zeigt sich, daß die inhaltlich motivierte Wahl gemäß Abschnitt 4.9 zu Intensitäten führt, die dem beobachteten Pfad bereits deutlich besser angepaßt sind als die Startparameter der Simulation. Die hatten wir koordinatenweise gleichverteilt zwischen 75% und 125% des wahren Wertes gewählt. So erreichen wir auf den realen Daten

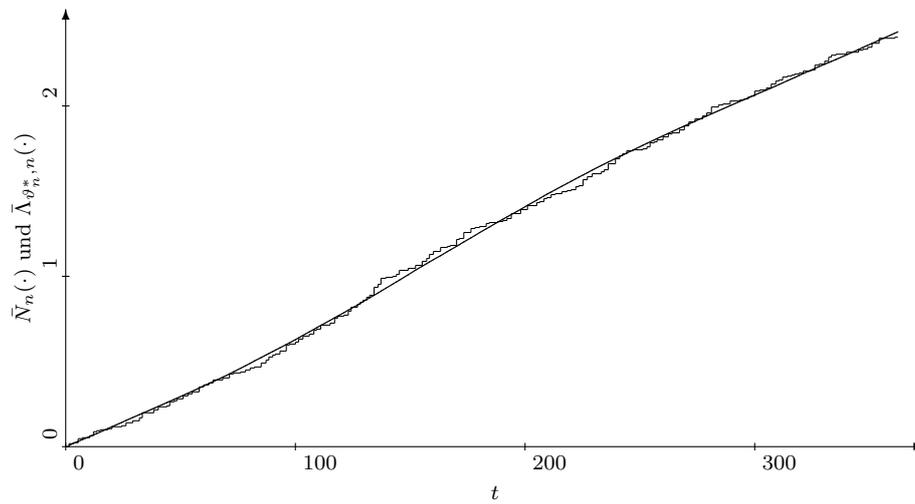
	$\ N_n(\cdot) - \Lambda_{\vartheta,n}(\cdot)\ _{N_n}^2$	$\ N_n(\cdot) - \Lambda_{\vartheta_n^*,n}(\cdot)\ _{N_n}^2$
Min.	0.6227	0.1705
1st Qu.	1.661	0.4589
Median	4.095	0.6158
Mean	5.734	0.7203
3rd Qu.	8.402	0.8355
Max.	22.25	2.17
Std. Dev.	5.664	0.49
CV	0.9878	0.6802

Tabelle 5: Cramér-von Mises-Distanzen für die Kaufverhaltensmodelle der sieben Haushaltsgruppen

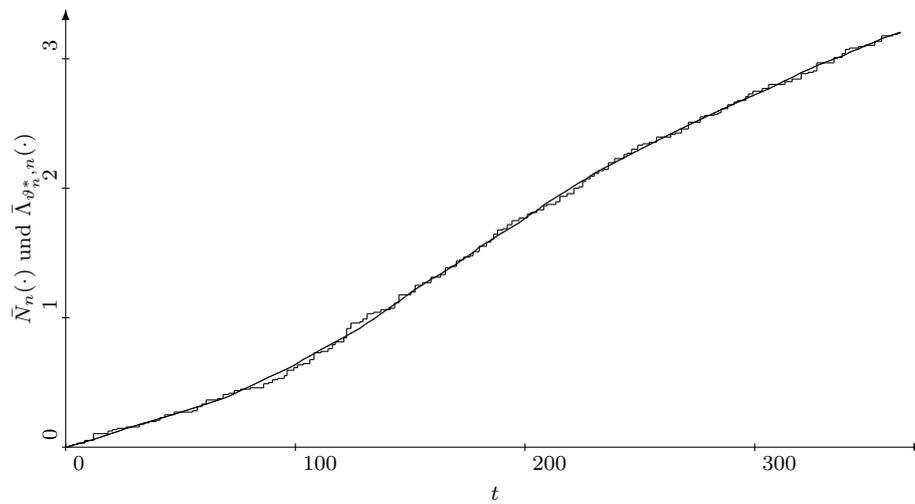
eine mittlere initiale Distanz von 5.734, der ein durchschnittlicher Abstand von 95.11 bei den Simulationen gegenübersteht. Wir können deshalb hoffen, daß die inhaltlich plausibel gewählten Startparameter nicht allzuweit von den KQS entfernt liegen. Inwieweit dies tatsächlich der Fall ist, untersuchen wir im nächsten Abschnitt.



(a) Haushaltsgruppe 1

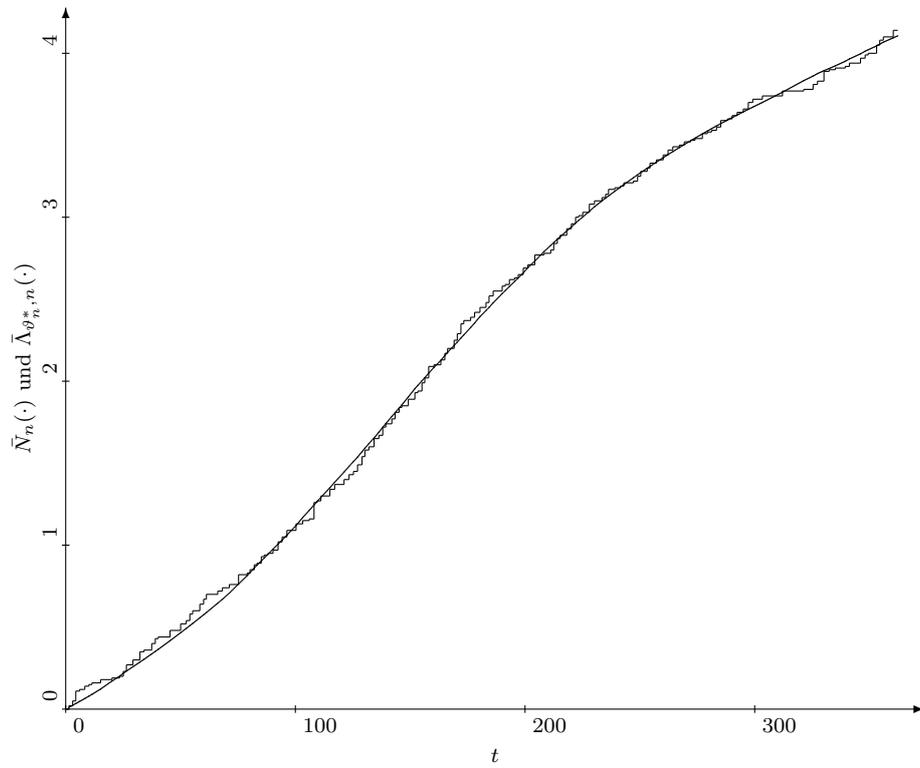


(b) Haushaltsgruppe 2

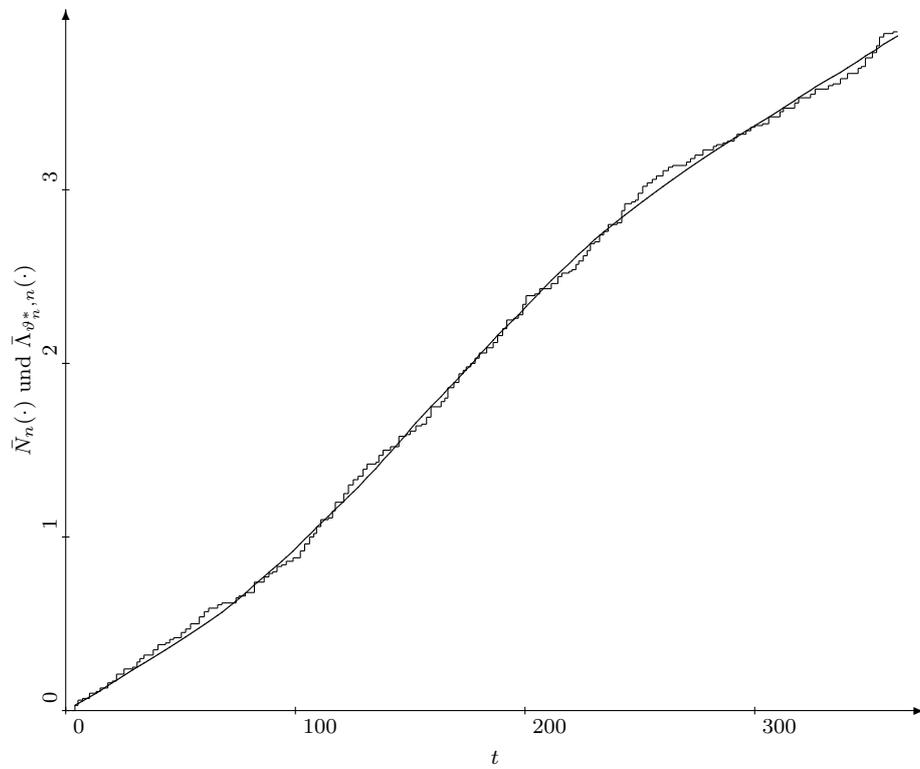


(c) Haushaltsgruppe 3

Abbildung 9: Der Zählprozeß $\bar{N}_n(\cdot)$ und die geschätzte Intensität $\bar{\Lambda}_{\theta_n^*, n}(\cdot)$

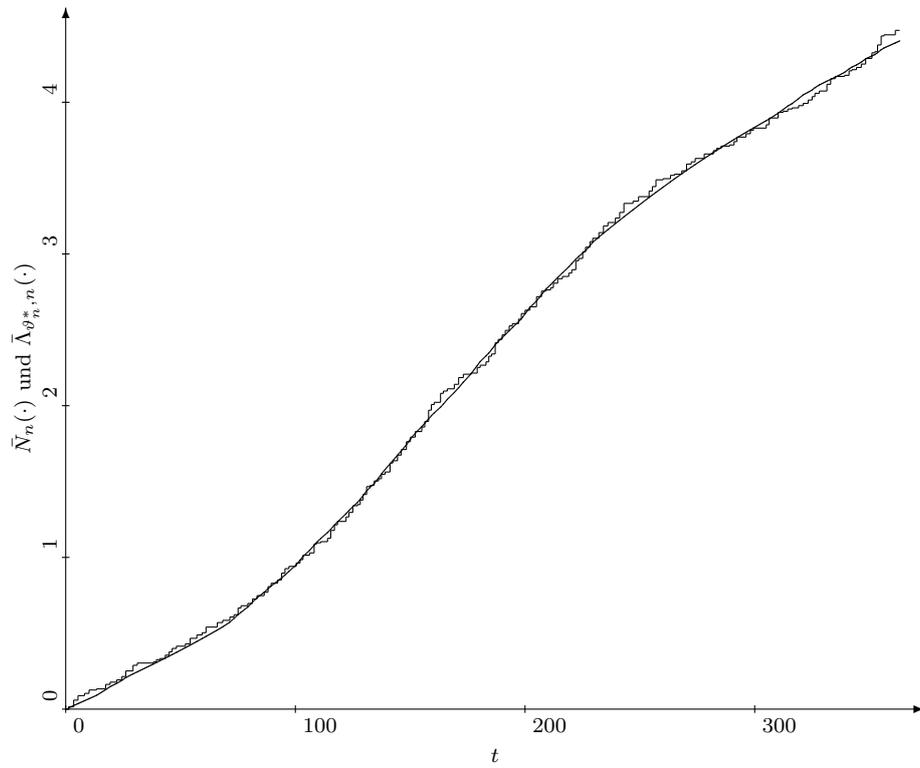


(a) Haushaltsgruppe 4

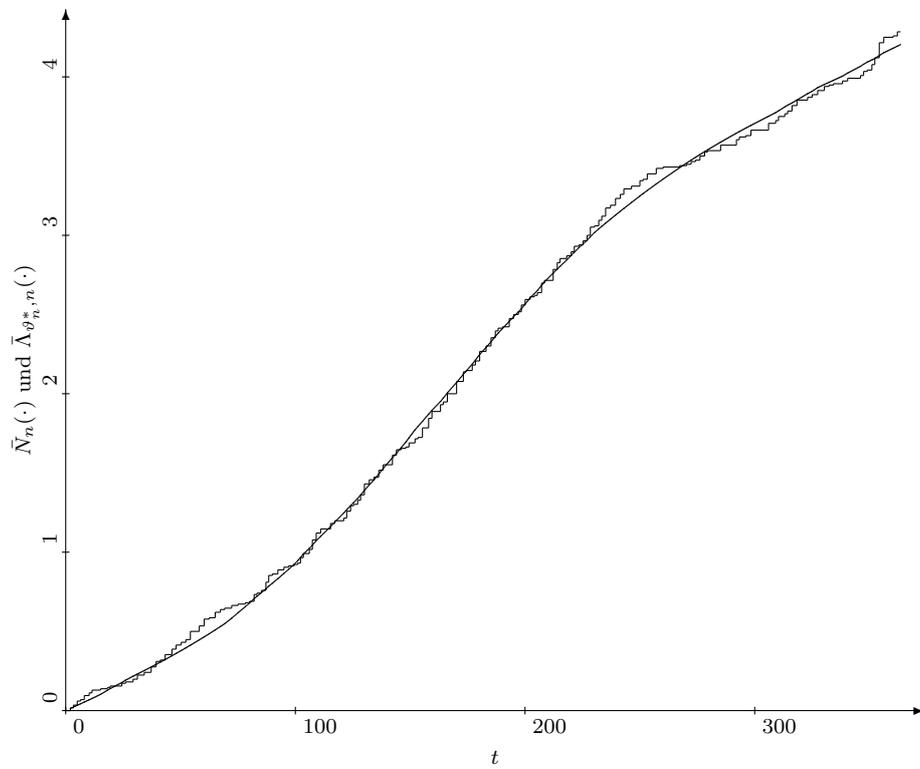


(b) Haushaltsgruppe 5

Abbildung 10: Der Zählprozess $\bar{N}_n(\cdot)$ und die geschätzte Intensität $\bar{\Lambda}_{\theta^*,n}(\cdot)$

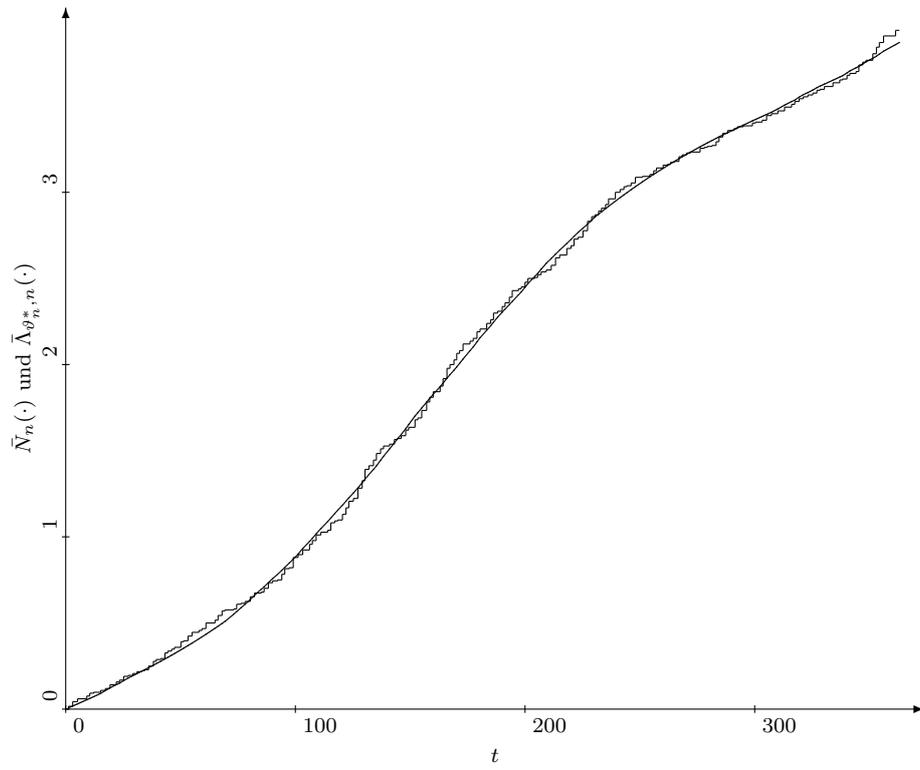


(a) Haushaltsgruppe 6

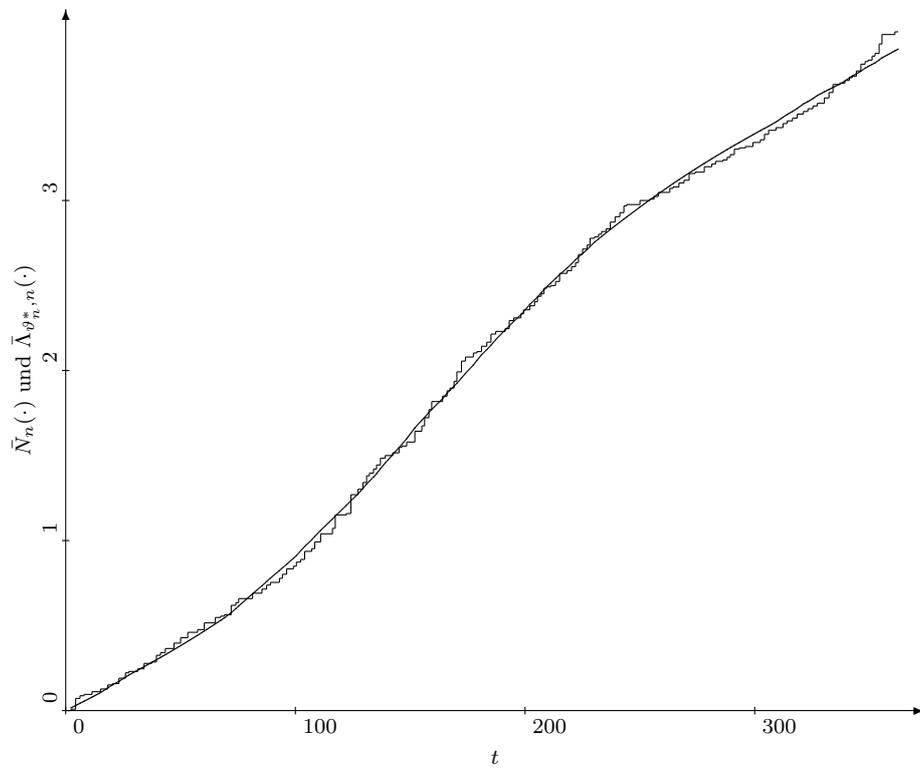


(b) Haushaltsgruppe 7

Abbildung 11: Der Zählprozess $\bar{N}_n(\cdot)$ und die geschätzte Intensität $\bar{\Lambda}_{\theta^*,n}(\cdot)$

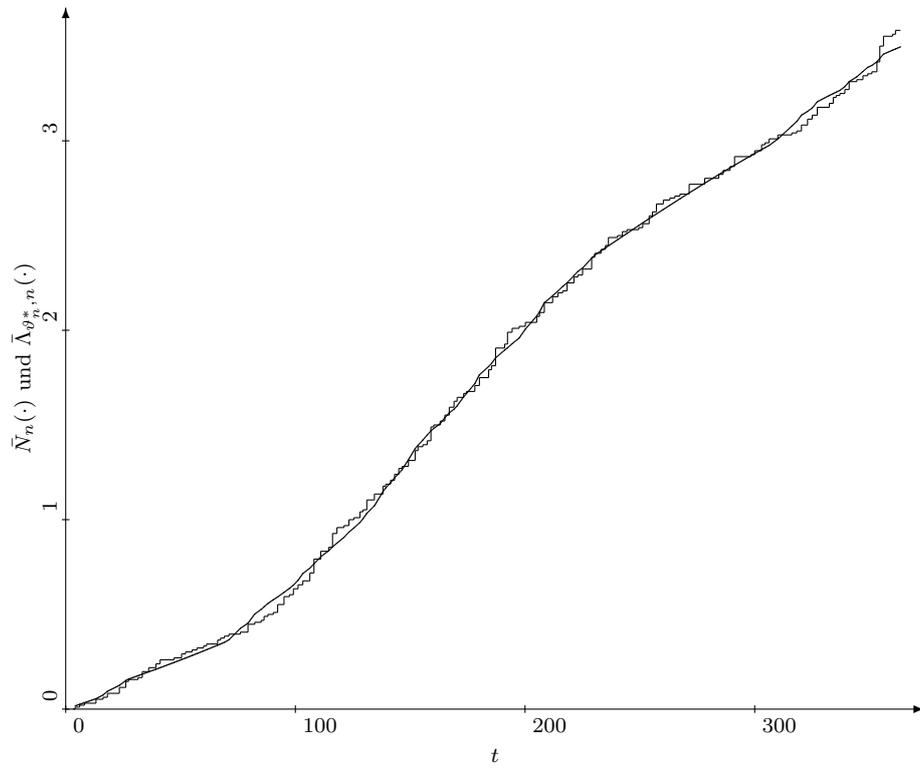


(a) Haushaltsgruppe 8

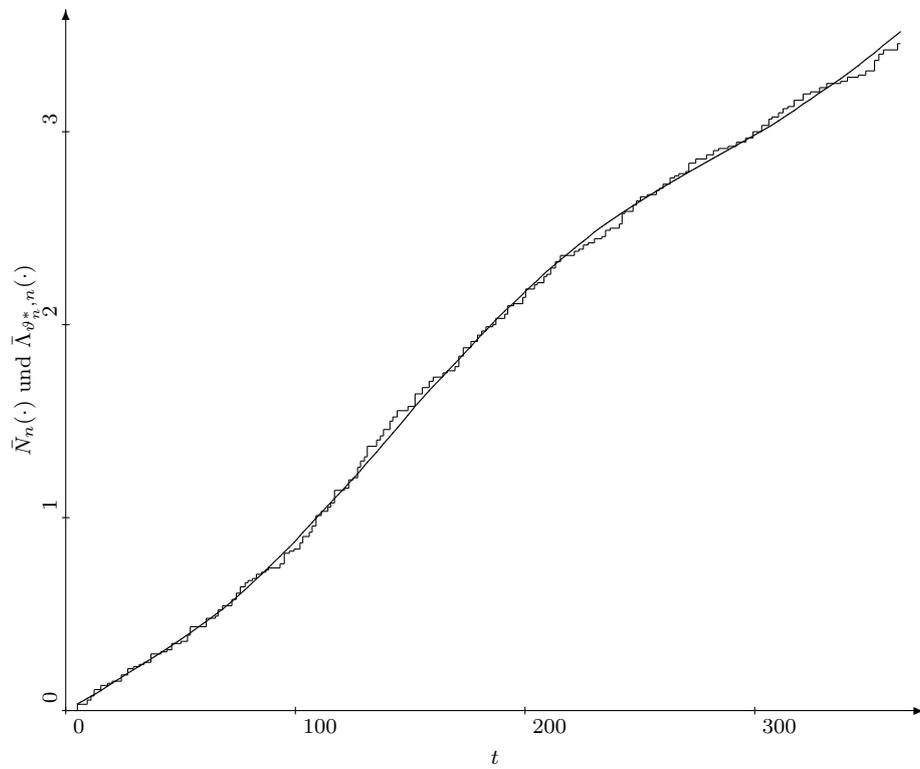


(b) Haushaltsgruppe 9

Abbildung 12: Der Zählprozess $\bar{N}_n(\cdot)$ und die geschätzte Intensität $\bar{\Lambda}_{n, n}^{g^*, n}(\cdot)$

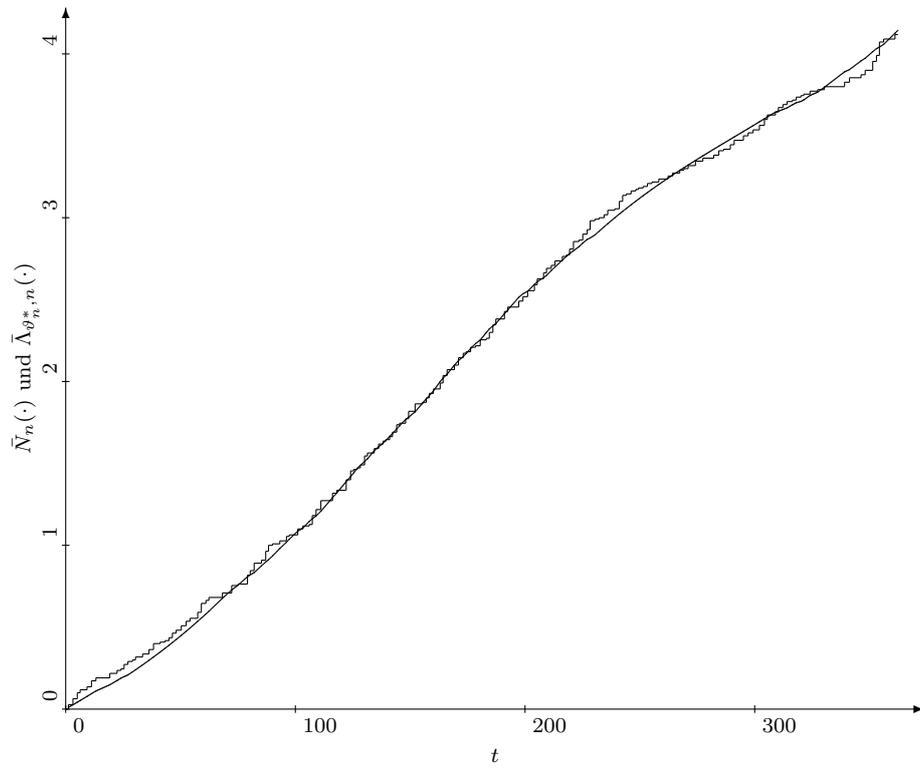


(a) Haushaltsgruppe 10

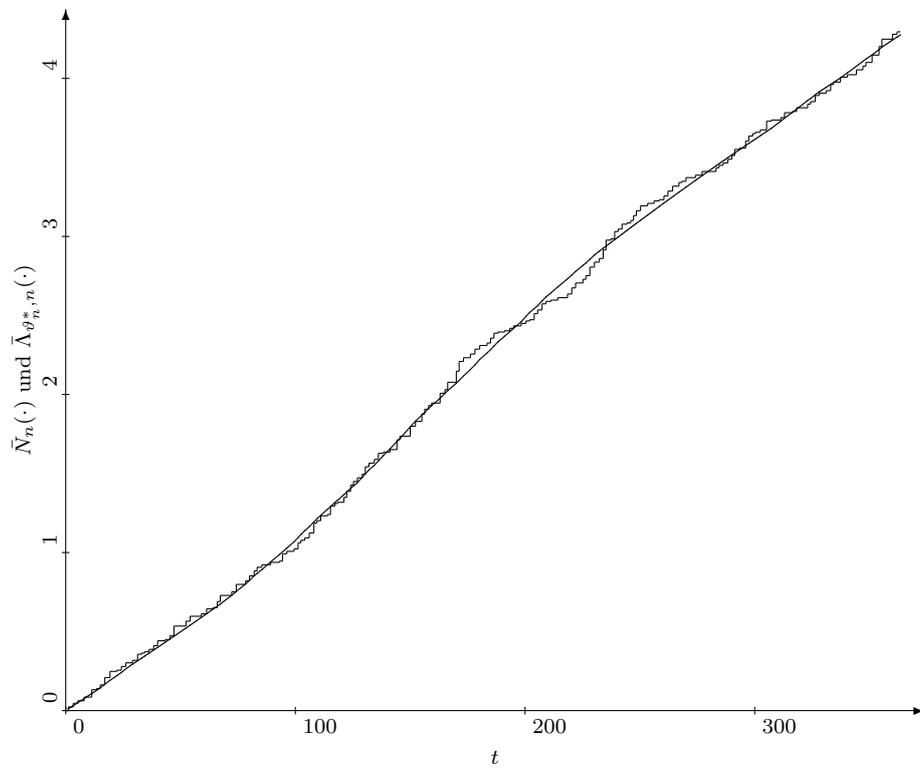


(b) Haushaltsgruppe 11

Abbildung 13: Der Zählprozess $\bar{N}_n(\cdot)$ und die geschätzte Intensität $\bar{\Lambda}_{\theta^*,n}(\cdot)$

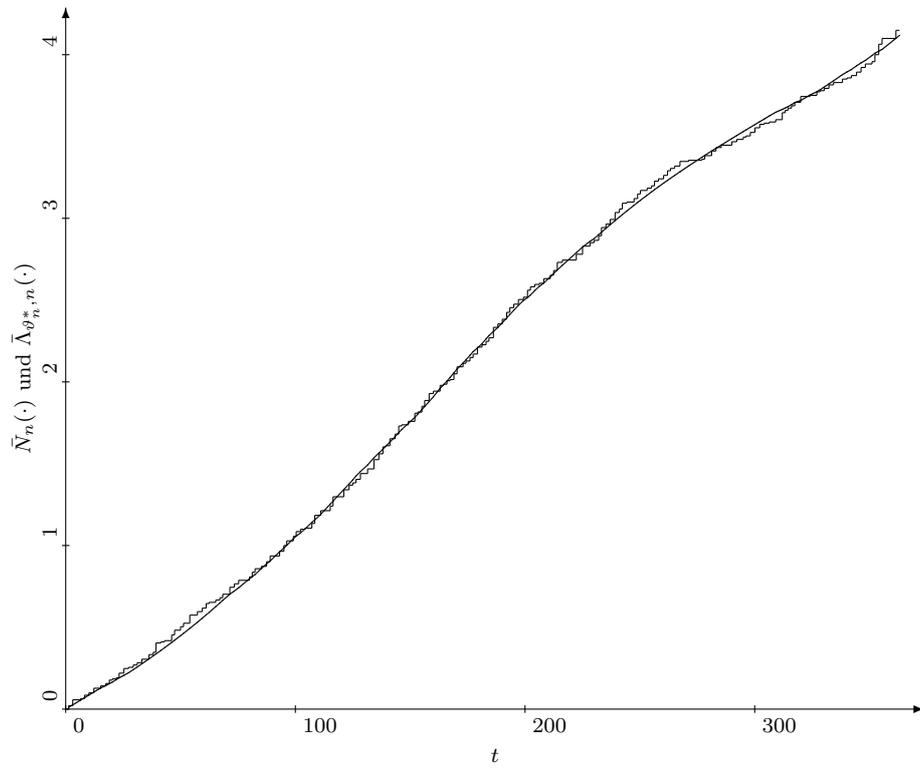


(a) Haushaltsgruppe 12

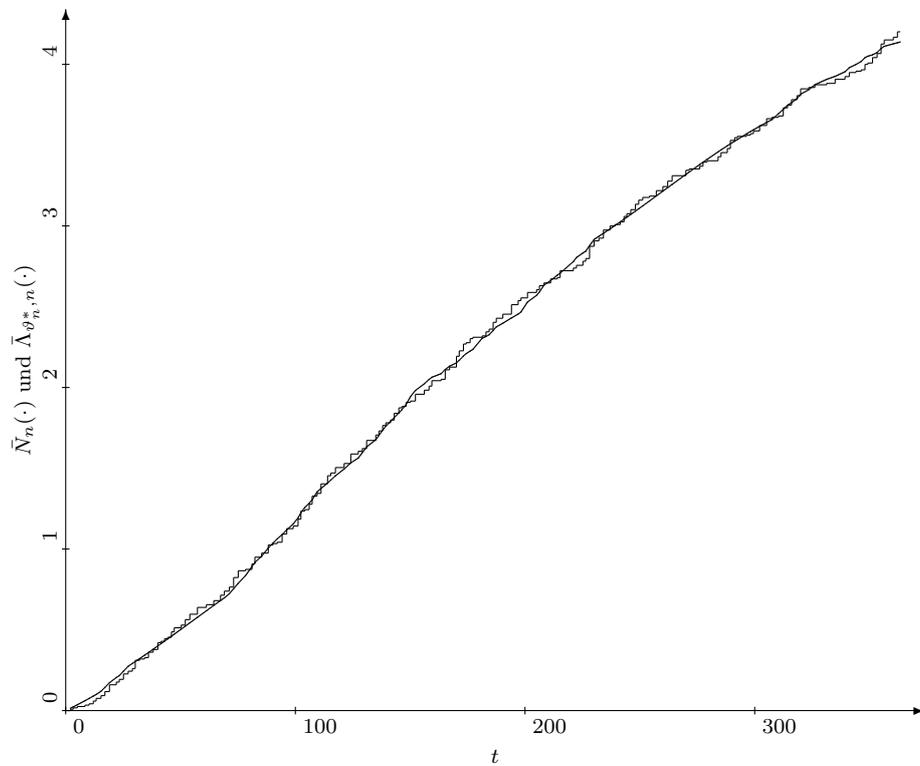


(b) Haushaltsgruppe 13

Abbildung 14: Der Zählprozeß $\bar{N}_n(\cdot)$ und die geschätzte Intensität $\bar{\Lambda}_{\theta^*,n}(\cdot)$

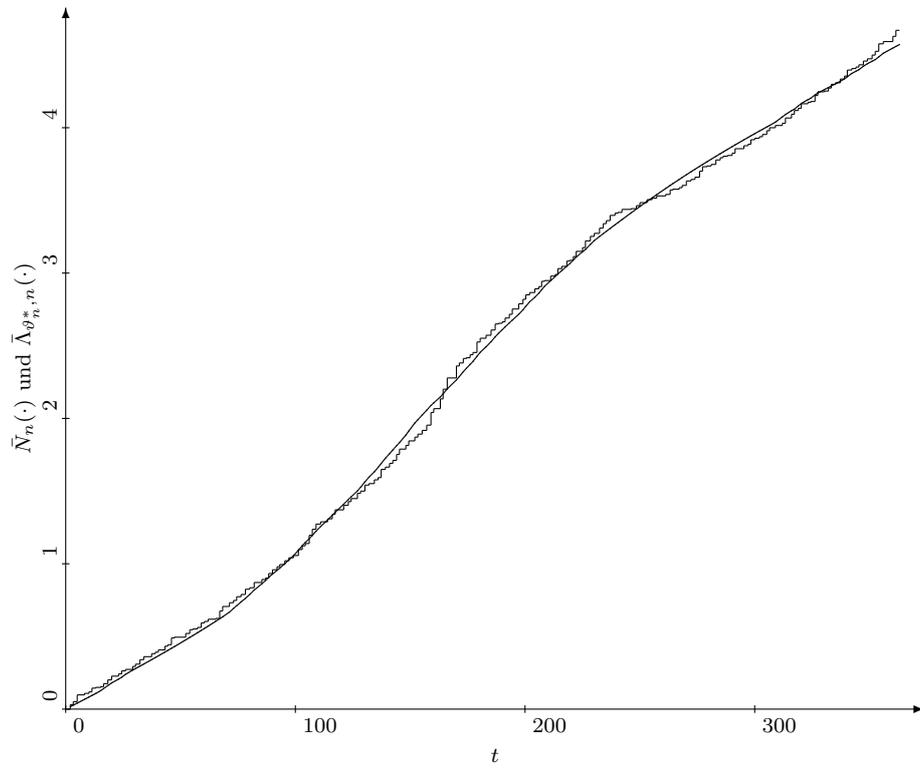


(a) Haushaltsgruppe 14

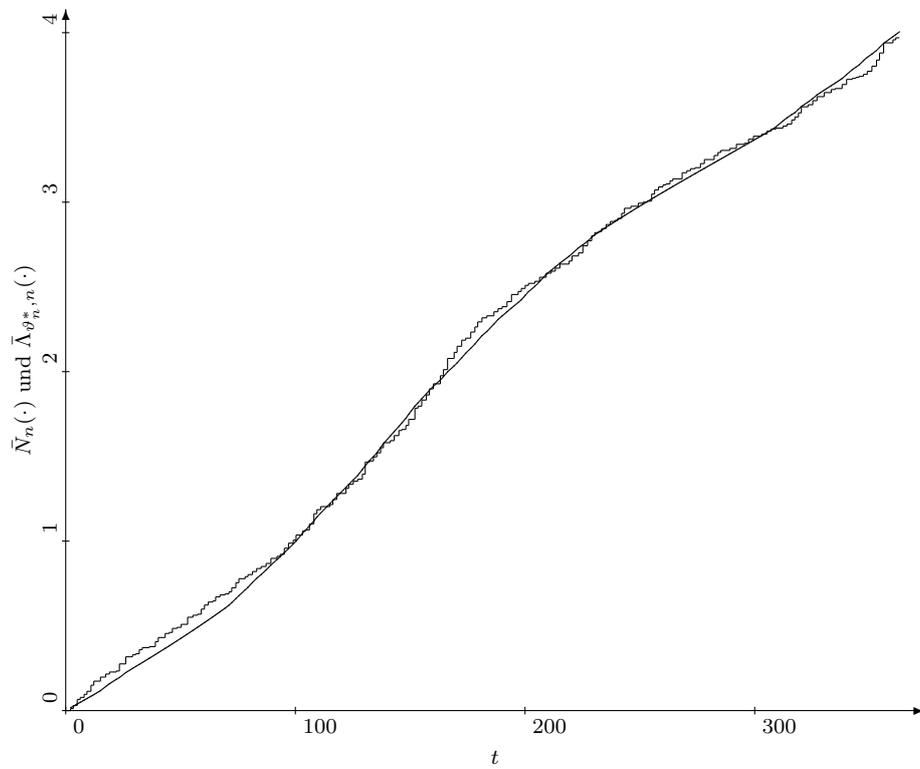


(b) Haushaltsgruppe 15

Abbildung 15: Der Zählprozeß $\bar{N}_n(\cdot)$ und die geschätzte Intensität $\bar{\Lambda}_{\theta_n^*, n}(\cdot)$



(a) Haushaltsgruppe 16



(b) Haushaltsgruppe 17

Abbildung 16: Der Zählprozeß $\bar{N}_n(\cdot)$ und die geschätzte Intensität $\bar{\Lambda}_{\theta^*,n}(\cdot)$

6.3 Die Parameterschätzung

In Tabelle 6 geben wir die Kleinst-Quadrate-Schätzungen ϑ_n^* für die siebzehn Subsample gemeinsam mit der erzielten quadrierten Cramér-von Mises-Distanz $\|\bar{N}_n(\cdot) - \bar{\Lambda}_{\vartheta_n^*, n}(\cdot)\|_{\bar{N}_n}^2$ an. Die Startwerte für die Schätzungen haben wir jeweils gemäß Abschnitt 4.9 aus den Daten bestimmt. Falls das Optimierungsverfahren den Startwert eines Parameters bei der Bestimmung des KQS um weniger als 1% verändert hat, so erscheint der entsprechende Wert in der Tabelle kursiv und eingeklammert.

Wir überprüfen zunächst die in Abschnitt 4.7 formulierten Erwartungen an die Parameterschätzungen. Die Ungleichung $\delta > \alpha > 0$ wurde in allen 17 Fällen erfüllt, $\lambda_{\vartheta_n^*, i}(\cdot)$ ist folglich für alle $i = 1, \dots, n$ strikt positiv und damit tatsächlich eine Intensität.

Die Plausibilitätskontrolle $\beta \approx 2\pi/365 \approx 0.0172$ wurde mit Ausnahme von Gruppe 15 bei allen Schätzern eingehalten. Das Modell hat in sechzehn Fällen die jährliche Saisonalität "erkannt". Bei der Platzierung der Absatzspitzen im Sommer hat es in zwei Fällen Probleme gegeben. Unsere Vermutung $\gamma \approx -\pi/2 \approx -1.57$ hat sich für die Gruppen 4 und 15 nicht bestätigt.

Die "Erholungsrate" ε der Kaufintensität nach einem Einkaufsakt ist immer positiv, die "Nachwirkungsrate" η eines Werbekontaktes immer negativ. Beides entspricht unseren Erwartungen.

Der initiale Werbewirkungseffekt ζ fällt in zwei Fällen negativ aus, und zwar für die Gruppen 12 und 14. Dies widerspricht unserer Vermutung. Allerdings hatten wir bereits in der Simulationsstudie gesehen, daß die Präzision dieser Parameterschätzung unter der Tatsache leidet, daß die Wirkung der saisonal geschalteten Werbung schwierig von dem direkten jahreszeitlichen Effekt zu trennen ist. Dasselbe Phänomen könnte auch hier aufgetreten sein; die beiden betroffenen Gruppen haben unter allen siebzehn geschätzten Saisonamplituden α die größten Werte.

Untersuchen wir nun, inwieweit das Optimierungsverfahren bei der Bestimmung der KQS die vorgegebenen Startwerte verändert hat. Dabei betrachten wir eine Differenz von mindestens einem Prozent des Ausgangswertes als Anzeichen dafür, daß die Modellanpassung durch diesen Parameter in relevantem Ausmaß beeinflußt wird. Es ergibt sich dasselbe Bild wie in der Simulationsstudie: Die Parameter γ , ε und η wurden in der Mehrzahl der Fälle, nämlich in 14, 15 bzw. 16 Gruppen, bei der Modellanpassung nur minimal verändert. Die Parameter α , β , δ und ζ hingegen wurden meistens deutlich variiert, und zwar bei der KQS-Bestimmung in 16, 16, 17 bzw. 11 Gruppen.

Gruppe	1	2	3	4	5
Anz. Pers.	1	1	1	2	2
Eink. (DM)	< 2000	2000-2999	≥ 3000	< 2500	2500-2999
α	0.00265	0.000864	0.00239	0.00426	0.00265
β	0.017	0.0193	0.0151	0.0157	0.0188
γ	(-1.57)	(-1.57)	(-1.57)	-0.663	(-1.57)
δ	0.00764	0.0066	0.00729	0.0103	0.0102
ε	(0.658)	(0.658)	(0.658)	0.667	(0.658)
ζ	0.00869	0.00407	0.0292	0.0105	(0.0148)
η	(-1.98)	(-2)	(-2)	(-2.01)	(-2)
CvM ²	0.626	0.239	0.171	0.544	0.495
Gruppe	6	7	8	9	10
Anz. Pers.	2	2	2	2	2
Eink. (DM)	3000-3499	3500-3999	4000-4999	5000-5999	≥ 6000
α	0.00354	0.00426	0.00444	(0.00306)	0.00211
β	(0.0172)	0.0187	0.0199	0.0188	0.0146
γ	(-1.57)	(-1.57)	(-1.57)	(-1.57)	(-1.57)
δ	0.0104	0.011	0.0104	0.01	0.00627
ε	(0.658)	(0.658)	(0.658)	(0.658)	(0.658)
ζ	0.029	(0.019)	0.017	(0.0222)	0.0846
η	(-2)	(-2)	(-2)	(-2)	(-2)
CvM ²	0.616	0.945	0.835	0.794	0.413
Gruppe	11	12	13	14	15
Anz. Pers.	≥ 3	≥ 3	≥ 3	≥ 3	≥ 3
Eink. (DM)	< 3000	3000-3499	3500-3999	4000-4499	4500-4999
α	0.00277	0.00607	0.00126	0.00485	0.00336
β	0.0218	0.0216	0.0185	0.0208	0.0306
γ	(-1.57)	(-1.56)	(-1.57)	-1.53	-0.0923
δ	0.00949	0.0145	0.0109	0.0134	0.00719
ε	(0.658)	(0.658)	(0.658)	(0.662)	0.729
ζ	0.011	-0.0335	(0.0203)	-0.0283	0.0804
η	(-2)	(-2)	(-2)	(-2)	-2.03
CvM ²	0.306	0.722	0.921	0.459	0.489
Gruppe	16	17			
Anz. Pers.	≥ 3	≥ 3			
Eink. (DM)	5000-5999	≥ 6000			
α	0.00298	0.00212			
β	0.0186	0.022			
γ	(-1.57)	(-1.57)			
δ	0.0113	0.0099			
ε	(0.658)	(0.658)			
ζ	(0.0387)	(0.0434)			
η	(-2)	(-2)			
CvM ²	2.17	1.5			

Tabelle 6: Parameterschätzungen und Cramér-von Mises-Distanzen der Kaufverhaltensmodelle, auf drei signifikante Stellen gerundet. Geschätzte Werte in Klammern unterscheiden sich vom Startwert um weniger als 1%.

6.4 Interpretation der Ergebnisse

Wir lassen die Ergebnisse für die Gruppe 15 außer acht, da sich die unplausibel geschätzte Saisonalität auf alle anderen Parameterschätzungen ausgewirkt haben kann.⁷

Zunächst bemerken wir, daß weder innerhalb noch zwischen den soziodemographischen Gruppen eindeutige Tendenzen der Parameterschätzungen auszumachen sind. Dies bestätigt die zurückhaltende Einschätzung soziodemographischer Effekte, die wir in Abschnitt 4.7 zitiert haben.

Wenn wir die Grundrate δ der Kaufneigung für die Gruppe 10 außer acht lassen, läßt sich für diesen Parameter zumindest feststellen, daß die Einpersonenhaushalte (Gruppe 1 bis 3) die niedrigsten Parameterwerte für δ zugewiesen bekommen haben. Für die Gruppen 1 bis 3 gilt $\delta \approx 0.007$, für die anderen Gruppen hingegen $\delta \approx 0.01$. Dies paßt sehr gut zu der betrachteten Warengruppe, die Speiseeishaltspackungen, also Süßspeisen in größeren Verpackungseinheiten umfaßt. Auf die Sonderrolle der Gruppe 10 kommen wir bei der Betrachtung der Werbewirkung ζ noch kurz zurück.

Ein anderer Parameter, der in der Simulationsstudie die erwartete Normalverteilung zeigt, ist die Saisonamplitude α . Die beiden höchsten Parameterschätzungen, Gruppe 12 und 14, betrachten wir wegen der vermuteten Überlagerung mit dem Werbeeffect ζ nicht genauer.

Die niedrigste Saisonausprägung zeigt die Gruppe 2, hier sind außerdem die Grundrate δ und die initiale Werbewirkung ζ minimal. Auch der beobachtete Zählprozeß in Abbildung 9 bestätigt den Eindruck, daß diese Gruppe von Einpersonenhaushalten eine niedrige und fast konstante Kaufneigung für die betrachtete Warengruppe über das gesamte Kalenderjahr hat.

Die anderen beiden Parameter zur Beschreibung der Saison geben ihre Länge und Position im Jahr an, es handelt sich um β und γ . Die Simulationsstudie hinterläßt den Eindruck, daß insbesondere β gut aus den Daten rekonstruiert werden kann. Betrachten wir die drei Gruppen mit der kürzesten Saison, also dem größten Wert β . Dies sind die Gruppen 11, 12 und 17. Für diese Haushalte gilt $\beta \cdot t_{\max} + \gamma > 2\pi$, d. h. am Jahresende ist der Saisoneffekt bereits wieder positiv. Wir erkennen dies auch an der modellierten Intensität, die gegen Ende des Untersuchungszeitraumes wieder steiler wird. Verantwortlich für dieses Verhalten ist vermutlich das Weihnachtsfest, daß bei diesen Haushaltsgruppen offenbar die Kaufintensität steigert, ohne daß andere Einflüsse dies in ausreichendem Maße erklären könnten.⁸

Die Saisonspitze γ unterscheidet sich zwischen den verschiedenen Gruppen kaum, lediglich Gruppe 4 fällt aus dem Rahmen. Für diese Haushalte hat das Modell die Saison ungefähr $|\gamma + \pi/2| \cdot 365/2\pi \approx 53$ Tage früher platziert, also fast zwei Monate.

Die Dämpfung ε der Kaufneigung nach einem erfolgten Einkaufsakt wurde bei

⁷Diese Gruppe ist die einzige, bei der jeder Startparameter während der KQS-Bestimmung um mehr als 1% seines Wertes verändert wurde.

⁸Über alle Haushalte betrachtet gibt es durchaus eine kleinere Werboffensive zu Weihnachten, vgl. Abbildung 3 in Abschnitt 4.6.

der KQS-Bestimmung für die Gruppen 4 und 14 leicht angepaßt, die resultierenden Schätzwerte haben jedoch kaum Auswirkungen auf die Modellintensität. Anstatt nach 7 Tagen erreichen diese Haushalte 99% ihrer vorherigen Intensität nach $\log 100/\gamma \approx 6.9$ bzw. ≈ 6.95 Tagen.

Wir beenden diesen Abschnitt mit der Betrachtung der Werbewirkungsparameter ζ und η . Die Gruppen 12 und 14 schließen wir dabei aus, da hier offenbar die Trennung des jahreszeitlichen Effekts fehlgeschlagen ist.

Der kleinste positiv geschätzte initiale Effekt ζ gehört zu Gruppe 2. Diese Einpersonenhaushalte waren uns bereits weiter oben wegen ihrer geringen Grundrate δ und der minimalen Saisonausprägung α aufgefallen. Vor diesem Hintergrund scheint eine geringe Werbewirkung plausibel. Wir erinnern nochmal daran, daß wir in unserem Modell eine starke Werbewirkung im Sinne des Abschnittes 1.2 untersuchen. Das bedeutet hier, daß auch Haushalte mit wenigen oder gar keinen Einkaufsaktens genauso zur Bestimmung der Werbewirkung beitragen wie Vielkäufer. Diese an sich naheliegende Eigenschaft unseres Modells wird von vielen Modellen der Literatur nicht geteilt, da dort der einzelne Einkaufsakt als Datenpunkt aufgefaßt wird und nicht die Kaufhistorie eines Haushaltes.

Der größte initiale Effekt ζ wurde Gruppe 10 zugeordnet. Diese wohlhabenden Zweipersonenhaushalte hatten wir bereits weiter oben hervorgehoben, und zwar wegen ihrer sehr niedrigen Grundrate δ . Tatsächlich handelt es sich hier um die einzige Gruppe von Mehrpersonenhaushalten, deren grundlegende Kaufneigung so niedrig geschätzt wird wie die der Einpersonenhaushalte. Womöglich haben wir in dieser soziodemographischen Gruppe in gewisser Weise ein Gegenstück zu der "insgesamt tragen" Gruppe 2 vor uns; zumindest scheinen die Einkaufsakte dieser Haushalte verstärkt kurz nach Werbekontakten aufzutreten und weniger einer allgemeinen Grundneigung zuzuschreiben zu sein.

Die Werbenachwirkung η hält keine Überraschungen bereit, die wenigen Abweichungen vom Startparameterwert führen ähnlich wie bei ε nicht zu deutlich anderen Modellintensitäten.

7 Beweis der starken Konsistenz des KQS

7.1 Einführung

In Abschnitt 3.2.5 haben wir im Zusammenhang mit dem Kleinst-Quadrat-Schätzer ϑ_n^* zwei Halbnormen und die zugehörigen Pseudometriken betrachtet. Mit (3.2.7) haben wir eine Bedingung an unser Modell formuliert, die eine Pseudometrik verwendet. Zum Nachweis von Theorem 3.3.1, der starken Konsistenz von ϑ_n^* , greifen wir nun diese Überlegungen wieder auf.

In Abschnitt 7.2 stellen wir wichtige technische Eigenschaften zum Beweis der starken Konsistenz zusammen. Dabei wird insbesondere das f. s. asymptotische Verhalten mehrerer arithmetischer Mittel betrachtet.

Der eigentliche Beweis des Theorems 3.3.1 wird in Abschnitt 7.3 geführt.

7.2 Technische Eigenschaften

Für $\vartheta \in \bar{\Theta}$, $\varepsilon \geq 0$ und $0 \leq r \leq r_o(\vartheta)$ bezeichne

$$B_r^\varepsilon(\vartheta) := \{\vartheta' \in \bar{\Theta} \setminus B_\varepsilon(\vartheta_o) : |\vartheta' - \vartheta| < r\}$$

den Schnitt der r -Kugel um ϑ mit dem Parameterteilraum $\bar{\Theta} \setminus B_\varepsilon(\vartheta_o)$.

Die Aussagen in den folgenden Lemmata betreffen Integrale der Form

$$\int_I |X_\vartheta(t) \cdot Y_\vartheta(t)| (Z_\vartheta^+(dt) + Z_\vartheta^-(dt)),$$

wobei $X_\vartheta, Y_\vartheta, Z_\vartheta^+$ und Z_ϑ^- durch $\vartheta \in \Theta$ parametrisierte Prozesse auf dem kompakten Intervall I seien. Z_ϑ^+ und Z_ϑ^- seien f. s. positiv, monoton wachsend und rechtsstetig. Aussagen der Gestalt

$$\mathbb{E} \left(\sup_{\vartheta' \in B_r^\varepsilon(\vartheta)} \int_I |X_{\vartheta'}(t) \cdot Y_{\vartheta'}(t)| (Z_{\vartheta'}^+(dt) + Z_{\vartheta'}^-(dt)) \right) < \infty \quad (7.2.1)$$

können in unserer Situation folgendermaßen sichergestellt werden: Der Prozeß $|X_\vartheta(t) \cdot Y_\vartheta(t)|$ sei durch einen positiven, auf I rechtsstetigen und monoton wachsenden Prozeß \tilde{X}_ϑ beschränkt:

$$|X_\vartheta(t) \cdot Y_\vartheta(t)| \leq \tilde{X}_\vartheta(t) \quad \forall t \in I, \vartheta \in \Theta$$

(7.2.1) folgt nun, sofern der neu eingeführte Prozeß folgende Bedingung am rechten Intervallende erfüllt:

$$\mathbb{E} \sup_{\vartheta' \in B_r^\varepsilon(\vartheta)} \left(\tilde{X}_{\vartheta'}(t_{\max}) \cdot (Z_{\vartheta'}^+(t_{\max}) + Z_{\vartheta'}^-(t_{\max})) \right) < \infty \quad (7.2.2)$$

Da das Z -Integral und der Erwartungswert monoton sind, gilt nämlich mit den soeben eingeführten Bezeichnungen:

$$\int_I |X_\vartheta(t) \cdot Y_\vartheta(t)| (Z_\vartheta^+(dt) + Z_\vartheta^-(dt)) \leq \tilde{X}_\vartheta(t_{\max}) \cdot (Z_\vartheta^+(t_{\max}) + Z_\vartheta^-(t_{\max})) \quad (7.2.3)$$

7.2.1 Lemma Seien $\vartheta' \in \Theta$ und $i = 1, \dots, n$. Wir betrachten die Prozesse bzw. Funktionen:

$$X_{\vartheta',i} \in \{N_i - \Lambda_{\vartheta'',i}, \mathbb{E} N_i - \mathbb{E} \Lambda_{\vartheta'',i}, \vartheta'' \in \Theta\}$$

Seien des weiteren $\vartheta \in \Theta$ und $j, k = 1, \dots, n$. Mit den Voraussetzungen (3.2.1) und (3.2.6) existieren die folgenden Erwartungswerte:

$$\mathbb{E} \left(\int_I |X_{\vartheta',i}(t)| |X_{\vartheta',j}(t)| N_k(dt) \right) < \infty \quad (7.2.4)$$

$$\mathbb{E} \left(\sup_{\vartheta' \in B_r^\varepsilon(\vartheta)} \int_I |X_{\vartheta',i}(t)| |X_{\vartheta',j}(t)| N_k(dt) \right) < \infty \quad (7.2.5)$$

Beweis:

Wir zeigen (7.2.5), daraus folgt dann (7.2.4). In (7.2.3) setzen wir:

$$\begin{aligned} X &= |N_i - \Lambda_{\vartheta',i}| & Y &= |N_j - \Lambda_{\vartheta',j}| \\ Z^+ &= N_k & Z^- &= 0 \\ \tilde{X} &= N_i N_j + N_i \Lambda_{\vartheta',j} + N_j \Lambda_{\vartheta',i} + \Lambda_{\vartheta',i} \Lambda_{\vartheta',j} \end{aligned}$$

Außerdem verwenden wir die Höldersche Ungleichung und die Voraussetzungen (3.2.1) und (3.2.6). \blacksquare

Die Aussagen des Lemmas 7.2.1 nutzen wir nun, um das f. s. asymptotische Verhalten mehrerer arithmetischer Mittel zu untersuchen:

7.2.2 Lemma *In der Situation von Lemma 7.2.1 gilt:*

$$\begin{aligned} & \frac{1}{n} \sum_{i=1}^n \sup_{\vartheta' \in B_r^\varepsilon(\vartheta)} \int_I (N_i(t) - \Lambda_{\vartheta',i}(t))^2 N_i(dt) \\ & \xrightarrow{\text{f. s.}} \mathbb{E} \sup_{\vartheta' \in B_r^\varepsilon(\vartheta)} \int_I (N(t) - \Lambda_{\vartheta'}(t))^2 N(dt) \end{aligned} \quad (7.2.6a)$$

$$\begin{aligned} & \frac{1}{n \cdot (n-1)} \sum_{\substack{i,j=1 \\ i \neq j}}^n \sup_{\vartheta' \in B_r^\varepsilon(\vartheta)} \int_I (N_i(t) - \Lambda_{\vartheta',i}(t))^2 N_j(dt) \\ & \xrightarrow{\text{f. s.}} \mathbb{E} \sup_{\vartheta' \in B_r^\varepsilon(\vartheta)} \int_I (N_1(t) - \Lambda_{\vartheta',1}(t))^2 N_2(dt) \end{aligned} \quad (7.2.6b)$$

$$\begin{aligned} & \frac{1}{n \cdot (n-1)} \sum_{\substack{i,j=1 \\ i \neq j}}^n \sup_{\vartheta' \in B_r^\varepsilon(\vartheta)} \int_I (N_i(t) - \Lambda_{\vartheta',i}(t)) \cdot (N_j(t) - \Lambda_{\vartheta',j}(t)) N_i(dt) \\ & \xrightarrow{\text{f. s.}} \mathbb{E} \sup_{\vartheta' \in B_r^\varepsilon(\vartheta)} \int_I (N_1(t) - \Lambda_{\vartheta',1}(t)) \cdot (N_2(t) - \Lambda_{\vartheta',2}(t)) N_1(dt) \end{aligned} \quad (7.2.6c)$$

$$\begin{aligned} & \frac{1}{n \cdot (n-1) \cdot (n-2)} \sum_{\substack{i,j,k=1 \\ i \neq j \neq k \neq i}}^n \sup_{\vartheta' \in B_r^\varepsilon(\vartheta)} \int_I (N_i(t) - \Lambda_{\vartheta',i}(t)) \cdot (N_j(t) - \Lambda_{\vartheta',j}(t)) N_k(dt) \\ & \xrightarrow{\text{f. s.}} \mathbb{E} \sup_{\vartheta' \in B_r^\varepsilon(\vartheta)} \int_I (N_1(t) - \Lambda_{\vartheta',1}(t)) \cdot (N_2(t) - \Lambda_{\vartheta',2}(t)) N_3(dt) \end{aligned} \quad (7.2.6d)$$

Die Aussagen (7.2.6a) bis (7.2.6d) gelten dabei auch punktweise im Parameter $\vartheta \in \Theta$.

Beweis:

(7.2.6a) folgt wegen (7.2.5) mit dem klassischen Gesetz der großen Zahlen. Die punktweise Fassung

$$\frac{1}{n} \sum_{i=1}^n \int_I (N_i(t) - \Lambda_{\vartheta,i}(t))^2 N_i(dt) \xrightarrow{\text{f. s.}} \mathbb{E} \int_I (N(t) - \Lambda_{\vartheta}(t))^2 N(dt)$$

folgt ebenso; dabei findet die punktweise Fassung (7.2.4) Verwendung.

Zum Nachweis von (7.2.6b) bis (7.2.6d) verwenden wir Satz A.4.1. Für (7.2.6b) setzen wir in den Bezeichnungen dieses Satzes:

$$\begin{aligned} m &:= 2 \\ X_i(t) &:= (N_i(t) - \Lambda_{\vartheta',i}(t))^2 \\ Y_j(t) &:= N_j(t) \\ h_2((X_i, Y_i), (X_j, Y_j)) &:= \sup_{\vartheta' \in B_r^{\varepsilon}(\vartheta)} \int_I X_i(t) Y_j(dt) \end{aligned}$$

Die Voraussetzung

$$\mathbb{E} h_2((X_1, Y_1), (X_2, Y_2)) = \mathbb{E} \sup_{\vartheta' \in B_r^{\varepsilon}(\vartheta)} \int_I (N_1(t) - \Lambda_{\vartheta',1}(t))^2 N_2(dt) < \infty$$

ist wieder wegen (7.2.5) erfüllt. Satz A.4.1 liefert (7.2.6b). Auch die punktweise Fassung von (7.2.6b) erhalten wir auf diese Weise. Wir verwenden dazu die punktweise Fassung (7.2.4) und setzen:

$$\begin{aligned} m &:= 2 \\ X_i(t) &:= (N_i(t) - \Lambda_{\vartheta,i}(t))^2 \\ Y_j(t) &:= N_j(t) \\ h_2((X_i, Y_i), (X_j, Y_j)) &:= \int_I X_i(t) Y_j(dt) \end{aligned}$$

Zum Nachweis von (7.2.6c) wenden wir Satz A.4.1 mit

$$\begin{aligned} m &:= 2 \\ X_i(t) &:= N_i(t) - \Lambda_{\vartheta',i}(t) \\ Y_j(t) &:= N_j(t) \\ h_2((X_i, Y_i), (X_j, Y_j)) &:= \sup_{\vartheta' \in B_r^{\varepsilon}(\vartheta)} \int_I X_i(t) \cdot X_j(t) Y_i(dt) \end{aligned}$$

an. Die Voraussetzung

$$\begin{aligned} &\mathbb{E} h_2((X_1, Y_1), (X_2, Y_2)) \\ &= \mathbb{E} \sup_{\vartheta' \in B_r^{\varepsilon}(\vartheta)} \int_I (N_1(t) - \Lambda_{\vartheta',1}(t)) \cdot (N_2(t) - \Lambda_{\vartheta',2}(t)) N_1(dt) < \infty \end{aligned}$$

ist wegen (7.2.5) sichergestellt.

Die Aussage (7.2.6d) folgt analog. ■

Nun verwenden wir Lemma 7.2.1, um den Limes aus der punktweisen Fassung von (7.2.6d) zu vereinfachen:

7.2.3 Lemma *In der Situation von Lemma 7.2.1 gilt:*

$$\mathbb{E} \int_I (N_1(t) - \Lambda_{\vartheta,1}(t)) \cdot (N_2(t) - \Lambda_{\vartheta,2}(t)) N_3(dt) = \|\mathbb{E} \Lambda_{\circ}(\cdot) - \mathbb{E} \Lambda_{\vartheta}(\cdot)\|_{\mathbb{E} \Lambda_{\circ}}^2$$

Beweis:

Wegen (7.2.4) ist zunächst Teil 1 des Lemmas A.2.3 zweimal anwendbar, so daß

$$\begin{aligned} & \mathbb{E} \int_I (N_1(t) - \Lambda_{\vartheta,1}(t)) \cdot (N_2(t) - \Lambda_{\vartheta,2}(t)) N_3(dt) \\ &= \mathbb{E} \int_I \mathbb{E}(N_1(t) - \Lambda_{\vartheta,1}(t)) \cdot \mathbb{E}(N_2(t) - \Lambda_{\vartheta,2}(t)) N_3(dt) \end{aligned}$$

folgt. Im Anschluß kann Teil 2 dieses Lemmas angewendet werden, woraus für den letzten Ausdruck folgt:

$$\begin{aligned} [\dots] &= \int_I \mathbb{E}(N_1(t) - \Lambda_{\vartheta,1}(t)) \cdot \mathbb{E}(N_2(t) - \Lambda_{\vartheta,2}(t)) \mathbb{E} N_3(dt) \\ &= \int_I \mathbb{E}^2(N(t) - \Lambda_{\vartheta}(t)) \mathbb{E} \Lambda_{\circ}(dt) \\ &= \|\mathbb{E} \Lambda_{\circ}(\cdot) - \mathbb{E} \Lambda_{\vartheta}(\cdot)\|_{\mathbb{E} \Lambda_{\circ}}^2 \end{aligned}$$

■

Nun werden wir zeigen, daß wir das mittlere Oszillationsverhalten der stochastischen Funktion

$$\vartheta \mapsto \int_I (N_1(t) - \Lambda_{\vartheta,1}(t)) \cdot (N_2(t) - \Lambda_{\vartheta,2}(t)) N_3(dt) \quad (7.2.7)$$

lokal kontrollieren können.

7.2.4 Lemma *Für beliebige positive Zahlen $\varepsilon > 0$, $\delta > 0$ und jeden Parameterwert $\vartheta \in \bar{\Theta}$ existiert ein positiver Radius $r(\vartheta) \equiv r > 0$, so daß*

$$\begin{aligned} & \left| \mathbb{E} \left(\sup_{\vartheta' \in B_r^{\varepsilon}(\vartheta)} \int_I (N_1(t) - \Lambda_{\vartheta',1}(t)) \cdot (N_2(t) - \Lambda_{\vartheta',2}(t)) N_3(dt) \right) \right. \\ & \quad \left. - \inf_{\vartheta' \in B_r^{\varepsilon}(\vartheta)} \|\mathbb{E} \Lambda_{\circ}(\cdot) - \mathbb{E} \Lambda_{\vartheta'}(\cdot)\|_{\mathbb{E} \Lambda_{\circ}}^2 \right| \leq \delta \end{aligned}$$

und

$$\begin{aligned} & \left| \mathbb{E} \left(\inf_{\vartheta' \in B_r^{\varepsilon}(\vartheta)} \int_I (N_1(t) - \Lambda_{\vartheta',1}(t)) \cdot (N_2(t) - \Lambda_{\vartheta',2}(t)) N_3(dt) \right) \right. \\ & \quad \left. - \|\mathbb{E} \Lambda_{\circ}(\cdot) - \mathbb{E} \Lambda_{\vartheta}(\cdot)\|_{\mathbb{E} \Lambda_{\circ}}^2 \right| \leq \delta \end{aligned}$$

gelten.

Beweis: Wir bemerken zunächst, daß die Funktion in (7.2.7) gemäß Voraussetzung (3.2.5) auf Θ f. s. stetig ist. Demnach gilt für jeden Parameterwert $\vartheta \in \Theta$:

$$\left| \sup_{\vartheta' \in B_r^\varepsilon(\vartheta)} \int_I (N_1(t) - \Lambda_{\vartheta',1}(t)) \cdot (N_2(t) - \Lambda_{\vartheta',2}(t)) N_3(dt) - \int_I (N_1(t) - \Lambda_{\vartheta,1}(t)) \cdot (N_2(t) - \Lambda_{\vartheta,2}(t)) N_3(dt) \right| \xrightarrow{\text{f. s.}} 0 \text{ für } r \searrow 0 \quad (7.2.8)$$

Der zweite Term im Betrag von (7.2.8) ist gemäß (7.2.4) integrabel. Auch der erste Term hat für beliebiges $r \leq r_o$ eine integrable Schranke. Dies folgt aus (7.2.5) unter der Annahme (3.2.6). Den Radius r_o haben wir in der genannten Voraussetzung eingeführt. Deshalb können wir aus der f. s. Konvergenz in (7.2.8) mit Lebesgues Lemma von der dominierten Konvergenz auf die Konvergenz der Erwartungswerte schließen:

$$\lim_{r \searrow 0} \mathbb{E} \left| \sup_{\vartheta' \in B_r^\varepsilon(\vartheta)} \int_I (N_1(t) - \Lambda_{\vartheta',1}(t)) \cdot (N_2(t) - \Lambda_{\vartheta',2}(t)) N_3(dt) - \int_I (N_1(t) - \Lambda_{\vartheta,1}(t)) \cdot (N_2(t) - \Lambda_{\vartheta,2}(t)) N_3(dt) \right| = 0$$

Daraus folgt

$$\lim_{r \searrow 0} \left| \mathbb{E} \left(\sup_{\vartheta' \in B_r^\varepsilon(\vartheta)} \int_I (N_1(t) - \Lambda_{\vartheta',1}(t)) \cdot (N_2(t) - \Lambda_{\vartheta',2}(t)) N_3(dt) \right) - \mathbb{E} \left(\int_I (N_1(t) - \Lambda_{\vartheta,1}(t)) \cdot (N_2(t) - \Lambda_{\vartheta,2}(t)) N_3(dt) \right) \right| = 0$$

Auf den zweiten Ausdruck im Betrag wenden wir Lemma 7.2.3 an. Die erste Aussage des Lemmas 7.2.4 folgt, wenn wir noch

$$\lim_{r \searrow 0} \left| \inf_{\vartheta' \in B_r^\varepsilon(\vartheta)} \left\| \mathbb{E} \Lambda_o(\cdot) - \mathbb{E} \Lambda_{\vartheta'}(\cdot) \right\|_{\mathbb{E} \Lambda_o}^2 - \left\| \mathbb{E} \Lambda_o(\cdot) - \mathbb{E} \Lambda_{\vartheta}(\cdot) \right\|_{\mathbb{E} \Lambda_o}^2 \right| = 0$$

zeigen. Dies gilt, da die deterministische Funktion

$$\vartheta \mapsto \left\| \mathbb{E} \Lambda_o(\cdot) - \mathbb{E} \Lambda_{\vartheta}(\cdot) \right\|_{\mathbb{E} \Lambda_o}^2$$

stetig ist. Die zweite Aussage des Lemmas kann ganz analog gezeigt werden. ■

7.3 Beweisschritte

Wir beweisen zunächst zwei f. s. Konvergenzaussagen, auf denen der Beweis der starken Konsistenz beruht:

7.3.1 Lemma *Für jeden fest gewählten Parameterwert $\vartheta \in \Theta$ gilt:*

$$\left\| \bar{N}_n(\cdot) - \bar{\Lambda}_{\vartheta,n}(\cdot) \right\|_{\bar{N}_n} \xrightarrow{\text{f. s.}} \left\| \mathbb{E} \Lambda_o(\cdot) - \mathbb{E} \Lambda_{\vartheta}(\cdot) \right\|_{\mathbb{E} \Lambda_o}$$

Beweis:

Es gilt:

$$\begin{aligned} & \|\bar{N}_n(\cdot) - \bar{\Lambda}_{\vartheta,n}(\cdot)\|_{\bar{N}_n}^2 \\ &= \int_I (\bar{N}_n(t) - \bar{\Lambda}_{\vartheta,n}(t))^2 \bar{N}_n(dt) \\ &= n^{-3} \sum_{i,j,k=1}^n \int_I (N_i(t) - \Lambda_{\vartheta,i}(t)) \cdot (N_j(t) - \Lambda_{\vartheta,j}(t)) N_k(dt) \\ &= n^{-3} \sum_{i=1}^n \int_I (N_i(t) - \Lambda_{\vartheta,i}(t))^2 N_i(dt) \end{aligned} \quad (7.3.1a)$$

$$+ n^{-3} \sum_{\substack{i,k=1 \\ i \neq k}}^n \int_I (N_i(t) - \Lambda_{\vartheta,i}(t))^2 N_k(dt) \quad (7.3.1b)$$

$$+ 2n^{-3} \sum_{\substack{i,j=1 \\ i \neq j}}^n \int_I (N_i(t) - \Lambda_{\vartheta,i}(t)) \cdot (N_j(t) - \Lambda_{\vartheta,j}(t)) N_i(dt) \quad (7.3.1c)$$

$$+ n^{-3} \sum_{\substack{i,j,k=1 \\ i \neq j \neq k \neq i}}^n \int_I (N_i(t) - \Lambda_{\vartheta,i}(t)) \cdot (N_j(t) - \Lambda_{\vartheta,j}(t)) N_k(dt) \quad (7.3.1d)$$

Auf (7.3.1a) können wir die punktweise Fassung von (7.2.6a) anwenden. Da in (7.3.1a) noch ein zusätzlicher Faktor n^{-2} auftritt, konvergiert die dortige Summe f. s. gegen Null.

Ganz entsprechend folgt die f. s. asymptotische Vernachlässigbarkeit von (7.3.1b) und (7.3.1c) aus der f. s. Konvergenz (7.2.6b) bzw. (7.2.6c) (punktweise im Parameter $\vartheta \in \Theta$) gegen endliche Werte, zusammen mit dem Verhalten der zusätzlichen Faktoren $(n-1)/n^2$ bzw. $2(n-1)/n^2$.

Auf (7.3.1d) wenden wir (7.2.6d) an; der Faktor $n(n-1)(n-2)/n^3$ stabilisiert sich in diesem Fall.

Insgesamt haben wir also bisher

$$\|\bar{N}_n(\cdot) - \bar{\Lambda}_{\vartheta,n}(\cdot)\|_{\bar{N}_n}^2 \xrightarrow{\text{f. s.}} \mathbb{E} \int_I (N_1(t) - \Lambda_{\vartheta,1}(t)) \cdot (N_2(t) - \Lambda_{\vartheta,2}(t)) N_3(dt)$$

erhalten. Auf die rechte Seite wenden wir Lemma 7.2.3 an und schließen so den Beweis ab. \blacksquare

Das folgende Lemma kann als gleichmäßige Version des vorherigen Lemmas 7.3.1 aufgefaßt werden.

7.3.2 Lemma Für jeden positiven Radius $\varepsilon > 0$ gilt:

$$\inf_{\vartheta: |\vartheta - \vartheta_0| \geq \varepsilon} \|\bar{N}_n(\cdot) - \bar{\Lambda}_{\vartheta,n}(\cdot)\|_{\bar{N}_n} \xrightarrow{\text{f. s.}} \inf_{\vartheta: |\vartheta - \vartheta_0| \geq \varepsilon} \|\mathbb{E} \Lambda_\circ(\cdot) - \mathbb{E} \Lambda_\vartheta(\cdot)\|_{\mathbb{E} \Lambda_\circ}$$

Beweis:

Wir werden zeigen:

$\forall \varepsilon > 0, \forall \delta > 0 :$

$$\limsup_{n \rightarrow \infty} \left| \inf_{\vartheta: |\vartheta - \vartheta_0| \geq \varepsilon} \|\bar{N}_n(\cdot) - \bar{\Lambda}_{\vartheta, n}(\cdot)\|_{\bar{N}_n}^2 - \inf_{\vartheta: |\vartheta - \vartheta_0| \geq \varepsilon} \|\mathbb{E} \Lambda_0(\cdot) - \mathbb{E} \Lambda_\vartheta(\cdot)\|_{\mathbb{E} \Lambda_0}^2 \right| \leq \delta \quad \text{f. s.} \quad (7.3.2)$$

Daraus folgt die Aussage des Lemmas dann ohne weiteres.

Wir fixieren in (7.3.2) sowohl $\varepsilon > 0$ als auch $\delta > 0$. Dann wählen wir $r(\vartheta) > 0$ gemäß der ersten Aussage des Lemmas 7.2.4 und erhalten die offene Überdeckung

$$\bigcup_{\vartheta \in \bar{\Theta} \setminus B_\varepsilon(\vartheta_0)} B_{r(\vartheta)}^\varepsilon(\vartheta) = \bar{\Theta} \setminus B_\varepsilon(\vartheta_0)$$

des Kompaktums $\bar{\Theta} \setminus B_\varepsilon(\vartheta_0)$. Darin findet sich die endliche Überdeckung

$$\bigcup_{i=1}^q B_{r_i}^\varepsilon(\vartheta_i) = \bar{\Theta} \setminus B_\varepsilon(\vartheta_0)$$

mit geeignetem $q \in \mathbb{N}$, $\vartheta_1, \dots, \vartheta_q \in \bar{\Theta} \setminus B_\varepsilon(\vartheta_0)$ und $r_i := r(\vartheta_i)$, $i = 1, \dots, q$.

Wir erhalten somit:

$$\begin{aligned} & \left| \inf_{\vartheta: |\vartheta - \vartheta_0| \geq \varepsilon} \|\bar{N}_n(\cdot) - \bar{\Lambda}_{\vartheta, n}(\cdot)\|_{\bar{N}_n}^2 - \inf_{\vartheta: |\vartheta - \vartheta_0| \geq \varepsilon} \|\mathbb{E} \Lambda_0(\cdot) - \mathbb{E} \Lambda_\vartheta(\cdot)\|_{\mathbb{E} \Lambda_0}^2 \right| \\ &= \left| \min_{1 \leq i \leq q} \inf_{\vartheta \in B_{r_i}^\varepsilon(\vartheta_i)} \|\bar{N}_n(\cdot) - \bar{\Lambda}_{\vartheta, n}(\cdot)\|_{\bar{N}_n}^2 - \min_{1 \leq i \leq q} \inf_{\vartheta \in B_{r_i}^\varepsilon(\vartheta_i)} \|\mathbb{E} \Lambda_0(\cdot) - \mathbb{E} \Lambda_\vartheta(\cdot)\|_{\mathbb{E} \Lambda_0}^2 \right| \\ &\leq \max_{1 \leq i \leq q} \left| \inf_{\vartheta \in B_{r_i}^\varepsilon(\vartheta_i)} \|\bar{N}_n(\cdot) - \bar{\Lambda}_{\vartheta, n}(\cdot)\|_{\bar{N}_n}^2 - \inf_{\vartheta \in B_{r_i}^\varepsilon(\vartheta_i)} \|\mathbb{E} \Lambda_0(\cdot) - \mathbb{E} \Lambda_\vartheta(\cdot)\|_{\mathbb{E} \Lambda_0}^2 \right| \end{aligned}$$

Zum Nachweis von (7.3.2) genügt es folglich,

$$\limsup_{n \rightarrow \infty} \left| \inf_{\vartheta \in B_{r_i}^\varepsilon(\vartheta_i)} \|\bar{N}_n(\cdot) - \bar{\Lambda}_{\vartheta, n}(\cdot)\|_{\bar{N}_n}^2 - \inf_{\vartheta \in B_{r_i}^\varepsilon(\vartheta_i)} \|\mathbb{E} \Lambda_0(\cdot) - \mathbb{E} \Lambda_\vartheta(\cdot)\|_{\mathbb{E} \Lambda_0}^2 \right| \leq \delta \quad \text{f. s.} \quad 1 \leq i \leq q$$

zu zeigen. Wir lösen den Betrag auf und weisen nacheinander

$$\limsup_{n \rightarrow \infty} \left(\inf_{\vartheta \in B_{r_i}^\varepsilon(\vartheta_i)} \|\bar{N}_n(\cdot) - \bar{\Lambda}_{\vartheta, n}(\cdot)\|_{\bar{N}_n}^2 - \inf_{\vartheta \in B_{r_i}^\varepsilon(\vartheta_i)} \|\mathbb{E} \Lambda_0(\cdot) - \mathbb{E} \Lambda_\vartheta(\cdot)\|_{\mathbb{E} \Lambda_0}^2 \right) \leq \delta \quad \text{f. s.} \quad 1 \leq i \leq q \quad (7.3.3)$$

und

$$\liminf_{n \rightarrow \infty} \left(\inf_{\vartheta \in B_{r_i}^\varepsilon(\vartheta_i)} \|\bar{N}_n(\cdot) - \bar{\Lambda}_{\vartheta, n}(\cdot)\|_{\bar{N}_n}^2 - \inf_{\vartheta \in B_{r_i}^\varepsilon(\vartheta_i)} \|\mathbb{E} \Lambda_0(\cdot) - \mathbb{E} \Lambda_\vartheta(\cdot)\|_{\mathbb{E} \Lambda_0}^2 \right) \geq -\delta \quad \text{f. s.} \quad 1 \leq i \leq q \quad (7.3.4)$$

nach.

Zunächst untersuchen wir (7.3.3). Es gilt

$$\begin{aligned}
& \inf_{\vartheta \in B_{r_i}^\varepsilon(\vartheta_i)} \|\bar{N}_n(\cdot) - \bar{\Lambda}_{\vartheta,n}(\cdot)\|_{\bar{N}_n}^2 \\
&= \inf_{\vartheta \in B_{r_i}^\varepsilon(\vartheta_i)} \int_I (\bar{N}_n(t) - \bar{\Lambda}_{\vartheta,n}(t))^2 \bar{N}_n(dt) \\
&= \inf_{\vartheta \in B_{r_i}^\varepsilon(\vartheta_i)} n^{-3} \sum_{j,k,l=1}^n \int_I (N_j(t) - \Lambda_{\vartheta,j}(t)) \cdot (N_k(t) - \Lambda_{\vartheta,k}(t)) N_l(dt) \\
&\leq n^{-3} \sum_{j,k,l=1}^n \sup_{\vartheta \in B_{r_i}^\varepsilon(\vartheta_i)} \int_I (N_j(t) - \Lambda_{\vartheta,j}(t)) \cdot (N_k(t) - \Lambda_{\vartheta,k}(t)) N_l(dt).
\end{aligned}$$

Damit erhalten wir für die linke Seite von (7.3.3) die folgende Abschätzung:

$$\begin{aligned}
& \limsup_{n \rightarrow \infty} \left(\inf_{\vartheta \in B_{r_i}^\varepsilon(\vartheta_i)} \|\bar{N}_n(\cdot) - \bar{\Lambda}_{\vartheta,n}(\cdot)\|_{\bar{N}_n}^2 - \inf_{\vartheta \in B_{r_i}^\varepsilon(\vartheta_i)} \|\mathbb{E} \Lambda_\circ(\cdot) - \mathbb{E} \Lambda_\vartheta(\cdot)\|_{\mathbb{E} \Lambda_\circ}^2 \right) \\
&\leq \limsup_{n \rightarrow \infty} \\
&\quad \left(n^{-3} \sum_{j,k,l=1}^n \sup_{\vartheta \in B_{r_i}^\varepsilon(\vartheta_i)} \int_I (N_j(t) - \Lambda_{\vartheta,j}(t)) \cdot (N_k(t) - \Lambda_{\vartheta,k}(t)) N_l(dt) \right. \\
&\quad \left. - \inf_{\vartheta \in B_{r_i}^\varepsilon(\vartheta_i)} \|\mathbb{E} \Lambda_\circ(\cdot) - \mathbb{E} \Lambda_\vartheta(\cdot)\|_{\mathbb{E} \Lambda_\circ}^2 \right)
\end{aligned}$$

Die rechte Seite zerlegen wir geeignet:

$$\begin{aligned}
& n^{-3} \sum_{j,k,l=1}^n \sup_{\vartheta \in B_{r_i}^\varepsilon(\vartheta_i)} \int_I (N_j(t) - \Lambda_{\vartheta,j}(t)) \cdot (N_k(t) - \Lambda_{\vartheta,k}(t)) N_l(dt) \\
&\quad - \inf_{\vartheta \in B_{r_i}^\varepsilon(\vartheta_i)} \|\mathbb{E} \Lambda_\circ(\cdot) - \mathbb{E} \Lambda_\vartheta(\cdot)\|_{\mathbb{E} \Lambda_\circ}^2 \\
&= n^{-3} \sum_{j=1}^n \sup_{\vartheta \in B_{r_i}^\varepsilon(\vartheta_i)} \int_I (N_j(t) - \Lambda_{\vartheta,j}(t))^2 N_j(dt) \\
&+ n^{-3} \sum_{\substack{j,k=1 \\ j \neq k}}^n \sup_{\vartheta \in B_{r_i}^\varepsilon(\vartheta_i)} \int_I (N_j(t) - \Lambda_{\vartheta,j}(t))^2 N_k(dt) \\
&+ 2n^{-3} \sum_{\substack{j,k=1 \\ j \neq k}}^n \sup_{\vartheta \in B_{r_i}^\varepsilon(\vartheta_i)} \int_I (N_j(t) - \Lambda_{\vartheta,j}(t)) \cdot (N_k(t) - \Lambda_{\vartheta,k}(t)) N_j(dt) \\
&+ n^{-3} \sum_{\substack{j,k,l=1 \\ j \neq k \neq l \neq j}}^n \sup_{\vartheta \in B_{r_i}^\varepsilon(\vartheta_i)} \int_I (N_j(t) - \Lambda_{\vartheta,j}(t)) \cdot (N_k(t) - \Lambda_{\vartheta,k}(t)) N_l(dt) \\
&\quad - \inf_{\vartheta \in B_{r_i}^\varepsilon(\vartheta_i)} \|\mathbb{E} \Lambda_\circ(\cdot) - \mathbb{E} \Lambda_\vartheta(\cdot)\|_{\mathbb{E} \Lambda_\circ}^2
\end{aligned}$$

Damit folgt (7.3.3) aus

$$\left| n^{-3} \sum_{j=1}^n \sup_{\vartheta \in B_{r_i}^\varepsilon(\vartheta_i)} \int_I (N_j(t) - \Lambda_{\vartheta,j}(t))^2 N_j(dt) \right| \xrightarrow{\text{f. s.}} 0, \quad (7.3.5a)$$

$$\left| n^{-3} \sum_{\substack{j,k=1 \\ j \neq k}}^n \sup_{\vartheta \in B_{r_i}^\varepsilon(\vartheta_i)} \int_I (N_j(t) - \Lambda_{\vartheta,j}(t))^2 N_k(dt) \right| \xrightarrow{\text{f. s.}} 0 \quad (7.3.5b)$$

sowie

$$\left| 2n^{-3} \sum_{\substack{j,k=1 \\ j \neq k}}^n \sup_{\vartheta \in B_{r_i}^\varepsilon(\vartheta_i)} \int_I (N_j(t) - \Lambda_{\vartheta,j}(t)) \cdot (N_k(t) - \Lambda_{\vartheta,k}(t)) N_j(dt) \right| \xrightarrow{\text{f. s.}} 0 \quad (7.3.5c)$$

und

$$\begin{aligned} & n^{-3} \sum_{\substack{j,k,l=1 \\ j \neq k \neq l \neq j}}^n \sup_{\vartheta \in B_{r_i}^\varepsilon(\vartheta_i)} \int_I (N_j(t) - \Lambda_{\vartheta,j}(t)) \cdot (N_k(t) - \Lambda_{\vartheta,k}(t)) N_l(dt) \\ & \xrightarrow{\text{f. s.}} \mathbb{E} \left(\sup_{\vartheta \in B_{r_i}^\varepsilon(\vartheta_i)} \int_I (N_1(t) - \Lambda_{\vartheta,1}(t)) \cdot (N_2(t) - \Lambda_{\vartheta,2}(t)) N_3(dt) \right). \end{aligned} \quad (7.3.5d)$$

Der Limes in (7.3.5d) hat gemäß der ersten Aussage des Lemmas 7.2.4 einen Abstand von höchstens δ zu $\inf_{\vartheta \in B_{r_i}^\varepsilon(\vartheta_i)} \|\mathbb{E} \Lambda_\circ(\cdot) - \mathbb{E} \Lambda_\vartheta(\cdot)\|_{\mathbb{E} \Lambda_\circ}^2$.

Wir müssen nun noch (7.3.5a) bis (7.3.5d) nachweisen. Dies gelingt völlig analog zum Nachweis des Konvergenzverhaltens von (7.3.1a) bis (7.3.1d) im Beweis des Lemmas 7.3.1, da die f. s. Konvergenzaussagen (7.2.6) auch in einer lokal gleichmäßigen Fassung zur Verfügung stehen.

Wir wenden uns dem Nachweis von (7.3.4) zu. Es gelten

$$\begin{aligned} & \inf_{\vartheta \in B_{r_i}^\varepsilon(\vartheta_i)} \|\bar{N}_n(\cdot) - \bar{\Lambda}_{\vartheta,n}(\cdot)\|_{\bar{N}_n}^2 \\ & = \inf_{\vartheta \in B_{r_i}^\varepsilon(\vartheta_i)} \int_I (\bar{N}_n(t) - \bar{\Lambda}_{\vartheta,n}(t))^2 \bar{N}_n(dt) \\ & = \inf_{\vartheta \in B_{r_i}^\varepsilon(\vartheta_i)} n^{-3} \sum_{j,k,l=1}^n \int_I (N_j(t) - \Lambda_{\vartheta,j}(t)) \cdot (N_k(t) - \Lambda_{\vartheta,k}(t)) N_l(dt) \\ & \geq n^{-3} \sum_{j,k,l=1}^n \inf_{\vartheta \in B_{r_i}^\varepsilon(\vartheta_i)} \int_I (N_j(t) - \Lambda_{\vartheta,j}(t)) \cdot (N_k(t) - \Lambda_{\vartheta,k}(t)) N_l(dt) \end{aligned}$$

und

$$\begin{aligned} 0 & \leq \inf_{\vartheta \in B_{r_i}^\varepsilon(\vartheta_i)} \|\mathbb{E} \Lambda_\circ(\cdot) - \mathbb{E} \Lambda_\vartheta(\cdot)\|_{\mathbb{E} \Lambda_\circ}^2 \\ & \leq \|\mathbb{E} \Lambda_\circ(\cdot) - \mathbb{E} \Lambda_{\vartheta_i}(\cdot)\|_{\mathbb{E} \Lambda_\circ}^2. \end{aligned}$$

Wir können die linke Seite von (7.3.4) also in folgender Weise abschätzen:

$$\begin{aligned} & \liminf_{n \rightarrow \infty} \left(\inf_{\vartheta \in B_{r_i}^\varepsilon(\vartheta_i)} \|\bar{N}_n(\cdot) - \bar{\Lambda}_{\vartheta,n}(\cdot)\|_{\bar{N}_n}^2 - \inf_{\vartheta \in B_{r_i}^\varepsilon(\vartheta_i)} \|\mathbb{E} \Lambda_\circ(\cdot) - \mathbb{E} \Lambda_\vartheta(\cdot)\|_{\mathbb{E} \Lambda_\circ}^2 \right) \\ & \geq \liminf_{n \rightarrow \infty} \\ & \quad \left(n^{-3} \sum_{j,k,l=1}^n \inf_{\vartheta \in B_{r_i}^\varepsilon(\vartheta_i)} \int_I (N_j(t) - \Lambda_{\vartheta,j}(t)) \cdot (N_k(t) - \Lambda_{\vartheta,k}(t)) N_l(dt) \right. \\ & \quad \left. - \|\mathbb{E} \Lambda_\circ(\cdot) - \mathbb{E} \Lambda_{\vartheta_i}(\cdot)\|_{\mathbb{E} \Lambda_\circ}^2 \right) \end{aligned}$$

Eine Zerlegung der rechten Seite liefert, völlig analog zum Nachweis von (7.3.3):

$$\begin{aligned} & n^{-3} \sum_{j,k,l=1}^n \inf_{\vartheta \in B_{r_i}^\varepsilon(\vartheta_i)} \int_I (N_j(t) - \Lambda_{\vartheta,j}(t)) \cdot (N_k(t) - \Lambda_{\vartheta,k}(t)) N_l(dt) \\ & \quad - \|\mathbb{E} \Lambda_\circ(\cdot) - \mathbb{E} \Lambda_{\vartheta_i}(\cdot)\|_{\mathbb{E} \Lambda_\circ}^2 \\ & = n^{-3} \sum_{j=1}^n \inf_{\vartheta \in B_{r_i}^\varepsilon(\vartheta_i)} \int_I (N_j(t) - \Lambda_{\vartheta,j}(t))^2 N_j(dt) \\ & \quad + n^{-3} \sum_{\substack{j,k=1 \\ j \neq k}}^n \inf_{\vartheta \in B_{r_i}^\varepsilon(\vartheta_i)} \int_I (N_j(t) - \Lambda_{\vartheta,j}(t))^2 N_k(dt) \\ & \quad + 2n^{-3} \sum_{\substack{j,k=1 \\ j \neq k}}^n \inf_{\vartheta \in B_{r_i}^\varepsilon(\vartheta_i)} \int_I (N_j(t) - \Lambda_{\vartheta,j}(t)) \cdot (N_k(t) - \Lambda_{\vartheta,k}(t)) N_j(dt) \\ & \quad + n^{-3} \sum_{\substack{j,k,l=1 \\ j \neq k \neq l \neq j}}^n \inf_{\vartheta \in B_{r_i}^\varepsilon(\vartheta_i)} \int_I (N_j(t) - \Lambda_{\vartheta,j}(t)) \cdot (N_k(t) - \Lambda_{\vartheta,k}(t)) N_l(dt) \\ & \quad - \|\mathbb{E} \Lambda_\circ(\cdot) - \mathbb{E} \Lambda_{\vartheta_i}(\cdot)\|_{\mathbb{E} \Lambda_\circ}^2 \end{aligned}$$

Also folgt (7.3.4) aus

$$\left| n^{-3} \sum_{j=1}^n \inf_{\vartheta \in B_{r_i}^\varepsilon(\vartheta_i)} \int_I (N_j(t) - \Lambda_{\vartheta,j}(t))^2 N_j(dt) \right| \xrightarrow{\text{f. s.}} 0, \quad (7.3.6a)$$

$$\left| n^{-3} \sum_{\substack{j,k=1 \\ j \neq k}}^n \inf_{\vartheta \in B_{r_i}^\varepsilon(\vartheta_i)} \int_I (N_j(t) - \Lambda_{\vartheta,j}(t))^2 N_k(dt) \right| \xrightarrow{\text{f. s.}} 0 \quad (7.3.6b)$$

sowie

$$\begin{aligned} & \left| 2n^{-3} \sum_{\substack{j,k=1 \\ j \neq k}}^n \inf_{\vartheta \in B_{r_i}^\varepsilon(\vartheta_i)} \int_I (N_j(t) - \Lambda_{\vartheta,j}(t)) \cdot (N_k(t) - \Lambda_{\vartheta,k}(t)) N_j(dt) \right| \\ & \quad \xrightarrow{\text{f. s.}} 0 \quad (7.3.6c) \end{aligned}$$

und

$$n^{-3} \sum_{\substack{j,k,l=1 \\ j \neq k \neq l \neq j}}^n \inf_{\vartheta \in B_{r_i}^\varepsilon(\vartheta_i)} \int_I (N_j(t) - \Lambda_{\vartheta,j}(t)) \cdot (N_k(t) - \Lambda_{\vartheta,k}(t)) N_l(dt) \\ \xrightarrow{\text{f. s.}} \mathbb{E} \left(\inf_{\vartheta \in B_{r_i}^\varepsilon(\vartheta_i)} \int_I (N_1(t) - \Lambda_{\vartheta,1}(t)) \cdot (N_2(t) - \Lambda_{\vartheta,2}(t)) N_3(dt) \right). \quad (7.3.6d)$$

Die vier Aussagen (7.3.6) folgen wie (7.3.5). (7.3.4) ergibt sich schließlich, da die rechte Seite von (7.3.6d) und $\| \mathbb{E} \Lambda_\circ(\cdot) - \mathbb{E} \Lambda_{\vartheta,i}(\cdot) \|_{\mathbb{E} \Lambda_\circ}^2$ weniger als δ voneinander entfernt sind. Dies ist die zweite Aussage des Lemmas 7.2.4. \blacksquare

Nach diesen Vorbereitungen wenden wir uns nun dem Beweis des Theorems 3.3.1 zu.

Beweis von Theorem 3.3.1:

Es ist zu zeigen, daß für beliebiges $\varepsilon > 0$ gilt:

$$\mathbb{P}(\limsup_{n \rightarrow \infty} \{ |\vartheta_n^* - \vartheta_\circ| \geq \varepsilon \}) = 0$$

Wir haben:

$$\begin{aligned} & \{ |\vartheta_n^* - \vartheta_\circ| \geq \varepsilon \} \\ & \subset \left\{ \inf_{\vartheta: |\vartheta - \vartheta_\circ| \geq \varepsilon} \|\bar{N}_n(\cdot) - \bar{\Lambda}_{\vartheta,n}(\cdot)\|_{\bar{N}_n} < \inf_{\vartheta: |\vartheta - \vartheta_\circ| < \varepsilon} \|\bar{N}_n(\cdot) - \bar{\Lambda}_{\vartheta,n}(\cdot)\|_{\bar{N}_n} \right\} \\ & \subset \left\{ \inf_{\vartheta: |\vartheta - \vartheta_\circ| \geq \varepsilon} \|\bar{N}_n(\cdot) - \bar{\Lambda}_{\vartheta,n}(\cdot)\|_{\bar{N}_n} < \|\bar{N}_n(\cdot) - \bar{\Lambda}_{\vartheta_\circ,n}(\cdot)\|_{\bar{N}_n} \right\} \end{aligned}$$

Es genügt zu zeigen, daß für beliebiges $\varepsilon > 0$ gilt:

$$\mathbb{P} \left(\limsup_{n \rightarrow \infty} \left\{ \inf_{\vartheta: |\vartheta - \vartheta_\circ| \geq \varepsilon} \|\bar{N}_n(\cdot) - \bar{\Lambda}_{\vartheta,n}(\cdot)\|_{\bar{N}_n} < \|\bar{N}_n(\cdot) - \bar{\Lambda}_{\vartheta_\circ,n}(\cdot)\|_{\bar{N}_n} \right\} \right) = 0 \quad (7.3.7)$$

Wir werden nachweisen, daß sich (7.3.7) aus den beiden folgenden Konvergenzaussagen ergibt, welche wir mit den Lemmata 7.3.1 und 7.3.2 gezeigt haben:

$$\|\bar{N}_n(\cdot) - \bar{\Lambda}_{\vartheta_\circ,n}(\cdot)\|_{\bar{N}_n} \xrightarrow{\text{f. s.}} 0 \quad (7.3.8a)$$

$$\inf_{\vartheta: |\vartheta - \vartheta_\circ| \geq \varepsilon} \|\bar{N}_n(\cdot) - \bar{\Lambda}_{\vartheta,n}(\cdot)\|_{\bar{N}_n} \xrightarrow{\text{f. s.}} \inf_{\vartheta: |\vartheta - \vartheta_\circ| \geq \varepsilon} \|\mathbb{E} \Lambda_\circ(\cdot) - \mathbb{E} \Lambda_\vartheta(\cdot)\|_{\mathbb{E} \Lambda_\circ} \quad (7.3.8b)$$

Dazu wählen wir

$$\delta := \frac{1}{2} \inf_{\vartheta: |\vartheta - \vartheta_\circ| \geq \varepsilon} \|\mathbb{E} \Lambda_\circ(\cdot) - \mathbb{E} \Lambda_\vartheta(\cdot)\|_{\mathbb{E} \Lambda_\circ} > 0.$$

Gemäß (3.2.7) handelt es sich bei δ um eine strikt positive Zahl.

Wir zerlegen die in (7.3.7) angegebene Menge mit Hilfe eines weiteren Ereignisses in zwei disjunkte Teile:

$$\begin{aligned} & \left\{ \inf_{\vartheta: |\vartheta - \vartheta_o| \geq \varepsilon} \|\bar{N}_n(\cdot) - \bar{\Lambda}_{\vartheta, n}(\cdot)\|_{\bar{N}_n} < \|\bar{N}_n(\cdot) - \bar{\Lambda}_{\vartheta_o, n}(\cdot)\|_{\bar{N}_n} \right\} \\ &= \left\{ \inf_{\vartheta: |\vartheta - \vartheta_o| \geq \varepsilon} \|\bar{N}_n(\cdot) - \bar{\Lambda}_{\vartheta, n}(\cdot)\|_{\bar{N}_n} < \|\bar{N}_n(\cdot) - \bar{\Lambda}_{\vartheta_o, n}(\cdot)\|_{\bar{N}_n}, \right. \\ & \quad \left. \|\bar{N}_n(\cdot) - \bar{\Lambda}_{\vartheta_o, n}(\cdot)\|_{\bar{N}_n} \geq \delta \right\} \\ \cup & \left\{ \inf_{\vartheta: |\vartheta - \vartheta_o| \geq \varepsilon} \|\bar{N}_n(\cdot) - \bar{\Lambda}_{\vartheta, n}(\cdot)\|_{\bar{N}_n} < \|\bar{N}_n(\cdot) - \bar{\Lambda}_{\vartheta_o, n}(\cdot)\|_{\bar{N}_n}, \right. \\ & \quad \left. \|\bar{N}_n(\cdot) - \bar{\Lambda}_{\vartheta_o, n}(\cdot)\|_{\bar{N}_n} < \delta \right\} \end{aligned}$$

Diese beiden Mengen ersetzen wir zunächst durch jeweils größere Mengen:

$$\begin{aligned} [\dots] & \subset \left\{ \|\bar{N}_n(\cdot) - \bar{\Lambda}_{\vartheta_o, n}(\cdot)\|_{\bar{N}_n} \geq \delta \right\} \\ & \quad \cup \left\{ \inf_{\vartheta: |\vartheta - \vartheta_o| \geq \varepsilon} \|\bar{N}_n(\cdot) - \bar{\Lambda}_{\vartheta, n}(\cdot)\|_{\bar{N}_n} < \delta \right\} \\ &= \left\{ \|\bar{N}_n(\cdot) - \bar{\Lambda}_{\vartheta_o, n}(\cdot)\|_{\bar{N}_n} \geq \delta \right\} \\ \cup & \left\{ \inf_{\vartheta: |\vartheta - \vartheta_o| \geq \varepsilon} \|\bar{N}_n(\cdot) - \bar{\Lambda}_{\vartheta, n}(\cdot)\|_{\bar{N}_n} - \inf_{\vartheta: |\vartheta - \vartheta_o| \geq \varepsilon} \|\mathbf{E} \Lambda_o(\cdot) - \mathbf{E} \Lambda_\vartheta(\cdot)\|_{\mathbf{E} \Lambda_o} \right. \\ & \quad \left. < \delta - \inf_{\vartheta: |\vartheta - \vartheta_o| \geq \varepsilon} \|\mathbf{E} \Lambda_o(\cdot) - \mathbf{E} \Lambda_\vartheta(\cdot)\|_{\mathbf{E} \Lambda_o} \right\} \end{aligned}$$

Wegen der speziellen Wahl von δ gilt weiter:

$$\begin{aligned} [\dots] &= \left\{ \|\bar{N}_n(\cdot) - \bar{\Lambda}_{\vartheta_o, n}(\cdot)\|_{\bar{N}_n} \geq \delta \right\} \\ \cup & \left\{ \inf_{\vartheta: |\vartheta - \vartheta_o| \geq \varepsilon} \|\bar{N}_n(\cdot) - \bar{\Lambda}_{\vartheta, n}(\cdot)\|_{\bar{N}_n} \right. \\ & \quad \left. - \inf_{\vartheta: |\vartheta - \vartheta_o| \geq \varepsilon} \|\mathbf{E} \Lambda_o(\cdot) - \mathbf{E} \Lambda_\vartheta(\cdot)\|_{\mathbf{E} \Lambda_o} < -\delta \right\} \\ \subset & \left\{ \|\bar{N}_n(\cdot) - \bar{\Lambda}_{\vartheta_o, n}(\cdot)\|_{\bar{N}_n} \geq \delta \right\} \\ \cup & \left\{ \left| \inf_{\vartheta: |\vartheta - \vartheta_o| \geq \varepsilon} \|\bar{N}_n(\cdot) - \bar{\Lambda}_{\vartheta, n}(\cdot)\|_{\bar{N}_n} \right. \right. \\ & \quad \left. \left. - \inf_{\vartheta: |\vartheta - \vartheta_o| \geq \varepsilon} \|\mathbf{E} \Lambda_o(\cdot) - \mathbf{E} \Lambda_\vartheta(\cdot)\|_{\mathbf{E} \Lambda_o} \right| > \delta \right\} \end{aligned}$$

Aus der soeben hergeleiteten Inklusion folgt:

$$\begin{aligned}
& \limsup_{n \rightarrow \infty} \left\{ \inf_{\vartheta: |\vartheta - \vartheta_o| \geq \varepsilon} \|\bar{N}_n(\cdot) - \bar{\Lambda}_{\vartheta, n}(\cdot)\|_{\bar{N}_n} \leq \|\bar{N}_n(\cdot) - \bar{\Lambda}_{\vartheta_o, n}(\cdot)\|_{\bar{N}_n} \right\} \\
& \subset \limsup_{n \rightarrow \infty} \left(\left\{ \|\bar{N}_n(\cdot) - \bar{\Lambda}_{\vartheta_o, n}(\cdot)\|_{\bar{N}_n} \geq \delta \right\} \right. \\
& \cup \left. \left\{ \left| \inf_{\vartheta: |\vartheta - \vartheta_o| \geq \varepsilon} \|\bar{N}_n(\cdot) - \bar{\Lambda}_{\vartheta, n}(\cdot)\|_{\bar{N}_n} \right. \right. \\
& \qquad \qquad \qquad \left. \left. - \inf_{\vartheta: |\vartheta - \vartheta_o| \geq \varepsilon} \|\mathbb{E} \Lambda_o(\cdot) - \mathbb{E} \Lambda_\vartheta(\cdot)\|_{\mathbb{E} \Lambda_o} \right| > \delta \right\} \\
& \subset \limsup_{n \rightarrow \infty} \left\{ \|\bar{N}_n(\cdot) - \bar{\Lambda}_{\vartheta_o, n}(\cdot)\|_{\bar{N}_n} \geq \delta \right\} \\
& \cup \limsup_{n \rightarrow \infty} \left\{ \left| \inf_{\vartheta: |\vartheta - \vartheta_o| \geq \varepsilon} \|\bar{N}_n(\cdot) - \bar{\Lambda}_{\vartheta, n}(\cdot)\|_{\bar{N}_n} \right. \right. \\
& \qquad \qquad \qquad \left. \left. - \inf_{\vartheta: |\vartheta - \vartheta_o| \geq \varepsilon} \|\mathbb{E} \Lambda_o(\cdot) - \mathbb{E} \Lambda_\vartheta(\cdot)\|_{\mathbb{E} \Lambda_o} \right| > \delta \right\}
\end{aligned}$$

Für die zugehörigen Wahrscheinlichkeiten gilt folglich:

$$\begin{aligned}
& \mathbb{P} \left(\limsup_{n \rightarrow \infty} \left\{ \inf_{\vartheta: |\vartheta - \vartheta_o| \geq \varepsilon} \|\bar{N}_n(\cdot) - \bar{\Lambda}_{\vartheta, n}(\cdot)\|_{\bar{N}_n} \leq \|\bar{N}_n(\cdot) - \bar{\Lambda}_{\vartheta_o, n}(\cdot)\|_{\bar{N}_n} \right\} \right) \\
& \leq \mathbb{P} \left(\limsup_{n \rightarrow \infty} \left\{ \|\bar{N}_n(\cdot) - \bar{\Lambda}_{\vartheta_o, n}(\cdot)\|_{\bar{N}_n} \geq \delta \right\} \right) \\
& + \mathbb{P} \left(\limsup_{n \rightarrow \infty} \left\{ \left| \inf_{\vartheta: |\vartheta - \vartheta_o| \geq \varepsilon} \|\bar{N}_n(\cdot) - \bar{\Lambda}_{\vartheta, n}(\cdot)\|_{\bar{N}_n} \right. \right. \right. \\
& \qquad \qquad \qquad \left. \left. - \inf_{\vartheta: |\vartheta - \vartheta_o| \geq \varepsilon} \|\mathbb{E} \Lambda_o(\cdot) - \mathbb{E} \Lambda_\vartheta(\cdot)\|_{\mathbb{E} \Lambda_o} \right| > \delta \right\} \right)
\end{aligned}$$

Die Gültigkeit von (7.3.8) vorausgesetzt, verschwinden daher beide Wahrscheinlichkeiten auf der rechten Seite. Insgesamt erhalten wir also:

$$\mathbb{P} \left(\limsup_{n \rightarrow \infty} \left\{ \inf_{\vartheta: |\vartheta - \vartheta_o| \geq \varepsilon} \|\bar{N}_n(\cdot) - \bar{\Lambda}_{\vartheta, n}(\cdot)\|_{\bar{N}_n} < \|\bar{N}_n(\cdot) - \bar{\Lambda}_{\vartheta_o, n}(\cdot)\|_{\bar{N}_n} \right\} \right) = 0$$

Damit folgt (7.3.7) aus (7.3.8) und wir haben Theorem 3.3.1 gezeigt. \blacksquare

8 Beweis der asymptotischen Normalität des KQS

8.1 Einführung

In Abschnitt 8.2 führen wir die Beweise einiger Hilfssätze. Dabei handelt es sich um C -Straffheitsaussagen der standardisierenden Matrix $\Phi_n(\vartheta)$ sowie von drei auf Θ definierten Hilfsprozessen.

Nach diesen Vorbereitungen folgt in Abschnitt 8.3 der eigentliche Beweis von Theorem 3.5.1. Er ist in mehrere Schritte untergliedert.

8.2 Technische Eigenschaften

8.2.1 Hilfsaussagen

Wir untersuchen zunächst einige relevante Ausdrücke im Hinblick auf ihre zweiten Momente. Dabei bezeichne $|\cdot|$ für Vektoren und Matrizen die Euklidische Norm.

8.2.1 Lemma *Wir betrachten die sechs Prozesse:*

$$\begin{aligned} X_1, X_2, Z_1, Z_2 &\in \{N_k, \Lambda_{\vartheta, k}, \mathbb{E} \Lambda_{\vartheta}; k = 1, \dots, n; \vartheta \in \Theta\} \\ Y_1, Y_2 &\in \left\{ \frac{\partial}{\partial \vartheta} \Lambda_{\vartheta, k}, \mathbb{E} \frac{\partial}{\partial \vartheta} \Lambda_{\vartheta}, L'_k, \mathbb{E} |L'|, \frac{\partial^2}{\partial \vartheta^2} \Lambda_{\vartheta, k}, \mathbb{E} \frac{\partial^2}{\partial \vartheta^2} \Lambda_{\vartheta}; \right. \\ &\quad \left. k = 1, \dots, n; \vartheta \in \Theta \right\} \end{aligned}$$

Unter den Voraussetzungen (3.4.1) bis (3.4.10) existieren die folgenden Erwartungswerte und sind endlich:

$$\begin{aligned} &\mathbb{E} \left(\int_I |X_1(t) \pm X_2(t)| |Y_1(t) \pm Y_2(t)| |Z_1(dt) \pm Z_2(dt)| \right)^2 < \infty \\ &\mathbb{E} \left(\int_I \int_{[t, t_{\max}]} |Y_1(x) \pm Y_2(x)| |X_1(dx) \pm X_2(dx)| |Z_1(dt) \pm Z_2(dt)| \right)^2 < \infty \end{aligned}$$

Beweis:

Nach Definition sind die Prozesse X_1, X_2, Z_1 und Z_2 positiv und monoton wachsend. Darüber hinaus haben sie gemäß Voraussetzung (3.4.6) bzw. (3.4.7) endliche sechste Momente. Die Prozesse Y_1, Y_2 sind u. U. Vektoren oder Matrizen. Zwischen den sechs Prozessen können stochastische Abhängigkeiten bestehen.

Wir zeigen die erste Aussage des Lemmas:

$$\begin{aligned} &\mathbb{E} \left(\int_I |X_1(t) \pm X_2(t)| |Y_1(t) \pm Y_2(t)| |Z_1(dt) \pm Z_2(dt)| \right)^2 \\ &\leq \mathbb{E} \left(\sum_{i,j,k=1}^2 \int_I X_i(t) |Y_j(t)| Z_k(dt) \right)^2 \\ &\leq 8 \sum_{i,j,k=1}^2 \mathbb{E} \left(\int_I X_i(t) |Y_j(t)| Z_k(dt) \right)^2 \\ &\leq 8 \sum_{i,j,k=1}^2 \mathbb{E} \left(\int_I X_i^2(t) Z_k(dt) \cdot \int_I |Y_j(t)|^2 Z_k(dt) \right) \\ &\leq 8 \sum_{i,j,k=1}^2 \mathbb{E} \left(X_i^2(t_{\max}) \cdot Z_k(t_{\max}) \cdot \int_I |Y_j(t)|^2 Z_k(dt) \right) \\ &\leq 8 \sum_{i,j,k=1}^2 \mathbb{E}^{1/3} (X_i^6(t_{\max})) \cdot \mathbb{E}^{1/6} (Z_k^6(t_{\max})) \cdot \mathbb{E}^{1/2} \left(\int_I |Y_j(t)|^4 Z_k(dt) \right) \\ &< \infty \end{aligned}$$

Die Endlichkeit des letzten Ausdrucks folgt je nach Wahl von Y_j und Z_k aus einer der Voraussetzungen (3.4.8) bis (3.4.10) und (3.4.3) bis (3.4.5). Wir wenden uns der zweiten Aussage des Lemmas zu:

$$\begin{aligned}
& \mathbb{E} \left(\int_I \int_{[t, t_{\max}]} |Y_1(x) \pm Y_2(x)| |X_1(dx) \pm X_2(dx)| |Z_1(dt) \pm Z_2(dt)| \right)^2 \\
& \leq \mathbb{E} \left(\sum_{i,j,k=1}^2 \int_I \int_{[t, t_{\max}]} |Y_j(x)| X_i(dx) Z_k(dt) \right)^2 \\
& \leq 8 \sum_{i,j,k=1}^2 \mathbb{E} \left(\int_I \int_{[t, t_{\max}]} |Y_j(x)|^2 X_i(dx) Z_k(dt) \right) \\
& \leq 8 \sum_{i,j,k=1}^2 \mathbb{E} \left(Z_k(t_{\max}) \cdot \int_I |Y_j(x)|^2 X_i(dx) \right) \\
& \leq 8 \sum_{i,j,k=1}^2 \mathbb{E}^{1/2} (Z_k^2(t_{\max})) \cdot \mathbb{E}^{1/2} \left(\int_I |Y_j(x)|^4 X_i(dx) \right) \\
& < \infty
\end{aligned}$$

■

Um die Aussage des nächsten Lemmas präzise formulieren zu können, beachten wir, daß die Größen $\frac{\partial}{\partial \vartheta} \Lambda_{\vartheta, i}(t) \in \mathbb{R}^d$ Vektoren sind. Wir fassen sie im folgenden als Spaltenvektoren auf. Zu einem Spaltenvektor $V \in \mathbb{R}^d$ sei V^\top der durch Transposition entstehende Zeilenvektor. In dieser Schreibweise bezeichnet $U^\top V \in \mathbb{R}$ für Vektoren $U, V \in \mathbb{R}^d$ also das Skalarprodukt.

8.2.2 Lemma Seien $i, j, k, l, p, q = 1, \dots, n$. Es gelte eine der folgenden Bedingungen:

$$i \neq p, k, q, l, j \quad (8.2.1a)$$

$$\text{oder } j \neq p, k, i, q, l \quad (8.2.1b)$$

$$\text{oder } k \neq p, i, q, l, j \quad (8.2.1c)$$

$$\text{oder } l \neq p, k, i, q, j \quad (8.2.1d)$$

Wenigstens einer der vier Indices i, j, k, l soll also einen Wert annehmen, der sich von denen der fünf anderen Indices unterscheidet.

Wir betrachten die Prozesse:

$$W_q \in \{ \Lambda_{\vartheta, q} - \Lambda_{\vartheta', q}, \mathbb{E} \Lambda_{\vartheta}, \vartheta, \vartheta' \in \Theta \}$$

$$X_i \in \{ \frac{\partial}{\partial \vartheta} \Lambda_{\vartheta, i} - \mathbb{E} \frac{\partial}{\partial \vartheta} \Lambda_{\vartheta, i}, (\frac{\partial}{\partial \vartheta} \Lambda_{\vartheta, i} - \mathbb{E} \frac{\partial}{\partial \vartheta} \Lambda_{\vartheta, i}) - (\frac{\partial}{\partial \vartheta} \Lambda_{\vartheta', i} - \mathbb{E} \frac{\partial}{\partial \vartheta} \Lambda_{\vartheta', i}), \vartheta, \vartheta' \in \Theta \}$$

$$Y_p \in \{ \frac{\partial}{\partial \vartheta} \Lambda_{\vartheta, p}, \frac{\partial}{\partial \vartheta} \Lambda_{\vartheta, p} - \frac{\partial}{\partial \vartheta} \Lambda_{\vartheta', p}, \mathbb{E} \frac{\partial}{\partial \vartheta} \Lambda_{\vartheta}, \frac{\partial^2}{\partial \vartheta^2} \Lambda_{\vartheta, p}, \mathbb{E} \frac{\partial^2}{\partial \vartheta^2} \Lambda_{\vartheta}, \vartheta, \vartheta' \in \Theta \}$$

$$Z_k \in \{ M_k, \Lambda_{\vartheta, k} - \mathbb{E} \Lambda_{\vartheta}, \vartheta \in \Theta \}$$

Dann verschwinden die folgenden, je nach Wahl von Y skalaren oder matrix-

wertigen Erwartungswerte:

$$\mathbb{E} \left(\iint_{I \times I} Z_i(s) Z_j(t) Y_p^\top(s) Y_q(t) Z_k(ds) Z_l(dt) \right) = 0 \quad (8.2.2)$$

$$\mathbb{E} \left(\iint_{I \times I} \left(\iint_{\substack{[x, t_{\max}] \times \\ [y, t_{\max}]}} Y_p^\top(s) Y_q(t) Z_k(ds) Z_l(dt) \right) M_i(dx) M_j(dy) \right) = 0 \quad (8.2.3)$$

$$\mathbb{E} \left(\iint_{I \times I} Z_i(s) Z_j(t) X_k^\top(s) X_l(t) W_p(ds) W_q(dt) \right) = 0 \quad (8.2.4)$$

$$\mathbb{E} \left(\iint_{I \times I} \left(\iint_{\substack{[x, t_{\max}] \times \\ [y, t_{\max}]}} X_k^\top(s) X_l(t) \mathbb{E} \Lambda_\circ(ds) \mathbb{E} \Lambda_\circ(dt) \right) M_i(dx) M_j(dy) \right) = 0 \quad (8.2.5)$$

$$\mathbb{E} \left(\iint_{I \times I} W_p(s) W_q(t) X_k^\top(s) X_l(t) Z_i(ds) Z_j(dt) \right) = 0 \quad (8.2.6)$$

Beweis:

Wir zeigen (8.2.2) unter der Voraussetzung (8.2.1a) für skalare und vektorielle Y . Falls Y eine Matrix ist, kann elementweise argumentiert werden. Dann treten einzelne Spalten und Zeilen von Y im Integranden auf, dieser Fall wird von der Behandlung vektorieller Y abgedeckt. Es gilt:

$$\begin{aligned} & \mathbb{E} \left(\iint_{I \times I} Z_i(s) Z_j(t) Y_p(s)^\top Y_q(t) Z_k(ds) Z_l(dt) \right) \\ &= \sum_{h=1}^d \mathbb{E} \left(\int_I Z_i(s) \left(\int_I Z_j(t) (Y_p(s))_h (Y_q(t))_h Z_l(dt) \right) Z_k(ds) \right) \end{aligned}$$

Wir zeigen, daß jeder einzelne Summand verschwindet. Dazu fixieren wir h und setzen:

$$\begin{aligned} X(s) &:= Z_i(s) \\ Y(s) &:= \int_I Z_j(t) (Y_p(s))_h (Y_q(t))_h Z_l(dt) \\ Z(s) &:= Z_k(s) \end{aligned}$$

Nach Voraussetzung an den Index i ist $X(s)$ von $Y(s)$ und $Z(s)$ unabhängig. Daraus folgt insbesondere (A.2.1). Auch die weiteren Bedingungen des Teils 1 von Lemma A.2.3 sind erfüllt. Wir erhalten also

$$\begin{aligned} & \mathbb{E} \left(\iint_{I \times I} Z_i(s) Z_j(t) Y_p(s)^\top Y_q(t) Z_k(ds) Z_l(dt) \right) \\ &= \mathbb{E} \left(\iint_{I \times I} \mathbb{E}(Z_i(s)) Z_j(t) Y_p(s)^\top Y_q(t) Z_k(ds) Z_l(dt) \right) \end{aligned}$$

und damit (8.2.2), denn $Z_i(s)$ ist zentriert.

Alle weiteren Aussagen des Lemmas folgen analog; sofern anstelle eines Faktors im Integranden ein integrierendes Maß stochastisch unabhängig vom restlichen Ausdruck ist, findet statt Teil 1 der Teil 2 von Lemma A.2.3 Verwendung. Die Unabhängigkeit des integrierenden Maßes vom *gesamten* Integranden stellt dabei die zusätzliche Voraussetzung (A.2.3) sicher. ■

8.2.3 Lemma *Unter den Voraussetzungen (3.4.11) und (3.4.12) ist $\Phi_\circ(\vartheta)$ in einer Umgebung $U(\vartheta_\circ) \subset \Theta$ des wahren Parameters ϑ_\circ positiv definit.*

Beweis:

Die Funktion

$$\vartheta \mapsto \Phi_\circ(\vartheta)$$

ist stetig. Folglich genügt es, den Nachweis für ϑ_\circ zu führen. $\Phi_\circ(\vartheta_\circ)$ hat die Darstellung:

$$\begin{aligned} \Phi_\circ(\vartheta_\circ) &= \frac{\partial}{\partial \vartheta} \int_I (\mathbb{E} \Lambda_\vartheta(t) - \mathbb{E} \Lambda_\circ(t)) \mathbb{E} \frac{\partial}{\partial \vartheta} \Lambda_\vartheta(t) \mathbb{E} \Lambda_\circ(dt) \Big|_{\vartheta=\vartheta_\circ} \\ &= \int_I \mathbb{E} \frac{\partial}{\partial \vartheta} \Lambda_\vartheta(t) \mathbb{E} \frac{\partial}{\partial \vartheta} \Lambda_\vartheta(t)^\top \mathbb{E} \Lambda_\circ(dt) \Big|_{\vartheta=\vartheta_\circ} \\ &\quad + \int_I (\mathbb{E} \Lambda_\vartheta(t) - \mathbb{E} \Lambda_\circ(t)) \mathbb{E} \frac{\partial^2}{\partial \vartheta^2} \Lambda_\vartheta(t) \mathbb{E} \Lambda_\circ(dt) \Big|_{\vartheta=\vartheta_\circ} \\ &= \int_I \mathbb{E} \frac{\partial}{\partial \vartheta} \Lambda_\vartheta(t) \mathbb{E} \frac{\partial}{\partial \vartheta} \Lambda_\vartheta(t)^\top \mathbb{E} \Lambda_\circ(dt) \Big|_{\vartheta=\vartheta_\circ} \end{aligned}$$

Wir fixieren $x \in \mathbb{R}^d$ und betrachten die Lebesgue-Approximation zu $x^\top \Phi_\circ(\vartheta_\circ)x$. Sie hat eine Darstellung der Form:

$$\begin{aligned} x^\top \Phi_\circ(\vartheta_\circ)x &= \lim_{n \rightarrow \infty} \sum_{i=1}^n x^\top \mathbb{E} \frac{\partial}{\partial \vartheta} \Lambda_\vartheta(t_i^n) \mathbb{E} \frac{\partial}{\partial \vartheta} \Lambda_\vartheta(t_i^n)^\top x \mathbb{E} \Lambda_\circ(B_i^n) \\ &= \lim_{n \rightarrow \infty} \sum_{i=1}^n |x^\top \mathbb{E} \frac{\partial}{\partial \vartheta} \Lambda_\vartheta(t_i^n)|^2 \mathbb{E} \Lambda_\circ(B_i^n) \end{aligned} \quad (8.2.7)$$

Dabei sind $t_i^n \in I$ und $B_i^n \subset I$, $i = 1, \dots, n$, geeignet gewählt. Die Mengen B_i^n , $i = 1, \dots, n$, sind meßbar und bilden eine Zerlegung des Intervalls I . Nach Voraussetzung (3.4.12) gilt

$$|x^\top \mathbb{E} \frac{\partial}{\partial \vartheta} \Lambda_\vartheta(t_x)|^2 > 0.$$

Da $t \mapsto \mathbb{E} \frac{\partial}{\partial \vartheta} \Lambda_\vartheta(t)$ stetig ist, existieren $\delta_x, \varepsilon_x > 0$ mit

$$|x^\top \mathbb{E} \frac{\partial}{\partial \vartheta} \Lambda_\vartheta(t)|^2 \geq \varepsilon_x \quad \forall t \in (t_x - \delta_x, t_x + \delta_x).$$

Die Summanden in (8.2.7) sind positiv, zusammen mit Voraussetzung (3.4.11) folgt

$$x^\top \Phi_\circ(\vartheta_\circ)x \geq \varepsilon_x \mathbb{E} \Lambda_\circ((t_x - \delta_x, t_x + \delta_x)) > 0. \quad \blacksquare$$

8.2.2 Straffheit des α -Prozesses

In diesem Abschnitt untersuchen wir den Prozeß

$$\alpha_n(\vartheta) := -n^{1/2} \int_I (\bar{\Lambda}_{\vartheta,n}(t) - \bar{\Lambda}_{\vartheta_\circ,n}(t)) \frac{\partial}{\partial \vartheta} \bar{\Lambda}_{\vartheta,n}(t) \bar{M}_n(dt) \quad (8.2.8)$$

Wir zeigen, daß $\alpha_n(\vartheta)$ auf kompakten Parameterteilmengen $\Theta_\circ \subset \Theta$ C -straff ist. Darüber hinaus erhalten wir eine approximative Integraldarstellung mit deterministischem Integranden:

8.2.4 Satz *Der Prozeß $\alpha_n(\vartheta)$ erlaubt die Darstellung*

$$\alpha_n(\vartheta) = -n^{1/2} \int_I (\mathbb{E} \Lambda_\vartheta(t) - \mathbb{E} \Lambda_{\vartheta_\circ}(t)) \mathbb{E} \frac{\partial}{\partial \vartheta} \Lambda_\vartheta(t) \bar{M}_n(dt) + o_P(1).$$

Auf kompakten Parameterteilmengen $\Theta_\circ \subset \Theta$ ist das führende Integral C -straff. Der Restterm ist dort ebenfalls C -straff und konvergiert gleichmäßig in $\vartheta \in \Theta_\circ$ gegen Null.

Die Straffheit von α_n folgt dann daraus, daß die Summe zweier C -straffer Folgen über einem Raum stetiger Funktionen selbst C -straff ist. Dies ergibt sich direkt aus der Definition des Stetigkeitsmoduls, siehe Lemma A.5.5.

Beweis:

Wir werden den Integranden in

$$\alpha_n(\vartheta) = -n^{1/2} \int_I (\bar{\Lambda}_{\vartheta,n}(t) - \bar{\Lambda}_{\vartheta_\circ,n}(t)) \frac{\partial}{\partial \vartheta} \bar{\Lambda}_{\vartheta,n}(t) \bar{M}_n(dt)$$

linear entwickeln. Um die Darstellung übersichtlich zu halten, unterdrücken wir in der folgenden Gleichungskette die Argumente der auftretenden Funktionen:

$$\begin{aligned} & (\bar{\Lambda}_{\vartheta,n} - \bar{\Lambda}_{\vartheta_\circ,n}) \cdot \frac{\partial}{\partial \vartheta} \bar{\Lambda}_{\vartheta,n} \\ &= (\bar{\Lambda}_{\vartheta,n} - \bar{\Lambda}_{\vartheta_\circ,n}) \cdot \mathbb{E} \frac{\partial}{\partial \vartheta} \Lambda_\vartheta &+ (\bar{\Lambda}_{\vartheta,n} - \bar{\Lambda}_{\vartheta_\circ,n}) \cdot (\frac{\partial}{\partial \vartheta} \bar{\Lambda}_{\vartheta,n} - \mathbb{E} \frac{\partial}{\partial \vartheta} \Lambda_\vartheta) \\ &= (\mathbb{E} \Lambda_\vartheta - \bar{\Lambda}_{\vartheta_\circ,n}) \cdot \mathbb{E} \frac{\partial}{\partial \vartheta} \Lambda_\vartheta &+ (\bar{\Lambda}_{\vartheta,n} - \mathbb{E} \Lambda_\vartheta) \cdot \mathbb{E} \frac{\partial}{\partial \vartheta} \Lambda_\vartheta \\ & &+ (\bar{\Lambda}_{\vartheta,n} - \bar{\Lambda}_{\vartheta_\circ,n}) \cdot (\frac{\partial}{\partial \vartheta} \bar{\Lambda}_{\vartheta,n} - \mathbb{E} \frac{\partial}{\partial \vartheta} \Lambda_\vartheta) \\ &= (\mathbb{E} \Lambda_\vartheta - \mathbb{E} \Lambda_{\vartheta_\circ}) \cdot \mathbb{E} \frac{\partial}{\partial \vartheta} \Lambda_\vartheta &- (\bar{\Lambda}_{\vartheta_\circ,n} - \mathbb{E} \Lambda_{\vartheta_\circ}) \cdot \mathbb{E} \frac{\partial}{\partial \vartheta} \Lambda_\vartheta \\ & &+ (\bar{\Lambda}_{\vartheta,n} - \mathbb{E} \Lambda_\vartheta) \cdot \mathbb{E} \frac{\partial}{\partial \vartheta} \Lambda_\vartheta \\ & &+ (\bar{\Lambda}_{\vartheta,n} - \bar{\Lambda}_{\vartheta_\circ,n}) \cdot (\frac{\partial}{\partial \vartheta} \bar{\Lambda}_{\vartheta,n} - \mathbb{E} \frac{\partial}{\partial \vartheta} \Lambda_\vartheta) \\ &= (\mathbb{E} \Lambda_\vartheta - \mathbb{E} \Lambda_{\vartheta_\circ}) \cdot \mathbb{E} \frac{\partial}{\partial \vartheta} \Lambda_\vartheta &- r_n^{(1)} + r_n^{(2)} + r_n^{(3)} \end{aligned}$$

Dabei haben wir

$$r_n^{(1)}(t, \vartheta) := (\bar{\Lambda}_{\vartheta_\circ,n}(t) - \mathbb{E} \Lambda_{\vartheta_\circ}(t)) \cdot \mathbb{E} \frac{\partial}{\partial \vartheta} \Lambda_\vartheta(t) \quad (8.2.9a)$$

$$r_n^{(2)}(t, \vartheta) := (\bar{\Lambda}_{\vartheta,n}(t) - \mathbb{E} \Lambda_\vartheta(t)) \cdot \mathbb{E} \frac{\partial}{\partial \vartheta} \Lambda_\vartheta(t) \quad (8.2.9b)$$

$$r_n^{(3)}(t, \vartheta) := (\bar{\Lambda}_{\vartheta,n}(t) - \bar{\Lambda}_{\vartheta_\circ,n}(t)) \cdot (\frac{\partial}{\partial \vartheta} \bar{\Lambda}_{\vartheta,n}(t) - \mathbb{E} \frac{\partial}{\partial \vartheta} \Lambda_\vartheta(t)) \quad (8.2.9c)$$

gesetzt. Wir erhalten also die Darstellung

$$\alpha_n(\vartheta) = -\tilde{\alpha}_n(\vartheta) + \alpha_n^{(1)}(\vartheta) - \alpha_n^{(2)}(\vartheta) - \alpha_n^{(3)}(\vartheta),$$

wobei wir die Abkürzungen

$$\tilde{\alpha}_n(\vartheta) := n^{1/2} \int_I (\mathbb{E} \Lambda_\vartheta(t) - \mathbb{E} \Lambda_{\vartheta_\circ}(t)) \mathbb{E} \frac{\partial}{\partial \vartheta} \Lambda_\vartheta(t) \bar{M}_n(dt) \quad (8.2.10a)$$

und

$$\alpha_n^{(i)}(\vartheta) := n^{1/2} \int_I r_n^{(i)}(t, \vartheta) \bar{M}_n(dt) \quad i = 1, 2, 3, \quad (8.2.10b)$$

verwendet haben.

Satz 8.2.4 ist vollständig bewiesen, wenn wir neben der Straffheit von $\tilde{\alpha}_n(\vartheta)$ und $\alpha_n^{(i)}(\vartheta)$, $i = 1, 2, 3$, auf dem Kompaktum Θ_\circ noch nachweisen, daß die drei Restterme in Θ *punktweise* verschwinden:

$$\alpha_n^{(i)}(\vartheta) = o_{\mathbb{P}}(1) \quad \forall \vartheta \in \Theta, \quad i = 1, 2, 3. \quad (8.2.11)$$

Wir werden (8.2.11) in Lemma 8.2.6 zeigen.

Angesichts von (8.2.11) und der Tatsache, daß

$$\tilde{\alpha}_n(\vartheta_\circ) = n^{1/2} \int_I (\mathbb{E} \Lambda_{\vartheta_\circ}(t) - \mathbb{E} \Lambda_{\vartheta_\circ}(t)) \cdot \mathbb{E} \frac{\partial}{\partial \vartheta} \Lambda_{\vartheta_\circ}(t) \bar{M}_n(dt) \equiv 0$$

gilt, bietet es sich an, die Straffheit von $\tilde{\alpha}_n(\vartheta)$ und $\alpha_n^{(i)}(\vartheta)$, $i = 1, 2, 3$, mit Satz A.5.2 nachzuweisen. Wir haben dann lediglich noch die Momentenbedingung zu prüfen. In der vorliegenden Notation lautet sie

$$\mathbb{E} |\tilde{\alpha}_n(\vartheta) - \tilde{\alpha}_n(\vartheta')|^2 \leq \text{const.} \cdot |\vartheta - \vartheta'|^2, \quad \vartheta, \vartheta' \in \Theta_\circ. \quad (8.2.12)$$

$$\mathbb{E} |\alpha_n^{(i)}(\vartheta) - \alpha_n^{(i)}(\vartheta')|^2 \leq \text{const.} \cdot |\vartheta - \vartheta'|^2, \quad \vartheta, \vartheta' \in \Theta_\circ. \quad (8.2.13)$$

für $i = 1, 2, 3$. Die beiden Konstanten haben wir dabei als $\alpha = \gamma = 2$ gewählt. Wir werden (8.2.12) und (8.2.13) in Lemma 8.2.5 bzw. 8.2.7 nachweisen.

Zuvor strukturieren wir noch die Integranden, die auf der linken Seite von (8.2.12) und (8.2.13) auftauchen. Da in $r_n^{(1)}(t, \vartheta)$ gemäß (8.2.9a) nur ein Faktor von ϑ abhängt, erhalten wir direkt

$$\begin{aligned} r_n^{(1)}(t, \vartheta) - r_n^{(1)}(t, \vartheta') \\ = (\bar{\Lambda}_{\vartheta_\circ, n}(t) - \mathbb{E} \Lambda_{\vartheta_\circ}(t)) \cdot (\mathbb{E} \frac{\partial}{\partial \vartheta} \Lambda_\vartheta(t) - \mathbb{E} \frac{\partial}{\partial \vartheta} \Lambda_{\vartheta'}(t)). \end{aligned} \quad (8.2.14a)$$

Die drei anderen Differenzterme zu $r_n^{(2)}$, $r_n^{(3)}$ und dem führenden Ausdruck haben gemäß (8.2.9) eine gemeinsame Struktur, nämlich:

$$a(\vartheta)b(\vartheta) - a(\vartheta')b(\vartheta')$$

Alle drei Ausdrücke entwickeln wir wie die Produktregel aus der Differentialrechnung:

$$\begin{aligned} & a(\vartheta)b(\vartheta) - a(\vartheta')b(\vartheta') \\ &= a(\vartheta)b(\vartheta) - a(\vartheta')b(\vartheta) + a(\vartheta')b(\vartheta) - a(\vartheta')b(\vartheta') \\ &= (a(\vartheta) - a(\vartheta'))b(\vartheta) + a(\vartheta')(b(\vartheta) - b(\vartheta')) \end{aligned}$$

Wir erhalten also für den zweiten Restterm

$$\begin{aligned} & r_n^{(2)}(t, \vartheta) - r_n^{(2)}(t, \vartheta') \tag{8.2.14b} \\ &= (\bar{\Lambda}_{\vartheta, n}(t) - \mathbb{E} \Lambda_{\vartheta}(t) - \bar{\Lambda}_{\vartheta', n}(t) + \mathbb{E} \Lambda_{\vartheta'}(t)) \cdot \mathbb{E} \frac{\partial}{\partial \vartheta} \Lambda_{\vartheta}(t) \\ &+ (\bar{\Lambda}_{\vartheta', n}(t) - \mathbb{E} \Lambda_{\vartheta'}(t)) \cdot (\mathbb{E} \frac{\partial}{\partial \vartheta} \Lambda_{\vartheta}(t) - \mathbb{E} \frac{\partial}{\partial \vartheta} \Lambda_{\vartheta'}(t)), \end{aligned}$$

für den dritten Restterm

$$\begin{aligned} & r_n^{(3)}(t, \vartheta) - r_n^{(3)}(t, \vartheta') \tag{8.2.14c} \\ &= (\bar{\Lambda}_{\vartheta, n}(t) - \bar{\Lambda}_{\vartheta', n}(t)) \cdot (\frac{\partial}{\partial \vartheta} \bar{\Lambda}_{\vartheta, n}(t) - \mathbb{E} \frac{\partial}{\partial \vartheta} \Lambda_{\vartheta}(t)) \\ &+ (\bar{\Lambda}_{\vartheta', n}(t) - \bar{\Lambda}_{\vartheta_{\circ}, n}(t)) \\ &\cdot (\frac{\partial}{\partial \vartheta} \bar{\Lambda}_{\vartheta, n}(t) - \mathbb{E} \frac{\partial}{\partial \vartheta} \Lambda_{\vartheta}(t) - \frac{\partial}{\partial \vartheta} \bar{\Lambda}_{\vartheta', n}(t) + \mathbb{E} \frac{\partial}{\partial \vartheta} \Lambda_{\vartheta'}(t)), \end{aligned}$$

und für den führenden Term

$$\begin{aligned} & (\mathbb{E} \Lambda_{\vartheta}(t) - \mathbb{E} \Lambda_{\vartheta_{\circ}}(t)) \cdot \mathbb{E} \frac{\partial}{\partial \vartheta} \Lambda_{\vartheta}(t) - (\mathbb{E} \Lambda_{\vartheta'}(t) - \mathbb{E} \Lambda_{\vartheta_{\circ}}(t)) \cdot \mathbb{E} \frac{\partial}{\partial \vartheta} \Lambda_{\vartheta'}(t) \\ &= (\mathbb{E} \Lambda_{\vartheta}(t) - \mathbb{E} \Lambda_{\vartheta'}(t)) \cdot \mathbb{E} \frac{\partial}{\partial \vartheta} \Lambda_{\vartheta}(t) \tag{8.2.14d} \\ &+ (\mathbb{E} \Lambda_{\vartheta'}(t) - \mathbb{E} \Lambda_{\vartheta_{\circ}}(t)) \cdot (\mathbb{E} \frac{\partial}{\partial \vartheta} \Lambda_{\vartheta}(t) - \mathbb{E} \frac{\partial}{\partial \vartheta} \Lambda_{\vartheta'}(t)). \end{aligned}$$

Es bleiben noch (8.2.11), (8.2.12) und (8.2.13) zu zeigen. Dies geschieht in den folgenden Lemmata. \blacksquare

Wir zeigen (8.2.12):

8.2.5 Lemma *Auf kompakten Parameterteilmengen $\Theta_{\circ} \subset \Theta$ gilt*

$$\mathbb{E} |\tilde{\alpha}_n(\vartheta) - \tilde{\alpha}_n(\vartheta')|^2 \leq \text{const.} \cdot |\vartheta - \vartheta'|^2, \quad \forall \vartheta, \vartheta' \in \Theta_{\circ}.$$

Beweis:

Mit (8.2.10a) und (8.2.14d) erhalten wir:

$$\begin{aligned} & \mathbb{E} |\tilde{\alpha}_n(\vartheta) - \tilde{\alpha}_n(\vartheta')|^2 \\ &= \mathbb{E} \left| n^{1/2} \int_I (\mathbb{E} \Lambda_{\vartheta}(t) - \mathbb{E} \Lambda_{\vartheta'}(t)) \cdot \mathbb{E} \frac{\partial}{\partial \vartheta} \Lambda_{\vartheta}(t) \right. \\ &\quad \left. + (\mathbb{E} \Lambda_{\vartheta'}(t) - \mathbb{E} \Lambda_{\vartheta_{\circ}}(t)) \cdot (\mathbb{E} \frac{\partial}{\partial \vartheta} \Lambda_{\vartheta}(t) - \mathbb{E} \frac{\partial}{\partial \vartheta} \Lambda_{\vartheta'}(t)) \bar{M}_n(dt) \right|^2 \end{aligned}$$

$$\begin{aligned}
&= n^{-1} \sum_{i,j=1}^n \mathbb{E} \left(\iint_{I \times I} \left((\mathbb{E} \Lambda_{\vartheta}(s) - \mathbb{E} \Lambda_{\vartheta'}(s)) \cdot \mathbb{E} \frac{\partial}{\partial \vartheta} \Lambda_{\vartheta}(s) \right. \right. \\
&\quad \left. \left. + (\mathbb{E} \Lambda_{\vartheta'}(s) - \mathbb{E} \Lambda_{\vartheta_{\circ}}(s)) \cdot (\mathbb{E} \frac{\partial}{\partial \vartheta} \Lambda_{\vartheta}(s) - \mathbb{E} \frac{\partial}{\partial \vartheta} \Lambda_{\vartheta'}(s)) \right)^{\top} \right. \\
&\quad \left. \left((\mathbb{E} \Lambda_{\vartheta}(t) - \mathbb{E} \Lambda_{\vartheta'}(t)) \cdot \mathbb{E} \frac{\partial}{\partial \vartheta} \Lambda_{\vartheta}(t) \right. \right. \\
&\quad \left. \left. + (\mathbb{E} \Lambda_{\vartheta'}(t) - \mathbb{E} \Lambda_{\vartheta_{\circ}}(t)) \cdot (\mathbb{E} \frac{\partial}{\partial \vartheta} \Lambda_{\vartheta}(t) - \mathbb{E} \frac{\partial}{\partial \vartheta} \Lambda_{\vartheta'}(t)) \right) M_i(ds) M_j(dt) \right)
\end{aligned}$$

Die Summanden für $i \neq j$ verschwinden. Dies liefert Teil 2 von Lemma A.2.3 mit

$$\begin{aligned}
X(t) &:= \int_I \left((\mathbb{E} \Lambda_{\vartheta}(s) - \mathbb{E} \Lambda_{\vartheta'}(s)) \cdot \mathbb{E} \frac{\partial}{\partial \vartheta} \Lambda_{\vartheta}(s) \right. \\
&\quad \left. + (\mathbb{E} \Lambda_{\vartheta'}(s) - \mathbb{E} \Lambda_{\vartheta_{\circ}}(s)) \cdot (\mathbb{E} \frac{\partial}{\partial \vartheta} \Lambda_{\vartheta}(s) - \mathbb{E} \frac{\partial}{\partial \vartheta} \Lambda_{\vartheta'}(s)) \right)^{\top} \\
&\quad \left((\mathbb{E} \Lambda_{\vartheta}(t) - \mathbb{E} \Lambda_{\vartheta'}(t)) \cdot \mathbb{E} \frac{\partial}{\partial \vartheta} \Lambda_{\vartheta}(t) \right. \\
&\quad \left. + (\mathbb{E} \Lambda_{\vartheta'}(t) - \mathbb{E} \Lambda_{\vartheta_{\circ}}(t)) \cdot (\mathbb{E} \frac{\partial}{\partial \vartheta} \Lambda_{\vartheta}(t) - \mathbb{E} \frac{\partial}{\partial \vartheta} \Lambda_{\vartheta'}(t)) \right) M_i(ds) \\
Y(t) &:= 1 \\
Z(t) &:= M_j(t)
\end{aligned}$$

wegen $\mathbb{E} M_j \equiv 0$. Die übrigen Summanden mit $i = j$ sind i. i. d., wir erhalten also:

$$\begin{aligned}
&\mathbb{E} |\tilde{\alpha}_n(\vartheta) - \tilde{\alpha}_n(\vartheta')|^2 \\
&= \mathbb{E} \left| \int_I (\mathbb{E} \Lambda_{\vartheta}(t) - \mathbb{E} \Lambda_{\vartheta'}(t)) \cdot \mathbb{E} \frac{\partial}{\partial \vartheta} \Lambda_{\vartheta}(t) \right. \\
&\quad \left. + (\mathbb{E} \Lambda_{\vartheta'}(t) - \mathbb{E} \Lambda_{\vartheta_{\circ}}(t)) \cdot (\mathbb{E} \frac{\partial}{\partial \vartheta} \Lambda_{\vartheta}(t) - \mathbb{E} \frac{\partial}{\partial \vartheta} \Lambda_{\vartheta'}(t)) M(dt) \right|^2 \\
&\leq \mathbb{E} \left(\int_I |\mathbb{E} \Lambda_{\vartheta}(t) - \mathbb{E} \Lambda_{\vartheta'}(t)| \cdot \mathbb{E} \left| \frac{\partial}{\partial \vartheta} \Lambda_{\vartheta}(t) \right| \right. \\
&\quad \left. + |\mathbb{E} \Lambda_{\vartheta'}(t) - \mathbb{E} \Lambda_{\vartheta_{\circ}}(t)| \cdot \left| \mathbb{E} \frac{\partial}{\partial \vartheta} \Lambda_{\vartheta}(t) - \mathbb{E} \frac{\partial}{\partial \vartheta} \Lambda_{\vartheta'}(t) \right| (N(dt) + \Lambda_{\circ}(dt)) \right)^2
\end{aligned}$$

Mit (3.2.5) folgt, daß die deterministischen Funktionen $\mathbb{E} \left| \frac{\partial}{\partial \vartheta} \Lambda_{\vartheta}(t) \right|$ und $|\mathbb{E} \Lambda_{\vartheta}(t) - \mathbb{E} \Lambda_{\vartheta_{\circ}}(t)|$ stetig von (t, ϑ) abhängen. Auf dem Kompaktum $I \times \Theta_{\circ}$ nehmen sie deshalb ihr Supremum an und sind beschränkt:

$$\begin{aligned}
[\dots] &\leq \text{const.} \cdot \mathbb{E} \left(\int_I |\mathbb{E} \Lambda_{\vartheta}(t) - \mathbb{E} \Lambda_{\vartheta'}(t)| \right. \\
&\quad \left. + \left| \mathbb{E} \frac{\partial}{\partial \vartheta} \Lambda_{\vartheta}(t) - \mathbb{E} \frac{\partial}{\partial \vartheta} \Lambda_{\vartheta'}(t) \right| (N(dt) + \Lambda_{\circ}(dt)) \right)^2
\end{aligned}$$

Aus (3.4.2), (3.4.3) und (3.4.4) folgt, daß das verbleibende Integral gegen $\text{const.} \cdot |\vartheta - \vartheta'|^2$ abgeschätzt werden kann. ■

Nun zeigen wir (8.2.11):

8.2.6 Lemma *Es gilt*

$$\alpha_n^{(i)}(\vartheta) = o_{\mathbb{P}}(1) \quad \forall \vartheta \in \Theta, \quad i = 1, 2, 3.$$

Beweis:

Gemäß Definition (8.2.10b) und (8.2.9) müssen wir

$$n^{1/2} \int_I (\bar{\Lambda}_{\vartheta_\circ, n}(t) - \mathbb{E} \Lambda_{\vartheta_\circ}(t)) \cdot \mathbb{E} \frac{\partial}{\partial \vartheta} \Lambda_{\vartheta}(t) \bar{M}_n(dt) \xrightarrow{\mathbb{P}} 0, \quad (8.2.15a)$$

$$n^{1/2} \int_I (\bar{\Lambda}_{\vartheta, n}(t) - \mathbb{E} \Lambda_{\vartheta}(t)) \cdot \mathbb{E} \frac{\partial}{\partial \vartheta} \Lambda_{\vartheta}(t) \bar{M}_n(dt) \xrightarrow{\mathbb{P}} 0 \quad (8.2.15b)$$

sowie

$$n^{1/2} \int_I (\bar{\Lambda}_{\vartheta, n}(t) - \bar{\Lambda}_{\vartheta_\circ, n}(t)) \cdot \left(\frac{\partial}{\partial \vartheta} \bar{\Lambda}_{\vartheta, n}(t) - \mathbb{E} \frac{\partial}{\partial \vartheta} \Lambda_{\vartheta}(t) \right) \bar{M}_n(dt) \xrightarrow{\mathbb{P}} 0 \quad (8.2.16)$$

zeigen. Dabei folgen (8.2.15a) und (8.2.15b) offenbar, sobald wir

$$n^{1/2} \int_I (\bar{\Lambda}_{\vartheta, n}(t) - \mathbb{E} \Lambda_{\vartheta}(t)) \cdot \mathbb{E} \frac{\partial}{\partial \vartheta} \Lambda_{\vartheta'}(t) \bar{M}_n(dt) \xrightarrow{\mathbb{P}} 0 \quad (8.2.17)$$

für $\vartheta, \vartheta' \in \Theta$ gezeigt haben.

Zum Beweis zerlegen wir (8.2.17) in zwei Teilsummen:

$$\begin{aligned} & n^{-3/2} \sum_{i,k=1}^n \int_I (\Lambda_{\vartheta, k}(t) - \mathbb{E} \Lambda_{\vartheta}(t)) \cdot \mathbb{E} \frac{\partial}{\partial \vartheta} \Lambda_{\vartheta'}(t) M_i(dt) \\ &= n^{-3/2} \sum_{i=1}^n \int_I (\Lambda_{\vartheta, i}(t) - \mathbb{E} \Lambda_{\vartheta}(t)) \cdot \mathbb{E} \frac{\partial}{\partial \vartheta} \Lambda_{\vartheta'}(t) M_i(dt) \end{aligned} \quad (8.2.18)$$

$$+ n^{-3/2} \sum_{\substack{i,k=1 \\ i \neq k}}^n \int_I (\Lambda_{\vartheta, k}(t) - \mathbb{E} \Lambda_{\vartheta}(t)) \cdot \mathbb{E} \frac{\partial}{\partial \vartheta} \Lambda_{\vartheta'}(t) M_i(dt) \quad (8.2.19)$$

Für das erste Moment von (8.2.18) erhalten wir

$$\begin{aligned} & \mathbb{E} \left| n^{-3/2} \sum_{i=1}^n \int_I (\Lambda_{\vartheta, i}(t) - \mathbb{E} \Lambda_{\vartheta}(t)) \cdot \mathbb{E} \frac{\partial}{\partial \vartheta} \Lambda_{\vartheta'}(t) M_i(dt) \right| \\ & \leq n^{-3/2} \sum_{i=1}^n \mathbb{E} \left| \int_I (\Lambda_{\vartheta, i}(t) - \mathbb{E} \Lambda_{\vartheta}(t)) \cdot \mathbb{E} \frac{\partial}{\partial \vartheta} \Lambda_{\vartheta'}(t) M_i(dt) \right| \end{aligned}$$

Dem Faktor $n^{-3/2}$ stehen also n Summanden gegenüber. Diese sind endlich, wie aus Lemma 8.2.1 folgt. Darüberhinaus haben sie alle denselben Wert, da die beteiligten Prozesse i. i. d. sind. Der Gesamtausdruck konvergiert folglich in L_1 gegen Null.

Wir wollen L_2 -Konvergenz für (8.2.19) zeigen, dazu muß der Ausdruck

$$n^{-3} \mathbb{E} \left| \sum_{\substack{i,k=1 \\ i \neq k}}^n \int_I (\Lambda_{\vartheta,k}(t) - \mathbb{E} \Lambda_{\vartheta}(t)) \cdot \mathbb{E} \frac{\partial}{\partial \vartheta} \Lambda_{\vartheta'}(t) M_i(dt) \right|^2$$

untersucht werden. Dazu setzen wir

$$U_{ki} := \int_I (\Lambda_{\vartheta,k}(t) - \mathbb{E} \Lambda_{\vartheta}(t)) \cdot \mathbb{E} \frac{\partial}{\partial \vartheta} \Lambda_{\vartheta'}(t) M_i(dt)$$

und wenden Lemma A.3.2 an. Dessen Voraussetzungen sind gemäß Lemma 8.2.1 und (8.2.2) erfüllt. Es folgt

$$\begin{aligned} n^{-3} \mathbb{E} \left| \sum_{\substack{i,k=1 \\ i \neq k}}^n \int_I (\Lambda_{\vartheta,k}(t) - \mathbb{E} \Lambda_{\vartheta}(t)) \cdot \mathbb{E} \frac{\partial}{\partial \vartheta} \Lambda_{\vartheta'}(t) M_i(dt) \right|^2 \\ \leq 2n^{-3} \sum_{\substack{i,k=1 \\ i \neq k}}^n \mathbb{E} \left| \int_I (\Lambda_{\vartheta,k}(t) - \mathbb{E} \Lambda_{\vartheta}(t)) \cdot \mathbb{E} \frac{\partial}{\partial \vartheta} \Lambda_{\vartheta'}(t) M_i(dt) \right|^2 \end{aligned}$$

Die entstandenen $O(n^2)$ Summanden haben wiederum eine gemeinsame, endliche Schranke, die Konvergenz der Summe folgt.

Wir zeigen nun (8.2.16). Es gilt:

$$\begin{aligned} \alpha_n^{(3)}(\vartheta) &= n^{-5/2} \sum_{p,k,i=1}^n \int_I (\Lambda_{\vartheta,p}(t) - \Lambda_{\vartheta_{\circ},p}(t)) \left(\frac{\partial}{\partial \vartheta} \Lambda_{\vartheta,k}(t) - \mathbb{E} \frac{\partial}{\partial \vartheta} \Lambda_{\vartheta,k}(t) \right) M_i(dt) \\ &= n^{-5/2} \sum_{i=1}^n \int_I (\Lambda_{\vartheta,i}(t) - \Lambda_{\vartheta_{\circ},i}(t)) \left(\frac{\partial}{\partial \vartheta} \Lambda_{\vartheta,i}(t) - \mathbb{E} \frac{\partial}{\partial \vartheta} \Lambda_{\vartheta,i}(t) \right) M_i(dt) \end{aligned} \quad (8.2.20)$$

$$+ n^{-5/2} \sum_{\substack{k,i=1 \\ k \neq i}}^n \int_I (\Lambda_{\vartheta,i}(t) - \Lambda_{\vartheta_{\circ},i}(t)) \left(\frac{\partial}{\partial \vartheta} \Lambda_{\vartheta,k}(t) - \mathbb{E} \frac{\partial}{\partial \vartheta} \Lambda_{\vartheta,k}(t) \right) M_i(dt) \quad (8.2.21)$$

$$+ n^{-5/2} \sum_{\substack{k,i=1 \\ k \neq i}}^n \int_I (\Lambda_{\vartheta,i}(t) - \Lambda_{\vartheta_{\circ},i}(t)) \left(\frac{\partial}{\partial \vartheta} \Lambda_{\vartheta,i}(t) - \mathbb{E} \frac{\partial}{\partial \vartheta} \Lambda_{\vartheta,i}(t) \right) M_k(dt) \quad (8.2.22)$$

$$+ n^{-5/2} \sum_{\substack{k,i=1 \\ k \neq i}}^n \int_I (\Lambda_{\vartheta,i}(t) - \Lambda_{\vartheta_{\circ},i}(t)) \left(\frac{\partial}{\partial \vartheta} \Lambda_{\vartheta,k}(t) - \mathbb{E} \frac{\partial}{\partial \vartheta} \Lambda_{\vartheta,k}(t) \right) M_k(dt) \quad (8.2.23)$$

$$+ n^{-5/2} \sum_{\substack{p,k,i=1 \\ p \neq k \neq i \neq p}}^n \int_I (\Lambda_{\vartheta,p}(t) - \Lambda_{\vartheta_{\circ},p}(t)) \left(\frac{\partial}{\partial \vartheta} \Lambda_{\vartheta,k}(t) - \mathbb{E} \frac{\partial}{\partial \vartheta} \Lambda_{\vartheta,k}(t) \right) M_i(dt) \quad (8.2.24)$$

In (8.2.20) bis (8.2.23) werden wir L_1 -Konvergenz gegen Null zeigen: Für (8.2.20) ergibt sich

$$\begin{aligned} & \mathbb{E} \left| n^{-5/2} \sum_{i=1}^n \int_I (\Lambda_{\vartheta,i}(t) - \Lambda_{\vartheta_0,i}(t)) \left(\frac{\partial}{\partial \vartheta} \Lambda_{\vartheta,i}(t) - \mathbb{E} \frac{\partial}{\partial \vartheta} \Lambda_{\vartheta,i}(t) \right) M_i(dt) \right| \\ & \leq n^{-5/2} \sum_{i=1}^n \mathbb{E} \left| \int_I (\Lambda_{\vartheta,i}(t) - \Lambda_{\vartheta_0,i}(t)) \left(\frac{\partial}{\partial \vartheta} \Lambda_{\vartheta,i}(t) - \mathbb{E} \frac{\partial}{\partial \vartheta} \Lambda_{\vartheta,i}(t) \right) M_i(dt) \right| \end{aligned}$$

Wie oben sind die n Summanden gleichmäßig beschränkt und die gesuchte Konvergenzaussage folgt direkt. Bei entsprechender Betrachtung der Terme (8.2.21) bis (8.2.23) treten $O(n^2)$ Summanden auf, das Konvergenzverhalten der Gesamtausdrücke läßt sich daraus entsprechend ableiten.

Das zweite Moment von (8.2.24) ist:

$$n^{-5} \mathbb{E} \left| \sum_{\substack{p,k,i=1 \\ p \neq k \neq i \neq p}}^n \int_I (\Lambda_{\vartheta,p}(t) - \Lambda_{\vartheta_0,p}(t)) \left(\frac{\partial}{\partial \vartheta} \Lambda_{\vartheta,k}(t) - \mathbb{E} \frac{\partial}{\partial \vartheta} \Lambda_{\vartheta,k}(t) \right) M_i(dt) \right|^2$$

Zu seiner Untersuchung setzen wir

$$U_{pki} := \int_I (\Lambda_{\vartheta,p}(t) - \Lambda_{\vartheta_0,p}(t)) \left(\frac{\partial}{\partial \vartheta} \Lambda_{\vartheta,k}(t) - \mathbb{E} \frac{\partial}{\partial \vartheta} \Lambda_{\vartheta,k}(t) \right) M_i(dt)$$

und wenden Lemma A.3.3 an. Dies wird durch Lemma 8.2.1 und (8.2.6) ermöglicht. Wir erhalten⁹

$$\begin{aligned} & n^{-5} \mathbb{E} \left| \sum_{\substack{p,k,i=1 \\ p \neq k \neq i \neq p}}^n \int_I (\Lambda_{\vartheta,p}(t) - \Lambda_{\vartheta_0,p}(t)) \left(\frac{\partial}{\partial \vartheta} \Lambda_{\vartheta,k}(t) - \mathbb{E} \frac{\partial}{\partial \vartheta} \Lambda_{\vartheta,k}(t) \right) M_i(dt) \right|^2 \\ & \leq 64n^{-5} \sum_{\substack{p,q,k,i=1 \\ i \leq k \leq p, q \leq i}}^n \mathbb{E} \left(\iint_{I \times I} (\Lambda_{\vartheta,p}(t) - \Lambda_{\vartheta_0,p}(t)) (\Lambda_{\vartheta,q}(s) - \Lambda_{\vartheta_0,q}(s)) \right. \\ & \quad \left. \left(\frac{\partial}{\partial \vartheta} \Lambda_{\vartheta,k}(t) - \mathbb{E} \frac{\partial}{\partial \vartheta} \Lambda_{\vartheta,k}(t) \right)^\top \left(\frac{\partial}{\partial \vartheta} \Lambda_{\vartheta,k}(s) - \mathbb{E} \frac{\partial}{\partial \vartheta} \Lambda_{\vartheta,k}(s) \right) M_i(dt) M_i(ds) \right) \end{aligned}$$

Die resultierenden $O(n^4)$ Terme sind endlich und haben eine gemeinsame Schranke, der Faktor n^{-5} liefert L_2 -Konvergenz. ■

Abschließend zeigen wir (8.2.13):

8.2.7 Lemma *Es gilt*

$$\begin{aligned} \mathbb{E} |\alpha_n^{(i)}(\vartheta) - \alpha_n^{(i)}(\vartheta')|^2 & \leq \text{const.} \cdot |\vartheta - \vartheta'|^2, \quad \vartheta, \vartheta' \in \Theta_0 \\ & \quad i = 1, 2, 3 \end{aligned}$$

⁹Für die Definition des Ausdrucks " $i \leq k \leq p, q \leq i$ " verweisen wir auf (A.3.2). Er stellt eine zusätzliche Bedingung an die Indices dar, welche die Anzahl von n^4 Summanden weiter reduziert.

Beweis:

Für Vektoren $a_1, \dots, a_l \in \mathbb{R}^d$, $l \in \mathbb{N}$, zeigt man durch vollständige Induktion

$$\left| \sum_{j=1}^l a_j \right|^2 \leq 2^l \cdot \sum_{j=1}^l |a_j|^2. \quad (8.2.25)$$

Deshalb können wir zum Nachweis des Lemmas den Ausdruck $\alpha_n^{(i)}(\vartheta) - \alpha_n^{(i)}(\vartheta')$ in Summanden zerlegen und für diese die Existenz entsprechender Schranken zeigen. Wir betrachten die Restterme für $i = 1, 2$ wieder gemeinsam. Gemäß (8.2.14b) genügt es, die zweiten Momente der Ausdrücke

$$n^{1/2} \int_I (\bar{\Lambda}_{\vartheta, n}(t) - \mathbb{E} \Lambda_{\vartheta}(t) - \bar{\Lambda}_{\vartheta', n}(t) + \mathbb{E} \Lambda_{\vartheta'}(t)) \mathbb{E} \frac{\partial}{\partial \vartheta} \Lambda_{\vartheta}(t) \bar{M}_n(dt) \quad (8.2.26)$$

und

$$n^{1/2} \int_I (\bar{\Lambda}_{\vartheta', n}(t) - \mathbb{E} \Lambda_{\vartheta'}(t)) (\mathbb{E} \frac{\partial}{\partial \vartheta} \Lambda_{\vartheta}(t) - \mathbb{E} \frac{\partial}{\partial \vartheta} \Lambda_{\vartheta'}(t)) \bar{M}_n(dt) \quad (8.2.27)$$

abzuschätzen. (8.2.26) zerlegen wir weiter:

$$\begin{aligned} & n^{1/2} \int_I (\bar{\Lambda}_{\vartheta, n}(t) - \mathbb{E} \Lambda_{\vartheta}(t) - \bar{\Lambda}_{\vartheta', n}(t) + \mathbb{E} \Lambda_{\vartheta'}(t)) \mathbb{E} \frac{\partial}{\partial \vartheta} \Lambda_{\vartheta}(t) \bar{M}_n(dt) \\ &= n^{-3/2} \sum_{k, i=1}^n \int_I (\Lambda_{\vartheta, k}(t) - \mathbb{E} \Lambda_{\vartheta}(t) - \Lambda_{\vartheta', k}(t) + \mathbb{E} \Lambda_{\vartheta'}(t)) \mathbb{E} \frac{\partial}{\partial \vartheta} \Lambda_{\vartheta}(t) M_i(dt) \\ &= n^{-3/2} \sum_{i=1}^n \int_I (\Lambda_{\vartheta, i}(t) - \mathbb{E} \Lambda_{\vartheta}(t) - \Lambda_{\vartheta', i}(t) + \mathbb{E} \Lambda_{\vartheta'}(t)) \mathbb{E} \frac{\partial}{\partial \vartheta} \Lambda_{\vartheta}(t) M_i(dt) \end{aligned} \quad (8.2.28)$$

$$+ n^{-3/2} \sum_{\substack{k, i=1 \\ k \neq i}}^n \int_I (\Lambda_{\vartheta, k}(t) - \mathbb{E} \Lambda_{\vartheta}(t) - \Lambda_{\vartheta', k}(t) + \mathbb{E} \Lambda_{\vartheta'}(t)) \mathbb{E} \frac{\partial}{\partial \vartheta} \Lambda_{\vartheta}(t) M_i(dt) \quad (8.2.29)$$

Das zweite Moment von (8.2.28) erlaubt die Abschätzung

$$\begin{aligned} & \mathbb{E} \left| n^{-3/2} \sum_{i=1}^n \int_I (\Lambda_{\vartheta, i}(t) - \mathbb{E} \Lambda_{\vartheta}(t) - \Lambda_{\vartheta', i}(t) + \mathbb{E} \Lambda_{\vartheta'}(t)) \mathbb{E} \frac{\partial}{\partial \vartheta} \Lambda_{\vartheta}(t) M_i(dt) \right|^2 \\ &= n^{-3} \mathbb{E} \left(\sum_{i, j=1}^n \iint_{I \times I} (\Lambda_{\vartheta, i}(t) - \mathbb{E} \Lambda_{\vartheta}(t) - \Lambda_{\vartheta', i}(t) + \mathbb{E} \Lambda_{\vartheta'}(t)) \right. \\ & \quad \left. \mathbb{E} \frac{\partial}{\partial \vartheta} \Lambda_{\vartheta}(t) \right)^{\top} \mathbb{E} \frac{\partial}{\partial \vartheta} \Lambda_{\vartheta}(s) (\Lambda_{\vartheta, j}(s) - \mathbb{E} \Lambda_{\vartheta}(s) - \Lambda_{\vartheta', j}(s) + \mathbb{E} \Lambda_{\vartheta'}(s)) \\ & \quad \left. M_i(dt) M_j(ds) \right) \end{aligned}$$

$$\begin{aligned}
&\leq n^{-3} \mathbb{E} \left(\sum_{i,j=1}^n \iint_{I \times I} |\Lambda_{\vartheta,i}(t) - \mathbb{E} \Lambda_{\vartheta}(t) - \Lambda_{\vartheta',i}(t) + \mathbb{E} \Lambda_{\vartheta'}(t)| \right. \\
&\quad \left. |\mathbb{E} \frac{\partial}{\partial \vartheta} \Lambda_{\vartheta}(t)| |\mathbb{E} \frac{\partial}{\partial \vartheta} \Lambda_{\vartheta}(s)| |\Lambda_{\vartheta,j}(s) - \mathbb{E} \Lambda_{\vartheta}(s) - \Lambda_{\vartheta',j}(s) + \mathbb{E} \Lambda_{\vartheta'}(s)| \right. \\
&\quad \left. (N_i(dt) + \Lambda_{\vartheta_{\circ},i}(dt)) (N_j(ds) + \Lambda_{\vartheta_{\circ},j}(ds)) \right) \\
&\leq n^{-3} |\vartheta - \vartheta'|^2 \sum_{i,j=1}^n \mathbb{E} \left(\iint_{I \times I} 4L^2(t) |\mathbb{E} \frac{\partial}{\partial \vartheta} \Lambda_{\vartheta}(t)| |\mathbb{E} \frac{\partial}{\partial \vartheta} \Lambda_{\vartheta}(s)| \right. \\
&\quad \left. (N_i(dt) + \Lambda_{\vartheta_{\circ},i}(dt)) (N_j(ds) + \Lambda_{\vartheta_{\circ},j}(ds)) \right)
\end{aligned}$$

Dabei bezeichnet $L(t)$ zu festem t eine Lipschitz-Konstante der bzgl. ϑ f. s. differenzierbaren Funktion $\Lambda_{\vartheta}(t)$. Die n^2 Erwartungswerte sind gemäß (3.4.3) und (3.4.4) endlich, weil die stetige Funktion $|\mathbb{E} \frac{\partial}{\partial \vartheta} \Lambda_{\vartheta}(t)|$ auf dem Kompaktum $I \times \Theta_{\circ}$ beschränkt ist. Da die beteiligten Prozesse i. i. d. sind, haben die n^2 Summanden sogar eine gemeinsame Schranke. Der Faktor n^{-3} sorgt dafür, daß die Summen einer von n unabhängigen Schranke unterliegen.

Nun wenden wir uns dem zweiten Moment von (8.2.29) zu. Es lautet:

$$n^{-3} \mathbb{E} \left| \sum_{\substack{k,i=1 \\ k \neq i}}^n \int_I (\Lambda_{\vartheta,k}(t) - \mathbb{E} \Lambda_{\vartheta}(t) - \Lambda_{\vartheta',k}(t) + \mathbb{E} \Lambda_{\vartheta'}(t)) \mathbb{E} \frac{\partial}{\partial \vartheta} \Lambda_{\vartheta}(t) M_i(dt) \right|^2$$

Wir setzen

$$U_{ki} := \int_I (\Lambda_{\vartheta,k}(t) - \mathbb{E} \Lambda_{\vartheta}(t) - \Lambda_{\vartheta',k}(t) + \mathbb{E} \Lambda_{\vartheta'}(t)) \mathbb{E} \frac{\partial}{\partial \vartheta} \Lambda_{\vartheta}(t) M_i(dt)$$

und wenden Lemma A.3.2 an. Dies wird durch Lemma 8.2.1 und (8.2.2) ermöglicht. Wir erhalten so die Abschätzung

$$\begin{aligned}
&n^{-3} \mathbb{E} \left| \sum_{\substack{k,i=1 \\ k \neq i}}^n \int_I (\Lambda_{\vartheta,k}(t) - \mathbb{E} \Lambda_{\vartheta}(t) - \Lambda_{\vartheta',k}(t) + \mathbb{E} \Lambda_{\vartheta'}(t)) \mathbb{E} \frac{\partial}{\partial \vartheta} \Lambda_{\vartheta}(t) M_i(dt) \right|^2 \\
&\leq 2n^{-3} \sum_{\substack{k,i=1 \\ k \neq i}}^n \mathbb{E} \left| \int_I (\Lambda_{\vartheta,k}(t) - \mathbb{E} \Lambda_{\vartheta}(t) - \Lambda_{\vartheta',k}(t) + \mathbb{E} \Lambda_{\vartheta'}(t)) \mathbb{E} \frac{\partial}{\partial \vartheta} \Lambda_{\vartheta}(t) M_i(dt) \right|^2 \\
&\leq 2n^{-3} \sum_{\substack{k,i=1 \\ k \neq i}}^n \mathbb{E} \left(\int_I |\Lambda_{\vartheta,k}(t) - \mathbb{E} \Lambda_{\vartheta}(t) - \Lambda_{\vartheta',k}(t) + \mathbb{E} \Lambda_{\vartheta'}(t)| |\mathbb{E} \frac{\partial}{\partial \vartheta} \Lambda_{\vartheta}(t)| \right. \\
&\quad \left. (N_i(dt) + \Lambda_{\circ}(dt)) \right)^2
\end{aligned}$$

Die Abschätzung wird wie im Beweis von (8.2.28) beendet.

Der Nachweis für (8.2.27) entspricht dem soeben für (8.2.26) geführten.

Die Aussage des Lemmas ist nun noch für $i = 3$ zu zeigen. Gemäß (8.2.14c) sind die beiden Ausdrücke

$$n^{1/2} \int_I (\bar{\Lambda}_{\vartheta,n}(t) - \bar{\Lambda}_{\vartheta',n}(t)) \left(\frac{\partial}{\partial \vartheta} \bar{\Lambda}_{\vartheta,n}(t) - \mathbb{E} \frac{\partial}{\partial \vartheta} \Lambda_{\vartheta}(t) \right) \bar{M}_n(dt) \quad (8.2.30)$$

und

$$n^{1/2} \int_I (\bar{\Lambda}_{\vartheta',n}(t) - \bar{\Lambda}_{\vartheta_\circ,n}(t)) \left(\frac{\partial}{\partial \vartheta} \bar{\Lambda}_{\vartheta,n}(t) - \mathbb{E} \frac{\partial}{\partial \vartheta} \Lambda_{\vartheta}(t) - \frac{\partial}{\partial \vartheta} \bar{\Lambda}_{\vartheta',n}(t) + \mathbb{E} \frac{\partial}{\partial \vartheta} \Lambda_{\vartheta'}(t) \right) \bar{M}_n(dt) \quad (8.2.31)$$

zu betrachten. Wir zerlegen (8.2.30) in Teilsommen:

$$\begin{aligned} & n^{1/2} \int_I (\bar{\Lambda}_{\vartheta,n}(t) - \bar{\Lambda}_{\vartheta',n}(t)) \left(\frac{\partial}{\partial \vartheta} \bar{\Lambda}_{\vartheta,n}(t) - \mathbb{E} \frac{\partial}{\partial \vartheta} \Lambda_{\vartheta}(t) \right) \bar{M}_n(dt) \\ &= n^{-5/2} \sum_{p,k,i=1}^n \int_I (\Lambda_{\vartheta,p}(t) - \Lambda_{\vartheta',p}(t)) \left(\frac{\partial}{\partial \vartheta} \Lambda_{\vartheta,k}(t) - \mathbb{E} \frac{\partial}{\partial \vartheta} \Lambda_{\vartheta}(t) \right) M_i(dt) \\ &= n^{-5/2} \sum_{i=1}^n \int_I (\Lambda_{\vartheta,i}(t) - \Lambda_{\vartheta',i}(t)) \left(\frac{\partial}{\partial \vartheta} \Lambda_{\vartheta,i}(t) - \mathbb{E} \frac{\partial}{\partial \vartheta} \Lambda_{\vartheta}(t) \right) M_i(dt) \quad (8.2.32) \end{aligned}$$

$$+ n^{-5/2} \sum_{\substack{i,k=1 \\ i \neq k}}^n \int_I (\Lambda_{\vartheta,i}(t) - \Lambda_{\vartheta',i}(t)) \left(\frac{\partial}{\partial \vartheta} \Lambda_{\vartheta,i}(t) - \mathbb{E} \frac{\partial}{\partial \vartheta} \Lambda_{\vartheta}(t) \right) M_k(dt) \quad (8.2.33)$$

$$+ n^{-5/2} \sum_{\substack{i,k=1 \\ i \neq k}}^n \int_I (\Lambda_{\vartheta,i}(t) - \Lambda_{\vartheta',i}(t)) \left(\frac{\partial}{\partial \vartheta} \Lambda_{\vartheta,k}(t) - \mathbb{E} \frac{\partial}{\partial \vartheta} \Lambda_{\vartheta}(t) \right) M_i(dt) \quad (8.2.34)$$

$$+ n^{-5/2} \sum_{\substack{i,k=1 \\ i \neq k}}^n \int_I (\Lambda_{\vartheta,i}(t) - \Lambda_{\vartheta',i}(t)) \left(\frac{\partial}{\partial \vartheta} \Lambda_{\vartheta,k}(t) - \mathbb{E} \frac{\partial}{\partial \vartheta} \Lambda_{\vartheta}(t) \right) M_k(dt) \quad (8.2.35)$$

$$+ n^{-5/2} \sum_{\substack{p,k,i=1 \\ p \neq k \neq i \neq p}}^n \int_I (\Lambda_{\vartheta,p}(t) - \Lambda_{\vartheta',p}(t)) \left(\frac{\partial}{\partial \vartheta} \Lambda_{\vartheta,k}(t) - \mathbb{E} \frac{\partial}{\partial \vartheta} \Lambda_{\vartheta}(t) \right) M_i(dt) \quad (8.2.36)$$

Das zweite Moment von (8.2.32) gestattet die Abschätzung:

$$\begin{aligned} & \mathbb{E} \left| n^{-5/2} \sum_{i=1}^n \int_I (\Lambda_{\vartheta,i}(t) - \Lambda_{\vartheta',i}(t)) \left(\frac{\partial}{\partial \vartheta} \Lambda_{\vartheta,i}(t) - \mathbb{E} \frac{\partial}{\partial \vartheta} \Lambda_{\vartheta}(t) \right) M_i(dt) \right|^2 \\ &= n^{-5} \mathbb{E} \left(\sum_{i,j=1}^n \iint_{I \times I} (\Lambda_{\vartheta,i}(t) - \Lambda_{\vartheta',i}(t)) (\Lambda_{\vartheta,j}(s) - \Lambda_{\vartheta',j}(s)) \right. \\ & \quad \left. \left(\frac{\partial}{\partial \vartheta} \Lambda_{\vartheta,i}(t) - \mathbb{E} \frac{\partial}{\partial \vartheta} \Lambda_{\vartheta}(t) \right)^\top \left(\frac{\partial}{\partial \vartheta} \Lambda_{\vartheta,j}(s) - \mathbb{E} \frac{\partial}{\partial \vartheta} \Lambda_{\vartheta}(s) \right) M_i(dt) M_j(ds) \right) \\ &\leq n^{-5} \mathbb{E} \left(\sum_{i,j=1}^n \iint_{I \times I} |\Lambda_{\vartheta,i}(t) - \Lambda_{\vartheta',i}(t)| |\Lambda_{\vartheta,j}(s) - \Lambda_{\vartheta',j}(s)| \right. \\ & \quad \left. \left| \frac{\partial}{\partial \vartheta} \Lambda_{\vartheta,i}(t) - \mathbb{E} \frac{\partial}{\partial \vartheta} \Lambda_{\vartheta}(t) \right| \left| \frac{\partial}{\partial \vartheta} \Lambda_{\vartheta,j}(s) - \mathbb{E} \frac{\partial}{\partial \vartheta} \Lambda_{\vartheta}(s) \right| \right. \\ & \quad \left. (N_i(dt) + \Lambda_{\vartheta_\circ,i}(dt)) (N_j(ds) + \Lambda_{\vartheta_\circ,j}(ds)) \right) \end{aligned}$$

$$\leq n^{-5} |\vartheta - \vartheta'|^2 \sum_{i,j=1}^n \mathbb{E} \left(\iint_{I \times I} L^2(t) \left| \frac{\partial}{\partial \vartheta} \Lambda_{\vartheta,i}(t) - \mathbb{E} \frac{\partial}{\partial \vartheta} \Lambda_{\vartheta}(t) \right| \right. \\ \left. \left| \frac{\partial}{\partial \vartheta} \Lambda_{\vartheta,j}(s) - \mathbb{E} \frac{\partial}{\partial \vartheta} \Lambda_{\vartheta}(s) \right| (N_i(dt) + \Lambda_{\vartheta_\circ,i}(dt)) (N_j(ds) + \Lambda_{\vartheta_\circ,j}(ds)) \right)$$

Die Doppelintegrale haben endliche Erwartungswerte. Dies folgt unter ausschließlicher Verwendung der Hölderschen Ungleichung mit (3.4.8), (3.4.9), (3.4.3) und (3.4.4). Mögliche stochastische Abhängigkeiten spielen deshalb keine Rolle. Da die beteiligten Prozesse i. i. d. sind, gibt es darüber hinaus eine gemeinsame Schranke für die Beträge der n^2 Summanden. Aufgrund des Faktors n^{-5} hat der gesamte Ausdruck eine Schranke, die nicht von n abhängt.

Die zweiten Momente der Terme (8.2.33) bis (8.2.35) gestatten eine ganz entsprechende Betrachtung, bei der schließlich $O(n^4)$ gleichmäßig beschränkte Summanden neben dem Faktor n^{-5} auftreten.

Der Ausdruck (8.2.36) muß gesondert untersucht werden. Sein zweites Moment lautet:

$$n^{-5} \mathbb{E} \left| \sum_{\substack{p,k,i=1 \\ p \neq k \neq i \neq p}}^n \int_I (\Lambda_{\vartheta,p}(t) - \Lambda_{\vartheta',p}(t)) \left(\frac{\partial}{\partial \vartheta} \Lambda_{\vartheta,k}(t) - \mathbb{E} \frac{\partial}{\partial \vartheta} \Lambda_{\vartheta}(t) \right) M_i(dt) \right|^2$$

Wir setzen

$$U_{pki} := \int_I (\Lambda_{\vartheta,p}(t) - \Lambda_{\vartheta',p}(t)) \left(\frac{\partial}{\partial \vartheta} \Lambda_{\vartheta,k}(t) - \mathbb{E} \frac{\partial}{\partial \vartheta} \Lambda_{\vartheta}(t) \right) M_i(dt)$$

und wenden Lemma A.3.3 an. Dazu beachten wir Lemma 8.2.1 und (8.2.6). Wir erhalten die Abschätzung:

$$n^{-5} \mathbb{E} \left| \sum_{\substack{p,k,i=1 \\ p \neq k \neq i \neq p}}^n \int_I (\Lambda_{\vartheta,p}(t) - \Lambda_{\vartheta',p}(t)) \left(\frac{\partial}{\partial \vartheta} \Lambda_{\vartheta,k}(t) - \mathbb{E} \frac{\partial}{\partial \vartheta} \Lambda_{\vartheta}(t) \right) M_i(dt) \right|^2 \\ \leq 64n^{-5} \sum_{\substack{p,q,k,i=1 \\ k \leq i \leq p, q \leq k}}^n \mathbb{E} \left| \iint_{I \times I} (\Lambda_{\vartheta,p}(t) - \Lambda_{\vartheta',p}(t)) (\Lambda_{\vartheta,q}(s) - \Lambda_{\vartheta',q}(s)) \right. \\ \left. \left(\frac{\partial}{\partial \vartheta} \Lambda_{\vartheta,k}(s) - \mathbb{E} \frac{\partial}{\partial \vartheta} \Lambda_{\vartheta}(s) \right) \left(\frac{\partial}{\partial \vartheta} \Lambda_{\vartheta,k}(t) - \mathbb{E} \frac{\partial}{\partial \vartheta} \Lambda_{\vartheta}(t) \right) M_i(ds) M_i(dt) \right|^2$$

Damit haben wir die Anzahl der Summanden auf $O(n^4)$ reduziert; der weitere Beweisgang entspricht den bereits vorgestellten.

Schließlich wird (8.2.31) entsprechend zu (8.2.30) bewiesen. \blacksquare

8.2.3 Straffheit des β -Prozesses

Analog zu Satz 8.2.4 zeigen wir in diesem Abschnitt die Straffheit des Prozesses

$$\beta_n(\vartheta) := n^{1/2} \int_I \bar{M}_n(t) \frac{\partial}{\partial \vartheta} \bar{\Lambda}_{\vartheta,n}(t) \bar{M}_n(dt) \quad (8.2.37)$$

auf Kompakta $\Theta_\circ \subset \Theta$:

8.2.8 Satz *Es gilt punktweise*

$$\beta_n(\vartheta) = o_{\mathbb{P}}(1).$$

Auf kompakten Parameterteilmengen $\Theta_o \subset \Theta$ ist $\beta_n(\vartheta)$ C -straff und konvergiert deswegen gleichmäßig in $\vartheta \in \Theta_o$ gegen Null.

Beweis:

Wie im Beweis des Satzes 8.2.4 genügt es, die punktweise Konvergenz

$$\beta_n(\vartheta) = o_{\mathbb{P}}(1) \quad \forall \vartheta \in \Theta \quad (8.2.38)$$

und die folgende gleichmäßige Schranke des zweiten Moments auf kompakten Parameterteilmengen $\Theta_o \subset \Theta$

$$\mathbb{E} |\beta_n(\vartheta) - \beta_n(\vartheta')|^2 \leq \text{const.} \cdot |\vartheta - \vartheta'|^2 \quad \forall \vartheta, \vartheta' \in \Theta_o. \quad (8.2.39)$$

nachzuweisen. Natürlich folgen die in Satz A.5.2 benötigten Momentenbedingungen für die Komponenten $(\beta_n(\vartheta))_j$ aus (8.2.39), da

$$\mathbb{E} (\beta_n(\vartheta) - \beta_n(\vartheta'))_j^2 \leq \mathbb{E} |\beta_n(\vartheta) - \beta_n(\vartheta')|^2 \quad j = 1, \dots, d$$

gilt.

Wir werden (8.2.38) und (8.2.39) in den beiden folgenden Lemmata zeigen. ■

Wir beginnen mit dem Nachweis von (8.2.38):

8.2.9 Lemma *Es gilt*

$$\beta_n(\vartheta) = o_{\mathbb{P}}(1) \quad \forall \vartheta \in \Theta.$$

Beweis:

Wir zeigen die stochastische Konvergenz als Folge der L_1 - oder L_2 -Konvergenz. Dafür zerlegen wir den Prozeß $\beta_n(\vartheta)$ zunächst geeignet:

$$\begin{aligned} \beta_n(\vartheta) &= n^{-5/2} \sum_{p,k,i=1}^n \int_I M_k(t) \frac{\partial}{\partial \vartheta} \Lambda_{\vartheta,p}(t) M_i(dt) \\ &= n^{-5/2} \sum_{i=1}^n \int_I M_i(t) \frac{\partial}{\partial \vartheta} \Lambda_{\vartheta,i}(t) M_i(dt) \end{aligned} \quad (8.2.40)$$

$$+ n^{-5/2} \sum_{\substack{i,k=1 \\ i \neq k}}^n \int_I M_i(t) \frac{\partial}{\partial \vartheta} \Lambda_{\vartheta,i}(t) M_k(dt) \quad (8.2.41)$$

$$+ n^{-5/2} \sum_{\substack{i,k=1 \\ i \neq k}}^n \int_I M_i(t) \frac{\partial}{\partial \vartheta} \Lambda_{\vartheta,k}(t) M_k(dt) \quad (8.2.42)$$

$$+ n^{-5/2} \sum_{\substack{i,k=1 \\ i \neq k}}^n \int_I M_i(t) \frac{\partial}{\partial \vartheta} \Lambda_{\vartheta,k}(t) M_i(dt) \quad (8.2.43)$$

$$+ n^{-5/2} \sum_{\substack{p,k,i=1 \\ p \neq k \neq i \neq p}}^n \int_I M_k(t) \frac{\partial}{\partial \vartheta} \Lambda_{\vartheta,p}(t) M_i(dt) \quad (8.2.44)$$

Wir betrachten das erste Moment von (8.2.40):

$$\begin{aligned} \mathbb{E} \left| n^{-5/2} \sum_{i=1}^n \int_I M_i(t) \frac{\partial}{\partial \vartheta} \Lambda_{\vartheta,i}(t) M_i(dt) \right| \\ \leq n^{-5/2} \sum_{i=1}^n \mathbb{E} \left| \int_I M_i(t) \frac{\partial}{\partial \vartheta} \Lambda_{\vartheta,i}(t) M_i(dt) \right| \end{aligned}$$

Es treten n Summanden auf. Sie sind endlich, wie aus Lemma 8.2.1 folgt. Da die beteiligten Prozesse i. i. d. sind, haben außerdem alle Summanden denselben Wert. Der Faktor $n^{-5/2}$ liefert das gesuchte L_1 -Konvergenzverhalten.

Für die Summen (8.2.41) bis (8.2.43) treten bei entsprechender Betrachtung $O(n^2)$ gleichmäßig beschränkte Summanden auf. Wieder sorgt $n^{-5/2}$ für ihr asymptotisches Verschwinden in L_1 .

Es bleibt (8.2.44) zu untersuchen. Für die L_2 -Konvergenz ist der Ausdruck

$$n^{-5} \mathbb{E} \left| \sum_{\substack{p,k,i=1 \\ p \neq k \neq i \neq p}}^n \int_I M_k(t) \frac{\partial}{\partial \vartheta} \Lambda_{\vartheta,p}(t) M_i(dt) \right|^2$$

relevant. Wir setzen

$$U_{pki} := \int_I M_k(t) \frac{\partial}{\partial \vartheta} \Lambda_{\vartheta,p}(t) M_i(dt)$$

und wenden Lemma A.3.3 an. Dies ist möglich, da die Voraussetzungen gemäß Lemma 8.2.1 und (8.2.2) erfüllt sind. Lemma A.3.3 liefert also für den Term (8.2.44) folgende obere Schranke:

$$\begin{aligned} n^{-5} \mathbb{E} \left| \sum_{\substack{p,k,i=1 \\ p \neq k \neq i \neq p}}^n \int_I M_k(t) \frac{\partial}{\partial \vartheta} \Lambda_{\vartheta,p}(t) M_i(dt) \right|^2 \\ \leq 64n^{-5} \sum_{\substack{p,q,k,i=1 \\ i \leq k \leq p, q \leq i}}^n \mathbb{E} \left(\iint_{I \times I} M_k(t) M_k(s) \frac{\partial}{\partial \vartheta} \Lambda_{\vartheta,p}(t)^\top \frac{\partial}{\partial \vartheta} \Lambda_{\vartheta,q}(s) M_i(dt) M_i(ds) \right) \end{aligned}$$

Die $O(n^4)$ Summanden dieses Terms sind wie die entsprechenden Terme aus (8.2.40) bis (8.2.43) endlich und besitzen ebenfalls eine gemeinsame obere Schranke, der Faktor n^{-5} führt zu asymptotischer Vernachlässigbarkeit des Gesamtausdrucks. \blacksquare

Wir wenden uns (8.2.39) zu.

8.2.10 Lemma Auf kompakten Parameter-teilmengen $\Theta_\circ \subset \Theta$ gilt

$$\mathbb{E} |\beta_n(\vartheta) - \beta_n(\vartheta')|^2 \leq \text{const.} \cdot |\vartheta - \vartheta'|^2, \quad \forall \vartheta, \vartheta' \in \Theta_\circ.$$

Beweis:

Wir beginnen mit der Darstellung:

$$\begin{aligned} \beta_n(\vartheta) - \beta_n(\vartheta') &= n^{-5/2} \sum_{p,k,i=1}^n \int_I M_k(t) \left(\frac{\partial}{\partial \vartheta} \Lambda_{\vartheta,p}(t) - \frac{\partial}{\partial \vartheta'} \Lambda_{\vartheta',p}(t) \right) M_i(dt) \\ &= n^{-5/2} \sum_{i=1}^n \int_I M_i(t) \left(\frac{\partial}{\partial \vartheta} \Lambda_{\vartheta,i}(t) - \frac{\partial}{\partial \vartheta'} \Lambda_{\vartheta',i}(t) \right) M_i(dt) \\ &\quad + n^{-5/2} \sum_{\substack{i,k=1 \\ i \neq k}}^n \int_I M_i(t) \left(\frac{\partial}{\partial \vartheta} \Lambda_{\vartheta,i}(t) - \frac{\partial}{\partial \vartheta'} \Lambda_{\vartheta',i}(t) \right) M_k(dt) \\ &\quad + n^{-5/2} \sum_{\substack{i,k=1 \\ i \neq k}}^n \int_I M_i(t) \left(\frac{\partial}{\partial \vartheta} \Lambda_{\vartheta,k}(t) - \frac{\partial}{\partial \vartheta'} \Lambda_{\vartheta',k}(t) \right) M_k(dt) \\ &\quad + n^{-5/2} \sum_{\substack{i,k=1 \\ i \neq k}}^n \int_I M_i(t) \left(\frac{\partial}{\partial \vartheta} \Lambda_{\vartheta,k}(t) - \frac{\partial}{\partial \vartheta'} \Lambda_{\vartheta',k}(t) \right) M_i(dt) \\ &\quad + n^{-5/2} \sum_{\substack{p,k,i=1 \\ p \neq k \neq i \neq p}}^n \int_I M_k(t) \left(\frac{\partial}{\partial \vartheta} \Lambda_{\vartheta,p}(t) - \frac{\partial}{\partial \vartheta'} \Lambda_{\vartheta',p}(t) \right) M_i(dt) \end{aligned}$$

Die angegebene Darstellung von $\beta_n(\vartheta) - \beta_n(\vartheta')$ mit $l = 5$ Summanden erlaubt gemäß (8.2.25) die folgende Abschätzung des zweiten Moments:

$$\begin{aligned} &\mathbb{E} |\beta_n(\vartheta) - \beta_n(\vartheta')|^2 \\ &= n^{-5} \mathbb{E} \left| \sum_{p,k,i=1}^n \int_I M_k(t) \left(\frac{\partial}{\partial \vartheta} \Lambda_{\vartheta,p}(t) - \frac{\partial}{\partial \vartheta'} \Lambda_{\vartheta',p}(t) \right) M_i(dt) \right|^2 \\ &\leq 32n^{-5} \mathbb{E} \left| \sum_{i=1}^n \int_I M_i(t) \left(\frac{\partial}{\partial \vartheta} \Lambda_{\vartheta,i}(t) - \frac{\partial}{\partial \vartheta'} \Lambda_{\vartheta',i}(t) \right) M_i(dt) \right|^2 \\ &\quad + 32n^{-5} \mathbb{E} \left| \sum_{\substack{i,k=1 \\ i \neq k}}^n \int_I M_i(t) \left(\frac{\partial}{\partial \vartheta} \Lambda_{\vartheta,i}(t) - \frac{\partial}{\partial \vartheta'} \Lambda_{\vartheta',i}(t) \right) M_k(dt) \right|^2 \\ &\quad + 32n^{-5} \mathbb{E} \left| \sum_{\substack{i,k=1 \\ i \neq k}}^n \int_I M_i(t) \left(\frac{\partial}{\partial \vartheta} \Lambda_{\vartheta,k}(t) - \frac{\partial}{\partial \vartheta'} \Lambda_{\vartheta',k}(t) \right) M_k(dt) \right|^2 \\ &\quad + 32n^{-5} \mathbb{E} \left| \sum_{\substack{i,k=1 \\ i \neq k}}^n \int_I M_i(t) \left(\frac{\partial}{\partial \vartheta} \Lambda_{\vartheta,k}(t) - \frac{\partial}{\partial \vartheta'} \Lambda_{\vartheta',k}(t) \right) M_i(dt) \right|^2 \end{aligned}$$

$$+ 32n^{-5} \mathbb{E} \left| \sum_{\substack{p,k,i=1 \\ p \neq k \neq i \neq p}}^n \int_I M_k(t) \left(\frac{\partial}{\partial \vartheta} \Lambda_{\vartheta,p}(t) - \frac{\partial}{\partial \vartheta'} \Lambda_{\vartheta',p}(t) \right) M_i(dt) \right|^2$$

In den ersten vier entstandenen Erwartungswerten ersetzen wir nun das integrierende Maß durch eine monoton wachsende obere Schranke:

$$|M_i(t)| = |N_i(t) - \Lambda_{\vartheta_\circ,i}(t)| \leq N_i(t) + \Lambda_{\vartheta_\circ,i}(t)$$

Im Integranden betrachten wir anstelle des signierten Faktors $M_i(t)$ seinen Betrag $|M_i(t)|$.

Im Anschluß können wir die lokale Lipschitzstetigkeit der abgeleiteten Intensität gemäß (3.4.2) benutzen, um eine weitere Abschätzung nach oben zu erhalten:

$$\begin{aligned} & \mathbb{E}(\beta_n(\vartheta) - \beta_n(\vartheta'))^2 \\ & \leq 32|\vartheta - \vartheta'|^2 n^{-5} \mathbb{E} \left(\sum_{i=1}^n \int_I |M_i(t)| L'_i(t) (N_i(dt) + \Lambda_{\vartheta_\circ,i}(dt)) \right)^2 \end{aligned} \quad (8.2.45)$$

$$+ 32|\vartheta - \vartheta'|^2 n^{-5} \mathbb{E} \left(\sum_{\substack{i,k=1 \\ i \neq k}}^n \int_I |M_i(t)| L'_i(t) (N_k(dt) + \Lambda_{\vartheta_\circ,k}(dt)) \right)^2 \quad (8.2.46)$$

$$+ 32|\vartheta - \vartheta'|^2 n^{-5} \mathbb{E} \left(\sum_{\substack{i,k=1 \\ i \neq k}}^n \int_I |M_i(t)| L'_k(t) (N_k(dt) + \Lambda_{\vartheta_\circ,k}(dt)) \right)^2 \quad (8.2.47)$$

$$+ 32|\vartheta - \vartheta'|^2 n^{-5} \mathbb{E} \left(\sum_{\substack{i,k=1 \\ i \neq k}}^n \int_I |M_i(t)| L'_k(t) (N_i(dt) + \Lambda_{\vartheta_\circ,i}(dt)) \right)^2 \quad (8.2.48)$$

$$+ 32n^{-5} \mathbb{E} \left| \sum_{\substack{p,k,i=1 \\ p \neq k \neq i \neq p}}^n \int_I M_k(t) \left(\frac{\partial}{\partial \vartheta} \Lambda_{\vartheta,p}(t) - \frac{\partial}{\partial \vartheta'} \Lambda_{\vartheta',p}(t) \right) M_i(dt) \right|^2 \quad (8.2.49)$$

Nun schätzen wir jeden der Summanden (8.2.45) bis (8.2.49) einzeln ab. Für (8.2.45) bekommen wir den Ausdruck

$$\begin{aligned} & 32|\vartheta - \vartheta'|^2 n^{-5} \mathbb{E} \left(\sum_{i=1}^n \int_I |M_i(t)| L'_i(t) (N_i(dt) + \Lambda_{\vartheta_\circ,i}(dt)) \right)^2 \\ & = 32|\vartheta - \vartheta'|^2 n^{-5} \sum_{i,j=1}^n \mathbb{E} \left(\iint_{I \times I} |M_i(t) M_j(s)| L'_i(t) L'_j(s) \right. \\ & \quad \left. (N_i(dt) + \Lambda_{\vartheta_\circ,i}(dt)) (N_j(ds) + \Lambda_{\vartheta_\circ,j}(ds)) \right) \end{aligned}$$

Die n^2 Summanden sind jeweils endlich, wie mit Hölder aus Lemma 8.2.1 folgt. Außerdem unterliegen sie einer gemeinsamen Schranke, da alle überhaupt auftretenden Erwartungswerte aufgrund der i. i. d.-Annahmen bereits für $i, j = 1, 2$ angenommen werden. Der Faktor n^{-5} liefert die benötigte, von n unabhängige Schranke für (8.2.45).

Eine entsprechende Überlegung liefert obere Schranken für die Terme (8.2.46) bis (8.2.48). In diesen Fällen treten $O(n^4)$ gleichmäßig beschränkte Summanden und ein Faktor n^{-5} auf.

Für die Untersuchung von (8.2.49) setzen wir

$$U_{pki} := \int_I M_k(t) \left(\frac{\partial}{\partial \vartheta} \Lambda_{\vartheta,p}(t) - \frac{\partial}{\partial \vartheta'} \Lambda_{\vartheta',p}(t) \right) M_i(dt)$$

und wenden Lemma A.3.3 an. Dazu benötigen wir Lemma 8.2.1 und (8.2.2). Lemma A.3.3 liefert für (8.2.49) also:

$$\begin{aligned} & 32n^{-5} \mathbb{E} \left| \sum_{\substack{p,k,i=1 \\ p \neq k \neq i \neq p}}^n \int_I M_k(t) \left(\frac{\partial}{\partial \vartheta} \Lambda_{\vartheta,p}(t) - \frac{\partial}{\partial \vartheta'} \Lambda_{\vartheta',p}(t) \right) M_i(dt) \right|^2 \\ & \leq 2^{11} n^{-5} \sum_{\substack{p,q,k,i=1 \\ i \leq k \leq p, q \leq i}}^n \mathbb{E} \left(\iint_{I \times I} M_k(t) M_k(s) \left(\frac{\partial}{\partial \vartheta} \Lambda_{\vartheta,p}(t) - \frac{\partial}{\partial \vartheta'} \Lambda_{\vartheta',p}(t) \right)^\top \right. \\ & \quad \left. \left(\frac{\partial}{\partial \vartheta} \Lambda_{\vartheta,q}(s) - \frac{\partial}{\partial \vartheta'} \Lambda_{\vartheta',q}(s) \right) M_i(dt) M_i(ds) \right) \end{aligned}$$

Nun führen wir im Integranden Beträge und als integrierendes Maß eine monoton wachsende Funktion ein, wie wir dies auch bereits bei (8.2.45) bis (8.2.48) getan haben. Außerdem schätzen wir das Skalarprodukt gegen das Produkt der Beträge der beteiligten Vektoren ab:

$$\begin{aligned} [\dots] & \leq 2^{11} n^{-5} \sum_{\substack{p,q,k,i=1 \\ p,q \neq k \neq i \neq p,q}}^n \mathbb{E} \left(\iint_{I \times I} |M_k(t) M_k(s)| \left| \frac{\partial}{\partial \vartheta} \Lambda_{\vartheta,p}(t) - \frac{\partial}{\partial \vartheta'} \Lambda_{\vartheta',p}(t) \right| \right. \\ & \quad \left. \left| \frac{\partial}{\partial \vartheta} \Lambda_{\vartheta,q}(s) - \frac{\partial}{\partial \vartheta'} \Lambda_{\vartheta',q}(s) \right| (N_i(dt) + \Lambda_{\vartheta_\circ,i}(dt)) (N_j(ds) + \Lambda_{\vartheta_\circ,j}(ds)) \right) \end{aligned}$$

Daraufhin können wir auch hier die Lipschitz-Stetigkeit (3.4.2) anwenden:

$$\begin{aligned} [\dots] & \leq 2^{11} n^{-5} |\vartheta - \vartheta'|^2 \sum_{\substack{p,q,k,i=1 \\ p,q \neq k \neq i \neq p,q}}^n \mathbb{E} \left(\iint_{I \times I} |M_k(t) M_k(s)| L'_p(t) \right. \\ & \quad \left. L'_q(s) (N_i(dt) + \Lambda_{\vartheta_\circ,i}(dt)) (N_i(ds) + \Lambda_{\vartheta_\circ,i}(ds)) \right) \end{aligned}$$

Die $O(n^4)$ Summanden unterliegen einer gemeinsamen Schranke, der Faktor n^{-5} besorgt das Übrige. ■

8.2.4 Straffheit des γ -Prozesses

In diesem Abschnitt behandeln wir den Prozeß

$$\gamma_n(\vartheta) := n^{1/2} \int_I \bar{M}_n(t) \frac{\partial}{\partial \vartheta} \bar{\Lambda}_{\vartheta,n}(t) \bar{\Lambda}_{\vartheta_\circ,n}(dt) \quad (8.2.50)$$

Die Darstellung ähnelt der in Abschnitt 8.2.2.

8.2.11 Satz Der Prozeß $\gamma_n(\vartheta)$ erlaubt die Darstellung

$$\gamma_n(\vartheta) = n^{1/2} \int_I \bar{M}_n(t) E \frac{\partial}{\partial \vartheta} \Lambda_\vartheta(t) E \Lambda_\circ(dt) + o_P(1).$$

Auf kompakten Parameterteilmengen $\Theta_\circ \subset \Theta$ sind sowohl das führende Integral als auch das Restglied C -straff.

Beweis:

Wir führen zunächst einige Bezeichnungen ein:

$$\begin{aligned} \tilde{\gamma}_n(\vartheta) &:= n^{1/2} \int_I \bar{M}_n(t) E \frac{\partial}{\partial \vartheta} \Lambda_\vartheta(t) E \Lambda_\circ(dt) \\ \gamma_n^{(1)}(\vartheta) &:= n^{1/2} \int_I \bar{M}_n(t) \left(\frac{\partial}{\partial \vartheta} \bar{\Lambda}_{\vartheta,n}(t) - E \frac{\partial}{\partial \vartheta} \Lambda_\vartheta(t) \right) E \Lambda_\circ(dt) \\ \gamma_n^{(2)}(\vartheta) &:= n^{1/2} \int_I \bar{M}_n(t) \frac{\partial}{\partial \vartheta} \bar{\Lambda}_{\vartheta,n}(t) \left(\bar{\Lambda}_{\vartheta_\circ,n}(dt) - E \Lambda_\circ(dt) \right) \end{aligned}$$

Dann gilt

$$\gamma_n(\vartheta) = \tilde{\gamma}_n(\vartheta) + \gamma_n^{(1)}(\vartheta) + \gamma_n^{(2)}(\vartheta)$$

und die Aussage des Satzes folgt aus der Straffheit von $\tilde{\gamma}_n(\vartheta)$, $\gamma_n^{(1)}(\vartheta)$ und $\gamma_n^{(2)}(\vartheta)$ sowie

$$\gamma_n^{(i)}(\vartheta) = o_P(1) \quad \forall \vartheta \in \Theta, \quad i = 1, 2. \quad (8.2.51)$$

Wie bereits im Beweis des Satzes 8.2.4 genügt es auch hier, neben der punktweisen asymptotischen Vernachlässigbarkeit von $\tilde{\gamma}_n(\vartheta)$ in jedem Parameterwert $\vartheta \in \Theta$ die Momentenbedingungen

$$E |\tilde{\gamma}_n(\vartheta) - \tilde{\gamma}_n(\vartheta')|^2 \leq \text{const.} \cdot |\vartheta - \vartheta'|^2 \quad \forall \vartheta, \vartheta' \in \Theta_\circ \quad (8.2.52)$$

$$E |\gamma_n^{(i)}(\vartheta) - \gamma_n^{(i)}(\vartheta')|^2 \leq \text{const.} \cdot |\vartheta - \vartheta'|^2 \quad \forall \vartheta, \vartheta' \in \Theta_\circ \quad i = 1, 2 \quad (8.2.53)$$

auf kompakten Parameterteilmengen $\Theta_\circ \subset \Theta$ zu prüfen.

Die punktweise Konvergenz von $\tilde{\gamma}_n(\vartheta)$ folgt mit dem klassischen ZGWS direkt aus der i. i. d.-Darstellung

$$\tilde{\gamma}_n(\vartheta) = n^{-1/2} \sum_{k=1}^n \int_I M_k(t) E \frac{\partial}{\partial \vartheta} \Lambda_\vartheta(t) E \Lambda_\circ(dt).$$

Die Summanden sind dabei zentriert, wie eine Anwendung des Satzes von Fubini zeigt.

Wir zeigen (8.2.51) bis (8.2.53) in den folgenden Lemmata. Im Unterschied zur Situation bei $\alpha_n(\vartheta)$ wird hier nicht bzgl. eines Martingals integriert, so daß wir – wie schon bei der Untersuchung von $\beta_n(\vartheta)$ – nur mit Lemma A.2.3 arbeiten können. ■

Wir beginnen mit dem Nachweis von (8.2.51):

8.2.12 Lemma *Es gilt*

$$\gamma_n^{(i)}(\vartheta) = o_{\mathbb{P}}(1), \quad \forall \vartheta \in \Theta, \quad i = 1, 2.$$

Beweis:

Wie bereits in den Beweisen früherer Lemmata zeigen wir statt der stochastischen Konvergenz die Konvergenz in L_1 oder L_2 . Es gilt

$$\begin{aligned} \gamma_n^{(1)}(\vartheta) &= n^{-3/2} \sum_{i,k=1}^n \int_I M_k(t) \left(\frac{\partial}{\partial \vartheta} \Lambda_{\vartheta,i}(t) - \mathbb{E} \frac{\partial}{\partial \vartheta} \Lambda_{\vartheta}(t) \right) \mathbb{E} \Lambda_{\circ}(dt) \\ &= n^{-3/2} \sum_{i=1}^n \int_I M_i(t) \left(\frac{\partial}{\partial \vartheta} \Lambda_{\vartheta,i}(t) - \mathbb{E} \frac{\partial}{\partial \vartheta} \Lambda_{\vartheta}(t) \right) \mathbb{E} \Lambda_{\circ}(dt) \end{aligned} \quad (8.2.54)$$

$$+ n^{-3/2} \sum_{\substack{i,k=1 \\ i \neq k}}^n \int_I M_k(t) \left(\frac{\partial}{\partial \vartheta} \Lambda_{\vartheta,i}(t) - \mathbb{E} \frac{\partial}{\partial \vartheta} \Lambda_{\vartheta}(t) \right) \mathbb{E} \Lambda_{\circ}(dt) \quad (8.2.55)$$

Die L_1 -Konvergenz von (8.2.54) folgt nun aus:

$$\begin{aligned} \mathbb{E} \left| n^{-3/2} \sum_{i=1}^n \int_I M_i(t) \left(\frac{\partial}{\partial \vartheta} \Lambda_{\vartheta,i}(t) - \mathbb{E} \frac{\partial}{\partial \vartheta} \Lambda_{\vartheta}(t) \right) \mathbb{E} \Lambda_{\circ}(dt) \right| \\ \leq n^{-3/2} \sum_{i=1}^n \mathbb{E} \left| \int_I M_i(t) \left(\frac{\partial}{\partial \vartheta} \Lambda_{\vartheta,i}(t) - \mathbb{E} \frac{\partial}{\partial \vartheta} \Lambda_{\vartheta}(t) \right) \mathbb{E} \Lambda_{\circ}(dt) \right| \end{aligned}$$

Die auftretenden n Summanden haben wieder alle denselben endlichen Wert. Der Faktor $n^{-3/2}$ sorgt für die asymptotische Vernachlässigbarkeit des Gesamtausdrucks.

Auf das zweite Moment der Summe (8.2.55) wenden wir Lemma A.3.2 an. Dabei setzen wir:

$$U_{ki} := \int_I M_k(t) \left(\frac{\partial}{\partial \vartheta} \Lambda_{\vartheta,i}(t) - \mathbb{E} \frac{\partial}{\partial \vartheta} \Lambda_{\vartheta}(t) \right) \mathbb{E} \Lambda_{\circ}(dt)$$

Die Voraussetzungen des Lemmas, sind nach Lemma 8.2.1 und (8.2.4) erfüllt. Lemma A.3.2 liefert damit für das zweite Moment von (8.2.55) die Schranke:

$$\begin{aligned} n^{-3} \mathbb{E} \left| \sum_{\substack{i,k=1 \\ i \neq k}}^n \int_I M_k(t) \left(\frac{\partial}{\partial \vartheta} \Lambda_{\vartheta,i}(t) - \mathbb{E} \frac{\partial}{\partial \vartheta} \Lambda_{\vartheta}(t) \right) \mathbb{E} \Lambda_{\circ}(dt) \right|^2 \\ \leq 2n^{-3} \sum_{\substack{i,k=1 \\ i \neq k}}^n \mathbb{E} \left| \int_I M_k(t) \left(\frac{\partial}{\partial \vartheta} \Lambda_{\vartheta,i}(t) - \mathbb{E} \frac{\partial}{\partial \vartheta} \Lambda_{\vartheta}(t) \right) \mathbb{E} \Lambda_{\circ}(dt) \right|^2 \end{aligned}$$

Die $O(n^2)$ Summanden haben eine gemeinsame Schranke, der Faktor n^{-3} sorgt für asymptotische Vernachlässigbarkeit. ■

Nun zeigen wir (8.2.52):

8.2.13 Lemma Auf kompakten Parameterteilmengen $\Theta_\circ \subset \Theta$ gilt

$$\mathbb{E} |\tilde{\gamma}_n(\vartheta) - \tilde{\gamma}_n(\vartheta')|^2 \leq \text{const.} \cdot |\vartheta - \vartheta'|^2, \quad \forall \vartheta, \vartheta' \in \Theta_\circ.$$

Beweis:

Aus der Darstellung:

$$\tilde{\gamma}_n(\vartheta) - \tilde{\gamma}_n(\vartheta') = n^{-1/2} \sum_{i=1}^n \int_I M_i(t) \mathbb{E} \left(\frac{\partial}{\partial \vartheta} \Lambda_\vartheta(t) - \frac{\partial}{\partial \vartheta} \Lambda_{\vartheta'}(t) \right) \mathbb{E} \Lambda_\circ(dt)$$

erhalten wir für das zweite Moment:

$$\begin{aligned} & \mathbb{E} |\tilde{\gamma}_n(\vartheta) - \tilde{\gamma}_n(\vartheta')|^2 \\ &= n^{-1} \mathbb{E} \left(\sum_{i,j=1}^n \iint_{I \times I} M_i(t) M_j(s) \mathbb{E} \left(\frac{\partial}{\partial \vartheta} \Lambda_\vartheta(t) - \frac{\partial}{\partial \vartheta} \Lambda_{\vartheta'}(t) \right)^\top \right. \\ & \qquad \qquad \qquad \mathbb{E} \left(\frac{\partial}{\partial \vartheta} \Lambda_\vartheta(s) - \frac{\partial}{\partial \vartheta} \Lambda_{\vartheta'}(s) \right) \mathbb{E} \Lambda_\circ(dt) \mathbb{E} \Lambda_\circ(ds) \Big) \\ &= n^{-1} \sum_{i,j=1}^n \iint_{I \times I} \mathbb{E} (M_i(t) M_j(s)) \mathbb{E} \left(\frac{\partial}{\partial \vartheta} \Lambda_\vartheta(t) - \frac{\partial}{\partial \vartheta} \Lambda_{\vartheta'}(t) \right)^\top \\ & \qquad \qquad \qquad \mathbb{E} \left(\frac{\partial}{\partial \vartheta} \Lambda_\vartheta(s) - \frac{\partial}{\partial \vartheta} \Lambda_{\vartheta'}(s) \right) \mathbb{E} \Lambda_\circ(dt) \mathbb{E} \Lambda_\circ(ds) \end{aligned}$$

Die Summanden mit $i \neq j$ verschwinden, da $M_i(\cdot)$ und $M_j(\cdot)$ unabhängig und jeweils zentriert sind. Wir erhalten:

$$\begin{aligned} & \mathbb{E} |\tilde{\gamma}_n(\vartheta) - \tilde{\gamma}_n(\vartheta')|^2 \\ &= n^{-1} \sum_{i=1}^n \iint_{I \times I} \mathbb{E} (M_i(t) M_i(s)) \mathbb{E} \left(\frac{\partial}{\partial \vartheta} \Lambda_\vartheta(t) - \frac{\partial}{\partial \vartheta} \Lambda_{\vartheta'}(t) \right)^\top \\ & \qquad \qquad \qquad \mathbb{E} \left(\frac{\partial}{\partial \vartheta} \Lambda_\vartheta(s) - \frac{\partial}{\partial \vartheta} \Lambda_{\vartheta'}(s) \right) \mathbb{E} \Lambda_\circ(dt) \mathbb{E} \Lambda_\circ(ds) \end{aligned}$$

Da die Prozesse $M_i(\cdot)$ i. i. d. sind, vereinfacht sich dieser Ausdruck zu:

$$\begin{aligned} [\dots] &= \iint_{I \times I} \mathbb{E} (M(t) M(s)) \mathbb{E} \left(\frac{\partial}{\partial \vartheta} \Lambda_\vartheta(t) - \frac{\partial}{\partial \vartheta} \Lambda_{\vartheta'}(t) \right)^\top \\ & \qquad \qquad \qquad \mathbb{E} \left(\frac{\partial}{\partial \vartheta} \Lambda_\vartheta(s) - \frac{\partial}{\partial \vartheta} \Lambda_{\vartheta'}(s) \right) \mathbb{E} \Lambda_\circ(dt) \mathbb{E} \Lambda_\circ(ds) \\ &\leq \iint_{I \times I} \mathbb{E} |M(t) M(s)| \mathbb{E} \left| \frac{\partial}{\partial \vartheta} \Lambda_\vartheta(t) - \frac{\partial}{\partial \vartheta} \Lambda_{\vartheta'}(t) \right| \\ & \qquad \qquad \qquad \mathbb{E} \left| \frac{\partial}{\partial \vartheta} \Lambda_\vartheta(s) - \frac{\partial}{\partial \vartheta} \Lambda_{\vartheta'}(s) \right| \mathbb{E} \Lambda_\circ(dt) \mathbb{E} \Lambda_\circ(ds) \end{aligned}$$

Nun setzen wir die Lipschitz-Abschätzung gemäß (3.4.2) ein:

$$[\dots] \leq |\vartheta - \vartheta'|^2 \iint_{I \times I} \mathbb{E} |M(t) M(s)| \mathbb{E} |L'(t)| \mathbb{E} |L'(s)| \mathbb{E} \Lambda_\circ(dt) \mathbb{E} \Lambda_\circ(ds)$$

Wegen Lemma 8.2.1 ist das Integral endlich. Da es zudem deterministisch ist und nicht von n abhängt, haben wir eine Darstellung gefunden, wie sie in der Aussage des Lemmas gefordert wird. ■

Wir beenden den Beweis des Satzes 8.2.11 durch den Nachweis von (8.2.53):

8.2.14 Lemma *Auf kompakten Parametermengen $\Theta_o \subset \Theta$ gilt*

$$\mathbb{E} |\gamma_n^{(i)}(\vartheta) - \gamma_n^{(i)}(\vartheta')|^2 \leq \text{const.} \cdot |\vartheta - \vartheta'|^2, \quad \forall \vartheta, \vartheta' \in \Theta_o, \quad i = 1, 2.$$

Beweis:

Wie im Beweis des Lemmas 8.2.10 beginnen wir mit einer Darstellung:

$$\begin{aligned} & \gamma_n^{(1)}(\vartheta) - \gamma_n^{(1)}(\vartheta') \\ &= n^{-3/2} \sum_{k,i=1}^n \int_I M_i(t) \\ & \quad \left(\frac{\partial}{\partial \vartheta} \Lambda_{\vartheta,k}(t) - \mathbb{E} \frac{\partial}{\partial \vartheta} \Lambda_{\vartheta}(t) - \frac{\partial}{\partial \vartheta} \Lambda_{\vartheta',k}(t) + \mathbb{E} \frac{\partial}{\partial \vartheta} \Lambda_{\vartheta'}(t) \right) \mathbb{E} \Lambda_o(dt) \\ &= n^{-3/2} \sum_{i=1}^n \int_I M_i(t) \\ & \quad \left(\frac{\partial}{\partial \vartheta} \Lambda_{\vartheta,i}(t) - \mathbb{E} \frac{\partial}{\partial \vartheta} \Lambda_{\vartheta}(t) - \frac{\partial}{\partial \vartheta} \Lambda_{\vartheta',i}(t) + \mathbb{E} \frac{\partial}{\partial \vartheta} \Lambda_{\vartheta'}(t) \right) \mathbb{E} \Lambda_o(dt) \\ &+ n^{-3/2} \sum_{\substack{k,i=1 \\ k \neq i}}^n \int_I M_i(t) \\ & \quad \left(\frac{\partial}{\partial \vartheta} \Lambda_{\vartheta,k}(t) - \mathbb{E} \frac{\partial}{\partial \vartheta} \Lambda_{\vartheta}(t) - \frac{\partial}{\partial \vartheta} \Lambda_{\vartheta',k}(t) + \mathbb{E} \frac{\partial}{\partial \vartheta} \Lambda_{\vartheta'}(t) \right) \mathbb{E} \Lambda_o(dt) \end{aligned}$$

Für das zweite Moment erhalten wir so:

$$\begin{aligned} & \mathbb{E} |\gamma_n^{(1)}(\vartheta) - \gamma_n^{(1)}(\vartheta')|^2 \\ &= n^{-3} \mathbb{E} \left| \sum_{k,i=1}^n \int_I M_i(t) \right. \\ & \quad \left. \left(\frac{\partial}{\partial \vartheta} \Lambda_{\vartheta,k}(t) - \mathbb{E} \frac{\partial}{\partial \vartheta} \Lambda_{\vartheta}(t) - \frac{\partial}{\partial \vartheta} \Lambda_{\vartheta',k}(t) + \mathbb{E} \frac{\partial}{\partial \vartheta} \Lambda_{\vartheta'}(t) \right) \mathbb{E} \Lambda_o(dt) \right|^2 \\ &\leq 2n^{-3} \mathbb{E} \left| \sum_{i=1}^n \int_I M_i(t) \right. \\ & \quad \left. \left(\frac{\partial}{\partial \vartheta} \Lambda_{\vartheta,i}(t) - \mathbb{E} \frac{\partial}{\partial \vartheta} \Lambda_{\vartheta}(t) - \frac{\partial}{\partial \vartheta} \Lambda_{\vartheta',i}(t) + \mathbb{E} \frac{\partial}{\partial \vartheta} \Lambda_{\vartheta'}(t) \right) \mathbb{E} \Lambda_o(dt) \right|^2 \\ &+ 2n^{-3} \mathbb{E} \left| \sum_{\substack{k,i=1 \\ k \neq i}}^n \int_I M_i(t) \right. \\ & \quad \left. \left(\frac{\partial}{\partial \vartheta} \Lambda_{\vartheta,k}(t) - \mathbb{E} \frac{\partial}{\partial \vartheta} \Lambda_{\vartheta}(t) - \frac{\partial}{\partial \vartheta} \Lambda_{\vartheta',k}(t) + \mathbb{E} \frac{\partial}{\partial \vartheta} \Lambda_{\vartheta'}(t) \right) \mathbb{E} \Lambda_o(dt) \right|^2 \end{aligned}$$

Zum Beweis des Lemmas für den ersten Restterm müssen wir nun also

$$\begin{aligned}
& n^{-3} \mathbb{E} \left| \sum_{i=1}^n \int_I M_i(t) \right. \\
& \quad \left. \left(\frac{\partial}{\partial \vartheta} \Lambda_{\vartheta, i}(t) - \mathbb{E} \frac{\partial}{\partial \vartheta} \Lambda_{\vartheta}(t) - \frac{\partial}{\partial \vartheta} \Lambda_{\vartheta', i}(t) + \mathbb{E} \frac{\partial}{\partial \vartheta} \Lambda_{\vartheta'}(t) \right) \mathbb{E} \Lambda_{\circ}(dt) \right|^2 \\
& \leq \text{const.} \cdot |\vartheta - \vartheta'|^2 \quad (8.2.56)
\end{aligned}$$

und

$$\begin{aligned}
& n^{-3} \mathbb{E} \left| \sum_{\substack{k, i=1 \\ k \neq i}}^n \int_I M_i(t) \right. \\
& \quad \left. \left(\frac{\partial}{\partial \vartheta} \Lambda_{\vartheta, k}(t) - \mathbb{E} \frac{\partial}{\partial \vartheta} \Lambda_{\vartheta}(t) - \frac{\partial}{\partial \vartheta} \Lambda_{\vartheta', k}(t) + \mathbb{E} \frac{\partial}{\partial \vartheta} \Lambda_{\vartheta'}(t) \right) \mathbb{E} \Lambda_{\circ}(dt) \right|^2 \\
& \leq \text{const.} \cdot |\vartheta - \vartheta'|^2 \quad (8.2.57)
\end{aligned}$$

zeigen. Wir beginnen mit dem Nachweis von (8.2.56), wobei wir im Integranden Beträge einführen und die Dreiecksungleichung benutzen. Im Anschluß können wir dann die Lipschitz-Stetigkeit gemäß (3.4.2) anwenden:

$$\begin{aligned}
& n^{-3} \mathbb{E} \left| \sum_{i=1}^n \int_I M_i(t) \right. \\
& \quad \left. \left(\frac{\partial}{\partial \vartheta} \Lambda_{\vartheta, i}(t) - \mathbb{E} \frac{\partial}{\partial \vartheta} \Lambda_{\vartheta}(t) - \frac{\partial}{\partial \vartheta} \Lambda_{\vartheta', i}(t) + \mathbb{E} \frac{\partial}{\partial \vartheta} \Lambda_{\vartheta'}(t) \right) \mathbb{E} \Lambda_{\circ}(dt) \right|^2 \\
& \leq n^{-3} \mathbb{E} \left(\sum_{i=1}^n \int_I |M_i(t)| \right. \\
& \quad \left. \left(\left| \frac{\partial}{\partial \vartheta} \Lambda_{\vartheta, i}(t) - \frac{\partial}{\partial \vartheta} \Lambda_{\vartheta', i}(t) \right| + \left| \mathbb{E} \frac{\partial}{\partial \vartheta} \Lambda_{\vartheta}(t) - \mathbb{E} \frac{\partial}{\partial \vartheta} \Lambda_{\vartheta'}(t) \right| \right) \mathbb{E} \Lambda_{\circ}(dt) \right)^2 \\
& \leq n^{-3} |\vartheta - \vartheta'|^2 \mathbb{E} \left(\sum_{i=1}^n \int_I |M_i(t)| (L'_i(t) + \mathbb{E} L'(t)) \mathbb{E} \Lambda_{\circ}(dt) \right)^2 \\
& = n^{-3} |\vartheta - \vartheta'|^2 \sum_{i, j=1}^n \mathbb{E} \left(\iint_{I \times I} |M_i(t) M_j(s)| (L'_i(t) + \mathbb{E} L'(t)) (L'_j(s) + \mathbb{E} L'(s)) \right. \\
& \quad \left. \mathbb{E} \Lambda_{\circ}(dt) \mathbb{E} \Lambda_{\circ}(ds) \right)
\end{aligned}$$

Die n^2 Summanden sind wegen Lemma 8.2.1 endlich. Ihre Beträge sind außerdem gleichmäßig beschränkt, da die beteiligten Prozesse i. i. d. in i und j sind und somit alle für $i, j = 1, \dots, n$ auftretenden Erwartungswerte bereits unter den vier Werten für $i, j = 1, 2$ auftauchen. Der Faktor n^{-3} liefert die gleichmäßige Schranke in n .

Wir wenden uns (8.2.57) zu. Es ist der Ausdruck

$$n^{-3} \mathbb{E} \left| \sum_{\substack{k,i=1 \\ k \neq i}}^n \int_I M_i(t) \left(\frac{\partial}{\partial \vartheta} \Lambda_{\vartheta,k}(t) - \mathbb{E} \frac{\partial}{\partial \vartheta} \Lambda_{\vartheta}(t) - \frac{\partial}{\partial \vartheta} \Lambda_{\vartheta',k}(t) + \mathbb{E} \frac{\partial}{\partial \vartheta} \Lambda_{\vartheta'}(t) \right) \mathbb{E} \Lambda_{\circ}(dt) \right|^2$$

zu untersuchen. Wir wenden Lemma A.3.2 mit

$$U_{ki} := \int_I M_i(t) \left(\frac{\partial}{\partial \vartheta} \Lambda_{\vartheta,k}(t) - \mathbb{E} \frac{\partial}{\partial \vartheta} \Lambda_{\vartheta}(t) - \frac{\partial}{\partial \vartheta} \Lambda_{\vartheta',k}(t) + \mathbb{E} \frac{\partial}{\partial \vartheta} \Lambda_{\vartheta'}(t) \right) \mathbb{E} \Lambda_{\circ}(dt)$$

an. Dazu benötigen wir Lemma 8.2.1 und (8.2.4). Lemma A.3.2 liefert für (8.2.57) damit:

$$\begin{aligned} & n^{-3} \mathbb{E} \left| \sum_{\substack{k,i=1 \\ k \neq i}}^n \int_I M_i(t) \left(\frac{\partial}{\partial \vartheta} \Lambda_{\vartheta,k}(t) - \mathbb{E} \frac{\partial}{\partial \vartheta} \Lambda_{\vartheta}(t) - \frac{\partial}{\partial \vartheta} \Lambda_{\vartheta',k}(t) + \mathbb{E} \frac{\partial}{\partial \vartheta} \Lambda_{\vartheta'}(t) \right) \mathbb{E} \Lambda_{\circ}(dt) \right|^2 \\ & \leq 2n^{-3} \sum_{\substack{k,i=1 \\ k \neq i}}^n \mathbb{E} \left| \int_I M_i(t) \left(\frac{\partial}{\partial \vartheta} \Lambda_{\vartheta,k}(t) - \mathbb{E} \frac{\partial}{\partial \vartheta} \Lambda_{\vartheta}(t) - \frac{\partial}{\partial \vartheta} \Lambda_{\vartheta',k}(t) + \mathbb{E} \frac{\partial}{\partial \vartheta} \Lambda_{\vartheta'}(t) \right) \mathbb{E} \Lambda_{\circ}(dt) \right|^2 \end{aligned}$$

Wie im Nachweis von (8.2.56) setzen wir wieder Beträge im Integranden ein und bringen im Anschluß die Dreiecksungleichung und die Lipschitz-Stetigkeit gemäß (3.4.2) ein:

$$\begin{aligned} [\dots] & \leq 2n^{-3} \sum_{\substack{k,i=1 \\ k \neq i}}^n \mathbb{E} \left(\int_I |M_i(t)| \left(\left| \frac{\partial}{\partial \vartheta} \Lambda_{\vartheta,k}(t) - \frac{\partial}{\partial \vartheta} \Lambda_{\vartheta',k}(t) \right| + \left| \mathbb{E} \frac{\partial}{\partial \vartheta} \Lambda_{\vartheta}(t) - \mathbb{E} \frac{\partial}{\partial \vartheta} \Lambda_{\vartheta'}(t) \right| \right) \mathbb{E} \Lambda_{\circ}(dt) \right)^2 \\ & \leq 2n^{-3} |\vartheta - \vartheta'|^2 \sum_{\substack{k,i=1 \\ k \neq i}}^n \mathbb{E} \left(\int_I |M_i(t)| (L'_k(t) + \mathbb{E} L'(t)) \mathbb{E} \Lambda_{\circ}(dt) \right)^2 \end{aligned}$$

Die gleichmäßig beschränkten $O(n^2)$ Summanden und der Faktor n^{-3} liefern die Aussage des Lemmas für den ersten Restterm.

Nun betrachten wir den zweiten Restterm:

$$\begin{aligned} & \gamma_n^{(2)}(\vartheta) - \gamma_n^{(2)}(\vartheta') \\ & = n^{-5/2} \sum_{p,k,i=1}^n \int_I M_k(t) \left(\frac{\partial}{\partial \vartheta} \Lambda_{\vartheta,p}(t) - \frac{\partial}{\partial \vartheta} \Lambda_{\vartheta',p}(t) \right) (\Lambda_{\vartheta_{\circ},i}(dt) - \mathbb{E} \Lambda_{\circ}(dt)) \end{aligned}$$

Dieser Ausdruck entspricht dem Term $\beta_n(\vartheta) - \beta_n(\vartheta')$, wenn wir das integrierende Maß $M_i(\cdot)$ durch $\Lambda_{\vartheta_o, i}(\cdot) - \mathbb{E} \Lambda_o(\cdot)$ ersetzen. Wir erhalten deshalb, entsprechend (8.2.45) bis (8.2.49):

$$\begin{aligned}
 & \mathbb{E} |\gamma_n^{(2)}(\vartheta) - \gamma_n^{(2)}(\vartheta')|^2 \\
 & \leq 32|\vartheta - \vartheta'|^2 n^{-5} \mathbb{E} \left(\sum_{i=1}^n \int_I |M_i(t)| L'_i(t) (\Lambda_{\vartheta_o, i}(dt) + \mathbb{E} \Lambda_o(dt)) \right)^2 \\
 & + 32|\vartheta - \vartheta'|^2 n^{-5} \mathbb{E} \left(\sum_{\substack{i,k=1 \\ i \neq k}}^n \int_I |M_i(t)| L'_i(t) (\Lambda_{\vartheta_o, k}(dt) + \mathbb{E} \Lambda_o(dt)) \right)^2 \\
 & + 32|\vartheta - \vartheta'|^2 n^{-5} \mathbb{E} \left(\sum_{\substack{i,k=1 \\ i \neq k}}^n \int_I |M_i(t)| L'_k(t) (\Lambda_{\vartheta_o, k}(dt) + \mathbb{E} \Lambda_o(dt)) \right)^2 \\
 & + 32|\vartheta - \vartheta'|^2 n^{-5} \mathbb{E} \left(\sum_{\substack{i,k=1 \\ i \neq k}}^n \int_I |M_i(t)| L'_k(t) (\Lambda_{\vartheta_o, i}(dt) + \mathbb{E} \Lambda_o(dt)) \right)^2 \\
 & + 32n^{-5} \mathbb{E} \left| \sum_{\substack{p,k,i=1 \\ p \neq k \neq i \neq p}}^n \int_I M_k(t) \left(\frac{\partial}{\partial \vartheta} \Lambda_{\vartheta, p}(t) - \frac{\partial}{\partial \vartheta} \Lambda_{\vartheta', p}(t) \right) (\Lambda_{\vartheta_o, i}(dt) - \mathbb{E} \Lambda_o(dt)) \right|^2
 \end{aligned}$$

Auch das weitere Vorgehen ist analog zum Beweis des Lemmas 8.2.10: Die Summanden der ersten vier Terme sind wegen Lemma 8.2.1 endlich sowie dem Betrage nach gleichmäßig beschränkt und werden deshalb durch den Faktor n^{-5} dominiert. Bevor wir die entsprechende Argumentation auf den fünften Term anwenden, reduzieren wir die Anzahl der Summanden durch Anwendung des Lemmas A.3.3 mit

$$U_{pki} := \int_I M_k(t) \left(\frac{\partial}{\partial \vartheta} \Lambda_{\vartheta, p}(t) - \frac{\partial}{\partial \vartheta} \Lambda_{\vartheta', p}(t) \right) (\Lambda_{\vartheta_o, i}(dt) - \mathbb{E} \Lambda_o(dt))$$

Die Voraussetzungen gelten nach Lemma 8.2.1 und (8.2.2). Lemma A.3.3 liefert also wie bereits im Beweis des Lemmas 8.2.10:

$$\begin{aligned}
 & 32n^{-5} \mathbb{E} \left| \sum_{\substack{p,k,i=1 \\ p \neq k \neq i \neq p}}^n \int_I M_k(t) \left(\frac{\partial}{\partial \vartheta} \Lambda_{\vartheta, p}(t) - \frac{\partial}{\partial \vartheta} \Lambda_{\vartheta', p}(t) \right) (\Lambda_{\vartheta_o, i}(dt) - \mathbb{E} \Lambda_o(dt)) \right|^2 \\
 & \leq 2^{11} n^{-5} \sum_{\substack{p,q,k,i=1 \\ i \leq k \leq p, q \leq i}}^n \mathbb{E} \left(\iint_{I \times I} M_k(t) M_k(s) \right. \\
 & \quad \left. \left(\frac{\partial}{\partial \vartheta} \Lambda_{\vartheta, p}(t) - \frac{\partial}{\partial \vartheta} \Lambda_{\vartheta', p}(t) \right)^\top \left(\frac{\partial}{\partial \vartheta} \Lambda_{\vartheta, q}(s) - \frac{\partial}{\partial \vartheta} \Lambda_{\vartheta', q}(s) \right) \right. \\
 & \quad \left. (\Lambda_{\vartheta_o, i}(dt) - \mathbb{E} \Lambda_o(dt)) (\Lambda_{\vartheta_o, i}(ds) - \mathbb{E} \Lambda_o(ds)) \right) \\
 & \leq 2^{11} n^{-5} |\vartheta - \vartheta'|^2 \sum_{\substack{p,q,k,i=1 \\ i \leq k \leq p, q \leq i}}^n \mathbb{E} \left(\iint_{I \times I} |M_k(t) M_k(s)| L'_p(t) \right. \\
 & \quad \left. L'_q(s) (\Lambda_{\vartheta_o, i}(dt) + \mathbb{E} \Lambda_o(dt)) (\Lambda_{\vartheta_o, i}(ds) + \mathbb{E} \Lambda_o(ds)) \right)
 \end{aligned}$$

Die $O(n^4)$ gleichmäßig beschränkten Summanden werden von n^{-5} unterhalb einer Konstante gehalten. ■

8.2.5 Straffheit der standardisierenden Matrix Φ_n

Wir schließen die Vorbereitungen zum Beweis der asymptotischen Normalität des KQS mit dem Nachweis ab, daß auch die standardisierende Matrix $\Phi_n(\cdot)$ C -straff ist:

8.2.15 Satz Die standardisierende Matrix $\Phi_n(\vartheta)$ hat die Darstellung

$$\Phi_n(\vartheta) = \Phi_\circ(\vartheta) + o_P(1).$$

Sie ist auf kompakten Teilmengen $\Theta_\circ \subset \Theta$ des Parameterraumes C -straff.

Beweis:

Es gilt:

$$\begin{aligned} \Phi_n(\vartheta_\circ) &= \frac{\partial}{\partial \vartheta} \int_I (\bar{\Lambda}_{\vartheta,n}(t) - \bar{\Lambda}_{\vartheta_\circ,n}(t)) \frac{\partial}{\partial \vartheta} \bar{\Lambda}_{\vartheta,n}(t) \bar{\Lambda}_{\vartheta_\circ,n}(dt) \\ &= \int_I (\bar{\Lambda}_{\vartheta,n}(t) - \bar{\Lambda}_{\vartheta_\circ,n}(t)) \frac{\partial^2}{\partial \vartheta^2} \bar{\Lambda}_{\vartheta,n}(t) \bar{\Lambda}_{\vartheta_\circ,n}(dt) \end{aligned} \tag{8.2.58}$$

$$+ \int_I \left| \frac{\partial}{\partial \vartheta} \bar{\Lambda}_{\vartheta,n}(t) \right|^2 \bar{\Lambda}_{\vartheta_\circ,n}(dt) \tag{8.2.59}$$

Wir können für beide Summanden die C -Straffheit einzeln zeigen. Die Beweise verlaufen wie in den vorherigen Abschnitten. (8.2.58) zerlegen wir in einen führenden und drei Restterme:

$$\begin{aligned} &\int_I (\bar{\Lambda}_{\vartheta,n}(t) - \bar{\Lambda}_{\vartheta_\circ,n}(t)) \frac{\partial^2}{\partial \vartheta^2} \bar{\Lambda}_{\vartheta,n}(t) \bar{\Lambda}_{\vartheta_\circ,n}(dt) \\ &= \int_I (E \Lambda_\vartheta(t) - E \Lambda_\circ(t)) E \frac{\partial^2}{\partial \vartheta^2} \Lambda_\vartheta(t) E \Lambda_\circ(dt) \\ &+ \int_I (\bar{\Lambda}_{\vartheta,n}(t) - E \Lambda_\vartheta(t) - \bar{\Lambda}_{\vartheta_\circ,n}(t) + E \Lambda_\circ(t)) E \frac{\partial^2}{\partial \vartheta^2} \Lambda_\vartheta(t) E \Lambda_\circ(dt) \\ &+ \int_I (\bar{\Lambda}_{\vartheta,n}(t) - \bar{\Lambda}_{\vartheta_\circ,n}(t)) \left(\frac{\partial^2}{\partial \vartheta^2} \bar{\Lambda}_{\vartheta,n}(t) - E \frac{\partial^2}{\partial \vartheta^2} \Lambda_\vartheta(t) \right) \bar{\Lambda}_{\vartheta_\circ,n}(dt) \\ &+ \int_I (\bar{\Lambda}_{\vartheta,n}(t) - \bar{\Lambda}_{\vartheta_\circ,n}(t)) \frac{\partial^2}{\partial \vartheta^2} \bar{\Lambda}_{\vartheta,n}(t) (\bar{\Lambda}_{\vartheta_\circ,n}(dt) - E \Lambda_\circ(dt)) \end{aligned}$$

Der führende Term ist deterministisch und hängt nicht von n ab. Seine C -Straffheit ist ohne weiteres einsichtig.

Der erste Restterm entspricht fast den ersten beiden Restgliedern $\alpha_n^{(1)}(\vartheta)$, $\alpha_n^{(2)}(\vartheta)$ aus (8.2.10b). Er kann analog behandelt werden.

Der zweite Restterm erlaubt eine Untersuchung, die weitestgehend der des dritten Restglieds $\alpha_n^{(3)}(\vartheta)$ aus (8.2.10b) gleicht.

Der dritte Restterm schließlich hat einen Aufbau wie $\gamma_n^{(2)}(\vartheta)$ im Beweis des Satzes 8.2.11 und kann ebenso bearbeitet werden.

Ganz entsprechende Schlüsse sind beim Nachweis der C -Straffheit von (8.2.59) möglich. Eine geeignete Zerlegung ist:

$$\begin{aligned}
& \int_I \left| \frac{\partial}{\partial \vartheta} \bar{\Lambda}_{\vartheta, n}(t) \right|^2 \bar{\Lambda}_{\vartheta_0, n}(dt) \\
&= \int_I \left| \mathbb{E} \frac{\partial}{\partial \vartheta} \Lambda_{\vartheta}(t) \right|^2 \mathbb{E} \Lambda_{\circ}(dt) \\
&+ \int_I \left(\mathbb{E} \frac{\partial}{\partial \vartheta} \Lambda_{\vartheta}(t) - \frac{\partial}{\partial \vartheta} \bar{\Lambda}_{\vartheta, n}(t) \right)^{\top} \mathbb{E} \frac{\partial}{\partial \vartheta} \Lambda_{\vartheta}(t) \bar{\Lambda}_{\vartheta_0, n}(dt) \\
&+ \int_I \frac{\partial}{\partial \vartheta} \bar{\Lambda}_{\vartheta, n}^{\top}(t) \left(\mathbb{E} \frac{\partial}{\partial \vartheta} \Lambda_{\vartheta}(t) - \frac{\partial}{\partial \vartheta} \bar{\Lambda}_{\vartheta, n}(t) \right) \bar{\Lambda}_{\vartheta_0, n}(dt) \\
&+ \int_I \left| \frac{\partial}{\partial \vartheta} \bar{\Lambda}_{\vartheta, n}(t) \right|^2 \left(\bar{\Lambda}_{\vartheta_0, n}(dt) - \mathbb{E} \Lambda_{\circ}(dt) \right)
\end{aligned}$$

■

8.3 Beweisschritte

Beweis von Theorem 3.5.1:

Der Beweis zur asymptotischen Normalität des KQS erfolgt in mehreren Schritten. Wir werden zunächst zeigen, daß für eine geeignete Hilfsfolge $\tilde{\vartheta}_n \xrightarrow{\text{f. s.}} \vartheta_0$ gilt:

$$\begin{aligned}
& n^{1/2} \Phi_n(\tilde{\vartheta}_n) \cdot (\vartheta_n^* - \vartheta_0) \\
&= n^{1/2} \frac{\partial}{\partial \vartheta} \int_I \left(\bar{\Lambda}_{\vartheta, n}(t) - \bar{\Lambda}_{\vartheta_0, n}(t) \right) \frac{\partial}{\partial \vartheta} \bar{\Lambda}_{\vartheta, n}(t) \bar{\Lambda}_{\vartheta_0, n}(dt) \Big|_{\vartheta=\tilde{\vartheta}_n} \cdot (\vartheta_n^* - \vartheta_0) \\
&= n^{1/2} \int_I \left(\bar{\Lambda}_{\vartheta, n}(t) - \bar{\Lambda}_{\vartheta_0, n}(t) \right) \frac{\partial}{\partial \vartheta} \bar{\Lambda}_{\vartheta, n}(t) \bar{\Lambda}_{\vartheta_0, n}(dt) \Big|_{\vartheta=\vartheta_n^*} \quad (8.3.1)
\end{aligned}$$

$$= n^{1/2} \int_I \left(\bar{\Lambda}_{\vartheta, n}(t) - \bar{\Lambda}_{\vartheta_0, n}(t) \right) \frac{\partial}{\partial \vartheta} \bar{\Lambda}_{\vartheta, n}(t) \bar{N}_n(dt) \Big|_{\vartheta=\vartheta_n^*} + o_{\mathbb{P}}(1) \quad (8.3.2)$$

$$= n^{1/2} \int_I \left(\bar{N}_n(t) - \bar{\Lambda}_{\vartheta_0, n}(t) \right) \frac{\partial}{\partial \vartheta} \bar{\Lambda}_{\vartheta, n}(t) \Big|_{\vartheta=\vartheta_n^*} \bar{N}_n(dt) + o_{\mathbb{P}}(1) \quad (8.3.3)$$

$$= n^{1/2} \int_I \left(\bar{N}_n(t) - \bar{\Lambda}_{\vartheta_0, n}(t) \right) \frac{\partial}{\partial \vartheta} \bar{\Lambda}_{\vartheta, n}(t) \Big|_{\vartheta=\vartheta_n^*} \bar{\Lambda}_{\vartheta_0, n}(dt) + o_{\mathbb{P}}(1) \quad (8.3.4)$$

$$= n^{1/2} \int_I (\bar{N}_n(t) - \bar{\Lambda}_{\vartheta_\circ, n}(t)) \frac{\partial}{\partial \vartheta} \bar{\Lambda}_{\vartheta, n}(t) \Big|_{\vartheta=\vartheta_\circ} \bar{\Lambda}_{\vartheta_\circ, n}(dt) + o_{\mathbb{P}}(1) \quad (8.3.5)$$

$$= n^{1/2} \iint_{I \times [x, t_{\max}]} \frac{\partial}{\partial \vartheta} \bar{\Lambda}_{\vartheta_\circ, n}(t) \bar{\Lambda}_{\vartheta_\circ, n}(dt) (\bar{N}_n(dx) - \bar{\Lambda}_{\vartheta_\circ, n}(dx)) + o_{\mathbb{P}}(1) \quad (8.3.6)$$

$$= n^{1/2} \iint_{I \times [x, t_{\max}]} \mathbb{E} \frac{\partial}{\partial \vartheta} \Lambda_{\vartheta_\circ}(t) \mathbb{E} \Lambda_\circ(dt) (\bar{N}_n(dx) - \bar{\Lambda}_{\vartheta_\circ, n}(dx)) + o_{\mathbb{P}}(1) \quad (8.3.7)$$

Der Nachweis der Schritte (8.3.1) bis (8.3.7) erfolgt direkt im Anschluß durch die Lemmata 8.3.1 bis 8.3.7. Das innere Integral in (8.3.7) ist deterministisch, wir haben also die folgende i. i. d.-Darstellung erhalten:

$$[\dots] = n^{-1/2} \sum_{k=1}^n \int_I \left(\int_{[x, t_{\max}]} \mathbb{E} \frac{\partial}{\partial \vartheta} \Lambda_{\vartheta_\circ}(t) \mathbb{E} \Lambda_\circ(dt) \right) (N_k(dx) - \Lambda_{\vartheta_\circ, k}(dx)) + o_{\mathbb{P}}(1) \quad (8.3.8)$$

Die integrierenden Maße sind zentriert, dasselbe gilt folglich für die Summanden. Deren Kovarianzmatrizen enthalten deshalb für $1 \leq i, j \leq d$ die Einträge:

$$\begin{aligned} & \text{Cov} \left(\int_I \left(\int_{[x, t_{\max}]} (\mathbb{E} \frac{\partial}{\partial \vartheta} \Lambda_{\vartheta_\circ}(t))_i \mathbb{E} \Lambda_\circ(dt) \right) (N(dx) - \Lambda_\circ(dx)), \right. \\ & \quad \left. \int_I \left(\int_{[x, t_{\max}]} (\mathbb{E} \frac{\partial}{\partial \vartheta} \Lambda_{\vartheta_\circ}(t))_j \mathbb{E} \Lambda_\circ(dt) \right) (N(dx) - \Lambda_\circ(dx)) \right) \\ &= \mathbb{E} \left(\int_I \left(\int_{[x, t_{\max}]} (\mathbb{E} \frac{\partial}{\partial \vartheta} \Lambda_{\vartheta_\circ}(t))_i \mathbb{E} \Lambda_\circ(dt) \right) (N(dx) - \Lambda_\circ(dx)) \right. \\ & \quad \cdot \left. \int_I \left(\int_{[x, t_{\max}]} (\mathbb{E} \frac{\partial}{\partial \vartheta} \Lambda_{\vartheta_\circ}(t))_j \mathbb{E} \Lambda_\circ(dt) \right) (N(dx) - \Lambda_\circ(dx)) \right) \quad (8.3.9) \end{aligned}$$

Die inneren Integrale sind deterministisch, gemäß [5, S. 71, Th. II.3.1 und (2.3.5)] handelt es sich bei den äußeren Integralen um quadrat-integrierte Martingale. Das Produkt in dem zu betrachtenden Erwartungswert kann durch seine vorhersehbare Kovariation ersetzt werden, wie direkt aus deren Definition folgt, vgl. [5, S. 68]. Die vorhersehbare Kovariation einer "Martingaltransformierten" können wir gemäß [5, S. 71, Th. II.3.1 und (2.3.5)] auf ein Integral bzgl. der vorhersehbaren Variation $\langle M, M \rangle(\cdot)$ des integrierenden Martingals zurückführen. Gemäß [5, S. 74, (2.4.3)] entspricht die vorhersehbare Variation $\langle M, M \rangle(\cdot)$ dem (stetigen) Kompensator $\Lambda_\circ(\cdot)$ und wir erhalten:

$$[\dots] = \mathbb{E} \left(\int_I \left(\int_{[x, t_{\max}]} (\mathbb{E} \frac{\partial}{\partial \vartheta} \Lambda_{\vartheta_\circ}(t))_i \mathbb{E} \Lambda_\circ(dt) \right) \cdot \left(\int_{[x, t_{\max}]} (\mathbb{E} \frac{\partial}{\partial \vartheta} \Lambda_{\vartheta_\circ}(t))_j \mathbb{E} \Lambda_\circ(dt) \right) \Lambda_\circ(dx) \right)$$

Weiterhin ist auch Teil 2 des Lemmas A.2.3 anwendbar:

$$[\dots] = \int_I \left(\int_{[x, t_{\max}]} (\mathbb{E} \frac{\partial}{\partial \vartheta} \Lambda_{\vartheta_o}(t))_i \mathbb{E} \Lambda_o(dt) \right) \cdot \left(\int_{[x, t_{\max}]} (\mathbb{E} \frac{\partial}{\partial \vartheta} \Lambda_{\vartheta_o}(t))_j \mathbb{E} \Lambda_o(dt) \right) \mathbb{E} \Lambda_o(dx) = C_{ij}(\vartheta_o)$$

Der klassische ZGWS liefert nun aufgrund der bisherigen Überlegungen das Resultat:

$$n^{1/2} \Phi_n(\tilde{\vartheta}_n) \cdot (\vartheta_n^* - \vartheta_o) \xrightarrow{\mathcal{D}} \mathbb{N}(0, C(\vartheta_o)) \quad \text{wobei} \quad \tilde{\vartheta}_n \xrightarrow{\text{f. s.}} \vartheta_o \quad (8.3.10)$$

Wir wollen in (8.3.10) die Parameterfolge $\tilde{\vartheta}_n$ durch ihren deterministischen Grenzwert ϑ_o ersetzen. Satz 8.2.15 und Lemma A.1.1 liefern

$$\Phi_n(\tilde{\vartheta}_n) \xrightarrow{\mathbb{P}} \Phi_o(\vartheta_o).$$

Wir sind an einer entsprechenden Aussage für die Inverse interessiert, sofern letztere existiert. Wegen Lemma 8.2.3 gilt

$$\det \Phi_n(\tilde{\vartheta}_n) \xrightarrow{\mathbb{P}} \det \Phi_o(\vartheta_o) > 0,$$

die standardisierende Matrix $\Phi_n(\tilde{\vartheta}_n)$ ist also für genügend großen Stichprobenumfang n mit beliebig großer Wahrscheinlichkeit invertierbar. Die Inverse der Matrix $\Phi_n(\tilde{\vartheta}_n)$ läßt sich als Quotient aus der adjungierten Matrix $\text{adj}(\Phi_n(\tilde{\vartheta}_n))$ und der Determinante $\det \Phi_n(\tilde{\vartheta}_n)$ schreiben, vgl. [56, S. 148, Sätze 4 und 5, dort wird die Adjungierte als komplementäre Matrix bezeichnet]. Die Adjungierte ist wie die Determinante stetig von ϑ abhängig, es folgt also:

$$\text{adj}(\Phi_n(\tilde{\vartheta}_n)) \xrightarrow{\mathbb{P}} \text{adj}(\Phi_o(\vartheta_o))$$

Slutskys Theorem [67, S. 19, (iii)] liefert

$$(\Phi_n(\tilde{\vartheta}_n))^{-1} = \frac{\text{adj}(\Phi_n(\tilde{\vartheta}_n))}{\det \Phi_n(\tilde{\vartheta}_n)} \xrightarrow{\mathbb{P}} (\Phi_o(\vartheta_o))^{-1},$$

wobei wir nochmals beachten, daß die linke Seite nicht immer, aber bei wachsendem n mit sehr hoher Wahrscheinlichkeit definiert ist. Eine weitere Anwendung von Slutskys Theorem liefert

$$\begin{aligned} & n^{1/2} \Phi_n(\vartheta_o) \cdot (\vartheta_n^* - \vartheta_o) \\ &= \left(\Phi_n(\vartheta_o) (\Phi_n(\tilde{\vartheta}_n))^{-1} \right) \left(n^{1/2} \Phi_n(\tilde{\vartheta}_n) \cdot (\vartheta_n^* - \vartheta_o) \right) \xrightarrow{\mathcal{D}} \mathbb{N}(0, C(\vartheta_o)), \end{aligned}$$

denn der erste Klammerterm konvergiert in Wahrscheinlichkeit gegen die Identität, so daß der zweite Klammerterm das asymptotische Verhalten des Gesamtausdrucks gemäß (8.3.10) bestimmt. Damit ist Theorem 3.5.1 gezeigt. ■

8.3.1 Lemma *Es gilt (8.3.1), d. h. es existiert eine Folge $\tilde{\vartheta}_n \xrightarrow{\text{f. s.}} \vartheta_o$ mit:*

$$\begin{aligned} & n^{1/2} \frac{\partial}{\partial \vartheta} \int_I (\bar{\Lambda}_{\vartheta, n}(t) - \bar{\Lambda}_{\vartheta_o, n}(t)) \frac{\partial}{\partial \vartheta} \bar{\Lambda}_{\vartheta, n}(t) \bar{\Lambda}_{\vartheta_o, n}(dt) \Big|_{\vartheta=\tilde{\vartheta}_n} \cdot (\vartheta_n^* - \vartheta_o) \\ &= n^{1/2} \int_I (\bar{\Lambda}_{\vartheta, n}(t) - \bar{\Lambda}_{\vartheta_o, n}(t)) \frac{\partial}{\partial \vartheta} \bar{\Lambda}_{\vartheta, n}(t) \bar{\Lambda}_{\vartheta_o, n}(dt) \Big|_{\vartheta=\vartheta_n^*} \end{aligned}$$

Beweis:

Die stochastische Funktion

$$\vartheta \mapsto n^{1/2} \int_I (\bar{\Lambda}_{\vartheta,n}(t) - \bar{\Lambda}_{\vartheta_0,n}(t)) \frac{\partial}{\partial \vartheta} \bar{\Lambda}_{\vartheta,n}(t) \bar{\Lambda}_{\vartheta_0,n}(dt) \quad (8.3.11)$$

ist f. s. stetig differenzierbar, wir sehen dies folgendermaßen ein: Zunächst kann das Integral pfadweise betrachtet werden, also für festes $\omega \in \Omega$. Dann tritt ϑ lediglich im Integranden auf. Dieser ist wegen (3.2.3) stetig von ϑ abhängig. Der grundlegende Satz über stetig parametrisierte Integrale kann angewendet werden, vgl. etwa [32, S: 82, Satz 1].

Wir betrachten die lineare Entwicklung von (8.3.11) um ϑ_0 : Bezeichne

$$\tilde{\vartheta}_n(\omega) \equiv \tilde{\vartheta}_n = (\tilde{\vartheta}_n^1, \dots, \tilde{\vartheta}_n^d) \in \Theta \subset \mathbb{R}^d,$$

einen Parameterwert, welcher koordinatenweise zwischen dem KQS

$$\vartheta_n^* = ((\vartheta_n^*)^1, \dots, (\vartheta_n^*)^d) \in \bar{\Theta} \subset \mathbb{R}^d$$

und dem wahren Parameter

$$\vartheta_0 = (\vartheta_0^1, \dots, \vartheta_0^d) \in \Theta \subset \mathbb{R}^d$$

liegt:

$$\tilde{\vartheta}_n^i \in [(\vartheta_n^*)^i, \vartheta_0^i] \quad \forall i = 1, \dots, d$$

$\tilde{\vartheta}_n$ kann so gewählt werden, daß

$$\begin{aligned} & n^{1/2} \int_I (\bar{\Lambda}_{\vartheta,n}(t) - \bar{\Lambda}_{\vartheta_0,n}(t)) \frac{\partial}{\partial \vartheta} \bar{\Lambda}_{\vartheta,n}(t) \bar{\Lambda}_{\vartheta_0,n}(dt) \Big|_{\vartheta=\vartheta_n^*} \\ &= n^{1/2} \int_I (\bar{\Lambda}_{\vartheta,n}(t) - \bar{\Lambda}_{\vartheta_0,n}(t)) \frac{\partial}{\partial \vartheta} \bar{\Lambda}_{\vartheta,n}(t) \bar{\Lambda}_{\vartheta_0,n}(dt) \Big|_{\vartheta=\vartheta_0} \\ &+ n^{1/2} \frac{\partial}{\partial \vartheta} \int_I (\bar{\Lambda}_{\vartheta,n}(t) - \bar{\Lambda}_{\vartheta_0,n}(t)) \frac{\partial}{\partial \vartheta} \bar{\Lambda}_{\vartheta,n}(t) \bar{\Lambda}_{\vartheta_0,n}(dt) \Big|_{\vartheta=\tilde{\vartheta}_n} \cdot (\vartheta_n^* - \vartheta_0) \end{aligned}$$

gilt. Das erste Integral auf der rechten Seite verschwindet, da sein Integrand verschwindet. \blacksquare

8.3.2 Lemma *Es gilt (8.3.2):*

$$\begin{aligned} & n^{1/2} \int_I (\bar{\Lambda}_{\vartheta,n}(t) - \bar{\Lambda}_{\vartheta_0,n}(t)) \frac{\partial}{\partial \vartheta} \bar{\Lambda}_{\vartheta,n}(t) \bar{\Lambda}_{\vartheta_0,n}(dt) \Big|_{\vartheta=\vartheta_n^*} \\ &= n^{1/2} \int_I (\bar{\Lambda}_{\vartheta,n}(t) - \bar{\Lambda}_{\vartheta_0,n}(t)) \frac{\partial}{\partial \vartheta} \bar{\Lambda}_{\vartheta,n}(t) \bar{N}_n(dt) \Big|_{\vartheta=\vartheta_n^*} + o_{\mathbb{P}}(1) \end{aligned}$$

Beweis:

Wir spalten das Integral auf:

$$\begin{aligned}
& n^{1/2} \int_I (\bar{\Lambda}_{\vartheta,n}(t) - \bar{\Lambda}_{\vartheta_\circ,n}(t)) \frac{\partial}{\partial \vartheta} \bar{\Lambda}_{\vartheta,n}(t) \bar{\Lambda}_{\vartheta_\circ,n}(dt) \Big|_{\vartheta=\vartheta_n^*} \\
&= n^{1/2} \int_I (\bar{\Lambda}_{\vartheta,n}(t) - \bar{\Lambda}_{\vartheta_\circ,n}(t)) \frac{\partial}{\partial \vartheta} \bar{\Lambda}_{\vartheta,n}(t) \bar{N}_n(dt) \Big|_{\vartheta=\vartheta_n^*} \\
&\quad - n^{1/2} \int_I (\bar{\Lambda}_{\vartheta,n}(t) - \bar{\Lambda}_{\vartheta_\circ,n}(t)) \frac{\partial}{\partial \vartheta} \bar{\Lambda}_{\vartheta,n}(t) (\bar{N}_n(dt) - \bar{\Lambda}_{\vartheta_\circ,n}(dt)) \Big|_{\vartheta=\vartheta_n^*} \\
&= n^{1/2} \int_I (\bar{\Lambda}_{\vartheta,n}(t) - \bar{\Lambda}_{\vartheta_\circ,n}(t)) \frac{\partial}{\partial \vartheta} \bar{\Lambda}_{\vartheta,n}(t) \bar{N}_n(dt) \Big|_{\vartheta=\vartheta_n^*} + \alpha_n(\vartheta_n^*)
\end{aligned}$$

In der letzten Gleichung haben wir die Definition (8.2.8) von α_n verwendet.

Das Lemma ist also bewiesen, wenn wir

$$\alpha_n(\vartheta_n^*) = o_{\mathbb{P}}(1)$$

zeigen.

Nach Definition gilt $\alpha_n(\vartheta_\circ) = 0$, es genügt also,

$$\alpha_n(\vartheta_n^*) - \alpha_n(\vartheta_\circ) \xrightarrow{\mathbb{P}} 0$$

nachzuweisen.

Nun gilt $\vartheta_n^* \xrightarrow{\mathbb{P}} \vartheta_\circ$ nach Theorem 3.3.1 und wegen Satz 8.2.4 ist $\alpha_n(\cdot)$ auf beliebigen, kompakten Parameterteilmengen Θ_\circ straff. Lemma A.1.1 liefert die gesuchte Aussage. \blacksquare

8.3.3 Lemma *Es gilt (8.3.3):*

$$\begin{aligned}
& n^{1/2} \int_I (\bar{\Lambda}_{\vartheta,n}(t) - \bar{\Lambda}_{\vartheta_\circ,n}(t)) \frac{\partial}{\partial \vartheta} \bar{\Lambda}_{\vartheta,n}(t) \bar{N}_n(dt) \Big|_{\vartheta=\vartheta_n^*} \\
&= n^{1/2} \int_I (\bar{N}_n(t) - \bar{\Lambda}_{\vartheta_\circ,n}(t)) \frac{\partial}{\partial \vartheta} \bar{\Lambda}_{\vartheta,n}(t) \Big|_{\vartheta=\vartheta_n^*} \bar{N}_n(dt)
\end{aligned}$$

Beweis:

Eine geeignete Aufspaltung des Integranden liefert:

$$\begin{aligned}
& n^{1/2} \int_I (\bar{\Lambda}_{\vartheta,n}(t) - \bar{\Lambda}_{\vartheta_\circ,n}(t)) \frac{\partial}{\partial \vartheta} \bar{\Lambda}_{\vartheta,n}(t) \bar{N}_n(dt) \Big|_{\vartheta=\vartheta_n^*} \\
&= n^{1/2} \int_I (\bar{\Lambda}_{\vartheta,n}(t) - \bar{N}_n(t)) \frac{\partial}{\partial \vartheta} \bar{\Lambda}_{\vartheta,n}(t) \bar{N}_n(dt) \Big|_{\vartheta=\vartheta_n^*} \\
&\quad + n^{1/2} \int_I (\bar{N}_n(t) - \bar{\Lambda}_{\vartheta_\circ,n}(t)) \frac{\partial}{\partial \vartheta} \bar{\Lambda}_{\vartheta,n}(t) \bar{N}_n(dt) \Big|_{\vartheta=\vartheta_n^*}
\end{aligned}$$

Der erste Summand verschwindet wegen

$$\begin{aligned} n^{1/2} \int_I (\bar{\Lambda}_{\vartheta,n}(t) - \bar{N}_n(t)) \frac{\partial}{\partial \vartheta} \bar{\Lambda}_{\vartheta,n}(t) \bar{N}_n(dt) \Big|_{\vartheta=\vartheta_n^*} \\ = -2 \frac{\partial}{\partial \vartheta} \|\bar{N}_n(\cdot) - \bar{\Lambda}_{\vartheta,n}(\cdot)\|_{\bar{N}_n}^2 \Big|_{\vartheta=\vartheta_n^*} = 0. \end{aligned}$$

Dies gilt nach Definition des KQS in Abschnitt 3.2.5 und der Forderung (3.2.4), daß Θ offen ist. ■

8.3.4 Lemma *Es gilt (8.3.4):*

$$\begin{aligned} n^{1/2} \int_I (\bar{N}_n(t) - \bar{\Lambda}_{\vartheta_o,n}(t)) \frac{\partial}{\partial \vartheta} \bar{\Lambda}_{\vartheta,n}(t) \Big|_{\vartheta=\vartheta_n^*} \bar{N}_n(dt) \\ = n^{1/2} \int_I (\bar{N}_n(t) - \bar{\Lambda}_{\vartheta_o,n}(t)) \frac{\partial}{\partial \vartheta} \bar{\Lambda}_{\vartheta,n}(t) \Big|_{\vartheta=\vartheta_n^*} \bar{\Lambda}_{\vartheta_o,n}(dt) + o_P(1) \end{aligned}$$

Beweis:

Gemäß (8.2.37) entspricht die Aussage des Lemmas der Behauptung

$$\beta_n(\vartheta_n^*) = o_P(1).$$

Dies gilt wegen Satz 8.2.8. ■

8.3.5 Lemma *Es gilt (8.3.5):*

$$\begin{aligned} n^{1/2} \int_I (\bar{N}_n(t) - \bar{\Lambda}_{\vartheta_o,n}(t)) \frac{\partial}{\partial \vartheta} \bar{\Lambda}_{\vartheta,n}(t) \Big|_{\vartheta=\vartheta_n^*} \bar{\Lambda}_{\vartheta_o,n}(dt) \\ = n^{1/2} \int_I (\bar{N}_n(t) - \bar{\Lambda}_{\vartheta_o,n}(t)) \frac{\partial}{\partial \vartheta} \bar{\Lambda}_{\vartheta,n}(t) \Big|_{\vartheta=\vartheta_o} \bar{\Lambda}_{\vartheta_o,n}(dt) + o_P(1) \end{aligned}$$

Beweis:

Nach Definition (8.2.50) genügt es,

$$\gamma_n(\vartheta_n^*) - \gamma_n(\vartheta_o) = o_P(1)$$

zu zeigen. Wegen Satz 8.2.11 ist $\gamma_n(\vartheta)$ auf kompakten Parameterteilmengen straff. Die gesuchte Aussage ergibt sich deshalb wie im Beweis zu Lemma 8.3.2 aus der Konsistenz des KQS. ■

8.3.6 Lemma *Es gilt (8.3.6):*

$$\begin{aligned} n^{1/2} \int_I (\bar{N}_n(t) - \bar{\Lambda}_{\vartheta_o,n}(t)) \frac{\partial}{\partial \vartheta} \bar{\Lambda}_{\vartheta_o,n}(t) \bar{\Lambda}_{\vartheta_o,n}(dt) \\ = n^{1/2} \int_I \left(\int_{[x, t_{\max}]} \frac{\partial}{\partial \vartheta} \bar{\Lambda}_{\vartheta_o,n}(t) \bar{\Lambda}_{\vartheta_o,n}(dt) \right) (\bar{N}_n(dx) - \bar{\Lambda}_{\vartheta_o,n}(dx)) \end{aligned}$$

Beweis:

Wir schreiben Teile des Integranden als eigenes Integral:

$$\begin{aligned}
& n^{1/2} \int_I (\bar{N}_n(t) - \bar{\Lambda}_{\vartheta_o, n}(t)) \frac{\partial}{\partial \vartheta} \bar{\Lambda}_{\vartheta_o, n}(t) \bar{\Lambda}_{\vartheta_o, n}(dt) \\
&= n^{1/2} \int_I \left(\int_I \mathbf{1}\{x \leq t\} (\bar{N}_n(dx) - \bar{\Lambda}_{\vartheta_o, n}(dx)) \right) \frac{\partial}{\partial \vartheta} \bar{\Lambda}_{\vartheta_o, n}(t) \bar{\Lambda}_{\vartheta_o, n}(dt) \\
&= n^{1/2} \iint_{I \times I} \mathbf{1}\{x \leq t\} \frac{\partial}{\partial \vartheta} \bar{\Lambda}_{\vartheta_o, n}(t) (\bar{N}_n(dx) - \bar{\Lambda}_{\vartheta_o, n}(dx)) \bar{\Lambda}_{\vartheta_o, n}(dt)
\end{aligned}$$

Nun wenden wir den Satz von Fubini an:

$$\begin{aligned}
[\dots] &= n^{1/2} \iint_{I \times I} \mathbf{1}\{x \leq t\} \frac{\partial}{\partial \vartheta} \bar{\Lambda}_{\vartheta_o, n}(t) \bar{\Lambda}_{\vartheta_o, n}(dt) (\bar{N}_n(dx) - \bar{\Lambda}_{\vartheta_o, n}(dx)) \\
&= n^{1/2} \int_I \left(\int_{[x, t_{\max}]} \frac{\partial}{\partial \vartheta} \bar{\Lambda}_{\vartheta_o, n}(t) \bar{\Lambda}_{\vartheta_o, n}(dt) \right) (\bar{N}_n(dx) - \bar{\Lambda}_{\vartheta_o, n}(dx))
\end{aligned}$$

Das Lemma ist bewiesen. ■

8.3.7 Lemma *Es gilt (8.3.7):*

$$\begin{aligned}
& n^{1/2} \int_I \left(\int_{[x, t_{\max}]} \frac{\partial}{\partial \vartheta} \bar{\Lambda}_{\vartheta_o, n}(t) \bar{\Lambda}_{\vartheta_o, n}(dt) \right) \bar{M}_n(dx) \\
&= n^{1/2} \int_I \left(\int_{[x, t_{\max}]} \mathbb{E} \frac{\partial}{\partial \vartheta} \Lambda_o(t) \mathbb{E} \Lambda_o(dt) \right) \bar{M}_n(dx) + o_{\mathbb{P}}(1)
\end{aligned}$$

Beweis:

Wir spalten die Differenz der beiden inneren Integrale geeignet auf:

$$\begin{aligned}
& \int_{[x, t_{\max}]} \frac{\partial}{\partial \vartheta} \bar{\Lambda}_{\vartheta_o, n}(t) \bar{\Lambda}_{\vartheta_o, n}(dt) - \int_{[x, t_{\max}]} \mathbb{E} \frac{\partial}{\partial \vartheta} \Lambda_o(t) \mathbb{E} \Lambda_o(dt) \\
&= \int_{[x, t_{\max}]} \frac{\partial}{\partial \vartheta} \bar{\Lambda}_{\vartheta_o, n}(t) (\bar{\Lambda}_{\vartheta_o, n}(dt) - \mathbb{E} \Lambda_o(dt)) \\
&+ \int_{[x, t_{\max}]} \left(\frac{\partial}{\partial \vartheta} \bar{\Lambda}_{\vartheta_o, n}(t) - \mathbb{E} \frac{\partial}{\partial \vartheta} \Lambda_o(t) \right) \mathbb{E} \Lambda_o(dt)
\end{aligned}$$

Das Lemma ist also bewiesen, sofern wir

$$n^{1/2} \int_I \left(\int_{[x, t_{\max}]} \frac{\partial}{\partial \vartheta} \bar{\Lambda}_{\vartheta_o, n}(t) (\bar{\Lambda}_{\vartheta_o, n}(dt) - \mathbb{E} \Lambda_o(dt)) \right) \bar{M}_n(dx) = o_{\mathbb{P}}(1) \quad (8.3.12)$$

und

$$n^{1/2} \int \left(\int_I \left(\frac{\partial}{\partial \vartheta} \bar{\Lambda}_{\vartheta_o, n}(t) - \mathbb{E} \frac{\partial}{\partial \vartheta} \Lambda_o(t) \right) E \Lambda_o(dt) \right) \bar{M}_n(dx) = o_{\mathbb{P}}(1) \quad (8.3.13)$$

zeigen.

Wir beginnen mit (8.3.12). Der Beweisgang entspricht weitgehend dem von Lemma 8.2.9:

$$\begin{aligned} & n^{1/2} \int \left(\int_I \left(\frac{\partial}{\partial \vartheta} \bar{\Lambda}_{\vartheta_o, n}(t) (\bar{\Lambda}_{\vartheta_o, n}(dt) - \mathbb{E} \Lambda_o(dt)) \right) \right) \bar{M}_n(dx) \\ &= n^{-5/2} \sum_{p, k, i=1}^n \int \left(\int_I \left(\frac{\partial}{\partial \vartheta} \Lambda_{\vartheta_o, p}(t) (\Lambda_{\vartheta_o, k}(dt) - \mathbb{E} \Lambda_o(dt)) \right) \right) M_i(dx) \\ &= n^{-5/2} \sum_{i=1}^n \int \left(\int_I \left(\frac{\partial}{\partial \vartheta} \Lambda_{\vartheta_o, i}(t) (\Lambda_{\vartheta_o, i}(dt) - \mathbb{E} \Lambda_o(dt)) \right) \right) M_i(dx) \quad (8.3.14) \end{aligned}$$

$$+ n^{-5/2} \sum_{\substack{k, i=1 \\ k \neq i}}^n \int \left(\int_I \left(\frac{\partial}{\partial \vartheta} \Lambda_{\vartheta_o, k}(t) (\Lambda_{\vartheta_o, k}(dt) - \mathbb{E} \Lambda_o(dt)) \right) \right) M_i(dx) \quad (8.3.15)$$

$$+ n^{-5/2} \sum_{\substack{k, i=1 \\ k \neq i}}^n \int \left(\int_I \left(\frac{\partial}{\partial \vartheta} \Lambda_{\vartheta_o, i}(t) (\Lambda_{\vartheta_o, k}(dt) - \mathbb{E} \Lambda_o(dt)) \right) \right) M_i(dx) \quad (8.3.16)$$

$$+ n^{-5/2} \sum_{\substack{k, i=1 \\ k \neq i}}^n \int \left(\int_I \left(\frac{\partial}{\partial \vartheta} \Lambda_{\vartheta_o, k}(t) (\Lambda_{\vartheta_o, i}(dt) - \mathbb{E} \Lambda_o(dt)) \right) \right) M_i(dx) \quad (8.3.17)$$

$$+ n^{-5/2} \sum_{\substack{p, k, i=1 \\ p \neq k \neq i \neq p}}^n \int \left(\int_I \left(\frac{\partial}{\partial \vartheta} \Lambda_{\vartheta_o, p}(t) (\Lambda_{\vartheta_o, k}(dt) - \mathbb{E} \Lambda_o(dt)) \right) \right) M_i(dx) \quad (8.3.18)$$

Wir betrachten wieder jeden Term (8.3.14) bis (8.3.18) einzeln. Im ersten Moment von (8.3.14) bis (8.3.17) treten n gleiche bzw. $O(n^2)$ gleichmäßig beschränkte Summanden auf, welche vom Faktor $n^{-5/2}$ dominiert werden.

Auf (8.3.18) wenden wir zunächst Lemma A.3.3 an, wobei wir

$$U_{pki} := \int \left(\int_I \left(\frac{\partial}{\partial \vartheta} \Lambda_{\vartheta_o, p}(t) (\Lambda_{\vartheta_o, k}(dt) - \mathbb{E} \Lambda_o(dt)) \right) \right) M_i(dx)$$

setzen. Dazu brauchen wir lediglich Lemma 8.2.1 und (8.2.3). Lemma A.3.3

liefert dann folgende Schranke für das zweite Moment von (8.3.18):

$$\begin{aligned} & \mathbb{E} \left| n^{-5/2} \sum_{\substack{p,k,i=1 \\ p \neq k \neq i \neq p}}^n \int_I \left(\int_{[x, t_{\max}]} \frac{\partial}{\partial \vartheta} \Lambda_{\vartheta_\circ, p}(t) (\Lambda_{\vartheta_\circ, k}(dt) - \mathbb{E} \Lambda_\circ(dt)) \right) M_i(dx) \right|^2 \\ & \leq 64n^{-5} \sum_{\substack{p,q,k,i=1 \\ p,q \leq k \leq i \leq p,q}}^n \mathbb{E} \left(\iint_{I \times I} \int_{[x, t_{\max}]} \int_{[y, t_{\max}]} \frac{\partial}{\partial \vartheta} \Lambda_{\vartheta_\circ, p}(t) \perp \frac{\partial}{\partial \vartheta} \Lambda_{\vartheta_\circ, q}(s) \right. \\ & \quad \left. (\Lambda_{\vartheta_\circ, k}(dt) - \mathbb{E} \Lambda_\circ(dt)) (\Lambda_{\vartheta_\circ, k}(ds) - \mathbb{E} \Lambda_\circ(ds)) M_i(dy) M_i(dx) \right) \end{aligned}$$

Die Beträge der verbleibenden $O(n^4)$ Summanden sind gleichmäßig beschränkt und werden deshalb vom Faktor n^{-5} dominiert.

Damit ist der Beweis von (8.3.12) beendet, wir wenden uns (8.3.13) zu:

$$\begin{aligned} & n^{1/2} \int_I \left(\int_{[x, t_{\max}]} \left(\frac{\partial}{\partial \vartheta} \bar{\Lambda}_{\vartheta_\circ, n}(t) - \mathbb{E} \frac{\partial}{\partial \vartheta} \Lambda_\circ(t) \right) \mathbb{E} \Lambda_\circ(dt) \right) \bar{M}_n(dx) \\ & = n^{-3/2} \sum_{k,i=1}^n \int_I \left(\int_{[x, t_{\max}]} \left(\frac{\partial}{\partial \vartheta} \Lambda_{\vartheta_\circ, k}(t) - \mathbb{E} \frac{\partial}{\partial \vartheta} \Lambda_\circ(t) \right) \mathbb{E} \Lambda_\circ(dt) \right) M_i(dx) \\ & = n^{-3/2} \sum_{i=1}^n \int_I \left(\int_{[x, t_{\max}]} \left(\frac{\partial}{\partial \vartheta} \Lambda_{\vartheta_\circ, i}(t) - \mathbb{E} \frac{\partial}{\partial \vartheta} \Lambda_\circ(t) \right) \mathbb{E} \Lambda_\circ(dt) \right) M_i(dx) \quad (8.3.19) \end{aligned}$$

$$+ n^{-3/2} \sum_{\substack{k,i=1 \\ k \neq i}}^n \int_I \left(\int_{[x, t_{\max}]} \left(\frac{\partial}{\partial \vartheta} \Lambda_{\vartheta_\circ, k}(t) - \mathbb{E} \frac{\partial}{\partial \vartheta} \Lambda_\circ(t) \right) \mathbb{E} \Lambda_\circ(dt) \right) M_i(dx) \quad (8.3.20)$$

Die L_1 -Konvergenz von (8.3.19) ergibt sich nach bekanntem Muster aus den n gleichen Summanden und dem Faktor $n^{-3/2}$. Für das zweite Moment von (8.3.20) liefert Lemma A.3.2 eine obere Schranke mit $O(n^2)$ gleichmäßig beschränkten Summanden, die von n^{-3} dominiert werden. Die Voraussetzungen des Lemmas sind nach Lemma 8.2.1 und (8.2.5) erfüllt. ■

Wir beweisen nun noch die asymptotische Normalität des KQS bzgl. bedingter Intensitäten, die wir in (3.5.1) eingeführt haben.

Beweis von Korollar 3.5.2:

Der Beweis des Theorems 3.5.1 kann bis auf die Berechnung der asymptotischen Kovarianz nach (8.3.9) wörtlich beibehalten werden, da bis dahin lediglich die Zentriertheit der Prozesse $N_k(\cdot) - \Lambda_{\vartheta_\circ, k}(\cdot)$ verwendet wird, jedoch nicht ihre Martingaleigenschaft. Das folgende Lemma 8.3.8 ermöglicht für $1 \leq i, j \leq d$ die Berechnung der gemischten Momente

$$\begin{aligned} & \mathbb{E} \left(\int_I \left(\int_{[x, t_{\max}]} \left(\mathbb{E} \frac{\partial}{\partial \vartheta} \Lambda_{\vartheta_\circ}(t) \right)_i \mathbb{E} \Lambda_\circ(dt) \right) (N(dx) - \Lambda_{\vartheta_\circ}^{\mathcal{C}}(dx)) \right. \\ & \quad \left. \cdot \int_I \left(\int_{[x, t_{\max}]} \left(\mathbb{E} \frac{\partial}{\partial \vartheta} \Lambda_{\vartheta_\circ}(t) \right)_j \mathbb{E} \Lambda_\circ(dt) \right) (N(dx) - \Lambda_{\vartheta_\circ}^{\mathcal{C}}(dx)) \right), \end{aligned}$$

welche bei Verwendung bedingter Intensitäten anstelle von (8.3.9) auftreten. Es schließt so den Beweis des Korollars ab. ■

8.3.8 Lemma Für \mathcal{A} -vorhersehbare Funktionen $f, g : I \rightarrow \mathbb{R}$ gilt:

$$\begin{aligned} & \mathbb{E} \left(\int_I f(t) (N(dt) - \Lambda_{\vartheta_\circ}^{\mathcal{C}}(dt)) \cdot \int_I g(t) (N(dt) - \Lambda_{\vartheta_\circ}^{\mathcal{C}}(dt)) \right) \\ &= \mathbb{E} \left(\int_I f(t) (N(dt) - \Lambda_{\vartheta_\circ}(dt)) \cdot \int_I g(t) (N(dt) - \Lambda_{\vartheta_\circ}(dt)) \right) \\ &+ \mathbb{E} \left(\int_I f(t) (\Lambda_{\vartheta_\circ}(dt) - \Lambda_{\vartheta_\circ}^{\mathcal{C}}(dt)) \cdot \int_I g(t) (\Lambda_{\vartheta_\circ}(dt) - \Lambda_{\vartheta_\circ}^{\mathcal{C}}(dt)) \right) \end{aligned}$$

Beweis:

Zunächst gilt:

$$\begin{aligned} & \mathbb{E} \left(\int_I f(t) (N(dt) - \Lambda_{\vartheta_\circ}^{\mathcal{C}}(dt)) \cdot \int_I g(t) (N(dt) - \Lambda_{\vartheta_\circ}^{\mathcal{C}}(dt)) \right) \\ &= \mathbb{E} \left(\int_I f(t) (N(dt) - \Lambda_{\vartheta_\circ}(dt) + \Lambda_{\vartheta_\circ}(dt) - \Lambda_{\vartheta_\circ}^{\mathcal{C}}(dt)) \right. \\ &\quad \cdot \left. \int_I g(t) (N(dt) - \Lambda_{\vartheta_\circ}(dt) + \Lambda_{\vartheta_\circ}(dt) - \Lambda_{\vartheta_\circ}^{\mathcal{C}}(dt)) \right) \\ &= \mathbb{E} \left(\int_I f(t) (N(dt) - \Lambda_{\vartheta_\circ}(dt)) \cdot \int_I g(t) (N(dt) - \Lambda_{\vartheta_\circ}(dt)) \right) \\ &+ \mathbb{E} \left(\int_I f(t) (\Lambda_{\vartheta_\circ}(dt) - \Lambda_{\vartheta_\circ}^{\mathcal{C}}(dt)) \cdot \int_I g(t) (\Lambda_{\vartheta_\circ}(dt) - \Lambda_{\vartheta_\circ}^{\mathcal{C}}(dt)) \right) \\ &+ \mathbb{E} \left(\int_I f(t) (N(dt) - \Lambda_{\vartheta_\circ}(dt)) \cdot \int_I g(t) (\Lambda_{\vartheta_\circ}(dt) - \Lambda_{\vartheta_\circ}^{\mathcal{C}}(dt)) \right) \quad (8.3.21) \end{aligned}$$

$$+ \mathbb{E} \left(\int_I f(t) (\Lambda_{\vartheta_\circ}(dt) - \Lambda_{\vartheta_\circ}^{\mathcal{C}}(dt)) \cdot \int_I g(t) (N(dt) - \Lambda_{\vartheta_\circ}(dt)) \right) \quad (8.3.22)$$

Es genügt also zu zeigen, daß die Ausdrücke (8.3.21) und (8.3.22) verschwinden. Wir betrachten lediglich (8.3.21):

$$\begin{aligned} & \mathbb{E} \left(\int_I f(t) (N(dt) - \Lambda_{\vartheta_\circ}(dt)) \cdot \int_I g(t) (\Lambda_{\vartheta_\circ}(dt) - \Lambda_{\vartheta_\circ}^{\mathcal{C}}(dt)) \right) \\ &= \mathbb{E} \left(\mathbb{E} \left[\int_I f(t) (N(dt) - \Lambda_{\vartheta_\circ}(dt)) \cdot \int_I g(t) (\Lambda_{\vartheta_\circ}(dt) - \Lambda_{\vartheta_\circ}^{\mathcal{C}}(dt)) \middle| \mathcal{A}(t_{\max}-) \right] \right) \\ &= \mathbb{E} \left(\mathbb{E} \left[\int_I f(t) (N(dt) - \Lambda_{\vartheta_\circ}(dt)) \middle| \mathcal{A}(t_{\max}-) \right] \cdot \int_I g(t) (\Lambda_{\vartheta_\circ}(dt) - \Lambda_{\vartheta_\circ}^{\mathcal{C}}(dt)) \right) \end{aligned}$$

Der bedingte Erwartungswert verschwindet, da es sich bei dem Integral um ein \mathcal{A} -Martingal handelt. ■

9 Regressionsmodelle für allgemeine Punktprozesse

9.1 Übersicht über die Literatur

[71] bietet im Rahmen eines Lehrbuches eine umfangreiche Darstellung angewandter Probleme, die eine Formalisierung mit Punktprozessen zulassen. Auf das Problem der Parameterschätzung für "self-exciting" Punktprozesse (ohne externe Einflußgrößen) geht [71, pp. 302] ein. Dabei wird ein Schwerpunkt auf die Formulierung des Maximum-Likelihood-Schätzers gelegt. Asymptotische Eigenschaften dieses Schätzers werden nicht untersucht.¹⁰ Im Rahmen gefilterter Poissonprozesse werden Resultate vorgestellt, die z. B. den Schluß von einem beobachteten Poissonprozeß auf einen Gaußprozeß zulassen, der selbst nicht beobachtet wurde, aber die Intensität des Poissonprozesses auf bekannte Weise beeinflusst, [71, p. 412, Theorem 7.4.4]. In dieser Situation wird ein Kleinst-Quadrate-Schätzer auf seine asymptotischen Eigenschaften hin untersucht.

Es fällt auf, wie stark diese theoretischen Resultate auf relativ spezifische Anwendungsprobleme zugeschnitten sind. Im Vorwort notieren die Autoren: "*The changes we have introduced in this revision are in large part an outgrowth of our research over the past seven years on the development of methods for quantitative image-formation.*" ([71, p. vi]). Allgemeine Punktprozesse scheinen nur in speziellen Anwendungen überhaupt in ausreichender Zahl beobachtet zu werden, um asymptotische Eigenschaften entsprechender Schätzer praxisrelevant zu machen. Jedenfalls rankt sich der überwiegende Teil der stochastischen Literatur zur Parameterschätzung in Punktprozessen um recht spezielle Typen von Punktprozessen. Neben Poissonprozessen werden vor allem empirische Prozesse betrachtet, zu deren Schätzung ja keine i. i. d. Beobachtungen von Punktprozessen, sondern von reellen Zufallsvariablen benötigt werden, siehe z. B. [72].

Betrachten wir kurz einige Veröffentlichungen der letzten zehn Jahre, um diesen Trend auch für die Teile der Forschung zu belegen, die keine empirischen Prozesse betreffen. [37] untersuchen starke Konsistenz und asymptotische Normalität für einen homogenen Poissonprozeß, dessen Beobachtung durch eine "Totzeit" nach jedem Sprung behindert wird.

[24] betrachten asymptotische Eigenschaften eines nichtparametrischen Maximum-Likelihood-Schätzers für die periodische Intensität eines heterogenen Poissonprozesses.

[42] untersuchen den Kleinst-Quadrate-Schätzer für heterogene Poissonprozesse, wiederum auf einer einzigen Realisation. Die Asymptotik besteht in all diesen Untersuchungen darin, das Beobachtungsfenster für den Pfad immer weiter auszuweiten. Offenbar sind diese theoretischen Untersuchungen von Anwendungen geprägt, in denen keine großen Stichprobenumfänge erhoben werden können.

[62] betrachten Maximum-Likelihood-Schätzer, um die Existenz eines change-points eines ansonsten homogenen Poissonprozesses zu belegen und Parameter

¹⁰Eine frühere Auflage des Buches stellte den Maximum-Likelihood-Schätzer auch für die von uns betrachtete Situation eines allgemeinen Punktprozesses vor. Eigenschaften dieses Schätzers blieben jedoch ebenfalls unberücksichtigt.

eines zugrundegelegten Modells zu schätzen. Wieder fällt die spezielle Formulierung des Modelles ins Auge und die Tatsache, daß lediglich eine Beobachtung verfügbar ist.

Eine andere Forschungsrichtung untersucht stationäre Poissonprozesse in höheren Dimensionen, siehe z. B. [70] und die dort angegebene Literatur.

Dieser Vielfalt an weiterführenden stochastischen Erkenntnissen in relativ speziellen Situationen steht nun die angewandte ökonomische Literatur gegenüber, die wir in den letzten Abschnitten besprochen haben. In diesen Anwendungen müssen vielfach scharfe Restriktionen klassischer Regressionsmodelle in Kauf genommen werden, um sich den zeitlichen Phänomenen des Kaufverhaltens überhaupt nähern zu können. Vor diesem Hintergrund erscheint die Untersuchung grundlegender asymptotischer Eigenschaften naheliegender Schätzer für parametrische Regressionsmodelle allgemeiner Punktprozesse geradezu überfällig. Neben dem bereits eingangs besprochenen Lehrbuch [71] bin ich jedoch auch bei intensiver Recherche in der Lehrbuchliteratur lediglich noch auf [51] gestoßen. Diese Monographie stellt umfassend die Eigenschaften parametrischer Regressionsmodelle vor, soweit sie sich auf Poissonprozesse beziehen. Wir betrachten die Ergebnisse von [51, pp. 58-68] im übernächsten Abschnitt. Diese Resultate stehen in Zusammenhang mit unseren eigenen Theoremen 3.3.1 und 3.5.1. In derselben Arbeitsgruppe sind noch einige Doktorarbeiten und andere Veröffentlichungen zu Regressionsmodellen erschienen, die sich jedoch ebenfalls explizit auf Poissonprozesse beschränken. Wir betrachten im nächsten Abschnitt [6], einen der neueren dieser Artikel. Für die älteren Veröffentlichungen verweisen wir auf die dortige Literaturübersicht.

9.2 Die Ergebnisse von [6]

[6] betrachtet das Beispiel einer speziellen differentiellen Intensität auf einem kompakten Intervall I , die sich in Anlehnung an die von uns in Kapitel 3 verwendeten Begriffe so darstellt, vgl. [6, Condition C_0 , p. 4]:

$$\lambda_{\vartheta}(t) := \vartheta_1 g(\vartheta_2 \cdot t) + c \quad (9.2.1)$$

Dabei ist $\vartheta \in \Theta \subset \mathbb{R}_+ \times \mathbb{R}$ ein zweidimensionaler Parameter und Θ ein beschränkter Parameterraum. Die Konstante $c \geq 0$ ist bekannt, ebenso die positive Funktion g . Diese Funktion ist stetig differenzierbar auf dem Intervall I , abgesehen von einem Sprung erster Ordnung an der bekannten Stelle $\tau \in I$.

Also parametrisiert ϑ die Position und Höhe des einzigen Sprunges von λ_{ϑ} . Die Intensität gehört zu einem Poissonprozeß, ist also deterministisch. Einflußgrößen werden nicht betrachtet.

[6] untersucht nun die asymptotischen Eigenschaften eines Kleinst-Quadrate-Schätzers bzgl. der Norm

$$\|f\| := \left(\int_I f^2(t) dt \right)^{1/2}$$

Die L_2 -Norm wird hier also anders als in Abschnitt 3.2.5 mit dem Lebesgue-Maß auf I definiert.

[6] zeigt die schwache Konsistenz ([6, Theorem 3.1, p. 5]) und asymptotische Normalität ([6, Theorem 4.1, p. 7]) des KQS in dieser Beispielsituation. Die Beweisstrukturen entsprechen grob unseren, die Betrachtung eines Poissonprozeßbeispiels führt zu drei Unterschieden:

1. Viele unserer Voraussetzungen können für das Beispiel als Lemmata formuliert werden.
2. Die Poissoneigenschaft vereinfacht den Beweis vieler notwendiger Hilfsresultate dramatisch.
3. Die Beschränkung auf ein vorgegebenes Beispiel liefert entsprechend konkretere Darstellungen der KQS-Momente.

Beginnen wir mit den Voraussetzungen zur schwachen Konsistenz bei [6]. Abgesehen von (3.2.3), (3.2.5) und (3.2.7) sind alle unseren Bedingungen aus Abschnitt 3.2 für das Beispiel (9.2.1) offensichtlich erfüllt. Die Differenzierbarkeitsbedingung (3.2.3) sehen wir für (9.2.1), wenn wir das Integral an der Sprungstelle τ/ϑ_2 des Integranden zerlegen und die Stammfunktionen für die beiden resultierenden Teilintegrale als Funktion von ϑ betrachten. Die zur Identifizierbarkeit benötigte Voraussetzung (3.2.7) wird in [6, Lemma 3.2, p. 6] bewiesen. Dabei wird ein Kompaktheitsargument verwendet, das (3.2.5) implizit voraussetzt.

Der Beweis [6, p. 5] beginnt mit Ungleichungen, die ungefähr den Inklusionen entsprechen, welche zu (7.3.7) in unserem Konsistenzbeweis führen. Da [6] lediglich die schwache Konsistenz zeigt, genügt dort im Anschluß Tschebyscheffs Ungleichung, um den Beweis zu beenden. Wir zeigen hingegen f. s. Verhalten der beteiligten Terme gemäß (7.3.8), und ziehen uns dabei letztlich auf das starke Gesetz gemäß Satz A.4.1 zurück.

Wir wenden uns den Voraussetzungen der asymptotischen Normalität zu. Die Differenzierbarkeitsbedingung (3.4.1) behandeln wir entsprechend wie (3.2.3). Unsere Lipschitzbedingungen (3.4.2) bis (3.4.5) werden in ähnlicher Form durch [6, Lemma A.3, p. 22] bereitgestellt. Unsere Momentenbedingungen (3.4.6) bis (3.4.10) sind für die deterministische Intensität (9.2.1) ohne weiteres erfüllt. Die kumulative Intensität ist außerdem streng monoton wachsend, da $g > 0$ gilt. Damit ist (3.4.11) gegeben. Unsere letzte Bedingung (3.4.12) benötigen wir, um in Lemma 8.2.3 zu zeigen, daß die standardisierende Matrix invertierbar ist. Eine entsprechende Annahme trifft [6] auf Seite 16 unten, wenn durch die Determinante D dividiert wird.

[6, p. 17, Lemma A.1] soll die Identifizierbarkeitsbedingung [6, Lemma 3.2, p. 6] verschärfen. Wir benötigen jedoch auch in unserem Beweis zur asymptotischen Normalität in Kapitel 8 lediglich die schwächere Voraussetzung (3.2.7). Es erscheint deshalb fraglich, ob Lemma A.1 in [6] tatsächlich notwendig ist. Der in [6, pp. 17] angegebene Beweis ist außerdem fehlerhaft. Beim Übergang zu Gleichung (A.1) auf Seite 18 wurde (in der dort verwendeten Notation) der additive Restterm $2(\Lambda(\vartheta_\circ + h, t) - \Lambda(\vartheta_\circ, t))\langle h, \dot{\Lambda}(\vartheta_\circ, t) \rangle$ der Größenordnung $o(|h|)$ vergessen. Damit schlägt aber die folgende Beschränkung auf Werte $h/|h|$ der Einheitskreislinie fehl. Es ist unklar, inwieweit das Kompaktheitsargument noch beibehalten werden kann, das bei (A.3) beginnt und direkt vor (A.4) endet.

Der Beweis zur asymptotischen Normalität kann in vier inhaltliche Schritte untergliedert werden, die gut mit unseren in Kapitel 8 vergleichbar sind.

1. "Bild des KQS unter einer linearen Transformation": Der Ausgangspunkt beider Beweise ist die Minimalitätseigenschaft des KQS. Daraus resultiert eine Bedingung an die erste Ableitung der Abstandsfunktion zwischen beobachteten Prozessen und Modellintensität. Diese wird sowohl von [6, (4.4), (4.5), p. 8] als auch von uns in Lemma 8.3.3 ausgenutzt. Wir erhalten so das Bild von $n^{1/2}(\vartheta_n^* - \vartheta_\circ)$ unter einer linearen Transformation. Dabei bezeichnet ϑ_n^* den KQS und ϑ_\circ den wahren Parameter. Allerdings hängen sowohl die lineare Transformation als auch das Bild zunächst weiterhin vom KQS ab, vgl. unsere Größe auf der rechten Seite von (8.3.3) sowie bei [6] die Gleichung (4.7) auf Seite 8 und die Gleichung über (4.20) auf Seite 13.
2. "Übergang vom KQS zum wahren Parameter": Im zweiten Beweisschritt wird gezeigt, daß der KQS in der eben angesprochenen Transformation und im Bild asymptotisch durch den wahren Parameterwert ϑ_\circ ersetzt werden darf. [6] zeigt dies in den Lemmata 4.2 bis 4.7 auf den Seiten 8 bis 15 und bekommt schließlich [6, (4.19) und (4.20), pp. 12]. Wir verwenden die Straffheit der entstehenden Restterme und erhalten so (8.3.5).
3. "Schwache Konvergenz gegen Normalverteilung": Wir verfügen nun über approximative Darstellungen der linearen Transformation und des Bildes von $n^{1/2}(\vartheta_n^* - \vartheta_\circ)$, die nicht mehr vom KQS selbst abhängen. Allerdings variieren sie weiterhin mit der Stichprobengröße n . [6, p. 17 oben] vollzieht den Grenzübergang in n . Dabei kommt wesentlich [6, p. 20, Lemma A.2] zum Einsatz. Dieses Lemma beschreibt die asymptotische Verteilung eines Integrals des Martingals, das in unserer Notation mit $\bar{N}_n(\cdot) - \bar{\Lambda}_{\vartheta_\circ, n}(\cdot)$ bezeichnet wird. Im Beweis wird mehrfach verwendet, daß die beobachteten Prozesse Poissonprozesse sind, und zwar sowohl die resultierende Unabhängigkeit der Zuwächse als auch die Tatsache, daß die Intensität deterministisch ist. Der Beweis beruht letztlich wie unser auf Straffheitsargumenten. [6, p. 21] zitiert sie aus [45, p. 380, Theorem 22]. Den Abschluß des Beweises zur asymptotischen Normalität bildet bei [6, p. 17] dann ein ZGWS für integrierte Brownsche Bewegungen. Die bei uns in (8.3.5) auftretenden Intensitäten sind zufällig, wir ersetzen sie in Schritt (8.3.7) durch deterministische Größen. Anstatt einen ZGWS für integrierte Prozesse zu verwenden, ziehen wir uns mit Fubini in (8.3.6) auf Größen zurück, die mit dem klassischen ZGWS für reelle Zufallsvariablen behandelt werden können.
4. "Auflösen nach dem KQS": Abschließend muß nachgewiesen werden, daß die lineare Transformation, der wir die Differenz $n^{1/2}(\vartheta_n^* - \vartheta_\circ)$ bisher unterworfen haben, invertierbar ist. Wir zeigen dies im Anschluß an (8.3.10). Beim Vergleich unserer Voraussetzungen mit denen von [6] hatten wir bereits darauf hingewiesen, daß [6, p. 16 unten] durch die Determinante der linearen Transformation dividiert und damit deren Invertierbarkeit voraussetzt.

Der Beweiskgang von [6] zeigt auch, daß die Sprungstelle der *differentiellen* Inten-

sität für den Kleinst-Quadrate-Schätzer eigentlich unbedeutend ist. Dies wird dadurch bekräftigt, daß die asymptotische Normalität dieses Schätzers weiterhin gegeben ist. Hierin unterscheidet sich der KQS vom Maximum-Likelihood-Schätzer und vom Bayes-Schätzer, vgl. [6, pp. 3] und die dort angeführte Literatur. Als Grund für dieses besondere Verhalten des KQS können wir ansehen, daß eine "einfache" Unstetigkeit der differentiellen Intensität in t relevante Glattheitseigenschaften der kumulativen Intensität nicht stört. Letztere ist aber entscheidend, denn der KQS minimiert *ihren* Abstand vom Stichprobenmittel.

9.3 Die Ergebnisse von [51, pp. 58-68]

Bei [51] handelt es sich um eine Monographie, die Regressionsmodelle für Poissonprozesse behandelt. In diesem engeren Rahmen geht sie weit über die von uns erzielten Ergebnisse hinaus. Neben dem KQS werden auch der Maximum-Likelihood-Schätzer und der Bayes-Schätzer behandelt. Der Konsistenz und asymptotischen Normalität werden asymptotische Entwicklungen mit Gliedern höherer Ordnung zur Seite gestellt. Außerdem werden fehlspezifizierte und nicht identifizierbare Modelle untersucht. Schließlich werden auch nichtparametrische Schätzverfahren angesprochen. Die von uns verallgemeinerten Resultate finden wir zu Beginn des Abschnitts 2.2, [51, pp. 58-68].

[51, p. 65] formuliert das Theorem 2.8, welches die schwache Konsistenz, die asymptotische Normalität sowie die Konvergenz der Momente für den KQS liefert. Die ersten beiden Resultate können mit unseren beiden Theoremen verglichen werden.

Der Beweisgang kann wieder in die vier prinzipiellen Schritte zerlegt werden, die wir bereits im letzten Abschnitt verwendet haben. Bei der folgenden Darstellung von [51, pp. 58-68] wollen wir zwei Aspekte untersuchen:

1. Die Voraussetzungen sind bei [51] stärker an den Notwendigkeiten der Beweistechnik orientiert, während wir unsere Bedingungen vorrangig für das Modell formuliert haben. Deshalb ist ein direkter Vergleich der Voraussetzungen schwierig. Wir werden stattdessen nachweisen, daß unsere Annahmen ähnliche Zwecke im Beweis erfüllen wie die von [51].
2. Da die ersten zwei Aussagen von [51, p. 65, Theorem 2.8] als Spezialfälle unserer Theoreme für Poissonprozesse angesehen werden können, liegt die Frage nahe, ob die Beweisideen von [51] in einfacher Weise verallgemeinert werden können, um unsere Ergebnisse zu erzielen. Wir hatten bereits angedeutet, daß der prinzipielle Beweisgang ähnliche Schritte vollzieht wie unser. Wir werden jedoch belegen, daß sich die eingesetzten Techniken einer offensichtlichen Übertragung auf "allgemeine Punktprozesse" in unserem Sinne widersetzen. Dies gilt in besonderem Maße, wenn der Übergang zu bedingten Intensitäten angestrebt wird, der die Martingaleigenschaft aufgibt.

Mit den vorbereitenden Resultaten für Poissonprozesse stellt [51, pp. 17-30] wesentliche Hilfsmittel für seine späteren Aussagen bereit, insbesondere eine Maximalungleichung ([51, pp. 22, Lemma 1.3 und Korollar]) und einen ZGWS

([51, pp. 24, Theoreme 1.1 und 1.2]) für Prozesse, die in unserer Notation $\int_0^t f(s)(N(ds) - \Lambda_\circ(ds))$ heißen. Die angegebenen, kurzen Beweise dieser Hilfsergebnisse beruhen ihrerseits auf den charakteristischen Funktionen [51, p. 18, (1.4) und (1.5)] der betroffenen Prozesse. An diese Größen gelangen wir jedoch nur, wenn wir die Verteilung der Zuwächse $N(t) - N(s)$ für beliebige $t > s$ angeben können. Dies ist für Poissonprozesse leicht möglich, für allgemeine Punktprozesse jedoch nicht mehr offensichtlich.

Für den Beweis der Maximalungleichung [51, pp. 22, (1.16) und (1.17)] ist wohl noch eine Verallgemeinerung möglich, die direkt auf Doob's Maximalungleichung [8, p. 404, (46.2')] aufbaut. Sobald wir jedoch bedingte Intensitäten anstelle von $\Lambda_\circ(\cdot)$ betrachten wollen, verlieren wir die Martingaleigenschaft von $N(\cdot) - \Lambda_\circ(\cdot)$ und damit auch dieses Hilfsmittel.

Eine ähnliche Einschätzung gilt für den ZGWS [51, pp. 24, Theoreme 1.1 und 1.2]. Diese Aussage läßt sich i. w. ebenfalls auf Punktprozesse mit stochastischen Kompensatoren übertragen, wenn auch mit deutlich mehr Aufwand, vgl. z. B. [31, p. 204, Theorem 5.1.1]. Allerdings wird dann von der Martingalstruktur essentiell Gebrauch gemacht, so daß bedingte Intensitäten nicht in offensichtlicher Weise erfaßt werden können.

Die Voraussetzung [51, p. 59, C1] zur schwachen Konsistenz übernimmt und verschärft die Rolle unserer Bedingung (3.2.7) zur Identifizierbarkeit. Auf Differenzierbarkeit der kumulativen Intensität verzichtet [51, p. 59] hingegen zunächst ausdrücklich.

Wir betrachten nun kurz den Beweis der schwachen Konsistenz gemäß [51, pp. 60, Theorem 2.6]. Er beginnt in ähnlicher Weise wie [6, p. 5] und unser Theorem 3.3.1 mit einer Ungleichungskette der abzuschätzenden Wahrscheinlichkeit. An ihrem Ende beruft sich [51] auf die bereits angesprochene Maximalungleichung [51, p. 23, (1.17)].

[51, pp. 60] verzichtet weiterhin auf Differenzierbarkeitsbedingungen. Er formuliert zunächst [51, p. 61, Theorem 2.7], das den "schwachen Konvergenzschritt" des Beweisganges vorwegnimmt, den wir in Abschnitt 9.2 als Schritt 3 betrachtet hatten. Seine zusätzlichen Annahmen C2 bis C4 sind an den technischen Notwendigkeiten dieses allgemein geführten Beweises ausgerichtet. Insbesondere [51, p. 61 (2.42) und C4] setzen exakt die Bedingungen voraus, die Straffheit der Prozeßfolge nach sich ziehen. [51] verwendet hierbei wie [6] ein Resultat aus [45], nämlich [45, Theorem 1.A.21], das er als [51, p. 254, Lemma A.1] in den Anhang aufnimmt.

Es ist bemerkenswert, daß die Momentenbedingung [51, p. 61 (2.42)] zur Straffheit mehrdimensionaler Prozesse von deren Dimension abhängt. Dies konnten wir in Satz A.5.2 vermeiden, indem wir den Beweisgang von [11] übertragen haben.

Die Voraussetzung C2, (2.40) übernimmt die Rolle unserer Lipschitzbedingungen an das Modell, C3, (2.41) quantifiziert die Identifizierbarkeitsbedingung. C4 schließlich fordert Konvergenz endlichdimensionaler Randverteilungen, wie sie zum Nachweis der Straffheit einer Prozeßfolge benötigt wird.

[51, p. 61, Theorem 2.7] zeigt insbesondere, daß aus der Straffheit der betrachte-

ten Prozeßfolge mehrere asymptotische Verhaltensweisen des Schätzers folgen: Erstens weist er die Konvergenz des Schätzers gegen den wahren Parameterwert in Wahrscheinlichkeit nach. Für die standardisierte Differenz von Schätzer und wahren Parameterwert wird die Verteilungskonvergenz und die Konvergenz der Momente gezeigt. Die Rolle der Grenzverteilung wird dabei von der Verteilung des Minimums des vorausgesetzten Grenzprozesses übernommen.

Wie bereits erwähnt, führt [51, pp. 62] im Beweis des Theorems 2.7 die Straffheit der betrachteten Prozeßfolge auf die Bedingungen von [51, p. 254, Lemma A.1] zurück. Diese Voraussetzungen weist er in den Lemmata 2.5 und 2.6 nach. Der Beweis des Lemmas 2.5 ist sehr kurz, da die Voraussetzung C2, (2.40) bereits für Produkte mittlerer Intensitäten formuliert ist. Eine adäquate Entsprechung für stochastische Intensitäten ist nicht offensichtlich, da dann insbesondere der Erwartungswert dieser Produkte nicht mehr trivial abgeschätzt werden kann. Vermutlich sind aufwendige Betrachtungen, wie wir sie in den Anhängen A.2 und A.3 ausgeführt haben, auch hier notwendig, um den Beweisgang auf allgemeine Punktprozesse zu übertragen.

Der Beweis des Lemmas 2.6 beruht wiederum auf der Maximalungleichung (1.17), die wir bereits als problematisches Hilfsmittel bei Beweisen für allgemeine Punktprozesse mit bedingten Intensitäten ausgemacht hatten.

Mit [51, p. 65, Theorem 2.8] wird der Fall "glatter" Intensitäten untersucht. Die Betrachtung beschränkt sich nicht auf eine Distanz für den KQS, sondern läßt eine von der Stichprobengröße abhängige Distanzenfamilie zu, die allerdings von einem endlichen Maß dominiert werden muß. Außerdem wird der Definitionsbereich der Poissonprozesse nicht auf reelle Intervalle eingeschränkt, auch mehrdimensionale Bereiche sind zugelassen. Die Berechnung des KQS erfolgt aber unter Verwendung einer "scanning family". Darunter versteht [51, p. 59] eine aufsteigende Familie von Mengen, die den Definitionsbereich der Prozesse ausschöpfen und ihrerseits mit $[0, 1]$ indiziert werden können. Damit wird der Beweis i. w. für kompakte reelle Intervalle geführt.

Die Bedingung C5 fordert stetige Differenzierbarkeit der Modellintensität nach dem Parameter. In der Folge wird aus den abgeleiteten Intensitäten die standardisierende Matrix $\phi_n(\vartheta)^{-1}$ konstruiert, deren positive Definitheit ebenfalls vorausgesetzt wird. Damit entfällt der Beweisschritt, den wir in Abschnitt 9.2 mit Schritt 4, "Auflösen nach dem KQS", bezeichnet hatten. (2.52) und (2.53) stellen letztlich weitere Forderungen an die Matrix $\phi_n(\vartheta)$, die sich auf gleichmäßige Betragsschranken über dem Parameterraum beziehen.

Die Bedingung C6 formuliert mit (2.54) und (2.55) Voraussetzungen an gewisse Grenzwerte, die gemäß [51, p. 67 oben] endliche Momente der standardisierten Differenz von Schätzer und wahren Parameterwert nach sich ziehen: "Condition (2.55) and Lemma 2.2 yield [...] the estimate (2.50) for the random vector [...]" Die Annahme (2.55) garantiert dabei die Identifizierbarkeit des wahren Parameters, die wir durch (3.2.7) gewährleisten. (2.54) ist eine Glattheitsbedingung an die abgeleiteten Intensitäten in der Nähe des wahren Parameters, die mit unserer Annahme (3.4.1) bzgl. der $C^2(\Theta)$ -Stetigkeit vergleichbar ist.

Die Bedingung C7 fordert die gleichmäßige Konvergenz einer gewissen, auf den Modellintensitäten basierenden Integralfolge gegen eine reguläre Matrix \mathbf{S}_ϑ . Die-

se Matrix wird die Rolle der asymptotischen Kovarianzmatrix übernehmen.

Die Aussagen des Theorems 2.8 haben nun dieselbe Gestalt wie die von Theorem 2.7, allerdings wird die Grenzverteilung jetzt mit $N(0, \mathbf{S}_\vartheta)$ spezifiziert. In der Notation von [51] ist ϑ der wahre Parameter. Da der Beweis relativ knapp vorgestellt wird, werden wir auch im weiteren die Originalnotation von [51] verwenden. Auf diese Weise sollten sich unsere Kommentare leichter zuordnen lassen.

Wie angekündigt, beginnt der Beweis mit einem Schritt, den wir in Abschnitt 9.2 als Schritt 1 mit "Bild des KQS unter einer linearen Transformation" bezeichnet hatten. In der Bezeichnung von [51] heißt die Funktion, die von $\mathbf{u}_n^* := \phi_n(\vartheta)^{-1}(\vartheta_n^* - \vartheta)$ minimiert wird, $\|Y_\vartheta^{(n)}(\mathbf{u}, \cdot)\|$. Diese Funktion wird nun um $\mathbf{u} = 0$, also um den wahren Parameterwert, entwickelt.

Der Schritt 2, "Übergang vom KQS zum wahren Parameter", wird zunächst in der Notation vorbereitet. Dabei wird statt $\|Y_\vartheta^{(n)}(\mathbf{u}, \cdot)\|$ das in \mathbf{u} linearisierte Optimierungsproblem betrachtet. Dies entsteht, wenn die Entwicklung von $\|Y_\vartheta^{(n)}(\mathbf{u}, \cdot)\|$ nach dem linearen Glied abgeschnitten wird. [51] bezeichnet es mit $\|Y_\vartheta^{(0)}(\mathbf{u}, \cdot)\|$. Seine Lösung kann explizit angegeben werden, vgl. [51, p. 65, unten]. Sie wird mit ζ_n notiert. Im Hauptteil des Beweises, [51, p. 66 und 67], wird nun gezeigt, daß die Regularitätsforderungen die L_2 -Konvergenz von $\mathbf{u}_n^* - \zeta_n$ gegen Null nach sich ziehen.

Unser Schritt 3, "Schwache Konvergenz gegen Normalverteilung", läuft damit auf den Nachweis hinaus, daß ζ_n in Verteilung gegen $N(0, \mathbf{S}_\vartheta)$ konvergiert. Dies gelingt [51, p. 68, oben] unter Verwendung des eingangs erwähnten, als Hilfsresultat für Poissonprozesse bereitgestellten ZGWS mit Hilfe der Ljapunov-Bedingung [51, p. 25, Remark 1.2] sehr einfach, da für ζ_n eine explizite Darstellung vorliegt, die sich mit Fubini als Integral bzgl. $\bar{N}_n(\cdot) - \bar{\Lambda}_{\vartheta, n}(\cdot)$ (in unserer Notation) darstellen läßt. Der Nachweis der Ljapunov-Bedingung wird sich für stochastische Intensitäten vermutlich auch deutlich aufwendiger gestalten.

Zusammengefaßt verwendet [51] zum Beweis der Theoreme 2.6 bis 2.8 bereits an sehr grundlegenden Stellen Ergebnisse und Techniken, die nur für Poissonprozesse in dieser Einfachheit zu erwarten sind. Die Verallgemeinerung auf stochastische Intensitäten im Rahmen der Martingalthorie scheint möglich, wird die Beweise aber deutlich verlängern. Sobald bedingte Intensitäten betrachtet werden sollen, verliert man mit diesen Beweistechniken jedoch endgültig die Flexibilität, die wir bewahren konnten.

Wir schließen die Diskussion von [51, pp. 58-68] mit einer Bemerkung zum Charakter der von [51, pp. 64] angegebenen Voraussetzungen C5 bis C7. Die von [51, p. 64] getroffene Einschätzung "We propose several sufficient conditions that can be verified in problems with smooth enough functions $\Lambda_\vartheta^{(n)}(\cdot)$ (with respect to ϑ)." mag zwar prinzipiell stimmen. Die Formulierung dieser Bedingungen ist jedoch eher abstrakt, an den Beweistechniken orientiert und wenig konstruktiv. Ich bin der Meinung, daß die in den Abschnitten 3.2 und 3.4 formulierten Bedingungen für konkret angegebene Intensitätsmodelle in Anwendungen jedenfalls erheblich einfacher zu überprüfen sind.

Übrigens weist [51, pp. 196, Remark 5.1] explizit darauf hin, daß das asympto-

tische Verhalten des KQS, wie es in [51, p. 61, Theorem 2.7] beschrieben ist, nicht durch Sprünge der differentiellen Intensität in t gestört wird. Dies hatten wir bereits am Ende des letzten Abschnittes plausibel gemacht.

10 Zusammenfassung und Ausblick

10.1 Zusammenfassung

Im einleitenden Abschnitt 1.1 haben wir als Ziel dieser Arbeit formuliert, eine Klasse von Regressionsmodellen zu diskutieren, die zur Modellierung moderner Kaufverhaltenstheorien in besonderem Maße geeignet erscheint. Diese Theorien legen großen Wert auf die Wechselwirkung etablierten Kaufverhaltens und velseitiger verkaufsfördernder Maßnahmen im zeitlichen Ablauf.

Da der theoretische Apparat äußerst umfangreich ist, haben wir uns von Anfang an darauf beschränkt, die Eignung der vorgestellten Modellklasse an einem Beispiel zu untersuchen. Auf ökonomischer Seite haben wir uns dabei auf die TV-Werbewirkung zurückgezogen, deren theoretische Grundlagen wir in Abschnitt 1.2 kurz präsentiert haben. Aus den Schwerpunkten dieser Theorien ergab sich, daß die Verwendung disaggregierter Daten Vorteile verspricht, vgl. Abschnitt 1.5. Aus modelltechnischer Sicht haben wir uns dann die in der angewandten Literatur verbreitete Zerlegung des Kaufverhaltens in Kaufzeitpunkt, Marken- und Mengenwahl zu eigen gemacht. Um unseren Beitrag übersichtlich zu halten, haben wir uns im Anschluß auf den ersten Aspekt beschränkt.

Eine detaillierte Durchsicht der veröffentlichten Kaufverhaltensmodelle ergab dann in Kapitel 2, daß sich die bisher verwendeten Modellklassen im Rahmen von Punktprozessen gut vergleichen lassen. Die verfügbaren Modellklassen schöpfen diesen Rahmen bei weitem nicht aus. Einerseits schränken sie den Anwender bei der Beschreibung der Einkaufsakte im Zeitablauf auf Poisson- oder Erneuerungsprozesse ein, deren Intensität entweder nur von der Zeit selbst oder nur vom zeitlichen Abstand zum letzten Einkaufsakt abhängt. Andererseits sind der Modellierung des Einflusses zeitveränderlicher Marketingaktivitäten enge Grenzen gesetzt.

Beide Restriktionen vertragen sich sehr schlecht mit den verhaltensorientierten Theorien, die von diesen Modellen abgebildet werden sollen.

Ein Blick in die verfügbare stochastische Literatur zur statistischen Modellierung mit Punktprozessen in Kapitel 9 hat uns den Eindruck vermittelt, daß die theoretische Forschung in dem Sinne speziell ist, als sie sich an verfügbaren Datenquellen in spezifischen Anwendungen orientiert. So haben wir z. B. asymptotische Ergebnisse gefunden, die auf der fortdauernden Beobachtung eines einzigen Punktprozesses beruhen, für den Stationarität oder Periodizität angenommen wird. Bei der Untersuchung gefilterter Poissonprozesse stellt die Theorie z. B. einen Rahmen zur Verfügung, um von einem beobachteten Punktprozeß mittels eines parametrischen Modells auf den Verlauf eines unbeobachteten Einflußprozesses zu schließen.

Unsere theoretischen Resultate hingegen betreffen grundlegende Eigenschaf-

ten der Regressionsmodelle für Punktprozesse, die sich stärker an "klassischer" asymptotischer Statistik orientieren. So beweisen wir mit unseren Theoremen 3.3.1 und 3.5.1 die starke Konsistenz und asymptotische Normalität eines Kleinst-Quadrate-Schätzers für ein parametrisches Modell der stochastischen Intensität von i. i. d. Punktprozessen, die gemeinsam mit ihren Einflußgrößen auf demselben, kompakten Intervall beobachtet werden. Die Asymptotik bezieht sich dabei auf den wachsenden Stichprobenumfang, wie aus Regressionsmodellen für reelle Zufallsgrößen bekannt. Damit verallgemeinern wir Erkenntnisse, die bisher nur für Poissonprozesse ([51]) und empirische Prozesse ([72]) vorlagen.

Zum Beweis der starken Konsistenz verwenden wir dabei an zentraler Stelle Satz A.4.1, ein starkes Gesetz für Funktionale stochastischer Prozesse, dessen Beweis für U-Statistiken seit langem Eingang in die Lehrbücher gefunden hat.

Der Beweis der asymptotischen Normalität beruht ebenfalls auf klassischen Ideen der asymptotischen Statistik. Wir nutzen hier die grundlegende Eigenschaft aus Lemma A.1.1, die aus der Straffheit der "Restprozesse" in der i. i. d.-Darstellung des Kleinst-Quadrate-Schätzers folgt. Die Straffheit dieser Prozesse zeigen wir mit der hinreichenden Eigenschaft aus Satz A.5.2. Dabei spielt die i. i. d.-Struktur *zwischen* den beobachteten Punktprozessen die entscheidende Rolle. Die Martingalstruktur, die für die Differenz jedes einzelnen Punktprozesses und seiner Intensität vorliegt, können wir völlig vernachlässigen; lediglich die Zentriertheit dieser Prozesse wird benötigt. Diesem Umstand verdanken wir letztlich die Gültigkeit unserer Theoreme auch bei der Modellierung bedingter Intensitäten, wie sie im Rahmen der Berücksichtigung "unbeobachteter Heterogenität" von vielen Anwendern in der Ökonometrie favorisiert wird.

Insgesamt wird aus angewandter Sicht deutlich, daß sich die verhaltens- und zeitorientierten Hypothesen zum Kaufverhalten in dem von uns untersuchten stochastischen Rahmen prinzipiell sehr natürlich formalisieren lassen. Wir mußten uns die relevanten stochastischen Eigenschaften dieser Modellklasse in der vorliegenden Arbeit allerdings von Grund auf erschließen. Daher geht unser eigenes Kaufverhaltensmodell in Kapitel 4 nur die ersten Schritte bei der Umsetzung der umfangreichen, theoretischen Literatur zur TV-Werbewirkung. Wir beschränken uns darauf, am Beispiel der saisonalen Warengruppe der Haushaltspackungen Speiseeis die naheliegendsten Effekte von Kaufverhalten, Jahreszeit und TV-Werbung von einem Modell "auseinanderdividieren" zu lassen. Dies gelingt auf Anhieb in plausibler Weise bei 14 der untersuchten 17 Haushaltsgruppen.

10.2 Ausblick

Umfangreiche Verfeinerungen unseres Modells im Sinne der Kaufverhaltenstheorie sind offensichtlich möglich. Zu den naheliegenden gehört die explizite Berücksichtigung weiterer verkaufsfördernder Maßnahmen neben der TV-Werbung und die Differenzierung der Wirkung einzelner Kontakte anhand ihrer Position in der "Werbekontakthistorie" einzelner Haushalte.

Der von uns erarbeitete theoretische Rahmen ermöglicht jedoch noch deutlich aufwendigere, praktische Verallgemeinerungen des Modells. So können wir z. B.

frühere Einkaufsakte heranziehen, etwa die Länge des letzten Kaufintervalls.

Desweiteren ist es prinzipiell möglich, Interaktionen zwischen der Vergangenheit der betrachteten Prozesse zu modellieren, auch unter Einbeziehung des beeinflussten Kaufzeitpunktprozesses selbst. Auf diese Weise ergeben sich direkt Möglichkeiten, z. B. die TV-Werbewirkung für Haushalte zu unterscheiden, die die beworbene Warengruppe zuvor häufig, selten oder gar nicht gekauft haben. Anders als bei den bisher veröffentlichten Modellen, die sich am einzelnen Kaufintervall statt an der gesamten Kaufhistorie eines Haushaltes orientieren, können wir insbesondere "starke TV-Werbewirkungstheorien" im Sinne des Abschnitts 1.2 untersuchen, da Haushalte keine Einkäufe zu tätigen brauchen, um in unserem Modell angemessen berücksichtigt zu werden.

Die weitreichendste theoretische Verallgemeinerung, die ohne substantielle Änderung der Beweisstrukturen möglich sein sollte, ist die Betrachtung markierter Punktprozesse mit Markenräumen \mathbb{N}^q oder sogar \mathbb{R}_+^q , $q \in \mathbb{N}$. Eine derartige Erweiterung sollte die theoretisch saubere und praktisch äußerst flexible, gemeinsame Modellierung von Kaufzeitpunkten, Markenwahl und sogar Mengenwahl in *einem* Regressionsmodell ermöglichen. Entsprechende Modelle sind augenblicklich nur als "Semi-Markov-Modelle", also markierte Erneuerungsprozesse im Rahmen der competing risks untersucht.

Selbstverständlich verlangt die praktische Schätzung dieser Modelle durch Computerprogramme einige Kapazität. Die Möglichkeiten des von uns vorgestellten R-Programms haben wir bei der Untersuchung von Gruppen mit etwa 150 bis 200 Haushalten auf einem Laptop bereits ausgeschöpft. Die schärfste Restriktion scheint darin zu bestehen, daß diese frei verfügbare Statistik-Software alle Datenobjekte im Hauptspeicher hält. Insofern müßte bereits die Verwendung kommerzieller Pakete, die die Auslagerung von Daten auf die Festplatte einfacher ermöglichen, hier deutliche Fortschritte bringen.

A Technische Lemmata

A.1 Konvergenz in Wahrscheinlichkeit

In diesem Abschnitt stellen wir eine relevante Eigenschaft der Konvergenz in Wahrscheinlichkeit bereit.

A.1.1 Lemma *Die stochastische Folge X_n von Zufallsvektoren in \mathbb{R}^q , $q \in \mathbb{N}$, konvergiere in Wahrscheinlichkeit gegen den Vektor $Y \in \mathbb{R}^q$:*

$$X_n \xrightarrow{P} Y$$

Es sei $f_n : \mathbb{R}^q \rightarrow \mathbb{R}$ eine Folge stochastischer Prozesse, die auf jeder kompakten Teilmenge $K \subset \mathbb{R}^q$ C-traff ist. Dann gilt:

$$f_n(X_n) - f_n(Y) \xrightarrow{P} 0$$

Beweis:

Wir haben

$$\forall \varepsilon > 0: \lim_{n \rightarrow \infty} \mathbb{P}(|f_n(X_n) - f_n(Y)| > \varepsilon) = 0$$

zu zeigen. Seien $\delta > 0$ und $\eta > 0$. Wir zeigen

$$\lim_{n \rightarrow \infty} \mathbb{P}(|f_n(X_n) - f_n(Y)| > \varepsilon) \leq \eta.$$

Sei $K_\eta \in \mathbb{R}^q$ eine kompakte Menge mit $\mathbb{P}(Y \notin K_\eta) \leq \eta$. Eine Zerlegung der Menge $\{|f_n(X_n) - f_n(Y)| > \varepsilon\}$ mit Hilfe der Ereignisse $\{|X_n - Y| > \delta\}$ und $\{Y \in K_\eta\}$ liefert:

$$\begin{aligned} \mathbb{P}(|f_n(X_n) - f_n(Y)| > \varepsilon) &= \mathbb{P}(|f_n(X_n) - f_n(Y)| > \varepsilon, |X_n - Y| \leq \delta, Y \in K_\eta) \\ &\quad + \mathbb{P}(|f_n(X_n) - f_n(Y)| > \varepsilon, |X_n - Y| \leq \delta, Y \notin K_\eta) \\ &\quad + \mathbb{P}(|f_n(X_n) - f_n(Y)| > \varepsilon, |X_n - Y| > \delta) \\ &\leq \mathbb{P}\left(\sup_{\substack{|x-y| \leq \delta \\ y \in K_\eta}} |f_n(x) - f_n(y)| > \varepsilon\right) \\ &\quad + \mathbb{P}(Y \notin K_\eta) \\ &\quad + \mathbb{P}(|X_n - Y| > \delta) \end{aligned}$$

Wegen der vorausgesetzten Straffheit der Folge f_n konvergiert der erste Term gegen Null, vgl. [11, S. 58, (8.11)].¹¹ Dasselbe folgt für den dritten Term aus der stochastischen Konvergenz von X_n gegen Y . Der zweite Term ist nach Konstruktion durch η nach oben beschränkt. ■

A.2 Pfadweise Integrale

In diesem Abschnitt arbeiten wir auf Lemma A.2.3 hin. Dieses Lemma ermöglicht die Berechnung von Erwartungswerten pfadweise gebildeter Integrale, sofern Integrand und integrierendes Maß gewissen Regularitätsbedingungen genügen.

Wir zeigen zunächst eine wichtige Hilfsaussage für Lemma A.2.3. Dabei geht es um die Approximation von Integralen mit Riemann-Summen.

A.2.1 Situation Seien $X(\cdot), Y(\cdot)$ und $Z(\cdot) \equiv Z^+(\cdot) - Z^-(\cdot)$ stochastische Prozesse auf dem kompakten Intervall $I = [t_{\min}, t_{\max}] \subset \mathbb{R}$. Die Prozesse $Z^+(\cdot)$ und $Z^-(\cdot)$ seien f. s. positiv, rechtsstetig und monoton wachsend, so daß wir sie als zufällige, unsignierte Maße auf I ansehen können. Die Prozesse $X(\cdot)$ und $Y(\cdot)$ seien f. s. außerhalb der endlichen stochastischen Mengen $J_X, J_Y \subset I$ stetig. Für $t \in J_X$ oder $t \in J_Y$ seien für $X(t)$ bzw. $Y(t)$ Sprünge erster Ordnung zugelassen. Dann existiert

$$\int_I |X(t) \cdot Y(t)| (Z^+(dt) + Z^-(dt)) \in \mathbb{R} \cup \{+\infty\}.$$

Zum Nachweis können

$$\int_I |X(t) \cdot Y(t)| Z^+(dt)$$

¹¹In [11] wird lediglich $K_\eta = [0, 1]$ betrachtet; der Beweis benutzt aber nur die Kompaktheit und kann deshalb direkt übertragen werden.

und

$$\int_I |X(t) \cdot Y(t)| Z^-(dt)$$

einzelnen betrachtet werden. In beiden Fällen existiert das (Lebesgue-)Integral, da der Integrand $|X(t) \cdot Y(t)|$ additiv in einen stetigen Teil und eine Treppenfunktion zerfällt.

Wir führen noch den Positiv- und Negativteil einer reellwertigen Funktion f ein. Dabei handelt es sich um positive, reellwertige Funktionen, die wie folgt definiert sind:

$$\begin{aligned} f_+ &:= \max(f, 0) \\ f_- &:= -\min(f, 0) \end{aligned}$$

Schließlich bezeichne

$$\{t_{in}, i = 1, \dots, q(n)\} \subset \{t_{i,n+1}, i = 1, \dots, q(n+1)\} \subset I \quad n \in \mathbb{N}$$

eine nicht notwendig deterministische, jedoch immer aufsteigende Folge von Zerlegungen des Intervalls I . Für jedes $n \in \mathbb{N}$ gelte dabei $q(n) \in \mathbb{N}$ und $t_{\min} = t_{1,n} < \dots < t_{q(n),n} = t_{\max}$.

A.2.2 Lemma *In der Situation A.2.1 gelte*

$$Z^+(J_X) = Z^-(J_X) = 0 \quad f. s. \quad (\text{A.2.1})$$

sowie

$$\mathbb{E} \int_I |X(t) \cdot Y(t)| (Z^+(dt) + Z^-(dt)) < \infty. \quad (\text{A.2.2})$$

Dann folgt

$$\begin{aligned} \lim_{n \rightarrow \infty} \mathbb{E} \sum_{i=2}^{q(n)} & \left(\inf_{t_{i-1,n} \leq t < t_{in}} X_+(t) - \inf_{t_{i-1,n} \leq t < t_{in}} X_-(t) \right) \\ & \left(\inf_{t_{i-1,n} \leq t < t_{in}} Y_+(t) - \inf_{t_{i-1,n} \leq t < t_{in}} Y_-(t) \right) (Z(t_{in}) - Z(t_{i-1,n})) \\ & = \mathbb{E} \int_I X(t) Y(t) Z(dt) \end{aligned}$$

mit einer Folge von Y -meßbaren Zerlegungen $\{t_{in}\} \subset I$. Wenn wir zusätzlich

$$Z^+(J_Y) = Z^-(J_Y) = 0 \quad f. s. \quad (\text{A.2.3})$$

voraussetzen, kann die Folge $\{t_{in}\}$ deterministisch gewählt werden.

Beweis:

Wir beweisen die Aussage des Lemmas zunächst unter (A.2.3). Der nach Voraussetzung (A.2.2) endliche Ausdruck

$$\begin{aligned} & \mathbb{E} \int_I |X(t)Y(t)| (Z^+(dt) + Z^-(dt)) \\ &= \mathbb{E} \int_I (X_+(t) + X_-(t))(Y_+(t) + Y_-(t)) (Z^+(dt) + Z^-(dt)) \end{aligned}$$

zerfällt durch Auflösen der drei Klammerterme in acht beschränkte Summanden. Wir untersuchen lediglich einen davon, für den wir

$$\begin{aligned} \lim_{n \rightarrow \infty} \mathbb{E} \sum_{i=1}^{q(n)} \left(\inf_{t_{i-1,n} \leq t < t_{in}} X_+(t) \inf_{t_{i-1,n} \leq t < t_{in}} Y_+(t) \right) (Z^+(t_{in}) - Z^+(t_{i-1,n})) \\ = \mathbb{E} \int_I X_+(t) Y_+(t) Z^+(dt) \quad (\text{A.2.4}) \end{aligned}$$

zeigen werden. Die anderen Summanden erlauben entsprechende Approximationen; die Aussage des Lemmas folgt daraus durch Linearkombination.

Zum Nachweis von (A.2.4) betrachten wir die Treppenfunktion

$$V^n(t) := \sum_{i=1}^{q(n)} \inf_{t_{i-1,n} \leq s < t_{in}} X_+(s) \inf_{t_{i-1,n} \leq s < t_{in}} Y_+(s) \mathbf{1}_{\{t_{i-1,n} \leq t < t_{in}\}}.$$

Die aufsteigende Folge $\{t_{in}, i = 1, \dots, q(n)\} \subset I$ Zerlegungen wählen wir deterministisch. Als Erstes zeigen wir, daß $V^n(t)$ punktweise monoton wächst:

$$0 \leq V^n(t) \leq V^{n+1}(t) \quad n \in \mathbb{N}, t \in I \quad \text{f. s.} \quad (\text{A.2.5})$$

Wir fixieren dazu $t \in I$ und wählen passende Indices $i = 2, \dots, q(n)$ sowie $j = 2, \dots, q(n+1)$, mit

$$\begin{aligned} t_{i-1,n} \leq t < t_{in}, \\ t_{j-1,n+1} \leq t < t_{j,n+1}. \end{aligned}$$

Nach Definition der Zerlegungen folgt

$$[t_{j-1,n+1}, t_{j,n+1}) \subset [t_{i-1,n}, t_{in})$$

und somit

$$\begin{aligned} 0 \leq \inf_{t \in [t_{i-1,n}, t_{in})} X_+(t) &\leq \inf_{t \in [t_{j-1,n+1}, t_{j,n+1})} X_+(t), \\ 0 \leq \inf_{t \in [t_{i-1,n}, t_{in})} Y_+(t) &\leq \inf_{t \in [t_{j-1,n+1}, t_{j,n+1})} Y_+(t). \end{aligned}$$

Insgesamt folgt

$$\begin{aligned} 0 \leq V^n(t) &= \inf_{t \in [t_{i-1,n}, t_{in})} X_+(t) \inf_{t \in [t_{i-1,n}, t_{in})} Y_+(t) \\ &\leq \inf_{t \in [t_{j-1,n+1}, t_{j,n+1})} X_+(t) \inf_{t \in [t_{j-1,n+1}, t_{j,n+1})} Y_+(t) \\ &= V^{n+1}(t). \end{aligned}$$

Damit ist (A.2.5) gezeigt.

Als Nächstes weisen wir nach, daß die Folge der $V^n(t)$ außerhalb der Z -Nullmenge $J_X \cup J_Y$ punktweise den Integranden $X_+(t)Y_+(t)$ approximiert:

$$n \rightarrow \infty : \quad V^n(t) \nearrow X_+(t)Y_+(t) \quad t \in I \setminus (J_X \cup J_Y) \quad \text{f. s.} \quad (\text{A.2.6})$$

Da die Prozesse X und Y außerhalb der abgeschlossenen Menge $J_X \cup J_Y$ stetig sind, gibt es zu jedem Punkt $t \in I \setminus (J_X(\omega) \cup J_Y(\omega))$ eine positive Zufallsvariable $\varepsilon(\omega) > 0$, so daß X_+ und Y_+ auf $(t - \varepsilon, t + \varepsilon)$ f. s. stetig sind. Dies hat zur Folge, daß sowohl X_+ als auch Y_+ auf $(t - \varepsilon, t + \varepsilon)$ jeweils ihr Infimum f. s. annehmen. Für genügend großes $n \in \mathbb{N}$ und geeignetes $i = i(n) \in \{2, \dots, q(n)\}$ gilt

$$t \in [t_{i-1,n}, t_{in}) \subset (t - \varepsilon, t + \varepsilon)$$

und deshalb

$$\underline{t}^n := \arg \inf_{t \in [t_{i-1,n}, t_{in})} X_+(t)Y_+(t) \in [t_{i-1,n}, t_{in}).$$

Für wachsendes $n \in \mathbb{N}$ zieht sich das Intervall $[t_{i-1,n}, t_{in})$ auf den Punkt t zusammen. Damit folgt

$$\underline{t}^n \rightarrow t.$$

Auf dem relevanten Teilintervall ist X_+Y_+ stetig. Deshalb erhalten wir

$$\begin{aligned} V^n(t) &= \inf_{t \in [t_{i-1,n}, t_{in})} X_+(t)Y_+(t) \\ &= X_+(\underline{t}^n)Y_+(\underline{t}^n) \\ &\nearrow X_+(t)Y_+(t). \end{aligned}$$

Damit ist (A.2.6) gezeigt.

Da es sich bei V^n um eine Z^+ -f. ü. monoton konvergente Folge integrierbarer Funktionen auf I handelt, liefert der Satz von Beppo Levi:

$$\sup_{n \in \mathbb{N}} \int_I V^n(t) Z^+(dt) = \int_I X_+(t)Y_+(t) Z^+(dt) < \infty,$$

vgl. z. B. [7, S. 68, Satz 11.4 sowie S. 70, Satz 11.6]. Insbesondere ist die Integralfolge $\int_I V^n(t) Z^+(dt)$ beschränkt. Wegen (A.2.5) und der Monotonie des Integrals ist sie zudem monoton wachsend. Insgesamt folgt

$$0 \leq \int_I V^n(t) Z^+(dt) \leq \int_I V^{n+1}(t) Z^+(dt) \quad n \in \mathbb{N} \quad \text{f. s.}$$

und

$$n \rightarrow \infty : \quad \int_I V^n(t) Z^+(dt) \nearrow \int_I X_+(t)Y_+(t) Z^+(dt) \quad \text{f. s.}$$

Wieder wenden wir den Satz von Beppo Levi an, diesmal auf die Integration nach P . Wir erhalten

$$\lim_{n \rightarrow \infty} \mathbb{E} \int_I V^n(t) Z^+(dt) = \mathbb{E} \int_I X_+(t)Y_+(t) Z^+(dt)$$

und damit (A.2.4). Der Beweis des Lemmas unter (A.2.3) ist damit beendet.

Wenn wir die Annahme (A.2.3) streichen, ist die Ausnahmemenge $J_X \cup J_Y$ in (A.2.6) keine Z -Nullmenge mehr. Wir können nun jedoch erreichen, daß (A.2.6) außerhalb der Z -Nullmenge J_X gilt, wenn wir die Folge $\{t_{in}\}$ Y -meßbar wählen und dabei $J_Y \subset \{t_{in}\}$ sicherstellen. Mit dieser Modifikation kann der restliche Beweis unverändert übernommen werden. ■

A.2.3 Lemma *Wir betrachten die Situation A.2.1 mit den beiden zusätzlichen Voraussetzungen (A.2.1) und (A.2.2).*

1. Sei $X(t)$ unabhängig von $\sigma(Y(s), Z(s), s \leq t)$ für alle $t \in I$. Zusätzlich seien die Bedingungen

$$\mathbb{E} \int_I |E(X(t)) \cdot Y(t)| (Z^+(dt) + Z^-(dt)) < \infty, \quad (\text{A.2.7})$$

$$\mathbb{E}(\sup_{t \in I} |X(t)|) < \infty$$

und

$$P(t \in J_X) = 0 \quad \forall t \in I$$

erfüllt. Dann folgt

$$\mathbb{E} \int_I X(t) \cdot Y(t) Z(dt) = \mathbb{E} \int_I E(X(t)) \cdot Y(t) Z(dt).$$

Die angegebenen Integrale existieren.

2. Sei $Z(t)$ unabhängig von $\sigma(X(s), Y(s), s \leq t)$ für alle $t \in I$. Außerdem gelte (A.2.3) sowie

$$\mathbb{E} \int_I |X(t) \cdot Y(t)| (E Z^+(dt) + E Z^-(dt)) < \infty, \quad (\text{A.2.8})$$

und

$$E Z(t) < \infty \quad \text{für } t \in I.$$

Dann folgt

$$\mathbb{E} \int_I X(t) \cdot Y(t) Z(dt) = \mathbb{E} \int_I X(t) \cdot Y(t) E(Z(dt)).$$

Die angegebenen Integrale existieren.

Natürlich spielt die Produktstruktur des Integranden für die Gültigkeit des Teils 2 keine Rolle. Die Darstellung wurde lediglich gewählt, um die Analogie zu Teil 1 zu verdeutlichen.

Um Voraussetzungen der Form (A.2.2), (A.2.7) oder (A.2.8) nachzuweisen, beachten wir (7.2.3).

Beweis von Lemma A.2.3:

Wir weisen Teil 1 des Lemmas nach. Wir zeigen zunächst, daß Lemma A.2.2 auch auf die Prozesse EX, Y und Z anwendbar ist.

$EX(t)$ ist stetig, denn zu festem $t \in I$ zieht jede Folge $s \rightarrow t$ wegen $P(t \in J_X) = 0$ einerseits eine f. s. konvergente Folge $X(s) \rightarrow X(t)$ und andererseits eine durch $E(\sup_{t \in I} |X(t)|) < \infty$ beschränkte Folge $EX(s)$ nach sich. Lebesgues Lemma (vgl. etwa [7, S. 95, Satz 15.6]) liefert

$$EX(s) \rightarrow EX(t).$$

Die Voraussetzung (A.2.7) entspricht (A.2.2).

Der weitere Beweisgang wird einen Zusammenhang zwischen den Gliedern der approximierenden Folgen

$$\begin{aligned} \lim_{n \rightarrow \infty} E \sum_{i=2}^{q(n)} & \left(\inf_{t_{i-1,n} \leq t < t_{in}} X_+(t) - \inf_{t_{i-1,n} \leq t < t_{in}} X_-(t) \right) \\ & \left(\inf_{t_{i-1,n} \leq t < t_{in}} Y_+(t) - \inf_{t_{i-1,n} \leq t < t_{in}} Y_-(t) \right) (Z(t_{in}) - Z(t_{i-1,n})) \\ & = E \int X(t)Y(t)Z(dt) \quad (\text{A.2.9}) \end{aligned}$$

mit Y -meßbaren Zerlegungen $\{t_{in}\}$ und

$$\begin{aligned} \lim_{n \rightarrow \infty} E \sum_{i=2}^{q(n)} & \left(\inf_{t_{i-1,n} \leq t < t_{in}} EX_+(t) - \inf_{t_{i-1,n} \leq t < t_{in}} EX_-(t) \right) \\ & \left(\inf_{t_{i-1,n} \leq t < t_{in}} Y_+(t) - \inf_{t_{i-1,n} \leq t < t_{in}} Y_-(t) \right) (Z(t_{in}) - Z(t_{i-1,n})) \\ & = E \int E(X(t))Y(t)Z(dt) \quad (\text{A.2.10}) \end{aligned}$$

gemäß Lemma A.2.2 nachweisen. Im ersten Schritt werden wir für die Glieder der ersten Folge

$$\begin{aligned} E \sum_{i=2}^{q(n)} & \left(\inf_{t_{i-1,n} \leq t < t_{in}} X_+(t) - \inf_{t_{i-1,n} \leq t < t_{in}} X_-(t) \right) \\ & \left(\inf_{t_{i-1,n} \leq t < t_{in}} Y_+(t) - \inf_{t_{i-1,n} \leq t < t_{in}} Y_-(t) \right) (Z(t_{in}) - Z(t_{i-1,n})) \\ & = E \sum_{i=2}^{q(n)} E \left[\inf_{t_{i-1,n} \leq t < t_{in}} X_+(t) - \inf_{t_{i-1,n} \leq t < t_{in}} X_-(t) \middle| \sigma(Y(s), Z(s), s \leq t_{in}) \right] \\ & \left(\inf_{t_{i-1,n} \leq t < t_{in}} Y_+(t) - \inf_{t_{i-1,n} \leq t < t_{in}} Y_-(t) \right) (Z(t_{in}) - Z(t_{i-1,n})) \quad (\text{A.2.11}) \end{aligned}$$

zeigen. Im zweiten Schritt weisen wir nach, daß in jedem Punkt $t \notin J_X$ bei Wahl der Y -meßbaren Indices

$$i = i(t, n) = 2, \dots, q(n) \quad \text{mit} \quad t_{i(t,n)-1,n} \leq t < t_{i(t,n),n}$$

die beiden Folgen

$$\mathbb{E} \left[\inf_{t_{i-1,n} \leq s < t_{i,n}} X_{+/-}(s) \middle| \sigma(Y(s), Z(s), s \leq t_{i,n}) \right] \nearrow \mathbb{E} X_{+/-}(t) \quad \forall t \notin J_X \quad (\text{A.2.12})$$

und

$$\inf_{t_{i-1,n} \leq s < t_{i,n}} \mathbb{E} X_{+/-}(s) \nearrow \mathbb{E} X_{+/-}(t) \quad \forall t \notin J_X \quad (\text{A.2.13})$$

für wachsendes $n \in \mathbb{N}$ monoton demselben Grenzwert zustreben. Also können wir in (A.2.11) den inneren Erwartungswert und das Infimum über X vertauschen sowie die bedingenden σ -Algebren streichen, ohne die monotone Konvergenz gemäß des Satzes von Beppo Levi zu beeinflussen. Insgesamt haben wir dann gezeigt, daß die beiden Folgen (A.2.9) und (A.2.10) denselben Limes haben. Daraus folgt Teil 1 des Lemmas.

Wir beginnen mit dem Nachweis von (A.2.11):

$$\begin{aligned} & \mathbb{E} \sum_{i=2}^{q(n)} \left(\inf_{t_{i-1,n} \leq t < t_{i,n}} X_+(t) - \inf_{t_{i-1,n} \leq t < t_{i,n}} X_-(t) \right) \\ & \quad \left(\inf_{t_{i-1,n} \leq t < t_{i,n}} Y_+(t) - \inf_{t_{i-1,n} \leq t < t_{i,n}} Y_-(t) \right) (Z(t_{i,n}) - Z(t_{i-1,n})) \\ &= \sum_{i=2}^{q(n)} \mathbb{E} \left(\left(\inf_{t_{i-1,n} \leq t < t_{i,n}} X_+(t) - \inf_{t_{i-1,n} \leq t < t_{i,n}} X_-(t) \right) \right. \\ & \quad \left. \left(\inf_{t_{i-1,n} \leq t < t_{i,n}} Y_+(t) - \inf_{t_{i-1,n} \leq t < t_{i,n}} Y_-(t) \right) (Z(t_{i,n}) - Z(t_{i-1,n})) \right) \end{aligned}$$

Nun fügen wir einen bedingten Erwartungswert ein:

$$\begin{aligned} [\dots] &= \sum_{i=2}^{q(n)} \mathbb{E} \left(\mathbb{E} \left[\left(\inf_{t_{i-1,n} \leq t < t_{i,n}} X_+(t) - \inf_{t_{i-1,n} \leq t < t_{i,n}} X_-(t) \right) \right. \right. \\ & \quad \left. \left(\inf_{t_{i-1,n} \leq t < t_{i,n}} Y_+(t) - \inf_{t_{i-1,n} \leq t < t_{i,n}} Y_-(t) \right) (Z(t_{i,n}) - Z(t_{i-1,n})) \right. \\ & \quad \left. \left. \middle| \sigma(Y(s), Z(s), s \leq t_{i,n}) \right] \right) \end{aligned}$$

Ein Teil des Ausdrucks ist bzgl. der bedingenden σ -Algebra meßbar und kann deshalb aus dem bedingten Erwartungswert gezogen werden. Wir beachten dabei die Y -Meßbarkeit der $\{t_{i,n}\}$:

$$\begin{aligned} [\dots] &= \sum_{i=2}^{q(n)} \mathbb{E} \left(\mathbb{E} \left[\inf_{t_{i-1,n} \leq t < t_{i,n}} X_+(t) - \inf_{t_{i-1,n} \leq t < t_{i,n}} X_-(t) \middle| \sigma(Y(s), Z(s), s \leq t_{i,n}) \right] \right. \\ & \quad \left. \left(\inf_{t_{i-1,n} \leq t < t_{i,n}} Y_+(t) - \inf_{t_{i-1,n} \leq t < t_{i,n}} Y_-(t) \right) \right. \\ & \quad \left. (Z(t_{i,n}) - Z(t_{i-1,n})) \right) \end{aligned}$$

Im Anschluß kann der unbedingte Erwartungswert aus der Summe gezogen werden. Damit ist (A.2.11) nachgewiesen.

Wir zeigen (A.2.12). Da t f. s. nicht in J_X liegt, gibt es eine stochastische Größe $n(\omega) \in \mathbb{N}$, so daß für $m \geq n$ und geeignetes $i = i(m) \in \{2, \dots, q(m)\}$ einerseits $t \in [t_{i-1,m}, t_{im})$ gilt und andererseits $X_{+/-}$ auf $[t_{i-1,m}, t_{im})$ f. s. stetig ist. Dann nimmt $X_{+/-}$ auf $[t_{i-1,m}, t_{im})$ sein Infimum an und wir haben

$$\underline{t}^m := \arg \inf_{t_{i-1,m} \leq s < t_{im}} X_{+/-}(s) \in [t_{i-1,m}, t_{im}).$$

Da sich $[t_{i-1,m}, t_{im})$ mit wachsendem m auf t zusammenzieht, gilt $\underline{t}^m \rightarrow t$ f. s.. Da $X_{+/-}$ f. s. im relevanten Teilintervall stetig ist, folgt

$$\inf_{t_{i-1,m} \leq s < t_{im}} X_{+/-}(s) = X_{+/-}(\underline{t}^m) \xrightarrow{\text{f. s.}} X_{+/-}(t).$$

Die Konvergenz der bedingten Erwartungswerte folgt wie oben mit Lebesgues Lemma. Da der Grenzwert von der bedingenden σ -Algebra unabhängig ist, kann letztere gestrichen werden. Dies schließt den Nachweis von (A.2.12) ab.

Schließlich ist noch (A.2.13) nachzuweisen. Da $E X_{+/-}$ stetig ist, erhalten wir sofort

$$\underline{t}^n := \arg \inf_{t_{i-1,n} \leq s < t_{in}} E X_{+/-}(s) \in [t_{i-1,n}, t_{in})$$

bei geeignetem $i = i(n) \in \{2, \dots, q(n)\}$. Wieder gilt $\underline{t}^n \rightarrow t$ und

$$\inf_{t_{i-1,n} \leq s < t_{in}} E X_{+/-}(s) = E X_{+/-}(\underline{t}^n) \rightarrow E X_{+/-}(t).$$

Wir haben (A.2.13) und damit Teil 1 des Lemmas gezeigt.

Wir wenden uns nun Teil 2 des Lemmas zu. Analog zum Beweis des Teils 1 beginnen wir mit dem Nachweis, daß Lemma A.2.2 auch auf die Prozesse X, Y und $E Z$ anwendbar ist.

$E Z$ ist rechtsstetig, denn wir haben die Zerlegung

$$\begin{aligned} E Z &= E(Z^+ - Z^-) \\ &= E Z^+ - E Z^-, \end{aligned} \tag{A.2.14}$$

in welcher $E Z^{+/-}$ wie $Z^{+/-}$ selbst positiv und monoton wachsend sind. Wegen (A.2.14) wird für jede Folge $s_n \searrow t \in I$ die Folge der Werte $E Z^{+/-}(s_n) \geq 0$ durch $E Z^{+/-}(s_1) < \infty$ dominiert. Lebesgues Lemma liefert

$$\lim_{s \searrow t} E Z^{+/-}(s) = E Z^{+/-}(t),$$

vgl. etwa [7, S. 95, Satz 15.6].

Weiterhin haben wir

$$E Z^{+/-}(J_X) = E Z^{+/-}(J_Y) = 0,$$

da dies f. s. für $Z^{+/-}$ gilt.

Schließlich entspricht die Voraussetzung (A.2.8) gerade (A.2.2).

Wir folgen weiterhin der Beweisidee des Teils 1 und zeigen, daß sich die Glieder der approximierenden Folgen

$$\begin{aligned} \lim_{n \rightarrow \infty} \mathbb{E} \sum_{i=2}^{q(n)} & \left(\inf_{t_{i-1,n} \leq t < t_{in}} X_+(t) - \inf_{t_{i-1,n} \leq t < t_{in}} X_-(t) \right) \\ & \left(\inf_{t_{i-1,n} \leq t < t_{in}} Y_+(t) - \inf_{t_{i-1,n} \leq t < t_{in}} Y_-(t) \right) (Z(t_{in}) - Z(t_{i-1,n})) \\ & = \mathbb{E} \int X(t)Y(t) Z(dt) \end{aligned}$$

und

$$\begin{aligned} \lim_{n \rightarrow \infty} \mathbb{E} \sum_{i=2}^{q(n)} & \left(\inf_{t_{i-1,n} \leq t < t_{in}} X_+(t) - \inf_{t_{i-1,n} \leq t < t_{in}} X_-(t) \right) \\ & \left(\inf_{t_{i-1,n} \leq t < t_{in}} Y_+(t) - \inf_{t_{i-1,n} \leq t < t_{in}} Y_-(t) \right) (\mathbb{E} Z(t_{in}) - \mathbb{E} Z(t_{i-1,n})) \\ & = \mathbb{E} \int X(t)Y(t) \mathbb{E} Z(dt) \end{aligned}$$

gemäß Lemma A.2.2 entsprechen. Dann stimmen wieder die beiden Grenzwerte überein und Teil 2 ist ebenfalls nachgewiesen. Die Zerlegungen $\{t_{in}\}$ dürfen aufgrund der zusätzlichen Voraussetzung (A.2.3) deterministisch gewählt werden.

Tatsächlich gilt

$$\begin{aligned} & \mathbb{E} \sum_{i=2}^{q(n)} \left(\inf_{t_{i-1,n} \leq t < t_{in}} X_+(t) - \inf_{t_{i-1,n} \leq t < t_{in}} X_-(t) \right) \\ & \quad \left(\inf_{t_{i-1,n} \leq t < t_{in}} Y_+(t) - \inf_{t_{i-1,n} \leq t < t_{in}} Y_-(t) \right) (Z(t_{in}) - Z(t_{i-1,n})) \\ & = \sum_{i=2}^{q(n)} \mathbb{E} \left(\inf_{t_{i-1,n} \leq t < t_{in}} X_+(t) - \inf_{t_{i-1,n} \leq t < t_{in}} X_-(t) \right) \\ & \quad \left(\inf_{t_{i-1,n} \leq t < t_{in}} Y_+(t) - \inf_{t_{i-1,n} \leq t < t_{in}} Y_-(t) \right) (Z(t_{in}) - Z(t_{i-1,n})) \end{aligned}$$

Wieder fügen wir einen bedingten Erwartungswert ein:

$$\begin{aligned} [\dots] & = \sum_{i=2}^{q(n)} \mathbb{E} \left(\mathbb{E} \left[\left(\inf_{t_{i-1,n} \leq t < t_{in}} X_+(t) - \inf_{t_{i-1,n} \leq t < t_{in}} X_-(t) \right) \right. \right. \\ & \quad \left. \left. \left(\inf_{t_{i-1,n} \leq t < t_{in}} Y_+(t) - \inf_{t_{i-1,n} \leq t < t_{in}} Y_-(t) \right) (Z(t_{in}) - Z(t_{i-1,n})) \right] \right. \\ & \quad \left. \left| \sigma(X(s), Y(s), s \leq t_{in}) \right] \right) \end{aligned}$$

Ein Teil des Ausdrucks ist wiederum bzgl. der bedingenden σ -Algebra meßbar

und kann deshalb aus dem bedingten Erwartungswert gezogen werden:

$$[\dots] = \sum_{i=2}^{q(n)} \mathbb{E} \left(\left(\inf_{t_{i-1,n} \leq t < t_{in}} X_+(t) - \inf_{t_{i-1,n} \leq t < t_{in}} X_-(t) \right) \right. \\ \left. \left(\inf_{t_{i-1,n} \leq t < t_{in}} Y_+(t) - \inf_{t_{i-1,n} \leq t < t_{in}} Y_-(t) \right) \right. \\ \left. \mathbb{E}[(Z(t_{in}) - Z(t_{i-1,n})) | \sigma(X(s), Y(s), s \leq t_{in})] \right)$$

Der Ausdruck, der im bedingten Erwartungswert verbleibt, ist anders als im Beweis des Teils 1 von der bedingenden σ -Algebra unabhängig. Letztere kann also gestrichen werden:

$$[\dots] = \sum_{i=2}^{q(n)} \mathbb{E} \left(\left(\inf_{t_{i-1,n} \leq t < t_{in}} X_+(t) - \inf_{t_{i-1,n} \leq t < t_{in}} X_-(t) \right) \right. \\ \left. \left(\inf_{t_{i-1,n} \leq t < t_{in}} Y_+(t) - \inf_{t_{i-1,n} \leq t < t_{in}} Y_-(t) \right) \right. \\ \left. (\mathbb{E} Z(t_{in}) - \mathbb{E} Z(t_{i-1,n})) \right) \\ = \mathbb{E} \sum_{i=2}^{q(n)} \left(\inf_{t_{i-1,n} \leq t < t_{in}} X_+(t) - \inf_{t_{i-1,n} \leq t < t_{in}} X_-(t) \right) \\ \left(\inf_{t_{i-1,n} \leq t < t_{in}} Y_+(t) - \inf_{t_{i-1,n} \leq t < t_{in}} Y_-(t) \right) \\ (\mathbb{E} Z(t_{in}) - \mathbb{E} Z(t_{i-1,n}))$$

Wie oben erläutert, haben wir Teil 2 damit nachgewiesen. \blacksquare

A.3 L_2 -Techniken

In diesem Abschnitt stellen wir Aussagen zusammen, die hilfreich sind, um für gewisse Summen die Konvergenz im quadratischen Mittel nachzuweisen. Dabei werden wir über spezielle Indexkombinationen summieren, für die wir zunächst Kurzschreibweisen definieren:

$$\{k \leq p, q\} := \{k < p; k < q\} \cup \{k > p; k > q\} \quad (\text{A.3.1})$$

$$\{i \leq k \leq p, q \leq i\} := \{p < k < i; q < k < i\} \cup \{p < i < k; q < i < k\} \\ \cup \{k < p < i; k < q < i\} \cup \{i < p < k; i < q < k\} \\ \cup \{k < i < p; k < i < q\} \cup \{i < k < p; i < k < q\} \quad (\text{A.3.2})$$

Wir veranschaulichen diese Definitionen kurz an einem Beispiel: In der Menge (A.3.1) ist etwa die Kombination $p < k < q$ ausgeschlossen.

Im folgenden sei $d \in \mathbb{N}$. Ein Vektor $U \in \mathbb{R}^d$ sei als Spaltenvektor zu verstehen, U^\top bezeichne den durch Transposition aus U entstehenden Zeilenvektor. Zu zwei Vektoren $U, V \in \mathbb{R}^d$ ist dann $U^\top V \in \mathbb{R}$ das Skalarprodukt.

A.3.1 Lemma Für die \mathbb{R}^d -wertigen Zufallsvariablen U_{pk} , $p, k = 1, \dots, n$, gelte

$$\mathbb{E}|U_{pk}|^2 < \infty$$

sowie

$$\begin{aligned} \mathbb{E}(U_{pk}^\top U_{ql}) = 0 & \quad \text{für } k \neq p, q, l \\ & \quad \text{oder } l \neq p, k, q. \end{aligned} \quad (\text{A.3.3})$$

Dann folgt

$$\mathbb{E} \left| \sum_{\substack{k,p=1 \\ k \neq p}}^n U_{pk} \right|^2 \leq 2 \sum_{\substack{k,p,q=1 \\ k \leq p,q}}^n \mathbb{E}(U_{pk}^\top U_{qk}).$$

Beweis:

Es gilt

$$\begin{aligned} \mathbb{E} \left| \sum_{k \neq p} U_{pk} \right|^2 &= \mathbb{E} \left| \sum_{k < p} U_{pk} + \sum_{k > p} U_{pk} \right|^2 \\ &\leq \mathbb{E} \left(\left| \sum_{k < p} U_{pk} \right| + \left| \sum_{k > p} U_{pk} \right| \right)^2 \\ &\leq 2 \mathbb{E} \left| \sum_{k < p} U_{pk} \right|^2 + 2 \mathbb{E} \left| \sum_{k > p} U_{pk} \right|^2 \end{aligned} \quad (\text{A.3.4})$$

Wir betrachten den ersten Erwartungswert:

$$\begin{aligned} \mathbb{E} \left| \sum_{k < p} U_{pk} \right|^2 &= \mathbb{E} \left(\left(\sum_{k < p} U_{pk} \right)^\top \left(\sum_{k < p} U_{pk} \right) \right) \\ &= \sum_{\substack{k < p \\ l < q}} \mathbb{E}(U_{pk}^\top U_{ql}) \\ &= \sum_{\substack{l < k < p \\ l < q}} \mathbb{E}(U_{pk}^\top U_{ql}) + \sum_{\substack{k < p \\ k < l < q}} \mathbb{E}(U_{pk}^\top U_{ql}) + \sum_{\substack{k < p \\ k < q}} \mathbb{E}(U_{pk}^\top U_{qk}) \end{aligned}$$

In den Summanden der ersten Summe unterscheidet sich der Wert des Index l von den drei anderen, in der zweiten Summe gilt das Entsprechende für den Index k . Gemäß (A.3.3) verschwinden beide Summen und wir erhalten

$$\mathbb{E} \left(\sum_{k < p} U_{pk} \right)^2 = \sum_{k < p, q} \mathbb{E}(U_{pk}^\top U_{qk}).$$

Ganz analog erhalten wir

$$\mathbb{E} \left(\sum_{k > p} U_{pk} \right)^2 = \sum_{k > p, q} \mathbb{E}(U_{pk}^\top U_{qk}).$$

Setzen wir beides in (A.3.4) ein, so erhalten wir die Aussage des Lemmas. \blacksquare

A.3.2 Lemma Für die \mathbb{R}^d -wertigen Zufallsvariablen U_{ki} , $k, i = 1, \dots, n$, gelte $\mathbb{E}|U_{ki}|^2 < \infty$ sowie

$$\begin{aligned} \mathbb{E}(U_{ki}^\top U_{lj}) &= 0 && \text{für } k \neq i, l, j \\ &&& \text{oder } i \neq k, l, j \\ &&& \text{oder } l \neq k, i, j \\ &&& \text{oder } j \neq k, i, l. \end{aligned} \quad (\text{A.3.5})$$

Dann gilt

$$\mathbb{E} \left| \sum_{\substack{k,i=1 \\ k \neq i}}^n U_{ki} \right|^2 \leq 2 \sum_{\substack{k,i=1 \\ k \neq i}}^n \mathbb{E} |U_{ki}|^2.$$

Beweis:

Wieder gilt

$$\begin{aligned} \mathbb{E} \left| \sum_{k \neq i} U_{ki} \right|^2 &= \mathbb{E} \left| \sum_{k < i} U_{ki} + \sum_{k > i} U_{ki} \right|^2 \\ &\leq 2 \mathbb{E} \left| \sum_{k < i} U_{ki} \right|^2 + 2 \mathbb{E} \left| \sum_{k > i} U_{ki} \right|^2 \end{aligned} \quad (\text{A.3.6})$$

Wir betrachten wiederum den ersten Erwartungswert:

$$\begin{aligned} \mathbb{E} \left| \sum_{k < i} U_{ki} \right|^2 &= \mathbb{E} \left(\left(\sum_{k < i} U_{ki} \right)^\top \left(\sum_{k < i} U_{ki} \right) \right) \\ &= \sum_{\substack{k < i \\ l < j}} \mathbb{E}(U_{ki}^\top U_{lj}) \\ &= \sum_{\substack{l < k < i \\ l < j}} \mathbb{E}(U_{ki}^\top U_{lj}) + \sum_{\substack{k < i \\ k < l < j}} \mathbb{E}(U_{ki}^\top U_{lj}) + \sum_{\substack{k < i \\ k < j}} \mathbb{E}(U_{ki}^\top U_{kj}) \\ &= \sum_{\substack{k < i \\ k < j}} \mathbb{E}(U_{ki}^\top U_{kj}) \\ &= \sum_{\substack{k < i \\ k < j < i}} \mathbb{E}(U_{ki}^\top U_{kj}) + \sum_{\substack{k < i < j \\ k < j}} \mathbb{E}(U_{ki}^\top U_{kj}) + \sum_{k < i} \mathbb{E} |U_{ki}|^2 \\ &= \sum_{k < i} \mathbb{E} |U_{ki}|^2 \end{aligned}$$

Dabei haben wir alle vier Teile der Voraussetzung (A.3.5) je einmal angewendet. Ganz analog erhalten wir

$$\mathbb{E} \left| \sum_{k > i} U_{ki} \right|^2 = \sum_{k > i} \mathbb{E} |U_{ki}|^2.$$

Die Aussage des Lemmas folgt, wenn wir beide Ausdrücke in (A.3.6) einsetzen. ■

A.3.3 Lemma Die \mathbb{R}^d -wertigen Zufallsvariablen U_{pki} seien i. i. d. für $p, k, i = 1, \dots, n$. Außerdem gelte $\mathbb{E}|U_{pki}|^2 < \infty$ sowie

$$\begin{aligned} \mathbb{E}(U_{pki}^\top U_{qlj}) &= 0 && \text{für } k \neq p, i, q, l, j \quad (\text{A.3.7}) \\ &&& \text{oder } i \neq p, k, q, l, j \\ &&& \text{oder } l \neq p, k, i, q, j \\ &&& \text{oder } j \neq p, k, i, q, l. \end{aligned}$$

Dann gilt

$$\mathbb{E} \left| \sum_{\substack{p, k, i=1 \\ p \neq k \neq i \neq p}}^n U_{pki} \right|^2 \leq 64 \sum_{\substack{p, q, k, i=1 \\ i \leq k \leq p, q \leq i}}^n \mathbb{E}(U_{pki}^\top U_{qki}).$$

Beweis:

Der Beweis verläuft im Prinzip genau wie der des Lemmas A.3.2, durch die Dreifachsumme verlängert er sich allerdings etwas. Wir starten mit der Abschätzung gemäß (8.2.25):

$$\begin{aligned} \mathbb{E} \left| \sum_{p \neq k \neq i \neq p} U_{pki} \right|^2 &= \mathbb{E} \left| \begin{aligned} &\sum_{p < k < i} U_{pki} && + \sum_{p < i < k} U_{pki} \quad (\text{A.3.8}) \\ &+ \sum_{k < p < i} U_{pki} && + \sum_{i < p < k} U_{pki} \\ &+ \sum_{k < i < p} U_{pki} && + \sum_{i < k < p} U_{pki} \end{aligned} \right|^2 \\ &\leq 64 \mathbb{E} \left| \sum_{p < k < i} U_{pki} \right|^2 && + 64 \mathbb{E} \left| \sum_{p < i < k} U_{pki} \right|^2 \\ &&& + 64 \mathbb{E} \left| \sum_{k < p < i} U_{pki} \right|^2 && + 64 \mathbb{E} \left| \sum_{i < p < k} U_{pki} \right|^2 \\ &&& + 64 \mathbb{E} \left| \sum_{k < i < p} U_{pki} \right|^2 && + 64 \mathbb{E} \left| \sum_{i < k < p} U_{pki} \right|^2 \end{aligned}$$

Die sechs Erwartungswerte werden wieder alle gleich behandelt, wir betrachten lediglich den ersten:

$$\begin{aligned} \mathbb{E} \left| \sum_{p < k < i} U_{pki} \right|^2 &= \mathbb{E} \left(\left(\sum_{p < k < i} U_{pki} \right)^\top \left(\sum_{p < k < i} U_{pki} \right) \right) \\ &= \mathbb{E} \left(\sum_{\substack{p < k < i \\ q < l < j}} U_{pki}^\top U_{qlj} \right) \\ &= \mathbb{E} \left(\sum_{\substack{p < k < i < j \\ q < l < j}} U_{pki}^\top U_{qlj} \right) + \mathbb{E} \left(\sum_{\substack{p < k < i \\ q < l < j < i}} U_{pki}^\top U_{qlj} \right) \\ &+ \mathbb{E} \left(\sum_{\substack{p < k < i \\ q < l < i}} U_{pki}^\top U_{qli} \right) \end{aligned}$$

Nun nutzen wir aus, daß in den Summanden der ersten Summe der Index j und in denen der zweiten der Index i einen Wert hat, der von keinem anderen Index angenommen wird. (A.3.7) liefert:

$$\begin{aligned} [\dots] &= \mathbb{E} \left(\sum_{\substack{p < k < i \\ q < l < i}} U_{pki}^\top U_{qli} \right) \\ &= \mathbb{E} \left(\sum_{\substack{p < k < l < i \\ q < l < i}} U_{pki}^\top U_{qli} \right) + \mathbb{E} \left(\sum_{\substack{p < k < i \\ q < l < k < i}} U_{pki}^\top U_{qli} \right) \\ &\quad + \mathbb{E} \left(\sum_{\substack{p < k < i \\ q < k < i}} U_{pki}^\top U_{qki} \right) \end{aligned}$$

Wieder ist in den ersten beiden Summen ein Index isoliert, diesmal handelt es sich um l bzw. k . Beide Summen verschwinden nach Voraussetzung und wir erhalten:

$$\mathbb{E} \left| \sum_{p < k < i} U_{pki} \right|^2 = \mathbb{E} \left(\sum_{p, q < k < i} U_{pki}^\top U_{qki} \right)$$

Für die anderen fünf Erwartungswerte aus (A.3.8) folgt ganz entsprechend:

$$\begin{aligned} \mathbb{E} \left| \sum_{p < i < k} U_{pki} \right|^2 &= \mathbb{E} \left(\sum_{p, q < i < k} U_{pki}^\top U_{qki} \right) \\ \mathbb{E} \left| \sum_{k < p < i} U_{pki} \right|^2 &= \mathbb{E} \left(\sum_{k < p, q < i} U_{pki}^\top U_{qki} \right) \\ \mathbb{E} \left| \sum_{i < p < k} U_{pki} \right|^2 &= \mathbb{E} \left(\sum_{i < p, q < k} U_{pki}^\top U_{qki} \right) \\ \mathbb{E} \left| \sum_{k < i < p} U_{pki} \right|^2 &= \mathbb{E} \left(\sum_{k < i < p, q} U_{pki}^\top U_{qki} \right) \\ \mathbb{E} \left| \sum_{i < k < p} U_{pki} \right|^2 &= \mathbb{E} \left(\sum_{i < k < p, q} U_{pki}^\top U_{qki} \right) \end{aligned}$$

Ersetzen wir alle sechs Ausdrücke in (A.3.8), so folgt die gesuchte Aussage. ■

A.4 Ein starkes Gesetz für Funktionale stochastischer Prozesse

Seien

$$X_k(\cdot) : (\Omega, \mathcal{A}, \mathbb{P}) \times I \rightarrow (\mathbb{R}, \mathcal{B}) \quad k \in \mathbb{N},$$

und

$$Y_k(\cdot) : (\Omega, \mathcal{A}, \mathbb{P}) \times I \rightarrow (\mathbb{R}, \mathcal{B}) \quad k \in \mathbb{N},$$

zwei i. i. d. Folgen stochastischer Prozesse auf dem kompakten Intervall $I \subset \mathbb{R}$. Außerdem seien $X_k(\cdot)$ und $Y_j(\cdot)$ für unterschiedliche Indices $k \neq j$ unabhängig.

Die Prozesse können auch als Zufallsvariablen

$$X_k(\cdot) : (\Omega, \mathcal{A}, \mathbb{P}) \rightarrow (\mathbb{R}^I, \mathcal{B}^I) \quad k \in \mathbb{N},$$

bzw.

$$Y_k(\cdot) : (\Omega, \mathcal{A}, \mathbb{P}) \rightarrow (\mathbb{R}^I, \mathcal{B}^I) \quad k \in \mathbb{N},$$

aufgefaßt werden, vgl. etwa [8, S. 301, Theorem 35.3]. Dabei sei für zwei Mengen A und B

$$B^A := \text{Abb}(A, B) := \{f : A \rightarrow B\}$$

die Menge aller Abbildungen von A nach B . Für einen topologischen Raum B sei

$$\mathcal{B}^A := \mathcal{B}(A, B) := \mathcal{B}(\{f : A \rightarrow B\})$$

die Borel- σ -Algebra auf B^A .

Bezeichne

$$\begin{aligned} h_m &: (\text{Abb}(I, \mathbb{R}^2)^{\mathbb{N}}, \mathcal{B}(I, \mathbb{R}^2)^{\mathbb{N}}) \rightarrow (\mathbb{R}, \mathcal{B}) \\ &((f_1, g_1), (f_2, g_2), \dots) \mapsto h_m((f_1, g_1), \dots, (f_m, g_m)) \end{aligned}$$

eine reellwertiges, meßbares Funktional für Folgen von Funktionen $(f_i, g_i) : I \rightarrow \mathbb{R}^2$, $i \in \mathbb{N}$, das jedoch nur von den ersten m Folgengliedern abhängt. Damit ist h_m meßbar bzgl. $\mathcal{B}(I, \mathbb{R}^2)^m \subset \mathcal{B}(I, \mathbb{R}^2)^{\mathbb{N}}$.

Im folgenden betrachten wir die Statistik

$$\begin{aligned} R_{m,n}((f_1, g_1), \dots, (f_n, g_n)) \\ := \frac{(n-m)!}{n!} \sum_{\substack{i_1, \dots, i_m=1 \\ i_j \neq i_k}}^n h_m((f_{i_1}, g_{i_1}), \dots, (f_{i_m}, g_{i_m})) \quad (\text{A.4.1}) \end{aligned}$$

für $n \geq m$ und reelle Funktionen $f_i, g_i : I \rightarrow \mathbb{R}$, $i \in \mathbb{N}$. Ausgewertet auf den Pfaden der eingangs definierten stochastischen Prozesse, ist $R_{m,n}((X_1, Y_1), \dots, (X_n, Y_n))$ vom Typ einer U-Statistik, allerdings nehmen die Vektoren (X_{i_j}, Y_{i_j}) stochastischer Prozesse den Platz der reellwertigen Zufallsvariablen ein, vgl. etwa [67, S. 172].

In diesem Abschnitt beweisen wir ein starkes Gesetz für $R_{m,n}((X_1, Y_1), \dots, (X_n, Y_n))$ in n :

A.4.1 Satz Der Erwartungswert

$$\mathbb{E} h_m((X_1, Y_1), \dots, (X_m, Y_m)) < \infty$$

existiere. Dann gilt

$$R_{m,n}((X_1, Y_1), \dots, (X_n, Y_n)) \xrightarrow{f.s./L_1} \mathbf{E} h_m((X_1, Y_1), \dots, (X_m, Y_m))$$

für $n \rightarrow \infty$.

Der Beweis des Satzes A.4.1 folgt der Beweisidee des entsprechenden Satzes für U-Statistiken, siehe etwa [67, S. 190 ff]. Eine zentrale Rolle spielt die Darstellung von $(R_{m,n}((X_1, Y_1), \dots, (X_n, Y_n)))_{n \geq m}$ als reverses Martingal.

Bevor wir den Beweis des Satzes A.4.1 führen, stellen wir einige hilfreiche Aussagen zusammen.

Sei (G_n, \circ) die Gruppe der Permutationen auf $\{1, \dots, n\}$. Wir betrachten diese Gruppe als Teilmenge von $\text{Abb}(\mathbb{N}, \mathbb{N})$: Für $\pi \in G_n$ und $k \geq n+1$ setzen wir dazu $\pi(k) := k$. Außerdem führen wir für $j, i_1, \dots, i_j \in \mathbb{N}$ die Kurzschreibweise

$$\pi((X_{i_1}, Y_{i_1}), \dots, (X_{i_j}, Y_{i_j})) := ((X_{\pi(i_1)}, Y_{\pi(i_1)}), \dots, (X_{\pi(i_j)}, Y_{\pi(i_j)}))$$

ein.

Es bezeichne $\tilde{\mathcal{H}}_n \subset \mathcal{B}(I, \mathbb{R}^2)^n$ das Mengensystem aller Ereignisse $A \in \mathcal{B}(I, \mathbb{R}^2)^n$, die bzgl. $(X_1, Y_1), \dots, (X_n, Y_n)$ "symmetrisch" sind, so daß also

$$\begin{aligned} \tilde{\mathcal{H}}_n &:= \{A \in \mathcal{B}(I, \mathbb{R}^2)^n : ((X_1, Y_1), \dots, (X_n, Y_n)) \in A \\ &\Leftrightarrow \pi((X_1, Y_1), \dots, (X_n, Y_n)) \in A \quad \forall \pi \in G_n\} \\ &\subset \mathcal{B}(I, \mathbb{R}^2)^n \subset \mathcal{B}(I, \mathbb{R}^2)^{\mathbb{N}} \end{aligned}$$

gilt.

A.4.2 Lemma $\tilde{\mathcal{H}}_n \subset \mathcal{B}(I, \mathbb{R}^2)^n$ ist eine σ -Algebra.

Beweis:

Zunächst einmal gilt $\{\emptyset, \text{Abb}(I, \mathbb{R}^2)^n\} \subset \tilde{\mathcal{H}}_n$: Für $A := \emptyset$ ist die Bedingung in der Definition von $\tilde{\mathcal{H}}_n$ leer. Für $A := \text{Abb}(I, \mathbb{R}^2)^n$ ist

$$\begin{aligned} ((X_1, Y_1), \dots, (X_n, Y_n)) &\in \text{Abb}(I, \mathbb{R}^2)^n \\ &\Leftrightarrow \pi((X_1, Y_1), \dots, (X_n, Y_n)) \in \text{Abb}(I, \mathbb{R}^2)^n \quad \forall \pi \in G_n \end{aligned}$$

ebenfalls erfüllt.

Abzählbare Schnitte $\bigcap_{n=1}^{\infty} A_n$ meßbarer Mengen $A_n \in \tilde{\mathcal{H}}_n$ erfüllen für alle $\pi \in G_n$ die Bedingung

$$((X_1, Y_1), \dots, (X_n, Y_n)) \in \bigcap_{n=1}^{\infty} A_n \Leftrightarrow \pi((X_1, Y_1), \dots, (X_n, Y_n)) \in \bigcap_{n=1}^{\infty} A_n,$$

da nach Voraussetzung

$$((X_1, Y_1), \dots, (X_n, Y_n)) \in A_n \Leftrightarrow \pi((X_1, Y_1), \dots, (X_n, Y_n)) \in A_n \quad \forall n \in \mathbb{N}$$

gilt. Folglich ist $\tilde{\mathcal{H}}_n$ gegen abzählbare Schnittbildung abgeschlossen.

Schließlich ist noch das Komplement \bar{A} einer Menge $A \in \tilde{\mathcal{H}}_n$ zu betrachten. Nun sind aber die Bedingungen

$$\begin{aligned} ((X_1, Y_1), \dots, (X_n, Y_n)) \in \bar{A} &\Leftrightarrow \pi((X_1, Y_1), \dots, (X_n, Y_n)) \in \bar{A} \quad \forall \pi \in G_n, \\ ((X_1, Y_1), \dots, (X_n, Y_n)) \notin A &\Leftrightarrow \pi((X_1, Y_1), \dots, (X_n, Y_n)) \notin A \quad \forall \pi \in G_n \end{aligned}$$

und

$$((X_1, Y_1), \dots, (X_n, Y_n)) \in A \Leftrightarrow \pi((X_1, Y_1), \dots, (X_n, Y_n)) \in A \quad \forall \pi \in G_n$$

gleichbedeutend. Die erste ist für $\bar{A} \in \tilde{\mathcal{H}}_n$ zu zeigen, die letzte gilt jedoch wegen $A \in \tilde{\mathcal{H}}_n$. ■

Das nächste Lemma charakterisiert die reellwertigen Funktionale auf $\text{Abb}(I, \mathbb{R}^2)^n$, die bzgl. $\tilde{\mathcal{H}}_n$ meßbar sind.

A.4.3 Lemma *Ein Funktional*

$$f : (\text{Abb}(I, \mathbb{R}^2)^n, \mathcal{B}(I, \mathbb{R}^2)^n) \rightarrow (\mathbb{R}, \mathcal{B})$$

ist genau dann meßbar bzgl. $\tilde{\mathcal{H}}_n$, wenn es meßbar bzgl. $\mathcal{B}(I, \mathbb{R}^2)^n$ ist und

$$f((X_1, Y_1), \dots, (X_n, Y_n)) = f \circ \pi((X_1, Y_1), \dots, (X_n, Y_n)) \quad \forall \pi \in G_n$$

erfüllt.

Beweis:

" \Rightarrow ": Sei $f : (\text{Abb}(I, \mathbb{R}^2)^n, \mathcal{B}(I, \mathbb{R}^2)^n) \rightarrow (\mathbb{R}, \mathcal{B})$ meßbar bzgl. $\tilde{\mathcal{H}}_n$. Dann ist f insb. meßbar bzgl. $\mathcal{B}(I, \mathbb{R}^2)^n \supset \tilde{\mathcal{H}}_n$ und es gilt

$$f^{-1}(B) \in \tilde{\mathcal{H}}_n \quad \forall B \in \mathcal{B},$$

also

$$\begin{aligned} ((X_1, Y_1), \dots, (X_n, Y_n)) &\in f^{-1}(B) \\ \Leftrightarrow \pi((X_1, Y_1), \dots, (X_n, Y_n)) &\in f^{-1}(B) \quad \forall \pi \in G_n. \end{aligned}$$

Dies ist gleichbedeutend mit

$$\begin{aligned} f((X_1, Y_1), \dots, (X_n, Y_n)) &\in B \\ \Leftrightarrow f \circ \pi((X_1, Y_1), \dots, (X_n, Y_n)) &\in B \quad \forall \pi \in G_n. \end{aligned}$$

Wählen wir nun die speziellen einelementigen Borelmengen $B := \{a\} \in \mathcal{B}$, $a \in \mathbb{R}$, so folgt die Behauptung.

" \Leftarrow ": Sei $f : (\text{Abb}(I, \mathbb{R}^2)^n, \mathcal{B}(I, \mathbb{R}^2)^n) \rightarrow (\mathbb{R}, \mathcal{B})$ nun eine $\mathcal{B}(I, \mathbb{R}^2)^n$ -meßbare Funktion, die die Symmetrie-Eigenschaft

$$f((X_1, Y_1), \dots, (X_n, Y_n)) = f \circ \pi((X_1, Y_1), \dots, (X_n, Y_n)) \quad \forall \pi \in G_n$$

erfüllt. Für eine beliebige Borelmenge $B \in \mathcal{B}$ müssen wir zeigen:

$$f^{-1}(B) \in \tilde{\mathcal{H}}_n$$

Dies ist nach Definition von $\tilde{\mathcal{H}}_n$ gleichbedeutend mit:

$$\begin{aligned} & ((X_1, Y_1), \dots, (X_n, Y_n)) \in f^{-1}(B) \\ \Leftrightarrow & \pi((X_1, Y_1), \dots, (X_n, Y_n)) \in f^{-1}(B) \quad \forall \pi \in G_n \end{aligned}$$

Diese Behauptung können wir schreiben als

$$\begin{aligned} & f((X_1, Y_1), \dots, (X_n, Y_n)) \in B \\ \Leftrightarrow & f \circ \pi((X_1, Y_1), \dots, (X_n, Y_n)) \in B \quad \forall \pi \in G_n \end{aligned}$$

und dies ist nach Voraussetzung erfüllt. \blacksquare

A.4.4 Korollar Die Statistik $R_{m,n} : (\text{Abb}(I, \mathbb{R}^2)^n, \mathcal{B}(I, \mathbb{R}^2)^n) \rightarrow (\mathbb{R}, \mathcal{B})$ ist meßbar bzgl. $\tilde{\mathcal{H}}_n$.

Beweis:

Nach Konstruktion (A.4.1) handelt es sich bei $R_{m,n}$ um eine endliche Linearkombination der $\mathcal{B}(I, \mathbb{R}^2)^m$ -meßbaren Funktionale h_m für $n \geq m$. Aus $\mathcal{B}(I, \mathbb{R}^2)^m \subset \mathcal{B}(I, \mathbb{R}^2)^n$ ergibt sich unmittelbar die $\mathcal{B}(I, \mathbb{R}^2)^n$ -Meßbarkeit von $R_{m,n}$.

Aus der Definition (A.4.1) folgt außerdem

$$\begin{aligned} R_{m,n}((X_1, Y_1), \dots, (X_n, Y_n)) &= R_{m,n}((X_{\pi(1)}, Y_{\pi(1)}), \dots, (X_{\pi(n)}, Y_{\pi(n)})) \\ &= R_{m,n} \circ \pi((X_1, Y_1), \dots, (X_n, Y_n)) \end{aligned}$$

für alle Permutationen $\pi \in G_n$. Lemma A.4.3 liefert die Behauptung. \blacksquare

Wir wollen einen absteigenden Filter von σ -Algebren in (Ω, \mathcal{A}) einführen, der auf der Folge der $\tilde{\mathcal{H}}_n \subset \mathcal{B}(I, \mathbb{R}^2)^n$ basiert.

Wir beachten hierzu, daß die Folge

$$((X_1, Y_1), (X_2, Y_2), \dots) : (\Omega, \mathcal{A}, \mathbb{P}) \rightarrow (\text{Abb}(I, \mathbb{R}^2)^{\mathbb{N}}, \mathcal{B}(I, \mathbb{R}^2)^{\mathbb{N}})$$

von Prozessen (X_i, Y_i) mit $i \in \mathbb{N}$ vermöge

$$\tilde{\mathcal{G}}_n := ((X_1, Y_1), (X_2, Y_2), \dots, (X_n, Y_n))^{-1}(\tilde{\mathcal{H}}_n)$$

eine σ -Algebra $\tilde{\mathcal{G}}_n \subset \mathcal{A}$ definiert, vgl. etwa [7, S. 3, (1.5)].

Nach Konstruktion gilt:

A.4.5 Lemma Ein Funktional

$$f : (\text{Abb}(I, \mathbb{R}^2)^n, \mathcal{B}(I, \mathbb{R}^2)^n) \rightarrow (\mathbb{R}, \mathcal{B})$$

ist genau dann $\tilde{\mathcal{H}}_n$ -meßbar, wenn die Komposition

$$f((X_1, Y_1), (X_2, Y_2), \dots, (X_n, Y_n)) : (\Omega, \mathcal{A}, \mathbb{P}) \rightarrow (\mathbb{R}, \mathcal{B})$$

meßbar bzgl. $\tilde{\mathcal{G}}_n$ ist.

A.4.6 Bemerkung Sei $\tilde{f} : (\Omega, \mathcal{A}, \mathbb{P}) \rightarrow (\mathbb{R}, \mathcal{B})$ bzgl. $\tilde{\mathcal{G}}_n$ meßbar. Nach Konstruktion von $\tilde{\mathcal{G}}_n$ gilt:

$$\tilde{\mathcal{G}}_n \subset \sigma((X_1, Y_1), (X_2, Y_2), \dots, (X_n, Y_n))$$

Deshalb existiert eine Faktorisierung

$$\tilde{f} = f \circ ((X_1, Y_1), (X_2, Y_2), \dots, (X_n, Y_n)).$$

Das derart definierte Funktional

$$f : (\text{Abb}(I, \mathbb{R}^2)^n, \mathcal{B}(I, \mathbb{R}^2)^n) \rightarrow (\mathbb{R}, \mathcal{B})$$

ist $\tilde{\mathcal{H}}_n$ -meßbar gemäß Lemma A.4.5.

Mit Hilfe von $\tilde{\mathcal{G}}_n$ definieren wir nun die Folgenglieder des absteigenden Filters:

$$\mathcal{G}_n := \tilde{\mathcal{G}}_n \vee \sigma((X_i, Y_i), i \geq n+1) \subset \mathcal{A} \quad (\text{A.4.2})$$

A.4.7 Lemma Es gilt $\mathcal{G}_{n+1} \subset \mathcal{G}_n$ für $n \in \mathbb{N}$.

Beweis:

Es gilt

$$\mathcal{G}_{n+1} = \tilde{\mathcal{G}}_{n+1} \vee \sigma((X_i, Y_i), i \geq n+2).$$

Deshalb genügt es,

$$\tilde{\mathcal{G}}_{n+1} \subset \tilde{\mathcal{G}}_n \vee \sigma(X_{n+1}, Y_{n+1}) \quad (\text{A.4.3})$$

$$\subset \mathcal{G}_n \quad (\text{A.4.4})$$

und

$$\sigma((X_i, Y_i), i \geq n+2) \subset \mathcal{G}_n \quad (\text{A.4.5})$$

zu zeigen. (A.4.4) und (A.4.5) folgen direkt aus der Definition (A.4.2) von \mathcal{G}_n .

Wir haben noch (A.4.3) zu zeigen. Hierzu genügt es, die Meßbarkeit der Indikatorfunktionen $\mathbf{1}_A$ für Mengen $A \in \tilde{\mathcal{G}}_{n+1}$ bzgl. der σ -Algebra $\tilde{\mathcal{G}}_n \vee \sigma(X_{n+1}, Y_{n+1})$ nachzuweisen. Dies gelingt uns gleich für beliebige $\tilde{\mathcal{G}}_{n+1}$ -meßbare Funktionen

$$\tilde{f} : (\Omega, \mathcal{A}, \mathbb{P}) \rightarrow (\mathbb{R}, \mathcal{B}).$$

Derartige Funktionen zeichnet gemäß Lemma A.4.3 und Bemerkung A.4.6 nämlich neben ihrer \mathcal{A} -Meßbarkeit die Symmetrie-Eigenschaft

$$f((X_1, Y_1), \dots, (X_{n+1}, Y_{n+1})) = f \circ \pi((X_1, Y_1), \dots, (X_{n+1}, Y_{n+1})) \quad \forall \pi \in G_{n+1}$$

der Faktorisierung $\tilde{f} = f \circ ((X_1, Y_1), \dots, (X_{n+1}, Y_{n+1}))$ aus. Für die nachzuweisende Meßbarkeit bzgl. $\tilde{\mathcal{G}}_n \vee \sigma(X_{n+1}, Y_{n+1})$ genügt jedoch die \mathcal{A} -Meßbarkeit zusammen mit der schwächeren Symmetrie-Eigenschaft

$$\begin{aligned} f((X_1, Y_1), \dots, (X_{n+1}, Y_{n+1})) \\ = f \circ \pi((X_1, Y_1), \dots, (X_{n+1}, Y_{n+1})) \quad \forall \pi \in G_n \times \{\text{id}\}. \end{aligned}$$

Dabei werden nur noch solche Permutationen $\pi \in G_{n+1}$ herangezogen, die $\pi(n+1) = n+1$ erfüllen. ■

Die bisher gewonnenen Ergebnisse führen wir nun zu der Aussage zusammen, die entscheidend für den Beweis des Satzes A.4.1 ist:

A.4.8 Lemma Für $n \geq m$ gilt

$$\begin{aligned} R_{m,n}((X_1, Y_1), (X_2, Y_2), \dots, (X_n, Y_n)) \\ = \mathbb{E}[h_m((X_1, Y_1), \dots, (X_m, Y_m)) | \mathcal{G}_n]. \end{aligned} \quad (\text{A.4.6})$$

$(R_{m,n}((X_1, Y_1), (X_2, Y_2), \dots, (X_n, Y_n)), \mathcal{G}_n)_{n \geq m}$ ist ein reverses Martingal.

Beweis:

Zunächst ist

$$(R_{m,n}((X_1, Y_1), (X_2, Y_2), \dots, (X_n, Y_n)))_{n \geq m}$$

an $(\mathcal{G}_n)_{n \geq m}$ adaptiert, denn erstens bildet die Folge $(\mathcal{G}_n)_{n \geq m}$ der σ -Algebren einen absteigenden Filter und zweitens ist das n -te Glied der Folge

$$(R_{m,n}((X_1, Y_1), (X_2, Y_2), \dots, (X_n, Y_n)))_{n \geq m}$$

gemäß Korollar A.4.4 und Bemerkung A.4.6 bzgl. $\tilde{\mathcal{G}}_n$ meßbar, einer Unter- σ -Algebra von \mathcal{G}_n nach Definition (A.4.2).

Die Martingaleigenschaft von

$$(R_{m,n}((X_1, Y_1), (X_2, Y_2), \dots, (X_n, Y_n)), \mathcal{G}_n)_{n \geq m}$$

folgt damit aus der im Lemma angegebenen Darstellung (A.4.6) von

$$R_{m,n}((X_1, Y_1), (X_2, Y_2), \dots, (X_n, Y_n))$$

als \mathcal{G}_n -bedingtem Erwartungswert der von n unabhängigen Größe

$$h_m((X_1, Y_1), \dots, (X_m, Y_m)),$$

vgl. etwa [8, Ex. 4, S. 136].

Die Gleichung (A.4.6) selbst weisen wir in zwei Schritten nach, nämlich:

$$\begin{aligned} R_{m,n}((X_1, Y_1), (X_2, Y_2), \dots, (X_n, Y_n)) \\ = \mathbb{E}[h_m((X_1, Y_1), \dots, (X_m, Y_m)) | \tilde{\mathcal{G}}_n] \end{aligned} \quad (\text{A.4.7})$$

$$= \mathbb{E}[h_m((X_1, Y_1), \dots, (X_m, Y_m)) | \mathcal{G}_n] \quad (\text{A.4.8})$$

Wir beginnen mit (A.4.8):

\mathcal{G}_n wird nach Definition (A.4.2) von den beiden unabhängigen σ -Algebren $\tilde{\mathcal{G}}_n$ und $\sigma((X_i, Y_i), i \geq n+1)$ erzeugt. Wegen $n \geq m$ ist die zu bedingende Größe $h_m((X_1, Y_1), \dots, (X_m, Y_m))$ ebenfalls von $\sigma((X_i, Y_i), i \geq n+1)$ unabhängig. (A.4.8) folgt mit [8, Th. 15.5 S. 119].

Zum Beweis von (A.4.7) bemerken wir zunächst, daß aufgrund der $\tilde{\mathcal{G}}_n$ -Meßbarkeit von $E[h_m((X_1, Y_1), \dots, (X_m, Y_m)) | \tilde{\mathcal{G}}_n]$ gemäß Lemma A.4.3 und Bemerkung A.4.6 für alle Permutationen $\pi \in G_n$ gilt:

$$\begin{aligned} & E[h_m((X_1, Y_1), \dots, (X_m, Y_m)) | \tilde{\mathcal{G}}_n] \\ &= E[h_m(\cdot) | \tilde{\mathcal{H}}_n] \circ ((X_1, Y_1), \dots, (X_m, Y_m)) \\ &= E[h_m(\cdot) | \tilde{\mathcal{H}}_n] \circ \pi((X_1, Y_1), \dots, (X_m, Y_m)) \\ &= E[h_m((X_{\pi(1)}, Y_{\pi(1)}), \dots, (X_{\pi(m)}, Y_{\pi(m)})) | \tilde{\mathcal{G}}_n] \end{aligned}$$

Mit den Bezeichnungen der Bemerkung A.4.6 ist $f := E[h_m(\cdot) | \tilde{\mathcal{H}}_n]$ dabei das $\tilde{\mathcal{H}}_n$ -meßbare Funktional auf $(\text{Abb}(I, \mathbb{R}^2)^{\mathbb{N}}, \mathcal{B}(I, \mathbb{R}^2)^{\mathbb{N}})$ zu dem $\tilde{\mathcal{G}}_n$ -meßbaren Funktional

$$\tilde{f} := E[h_m((X_1, Y_1), \dots, (X_m, Y_m)) | \tilde{\mathcal{G}}_n]$$

auf $(\Omega, \mathcal{A}, \mathbb{P})$.

Insbesondere folgt

$$\begin{aligned} & E[h_m((X_1, Y_1), \dots, (X_m, Y_m)) | \tilde{\mathcal{G}}_n] \\ &= \frac{(n-m)!}{n!} \sum_{\substack{i_1, \dots, i_m=1 \\ i_j \neq i_k}}^n E[h_m((X_{i_1}, Y_{i_1}), \dots, (X_{i_m}, Y_{i_m})) | \tilde{\mathcal{G}}_n] \\ &= E[R_{m,n}((X_1, Y_1), \dots, (X_n, Y_n)) | \tilde{\mathcal{G}}_n] = R_{m,n}((X_1, Y_1), \dots, (X_n, Y_n)), \end{aligned}$$

die letzten beiden Gleichungen gelten nach Definition (A.4.1) bzw. $\tilde{\mathcal{G}}_n$ -Meßbarkeit von $R_{m,n}((X_1, Y_1), \dots, (X_n, Y_n))$ gemäß Korollar A.4.4 und Bemerkung A.4.6. ■

Nun wenden wir uns dem Beweis des Satzes A.4.1 zu.

Beweis von Satz A.4.1:

Es bezeichne

$$\mathcal{G}_\infty := \bigcap_{n \in \mathbb{N}} \mathcal{G}_n$$

die abschließende σ -Algebra der Folge $(\mathcal{G}_n)_{n \in \mathbb{N}}$. Da

$$(R_{m,n}((X_1, Y_1), \dots, (X_n, Y_n)), \mathcal{G}_n)_{n \geq m}$$

ein reverses Martingal ist, gilt

$$R_{m,n}((X_1, Y_1), \dots, (X_n, Y_n)) \xrightarrow{\text{f. s.} / L_1} E h_m[(X_1, Y_1), \dots, (X_m, Y_m) | \mathcal{G}_\infty]$$

für $n \rightarrow \infty$, vgl. etwa [8, Cor. 19.10, S. 167 und (19.12), S. 165]. Um die Aussage des Satzes zu zeigen, genügt deshalb der Nachweis, daß das Funktional

$$E h_m[(X_1, Y_1), \dots, (X_m, Y_m) | \mathcal{G}_\infty] : (\Omega, \mathcal{A}, \mathbb{P}) \rightarrow (\mathbb{R}, \mathcal{B}) \quad \text{f. s. const.} \quad (\text{A.4.9})$$

ist. Dann folgt nämlich

$$E h_m[(X_1, Y_1), \dots, (X_m, Y_m) | \mathcal{G}_\infty] = E h_m((X_1, Y_1), \dots, (X_m, Y_m)) \quad \text{f. s.}$$

Zum Beweis von (A.4.9) zeigen wir zunächst, daß alle Ereignisse $A \in \mathcal{G}_\infty$ bzgl. der Folge $(X_n, Y_n)_{n \in \mathbb{N}}$ "endlich vertauschbar" im Sinne von [8, Def. 11.5 und

(11.8'), S. 75] sind: Vertauschen wir nämlich eine endliche Anzahl Pfade der Prozesse $(X_n, Y_n)_{n \in \mathbb{N}}$ durch Anwendung einer Permutation aus G_k , so hat dies nach Konstruktion (A.4.2) der Folge $(\mathcal{G}_n)_{n \in \mathbb{N}}$ keine Auswirkung auf die σ -Algebren

$$\mathcal{G}_{k+1} \supset \mathcal{G}_{k+2} \supset \cdots \supset \mathcal{G}_\infty.$$

Also ist insb. das \mathcal{G}_∞ -meßbare Funktional $E h_m[(X_1, Y_1), \dots, (X_m, Y_m) | \mathcal{G}_\infty]$ endlich vertauschbar bzgl. $(X_n, Y_n)_{n \in \mathbb{N}}$. Da die Folge $(X_n, Y_n)_{n \in \mathbb{N}}$ der von uns betrachteten stochastischen Prozesse zudem i. i. d. ist, folgt (A.4.9) wörtlich wie [8, Cor. 11.7, S. 78] aus dem 0-1 Gesetz von Hewitt und Savage. Im Beweis von [8, Cor. 11.7, S. 78] wird eine Fassung des Satzes von Hewitt und Savage zugrundegelegt, die nur für Folgen reeller i. i. d. Zufallsvariablen gilt, nämlich [8, Th. 11.6, S. 76]. Im Original [43, Th. 11.3, S. 493] wird dieses Gesetz in der von uns benötigten Allgemeinheit ohne einschränkende Annahmen über den Bildraum der Zufallsvariablen gezeigt; und der Beweis von [8, Cor. 11.7, S. 78] ist wörtlich auf unsere Situation übertragbar. ■

A.5 Straffheit bei stetigen Funktionen mit mehreren Argumenten

In Übereinstimmung mit den Bezeichnungen aus Abschnitt 3.2.2 sei $\Theta \subset \mathbb{R}^d$, $d \in \mathbb{N}$, eine offene, beschränkte Teilmenge des \mathbb{R}^d . Mit $C(\bar{\Theta}, \mathbb{R}^c)$, $c \in \mathbb{N}$, bezeichnen wir die Menge stetiger, \mathbb{R}^c -wertiger Funktionen auf dem kompakten Abschluß $\bar{\Theta}$. Bekanntlich macht die Supremumsnorm

$$\sup_{\vartheta \in \bar{\Theta}} |f(\vartheta) - g(\vartheta)| \quad f, g \in C(\bar{\Theta}, \mathbb{R}^c)$$

diese Menge zu einem vollständigen, separablen, metrischen Raum.

Die stochastischen Prozesse $\alpha_n(\vartheta)$, $\beta_n(\vartheta)$ und $\gamma_n(\vartheta)$ des Abschnitts 8.2 können als Wahrscheinlichkeitsmaße auf $C(\bar{\Theta}, \mathbb{R}^d)$ interpretiert werden. Als σ -Algebra findet dabei die von der Supremumsnorm erzeugte Borel- σ -Algebra Verwendung.

Für den Beweis des Theorems 3.5.1 ist die Straffheit dieser Prozesse entscheidend. Sie wird in den Sätzen 8.2.4, 8.2.8 und 8.2.11 nachgewiesen. Ihre Bedeutung resultiert aus folgender Aussage, die für den eindimensionalen Fall $c = d = 1$ in [11, p. 54, Theorem 8.1] formuliert wird und deren Beweis wörtlich auf unsere Situation übertragbar ist:

A.5.1 Satz *Es bezeichnen $X_n(\cdot)$, $X(\cdot)$ stochastische Prozesse auf $C(\bar{\Theta}, \mathbb{R}^c)$. Die endlichdimensionalen Randverteilungen $(X_n(\vartheta_1), \dots, X_n(\vartheta_q))$ zu $q \in \mathbb{N}$ und $\vartheta_1, \dots, \vartheta_q \in \bar{\Theta}$ seien schwach konvergent gegen die entsprechenden Randverteilungen $(X(\vartheta_1), \dots, X(\vartheta_q))$ von $X(\cdot)$. Ist die Folge $X_n(\cdot)$ straff, so konvergiert sie dann selbst schwach gegen $X(\cdot)$.*

Im vorliegenden Abschnitt leiten wir die Charakterisierung der Straffheit her, die in den Beweisen der Sätze 8.2.4, 8.2.8 und 8.2.11 verwendet wird:

A.5.2 Satz *Die Folge stochastischer Prozesse*

$$X_n(\cdot) = (X_n^1(\cdot), \dots, X_n^c(\cdot)) : \bar{\Theta} \rightarrow \mathbb{R}^c$$

ist straff, wenn sie den folgenden zwei Bedingungen genügt:

1. Die Folge $X_n(\vartheta)$ ist in einem festen Punkt $\vartheta \in \bar{\Theta}$ straff.
2. Es existieren Konstanten $\alpha > 1$ und $\gamma \geq 0$ und ein Homöomorphismus $F : \bar{\Theta} \rightarrow [0, 1]^d$ auf den d -dimensionalen Einheitskubus, so daß jede Komponente $n \in \mathbb{N}$, $i = 1, \dots, c$, der Momentenbedingung

$$\mathbb{E}(X_n^i(\vartheta) - X_n^i(\vartheta'))^\gamma \leq \text{const.} \cdot |F(\vartheta) - F(\vartheta')|^\alpha \quad \vartheta, \vartheta' \in \bar{\Theta},$$

genügt.

Der Abschluß der Parametermenge ist aufgrund der Charakterisierung in Abschnitt 3.4.2 einfach zusammenhängend und damit homöomorph zu $[0, 1]^d$.

Die Aussage des Satzes A.5.2 ist an [11, S. 95, Theorem 12.3 und (12.51)] angelehnt. In diesem Theorem wird lediglich der eindimensionale Fall $\bar{\Theta} = [0, 1]$ untersucht, der Beweisgang läßt sich jedoch ohne größere Änderungen auf unsere Situation übertragen. Für andere Teile des Buches [11] ist dies bereits durch [10] geschehen.

Zum Beweis des Satzes A.5.2 stellen wir zunächst fest, daß es genügt, den Fall des eindimensionalen Bildraumes zu untersuchen. Dies liegt an dem folgenden

A.5.3 Lemma *Der stochastische Prozeß $X(\cdot) : \bar{\Theta} \rightarrow \mathbb{R}^c$ ist genau dann straff, wenn dies für seine Komponenten $X^i(\cdot) : \bar{\Theta} \rightarrow \mathbb{R}$, $i = 1, \dots, c$, gilt.*

Beweis:

Der Raum $C(\bar{\Theta}, \mathbb{R}^c)$ ist separabel. Das Wahrscheinlichkeitsmaß, welches der Prozeß $X(\cdot)$ auf $C(\bar{\Theta}, \mathbb{R}^c)$ induziert, ist das Produktmaß der Wahrscheinlichkeitsmaße, welche seine Komponenten $X^i(\cdot)$ auf $C(\bar{\Theta})$ erzeugen. [11, S. 21, Theorem 3.2] liefert die Behauptung. ■

Wir beschränken uns für den Rest dieses Abschnittes also auf den Fall $c = 1$. Außerdem werden wir uns für den Rest dieses Abschnittes auf den technisch günstigen Fall $\bar{\Theta} = [0, 1]^d$, $F = \text{id}$ beziehen. Dies stellt keine echte Einschränkung dar, weil F als stetige Abbildung auf dem Kompaktum $\bar{\Theta}$ sofort gleichmäßig stetig ist. Folglich kann die Momentenbedingung des Satzes A.5.2 durch

$$\mathbb{E}(X_n^i(\vartheta) - X_n^i(\vartheta'))^\gamma \leq \text{const.}' \cdot |\vartheta - \vartheta'|^\alpha \quad \vartheta, \vartheta' \in \bar{\Theta}, i = 1, \dots, c,$$

ersetzt werden, sofern nur die Konstante auf der rechten Seite geeignet angepaßt wird.

Wir stellen einige Aussagen zusammen, die teilweise bereits von [11] in der nun benötigten Allgemeinheit gezeigt werden:

A.5.4 Lemma *Jeder Prozess auf $C(\bar{\Theta})$ ist straff.*

Beweis:

Lemma A.5.4 folgt aus [11, p. 10, Theorem 1.4]. ■

Die Definition der Straffheit von Prozessen auf $C(\bar{\Theta})$ benutzt bekanntlich kompakte Teilmengen dieses Raumes. Im folgenden Lemma wird hierfür die Charakterisierung nach Arzelà-Ascoli eingesetzt. Zur Abkürzung führen wir den Stetigkeitsmodul

$$w(X, \delta) := \sup_{|\vartheta - \vartheta'| < \delta} |X(\vartheta) - X(\vartheta')| \quad \delta > 0$$

für Prozesse $X(\cdot)$ auf $C(\bar{\Theta})$ ein.

A.5.5 Lemma *Die Folge stochastischer Prozesse*

$$X_n(\cdot) : \bar{\Theta} \rightarrow \mathbb{R}$$

ist genau dann straff, wenn sie den folgenden zwei Bedingungen genügt:

1. Wir fixieren ein $\vartheta \in \bar{\Theta}$. Zu jedem $\eta > 0$ existiere dann ein $a > 0$, so daß gilt:

$$P(X_n(\vartheta) > a) \leq \eta \quad n \in \mathbb{N}$$

2. Zu jedem $\varepsilon > 0$ und $\eta > 0$ existiere $\delta > 0$ und $n_o \in \mathbb{N}$, so daß gilt:

$$P(w(X_n, \delta) \geq \varepsilon) \leq \eta \quad n \geq n_o$$

Beweis:

Der Beweis des eindimensionalen Falles kann wörtlich von [11, p. 55, Theorem 8.2] übernommen werden. Dort wird neben Lemma A.5.4 lediglich der Satz von Arzelà-Ascoli benutzt. Letzterer gilt auch in unserer mehrdimensionalen Situation. ■

Wie bei [11, p. 55, Theorem 8.2] folgt insbesondere, daß in der zweiten Bedingung des Lemmas ohne Einschränkung $n_o = 1$ gesetzt werden darf.

Wir folgen weiterhin der Darstellung in [11] und ersetzen entsprechend [11, p. 56, Corollary] die zweite Bedingung des Lemmas A.5.5 durch eine schärfere. Wie [11, p. 57] ausführt, stellt die neue Forderung für unabhängige Prozesse jedoch keine echte Verschärfung dar.

Die kompakte Menge $\bar{\Theta}$ wird für jedes $k \in \mathbb{N}$ mit einem d -dimensionalen Gitter der Kantenlänge $(2k+1)^{-1} > 0$ überdeckt:

Zu $v = (v_1, \dots, v_d) \in \{l/(2k+1), l = 0, \dots, 2k\}^d$ betrachten wir die abgeschlossenen Kuben

$$\{\vartheta = (\vartheta_1, \dots, \vartheta_d) \in \mathbb{R}^d : 0 \leq \vartheta_j - v_j \leq (2k+1)^{-1} \quad \forall j = 1, \dots, d\} \subset \mathbb{R}^d.$$

Zwei Punkte desselben Kubus haben voneinander höchstens den Abstand

$$\delta_k := d^{1/2}(2k+1)^{-1}, \quad k \in \mathbb{N}.$$

Insgesamt besteht das k -te Gitter aus

$$m(\delta_k) := (2k+1)^d$$

Kuben, die $\bar{\Theta}$ genau überdecken. Wir bezeichnen sie mit $K_i^{\delta_k} \subset \mathbb{R}^d$, $i = 1, \dots, m(\delta_k)$ und erhalten also:

$$\bar{\Theta} = \bigcup_{i=1}^{m(\delta_k)} K_i^{\delta_k}.$$

Sei für den Rest des Abschnittes $\delta > 0$ ein Element der positiven Nullfolge $\delta_k \searrow 0$.

Bei geeigneter Numerierung der Kuben ist es möglich, je zwei Ecken $v^i, w^i \in K_i^\delta$ mit $|v^i - w^i| = \delta_k$ zu fixieren, die zudem $w^i = v^{i+1}$, $i = 1, \dots, m(\delta) - 1$ erfüllen. Dies liegt insbesondere daran, daß in jeder Dimension eine ungerade Anzahl Kuben nebeneinander liegt. Zusätzlich können wir erreichen, daß Start- und Endpunkt v^1 bzw. $w^{m(\delta)}$ der Folge nicht von δ abhängen.

A.5.6 Lemma *Mit den soeben eingeführten Bezeichnungen gilt für jede Abbildung $X : \bar{\Theta} \rightarrow \mathbb{R}$:*

$$\mathbb{P}(w(X, \delta) \geq \varepsilon) \leq \sum_{i=1}^{m(\delta)} \mathbb{P}\left(\sup_{\vartheta \in K_i^\delta} |X(\vartheta) - X(v^i)| \geq \varepsilon/4\right)$$

Beweis:

Wir fixieren einen Pfad $\omega \in \Omega$, so daß $\vartheta, \vartheta' \in \bar{\Theta}$ mit $|\vartheta - \vartheta'| < \delta$ und $|X(\vartheta) - X(\vartheta')| \geq \varepsilon > 0$ existieren. Ist dies nicht möglich, so verschwindet die Wahrscheinlichkeit auf der linken Seite der Aussage ohnehin. Die Indices i, j seien nun so gewählt, daß $\vartheta \in K_i^\delta$ und $\vartheta' \in K_j^\delta$ gilt. Aufgrund des geringen Abstandes von ϑ und ϑ' müssen diese beiden Kuben entweder übereinstimmen oder benachbart sein. In jedem Fall gilt $K_i^\delta \cap K_j^\delta \neq \emptyset$; es sei $u \in K_i^\delta \cap K_j^\delta$. Wir erhalten dann:

$$\begin{aligned} \varepsilon &\leq |X(\vartheta) - X(\vartheta')| \\ &\leq |X(\vartheta) - X(v^i)| + |X(v^i) - X(u)| + |X(u) - X(v^j)| + |X(v^j) - X(\vartheta')| \\ &\leq 2 \sup_{\vartheta \in K_i^\delta} |X(\vartheta) - X(v^i)| + 2 \sup_{\vartheta \in K_j^\delta} |X(\vartheta) - X(v^j)| \end{aligned}$$

Deshalb folgt

$$w(X, \delta) \geq \varepsilon \Rightarrow \exists i \leq m(\delta) : \sup_{\vartheta \in K_i^\delta} |X(\vartheta) - X(v^i)| \geq \varepsilon/4$$

sowie

$$\begin{aligned} \mathbb{P}(w(X, \delta) \geq \varepsilon) &\leq \mathbb{P}\left(\bigcup_{i=1}^{m(\delta)} \sup_{\vartheta \in K_i^\delta} |X(\vartheta) - X(v^i)| \geq \varepsilon/4\right) \\ &\leq \sum_{i=1}^{m(\delta)} \mathbb{P}\left(\sup_{\vartheta \in K_i^\delta} |X(\vartheta) - X(v^i)| \geq \varepsilon/4\right), \end{aligned}$$

die Aussage des Lemmas. ■

Schließlich benötigen wir eine Aussage über das Oszillationsverhalten zufälliger Partialsummen. Sie wird als [11, p. 94, Theorem 12.2] bereits in der von uns benötigten Allgemeinheit gezeigt und deshalb hier ohne erneuten Beweis zitiert.

Es seien Y_1, \dots, Y_m reellwertige Zufallsvariablen, über deren Verteilung oder Abhängigkeitsstruktur wir zunächst keine Annahmen treffen. Die Partialsummen seien mit

$$S_k := \sum_{i=1}^k Y_i, \quad k = 0, \dots, m,$$

bezeichnet. Den Wert der dem Betrage nach größten Partialsumme kürzen wir mit

$$M_m := \max_{0 \leq k \leq m} |S_k|$$

ab.

Wir nehmen nun an, daß Zahlen $u_1, \dots, u_m \geq 0$ existieren, die

$$\mathbb{E}(|S_i - S_j|^\gamma) \leq \left(\sum_{l=i+1}^j u_l \right)^\alpha \quad 0 \leq i \leq j \leq m \quad (\text{A.5.1})$$

erfüllen. Dabei seien $\alpha > 1$ und $\gamma \geq 0$ die Konstanten aus Satz A.5.2. Dann gilt:

A.5.7 Lemma *Mit den eingeführten Bezeichnungen und unter (A.5.1) existiert ein $b > 0$, so daß für alle $a > 0$ gilt:*

$$\mathbb{P}(M_m \geq a) \leq \frac{b}{a^\gamma} \left(\sum_{l=1}^m u_l \right)^\alpha$$

Beweis:

Die Aussage des Lemmas folgt aus [11, p. 94, Theorem 12.2], wenn wir dort

$$K'_{\gamma, \alpha} := b \quad \lambda := a$$

setzen und statt [11, p. 94, (12.43)] die schärfere Voraussetzung [11, p. 94, (12.42)] anwenden. ■

Wir beweisen nun das eingangs formulierte Straffheitskriterium.

Beweis von Satz A.5.2:

Es genügt, die zweite Bedingung des Lemmas A.5.5 nachzuweisen. Wegen Lemma A.5.6 ist es dazu wiederum ausreichend, zu beliebigem $\varepsilon > 0$ und $\eta > 0$ ein $\delta > 0$ anzugeben, welches

$$\sum_{i=1}^{m(\delta)} \mathbb{P} \left(\sup_{\vartheta \in K_i^\delta} |X_n(\vartheta) - X_n(v^i)| \geq \varepsilon/4 \right) \leq \eta \quad n \in \mathbb{N} \quad (\text{A.5.2})$$

erfüllt.

Wir ersetzen das Supremum zunächst durch das Maximum über ein endliches, aber feines Punkteraster innerhalb des Kubus K_i^δ . Dabei wiederholen wir sinngemäß die Konstruktion, die wir bereits für $[0, 1]^d$ verwendet haben. Für $h \in \mathbb{N}$ führen wir die $(2h + 1)^d$ gleichmäßig in K_i^δ verteilten Punkte

$$v^i(j_1, \dots, j_d) := v^i + \sum_{l=1}^d \frac{j_l}{2h(2k+1)} \mathbf{e}_l \in K_i^\delta, \quad j_l = 0, \dots, 2h, \quad l = 1, \dots, d,$$

ein. Dabei bezeichne $\mathbf{e}_l \in \mathbb{R}^d$ den l -ten Einheitsvektor. Die neu eingeführte Punktmenge läßt sich in einer Reihenfolge durchlaufen, so daß je zwei aufeinanderfolgende Punkte den Abstand $\frac{\delta}{2h}$ voneinander haben. Eine ganz entsprechende Aussage hatten wir bereits für die Folge der Kuben K_i^δ , $i = 1, \dots, m(\delta)$, formuliert. Anhand dieser Reihenfolge numerieren wir die Punkte als $v^i(j) \in K_i^\delta$, $j = 1, \dots, (2h + 1)^d$ neu durch. Wir können außerdem erreichen, daß

$$\begin{aligned} v^i(1) &= v^i & i &= 1, \dots, m(\delta) \\ v^i((2h + 1)^d) &= w^i & i &= 1, \dots, m(\delta) \end{aligned}$$

gilt. Damit haben wir insbesondere den letzten Punkt

$$v^{m(\delta)}((2h + 1)^d) = w^{m(\delta)}$$

des letzten Kubus unabhängig von δ und h fixiert.

Nach dieser Vorbereitung betrachten wir nun die reellwertigen Zufallsvariablen

$$Y_j := X_n(v^i(j)) - X_n(v^i(j-1)), \quad j = 2, \dots, (2h + 1)^d$$

und zeigen zunächst, daß sie (A.5.1) genügen:

$$\begin{aligned} \mathbb{E}(|S_l - S_j|^\gamma) &= \mathbb{E} \left| \sum_{p=j+1}^l Y_p \right|^\gamma \\ &= \mathbb{E} |X_n(v^i(l)) - X_n(v^i(j))|^\gamma \\ &\leq \text{const.} \cdot |v^i(l) - v^i(j)|^\alpha \\ &= \left(\sum_{p=j+1}^l \text{const.}^{1/\alpha} \cdot (v^i(p) - v^i(p-1)) \right)^\alpha \\ &= \left(\sum_{p=j+1}^l u_p \right)^\alpha \end{aligned}$$

mit

$$u_p := \text{const.}^{1/\alpha} \cdot (v^i(p) - v^i(p-1))$$

Die verwendete Ungleichung gilt dabei nach der zweiten Voraussetzung des Satzes A.5.2 in dem von uns betrachteten Spezialfall $\bar{\Theta} = [0, 1]^d$, $F = \text{id}$.

Wir können also Lemma A.5.7 anwenden. Damit folgt:

$$\begin{aligned}
& \mathbb{P}\left(\max_{2 \leq j \leq (2h+1)^d} |X_n(v^i(j)) - X_n(v^i(1))| \geq \varepsilon/4\right) \\
&= \mathbb{P}\left(\max_{2 \leq j \leq (2h+1)^d} \left|\sum_{l=2}^j Y_l\right| \geq \varepsilon/4\right) \\
&= \mathbb{P}\left(\max_{2 \leq j \leq (2h+1)^d} |S_j| \geq \varepsilon/4\right) \\
&\leq \frac{4^\gamma b}{\varepsilon^\gamma} \cdot \left(\sum_{l=2}^{(2h+1)^d} u_l\right)^\alpha \\
&= \frac{4^\gamma b \cdot \text{const.}}{\varepsilon^\gamma} \cdot \left(\sum_{l=2}^{(2h+1)^d} (v^i(l) - v^i(l-1))\right)^\alpha \\
&= \frac{4^\gamma b \cdot \text{const.}}{\varepsilon^\gamma} \cdot (v^i((2h+1)^d) - v^i(1))^\alpha \\
&= \frac{4^\gamma b \cdot \text{const.}}{\varepsilon^\gamma} \cdot (w^i - v^i)^\alpha
\end{aligned}$$

Der Prozeß $X_n(\cdot)$ hat stetige Pfade. Für wachsendes $h \rightarrow \infty$ konvergiert das Maximum auf der linken Seite der Ungleichung deshalb gegen das Supremum in (A.5.2), die rechte Seite bleibt erhalten:

$$\mathbb{P}\left(\sup_{\vartheta \in K_i^\delta} |X_n(\vartheta) - X_n(v^i)| \geq \varepsilon/4\right) \leq \frac{4^\gamma b \cdot \text{const.}}{\varepsilon^\gamma} \cdot (w^i - v^i)^\alpha$$

Nun summieren wir auf beiden Seiten über die Kuben auf:

$$\begin{aligned}
& \sum_{i=1}^{m(\delta)} \mathbb{P}\left(\sup_{\vartheta \in K_i^\delta} |X_n(\vartheta) - X_n(v^i)| \geq \varepsilon/4\right) \\
&\leq \frac{4^\gamma b \cdot \text{const.}}{\varepsilon^\gamma} \sum_{i=1}^{m(\delta)} (w^i - v^i)^\alpha \\
&\leq \frac{4^\gamma b \cdot \text{const.}}{\varepsilon^\gamma} \max_{1 \leq i \leq m(\delta)} (w^i - v^i)^{\alpha-1} \sum_{i=1}^{m(\delta)} (w^i - v^i) \\
&= \frac{4^\gamma b \cdot \text{const.}}{\varepsilon^\gamma} (w^{m(\delta)} - v^1) \max_{1 \leq i \leq m(\delta)} (w^i - v^i)^{\alpha-1} \\
&= \frac{4^\gamma b \cdot \text{const.}}{\varepsilon^\gamma} (w^{m(\delta)} - v^1) \delta^{\alpha-1}
\end{aligned}$$

Bei der Konstruktion der Kuben haben wir darauf hingewiesen, daß weder v^1 noch $w^{m(\delta)}$ von δ abhängen. Also folgt insgesamt mit einem neuen Faktor $c(\varepsilon)$, der lediglich von ε abhängt:

$$\sum_{i=1}^{m(\delta)} \mathbb{P}\left(\sup_{\vartheta \in K_i^\delta} |X_n(\vartheta) - X_n(v^i)| \geq \varepsilon/4\right) \leq c(\varepsilon) \cdot \delta^{\alpha-1}$$

Deshalb ist es möglich, zu beliebigem $\varepsilon > 0$ sowie zu einer beliebigen Schranke $\eta > 0$ für die Summe auf der rechten Seite ein $\delta > 0$ anzugeben, welches (A.5.2) liefert. \blacksquare

B Implementation des KQS

B.1 Einführung

In diesem Anhang beschäftigen wir uns mit der Implementation des Modells aus Kapitel 4. In Abschnitt B.2 stellen wir zu diesem Zweck alle wichtigen Terme in einer Form bereit, die die direkte Berechnung ermöglicht. In Abschnitt B.3 folgt Programmcode in der Sprache R, mit dem wir den Kleinst-Quadrate-Schätzer und seine Varianz bei gegebenen Daten für das Modell aus Kapitel 4 bestimmen können.

B.2 Darstellungen relevanter Größen durch die Sprünge der Punktprozesse

B.2.1 Der Kleinst-Quadrate-Schätzer

Die Notation der Prozesse aus Abschnitt 4.2.2 gestattet eine Darstellung des KQS, welche für seine Bestimmung in Anwendungen günstig ist: ϑ_n^* aus Abschnitt 3.2.5 minimiert die Funktion

$$\vartheta \mapsto \sum_{j=1}^n \sum_{k=1}^{K_j} (\bar{N}_n(Y_{jk}) - \bar{\Lambda}_{\vartheta,n}(Y_{jk}))^2.$$

Um ihn auf gegebenen Daten numerisch zu approximieren, müssen wir den Ausdruck auf der rechten Seite für variierende Werte von ϑ effizient berechnen können. Dieser Aufgabe wenden wir uns in den folgenden Abschnitten zu.

B.2.2 Der mittlere Pfad

Für den Wert des mittleren Pfades \bar{N}_n an den Sprungstellen der Stichprobe erhalten wir ohne weiteres die Darstellung:

$$\bar{N}_n(Y_{jk}) = \frac{1}{n} \sum_{i=1}^n N_i(Y_{jk}) \quad (\text{B.2.1})$$

Eine entsprechende Formel für die mittlere Intensität an den Sprungstellen des mittleren Pfades leiten wir im nächsten Abschnitt her.

B.2.3 Die mittlere Intensität

Bereits in Abschnitt 4.8 hatten wir mit (4.8.3) eine additive Zerlegung der Modellintensität angegeben. Für die mittlere Intensität an den Sprungstellen

Y_{jk} , $j = 1, \dots, n$; $k = 1, \dots, K_j$, der Stichprobe folgt damit:

$$\begin{aligned}
\bar{\Lambda}_{\vartheta,n}(Y_{jk}) &= \frac{\alpha}{\beta} (\cos \gamma - \cos(\beta Y_{jk} + \gamma)) + \delta Y_{jk} \\
&+ \frac{\delta}{n\varepsilon} \sum_{i=1}^n \sum_{h=2}^{N_i(Y_{jk-})} (e^{-\varepsilon \cdot (Y_{ih} - Y_{iN_i(Y_{ih-})})} - 1) \\
&+ \frac{\delta}{n\varepsilon} \sum_{i=1}^n (e^{-\varepsilon \cdot (Y_{jk} - Y_{iN_i(Y_{jk-})})} - 1) \cdot \mathbf{1}\{Y_{jk} > Y_{i1}\} \\
&+ \frac{\alpha}{n(\beta^2 + \varepsilon^2)} \\
&\sum_{i=1}^n \sum_{h=2}^{N_i(Y_{jk-})} \left(e^{-\varepsilon \cdot (Y_{ih} - Y_{iN_i(Y_{ih-})})} (\varepsilon \sin(\beta Y_{ih} + \gamma) + \beta \cos(\beta Y_{ih} + \gamma)) \right. \\
&\quad \left. - (\varepsilon \sin(\beta Y_{iN_i(Y_{ih-})} + \gamma) + \beta \cos(\beta Y_{iN_i(Y_{ih-})} + \gamma)) \right) \\
&+ \frac{\alpha}{n(\beta^2 + \varepsilon^2)} \\
&\sum_{i=1}^n \left(e^{-\varepsilon \cdot (Y_{jk} - Y_{iN_i(Y_{jk-})})} (\varepsilon \sin(\beta Y_{jk} + \gamma) + \beta \cos(\beta Y_{jk} + \gamma)) \right. \\
&\quad \left. - (\varepsilon \sin(\beta Y_{iN_i(Y_{jk-})} + \gamma) + \beta \cos(\beta Y_{iN_i(Y_{jk-})} + \gamma)) \right) \cdot \mathbf{1}\{Y_{jk} > Y_{i1}\} \\
&+ \frac{\zeta}{n\eta} \sum_{i=1}^n \sum_{h=1}^{W_i(Y_{jk})} (e^{\eta \cdot (Y_{jk} - X_{ih})} - 1) \tag{B.2.2}
\end{aligned}$$

Wir haben dabei außerdem von der Beziehung

$$N_i(Y_{ih-}) = h - 1 \quad i = 1, \dots, n; \quad h = 1, \dots, K_i$$

Gebrauch gemacht. Dieser Zusammenhang wird bei der Datenaufbereitung, die wir im nächsten Abschnitt behandeln, von Nutzen sein.

B.2.4 Die standardisierende Matrix

Aus der Definition der Matrix $\Phi_n(\vartheta)$ in Abschnitt 3.4.3 folgt:

$$\Phi_n(\vartheta) = \frac{1}{2} \frac{\partial^2}{\partial \vartheta^2} \int_I (\bar{\Lambda}_{\vartheta,n}(t) - \bar{\Lambda}_{\vartheta_0,n}(t))^2 \cdot \bar{\lambda}_{\vartheta_0,n}(t) dt$$

In unserer Anwendung können wir also die Näherung

$$\Phi_n(\vartheta_0) \approx \frac{1}{2} \frac{\partial^2}{\partial \vartheta^2} \int_I (\bar{\Lambda}_{\vartheta,n}(t) - \bar{\Lambda}_{\vartheta_n^*,n}(t))^2 \cdot \bar{\lambda}_{\vartheta_n^*,n}(t) dt \Big|_{\vartheta=\vartheta_n^*} \tag{B.2.3}$$

verwenden. Sowohl die Integration als auch die partielle Ableitung werden wir numerisch durchführen. Eine analytische Berechnung ist für das in (4.8.1) und (4.8.3) definierte Modell zwar prinzipiell möglich, jedoch sehr aufwendig.

B.2.5 Die i. i. d. Darstellung des KQS

Wir werden die in Abschnitt 3.4.4 eingeführte asymptotische Kovarianzmatrix nicht direkt berechnen. Stattdessen verwenden wir zur Bestimmung der asymptotischen Kovarianz die i. i. d. Terme

$$T_k := \int_I \left(\int_{[x, t_{\max}]} \mathbb{E} \frac{\partial}{\partial \vartheta} \Lambda_{\vartheta_{\circ}}(t) \mathbb{E} \Lambda_{\circ}(dt) \right) (N_k(dx) - \Lambda_{\vartheta_{\circ}, k}(dx)) \quad (\text{B.2.4})$$

für $k = 1, \dots, n$ gemäß (8.3.8). Dort haben wir

$$\Phi_n(\vartheta_{\circ}) \cdot (\vartheta_n^* - \vartheta_{\circ}) = n^{-1/2} \sum_{k=1}^n T_k + o_{\mathbb{P}}(1) \quad (\text{B.2.5})$$

erhalten, woraus wir nun

$$\text{Var}(\vartheta_n^*) \approx (\Phi_n^{-1}(\vartheta_{\circ}))^{\top} \text{Var}(T_1) \Phi_n^{-1}(\vartheta_{\circ}) \in \mathbb{R}^{d \times d} \quad (\text{B.2.6})$$

folgern können. Zur Berechnung der rechten Seite verwenden wir die Näherung (B.2.3), dann fehlt uns noch $\text{Var}(T_1)$. Diese Größe können wir z. B. mit der Stichprobenvarianz

$$\frac{1}{n-1} \sum_{k=1}^n \left(T_k - \frac{1}{n} \sum_{i=1}^n T_i \right) \left(T_k - \frac{1}{n} \sum_{i=1}^n T_i \right)^{\top} \quad (\text{B.2.7})$$

abschätzen. Die Terme T_k ersetzen wir dabei ihrerseits durch Näherungen, denn die analytische Berechnung der in (B.2.4) auftretenden Integrale und Ableitungen ist noch aufwendiger als in (B.2.3). Als Erstes formen wir das innere Integral um. Mit der Kettenregel folgt zunächst

$$\frac{1}{2} \mathbb{E}^2 \Lambda_{\vartheta}(t) \Big|_x^{t_{\max}} = \int_x^{t_{\max}} \mathbb{E} \Lambda_{\vartheta}(t) \mathbb{E} \lambda_{\vartheta}(t) dt.$$

Die Produktregel liefert im Anschluß

$$\begin{aligned} \frac{1}{2} \frac{\partial}{\partial \vartheta} \mathbb{E}^2 \Lambda_{\vartheta}(t) \Big|_x^{t_{\max}} &= \int_x^{t_{\max}} \mathbb{E} \frac{\partial}{\partial \vartheta} \Lambda_{\vartheta}(t) \mathbb{E} \lambda_{\vartheta}(t) dt \\ &+ \int_x^{t_{\max}} \mathbb{E} \Lambda_{\vartheta}(t) \mathbb{E} \frac{\partial}{\partial \vartheta} \lambda_{\vartheta}(t) dt. \end{aligned}$$

Nun machen wir uns die Näherung $E \Lambda_\circ(\cdot) \approx \bar{N}_n(\cdot)$ zu Nutze:

$$\begin{aligned}
& \int_{[x, t_{\max}]} E \frac{\partial}{\partial \vartheta} \Lambda_{\vartheta_\circ}(t) E \Lambda_{\vartheta_\circ}(dt) \\
&= \frac{1}{2} \frac{\partial}{\partial \vartheta} E^2 \Lambda_{\vartheta_\circ}(t) \Big|_x^{t_{\max}} - \int_x^{t_{\max}} E \Lambda_{\vartheta_\circ}(t) E \frac{\partial}{\partial \vartheta} \lambda_{\vartheta_\circ}(t) dt \\
&\approx \frac{1}{2} \frac{\partial}{\partial \vartheta} E^2 \Lambda_{\vartheta_\circ}(t) \Big|_x^{t_{\max}} - \int_x^{t_{\max}} \bar{N}_n(t) E \frac{\partial}{\partial \vartheta} \lambda_{\vartheta_\circ}(t) dt \\
&= \frac{1}{2} \frac{\partial}{\partial \vartheta} E^2 \Lambda_{\vartheta_\circ}(t) \Big|_x^{t_{\max}} - \frac{1}{n} \sum_{j=1}^n \sum_{k=1}^{K_j} \int_{\max(x, Y_{jk})}^{t_{\max}} E \frac{\partial}{\partial \vartheta} \lambda_{\vartheta_\circ}(t) dt \\
&= \frac{\partial}{\partial \vartheta} \left(\frac{1}{2} E^2 \Lambda_{\vartheta_\circ}(t) \Big|_x^{t_{\max}} - \frac{1}{n} \sum_{j=1}^n \sum_{k=1}^{K_j} E \Lambda_{\vartheta_\circ}(t) \Big|_{\max(x, Y_{jk})}^{t_{\max}} \right)
\end{aligned}$$

Schließlich ersetzen wir die Erwartungswerte durch die entsprechenden empirischen Mittel und den wahren Parameterwert durch den KQS. Dies liefert die Näherung:

$$\begin{aligned}
T_i &\approx \frac{\partial}{\partial \vartheta} \int_I \left(\frac{1}{2} \bar{\Lambda}_{\vartheta_n^*, n}^2(t) \Big|_x^{t_{\max}} - \frac{1}{n} \sum_{j=1}^n \sum_{k=1}^{K_j} \bar{\Lambda}_{\vartheta_n^*, n}(t) \Big|_{\max(x, Y_{jk})}^{t_{\max}} \right) (N_i(dx) - \Lambda_{\vartheta_n^*, i}(dx)) \\
&= \frac{\partial}{\partial \vartheta} \left[\frac{1}{n} \sum_{h=1}^{K_i} \left(\frac{1}{2} \bar{\Lambda}_{\vartheta_n^*, n}^2(t) \Big|_{Y_{ih}}^{t_{\max}} - \frac{1}{n} \sum_{j=1}^n \sum_{k=1}^{K_j} \bar{\Lambda}_{\vartheta_n^*, n}(t) \Big|_{\max(Y_{ih}, Y_{jk})}^{t_{\max}} \right) \right. \\
&\quad \left. - \int_I \left(\frac{1}{2} \bar{\Lambda}_{\vartheta_n^*, n}^2(t) \Big|_x^{t_{\max}} - \frac{1}{n} \sum_{j=1}^n \sum_{k=1}^{K_j} \bar{\Lambda}_{\vartheta_n^*, n}(t) \Big|_{\max(x, Y_{jk})}^{t_{\max}} \right) \lambda_{\vartheta_n^*, i}(x) dx \right] \\
&= \frac{\partial}{\partial \vartheta} \left[\frac{K_i}{2n} \bar{\Lambda}_{\vartheta_n^*, n}^2(t_{\max}) - \frac{1}{2n} \sum_{h=1}^{K_i} \bar{\Lambda}_{\vartheta_n^*, n}^2(Y_{ih}) - \frac{K K_i}{n^2} \bar{\Lambda}_{\vartheta_n^*, n}(t_{\max}) \right. \\
&\quad \left. + \frac{1}{n^2} \sum_{h=1}^{K_i} \left(\bar{\Lambda}_{\vartheta_n^*, n}(Y_{ih}) \sum_{j=1}^n N_j(Y_{ih}) \right) + \frac{1}{n^2} \sum_{j=1}^n \sum_{k=1}^{K_j} \bar{\Lambda}_{\vartheta_n^*, n}(Y_{jk}) N_i(Y_{jk}) \right. \\
&\quad \left. - \int_I \left(\frac{1}{2} \bar{\Lambda}_{\vartheta_n^*, n}^2(t_{\max}) - \frac{1}{2} \bar{\Lambda}_{\vartheta_n^*, n}^2(x) - \frac{K}{n} \bar{\Lambda}_{\vartheta_n^*, n}(t_{\max}) \right) \lambda_{\vartheta_n^*, i}(x) dx \right] \tag{B.2.8} \\
&\quad \left. + \frac{1}{n} \sum_{j=1}^n \sum_{k=1}^{K_j} \max(\bar{\Lambda}_{\vartheta_n^*, n}(x), \bar{\Lambda}_{\vartheta_n^*, n}(Y_{jk})) \lambda_{\vartheta_n^*, i}(x) dx \right]
\end{aligned}$$

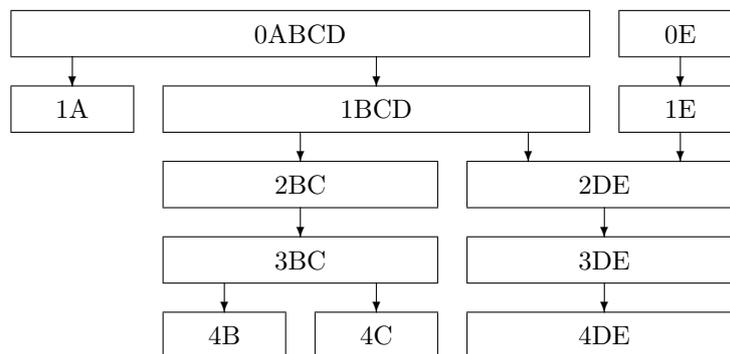
Die verbleibende Integration und Differentiation werden wir numerisch durchführen.

B.3 Ein R-Algorithmus zur KQS-Bestimmung

Wir geben einen Algorithmus für das Statistikprogramm R an, mit welchem der Kleinst-Quadrate-Schätzer aus Abschnitt 3.2.5 für das Modell (4.8.3) auf gegebenen Daten bestimmt werden kann. Da wir keine explizite Darstellung des KQS haben, werden wir die R-Routine `nlm` zur Lösung nichtlinearer Optimierungsprobleme einsetzen. `nlm` verwendet lokale Suchverfahren und muß deshalb den quadratischen Abstand von (B.2.1) und (B.2.2) bei festem Datensatz für viele unterschiedliche Werte von ϑ bestimmen. Wir werden uns im folgenden zunächst damit beschäftigen, den Datensatz in einer Form aufzubereiten, der diese Berechnungen effizient ermöglicht.

B.3.1 Datenaufbereitung

Im Zuge der Datenaufbereitung werden einige Hilfsgrößen bestimmt, die von `nlm` selbst nicht benötigt werden. Da diese Variablen mitunter beträchtlichen Speicherplatz belegen, ist es wichtig zu wissen, wann sie gelöscht werden können. Um derartige Abhängigkeiten unter den insgesamt zwölf verschiedenen Schritten des Programmes deutlich zu machen, kennzeichnen wir letztere mit Nummern und Buchstaben anhand des folgenden Schemas:



Mit den beiden Schritten 0ABCD und 0E startet das Programm. Weitere Schritte können erst ausgeführt werden, wenn zuvor alle Programmteile abgelaufen sind, welche eine niedrigere Nummer und mindestens einen der Buchstaben in ihrer Bezeichnung tragen: So benötigt z. B. Schritt 2DE Ergebnisse aus den Schritten 1BCD und 1E, welche ihrerseits auf 0ABCD bzw. 0E angewiesen sind. Die anderen sieben Programmteile hingegen sind für 2DE nicht nötig. 3DE und 4DE können ihrerseits sogar erst nach 2DE arbeiten, zu den verbleibenden fünf Schritten 1A, 2BC, 3BC, 4B und 4C steht 2DE in keiner Vorrangbeziehung.

Wir beschreiben nun nacheinander die zwölf Schritte der Datenaufbereitung. Dabei definieren wir jeweils neu eingeführte Größen, zählen notwendige Variablen aus vorangegangenen Schritten auf und weisen darauf hin, wann derartige Variablen gelöscht werden können. Wir geben zudem jeweils Programmcode in R an, welcher die neuen Variablen entsprechend belegt.

0ABCD. Inputgrößen: Dieser Schritt stellt die Inputgrößen

$$n \in \mathbb{N}$$

$$n_N := \sum_{i=1}^n \mathbf{1}\{K_i > 0\} \in \mathbb{N}$$

$$\text{PathLengthN} := (K_1 + 1, \dots, K_n + 1) \in \mathbb{N}^{n_N}, \quad K_i > 0$$

$$\text{Naug} := (0, Y_{11}, Y_{12}, Y_{13}, \dots, Y_{1K_1}, 0, Y_{21}, \dots, Y_{nK_n}) \in [0, t_{\max})^{K+n_N}$$

als Zahlen bzw. Vektoren in \mathbb{R} bereit. Die beiden Vektoren sollen dabei nur Informationen zu Pfaden enthalten, die mindestens einen Sprung im Untersuchungszeitraum haben. Dem Vektor **Naug** ist für derartige Pfade außerdem ein künstlicher Sprung an der Stelle $t_{\min} = 0$ vorangestellt.

0E. Inputgrößen: Ganz entsprechend werden in diesem Schritt die Inputgrößen

$$n_W := \sum_{i=1}^n \mathbf{1}\{J_i > 0\} \in \mathbb{N}$$

$$\text{PathLengthW} := (J_1 + 1, \dots, J_n + 1) \in \mathbb{N}^{n_W}, \quad J_i > 0$$

$$\text{Waug} := (0, X_{11}, X_{12}, X_{13}, \dots, X_{1J_1}, 0, X_{21}, \dots, X_{nK_n}) \in [0, t_{\max})^{J+n_W}$$

bereitgestellt. Sie beschreiben die Verläufe der beeinflussenden Pfade.

1A. Die Hilfsmatrix EntriesPathNi: Völlig unabhängig von der restlichen Datenaufbereitung kann die Hilfsmatrix

$$\text{EntriesPathNi} = \begin{pmatrix} 1 & \mathbf{1}\{K_1 > 0\} \\ \vdots & \vdots \\ K_1 + \mathbf{1}\{K_1 > 0\} & \mathbf{1}\{K_1 > 0\} \\ K_1 + \mathbf{1}\{K_1 > 0\} + 1 & \sum_{i=1}^2 \mathbf{1}\{K_i > 0\} \\ \vdots & \vdots \\ \sum_{i=1}^2 (K_i + \mathbf{1}\{K_i > 0\}) & \sum_{i=1}^2 \mathbf{1}\{K_i > 0\} \\ \sum_{i=1}^2 (K_i + \mathbf{1}\{K_i > 0\}) + 1 & \sum_{i=1}^3 \mathbf{1}\{K_i > 0\} \\ \vdots & \vdots \\ \sum_{i=1}^{n-1} (K_i + \mathbf{1}\{K_i > 0\}) & n_N \\ \vdots & \vdots \\ K + n_N & n_N \end{pmatrix} \in \mathbb{N}^{(K+n_N) \times 2}$$

erstellt werden. Sie wird erst wieder im Optimierungsschritt benötigt. Ihre Bedeutung ergibt sich aus der weiteren Datenaufbereitung in den folgenden Schritten. Dort werden wir zur Beschreibung des Datensatzes einige $(K + n_N) \times n_N$ -Matrizen konstruieren, z. B. die Matrix mit den Einträgen $N_i(Y_{jk-})$ in Schritt 2BC. $i = 1, \dots, n_N$ ist dabei der Spaltenindex. Den Zeilenindex $l = 1, \dots, K + n_N$ erhalten wir zu gegebenen Werten j, k als

$$l = \sum_{h=1}^{j-1} K_h + k.$$

Anschaulich beziehen sich die Zeilen der Matrix in der Reihenfolge auf die Sprünge des Datensatzes, die bereits bei `Naug` zugrundegelegt wurde. Dies deuten wir dadurch an, daß wir sowohl j als auch k zur Indizierung der Zeilen verwenden. Die Zeilen von `EntriesPathNi` liefern nun gerade die Kombinationen aus Zeilen- und Spaltenindex in den $(K + n_N) \times n_N$ -Matrizen, an denen $i = j$ gilt.

In R läßt sich die Matrix `EntriesPathNi` z. B. in folgender Weise berechnen:

```
EntriesPathNi <- cbind(1:sum(PathLengthN),
  rep(1:nN,times=PathLengthN))
```

1BCD. Pfadpositionen in Naug: Wir bestimmen den Hilfsvektor

$$\text{StartPosNaug} := \left(1, K_1 + 2, K_1 + K_2 + 3, \dots, \sum_{i=1}^{n_N-1} K_i + n_N \right) \in \mathbb{N}^{n_N},$$

welcher die Startpositionen der nichtleeren Pfade N_i in `Naug` enthält. Dazu benötigen wir offenbar die Größen n_N und `PathLengthN`:

```
StartPosNaug <- c(0,cumsum(PathLengthN)[-nN])+1
```

1E. Pfadpositionen in Waug: Der Hilfsvektor

$$\text{StartPosWaug} := \left(1, J_1 + 2, J_1 + J_2 + 3, \dots, \sum_{i=1}^{n-1} J_i + n_W \right) \in \mathbb{N}^{n_W},$$

entspricht dem Vektor `StartPosNaug`, beschreibt jedoch die (nichtleeren) Pfade der Prozesse W_i . Zu seiner Berechnung werden n_W und `PathLengthW` benötigt:

```
StartPosWaug <- c(0,cumsum(PathLengthW)[-nW])+1
```

2BC. Die Matrix $N_i(Y_{jk-})$: Aus den Größen n_N , `PathLengthN` und `StartPosNaug` läßt sich die Matrix

$$\text{NoJumpsNiBeforeYjk} := \begin{pmatrix} \vdots \\ \cdots N_i(Y_{jk-}) \cdots \\ \vdots \end{pmatrix}_{j,k} \in \mathbb{N}^{(K+n_N) \times n_N}$$

berechnen. Der Spaltenindex i durchläuft dabei nur nichtleere Pfade ($K_i > 0$). Die Zeilen sind durch die beiden Größen $j = 1, \dots, n$ (mit $K_j > 0$) und $k = 1, \dots, K_j$ indiziert.

```
## the following for-loop is used deliberately to avoid the
## construction of ONE large matrix which is known
## to cause runtime errors already for medium-sized samples
```

```

NoJumpsNiBeforeYjk <- NULL

## we need to ensure that both paths Ni and Nj have actual jumps.
## Otherwise, the ''explicit'' version
## of the ''outer''-statement we employ within the body of the loop
## will cause an error

for (i in 1:nN) {
  PathPosNaugI <- StartPosNaug[i]+(0:(PathLengthN[i]-1))
  JumpRelI <- NULL
  for (j in 1:nN) {
    PathPosNaugJ <-
      StartPosNaug[j]+(0:(PathLengthN[j]-1))

    ## the following two statements compute
    ## ''JumpRelIJ <- outer( Naug[PathPosNaugI],
    ## Naug[PathPosNaugJ], ''<'')'' without
    ## duplication of the arguments to save memory space

    JumpRelIJ <- 1 * (rep(Naug[PathPosNaugI],
      length.out=PathLengthN[i]*PathLengthN[j])
      < rep.int(Naug[PathPosNaugJ],rep.int(PathLengthN[i],
      PathLengthN[j])))
    dim(JumpRelIJ) <- c(PathLengthN[i],PathLengthN[j])

    ## end of ''JumpRelIJ <- outer( Naug[PathPosNaugI],
    ## Naug[PathPosNaugJ], ''<'')''
    ## construct the next entries of JumpRelI of the final
    ## matrix JumpRel by adding elements of JumpRelIJ within
    ## columns:
    ## this yields for each jump of path j the number of
    ## jumps in path i being strictly earlier. Since we
    ## have an artificial first jump at tmin to count,
    ## the values of JumpRel equal  $N_i(Y_{\{jk\}})+1$ 
    ## and we have to subtract 1 for the actual jumps

    if (is.null(JumpRelI)) JumpRelI <-
      pmax(colSums(JumpRelIJ)-1,0)
    else JumpRelI <- c(JumpRelI,pmax(colSums(JumpRelIJ)-1,0))
  } ## next (j in 1:n)
  ## now, JumpRelI holds  $N_i(Y_{\{jk\}})$  for all  $j=1,\dots,n$ ;
  ##  $k=1,\dots,K_j$ . These are  $K$  entries, regardless of  $i$ 
  ## JumpRelI is the next column of JumpRel

  if (is.null(NoJumpsNiBeforeYjk)) NoJumpsNiBeforeYjk <-
    JumpRelI
  else NoJumpsNiBeforeYjk <-
    cbind(NoJumpsNiBeforeYjk,JumpRelI)
} ##

```

```
next (i in 1:n)
rm(PathPosNaugI, PathPosNaugJ, JumpRelIJ, JumpRelI)
```

2DE. Die Matrix $W_i(Y_{jk})$: Die hier zu bestimmende Matrix entspricht ungefähr jener aus Schritt 2BC, allerdings werden hier die Verläufe der Prozesse N_j und W_i miteinander verglichen. Deswegen sind relativ viele Hilfsgrößen nötig, nämlich n_N , PathLengthN , StartPosNaug , n_W , PathLengthW und StartPosWaug . Mit ihrer Hilfe lassen sich die Elemente der Matrix

$$\text{NoJumpsWiBeforeYjk} := \begin{pmatrix} \vdots \\ \cdots W_i(Y_{jk}) \cdots \\ \vdots \end{pmatrix}_{j,k}^i \in \mathbb{N}^{n_W \times (K+n_N)}$$

bestimmen. Nur nichtleere Pfade N_j und W_i , also solche mit $K_j > 0$ bzw. $J_i > 0$, tragen zu $\text{NoJumpsWiBeforeYjk}$ bei:

```
NoJumpsWiBeforeYjk <- NULL
for (i in 1:nW) {
  PathPosWaugI <-
    StartPosWaug[i]+(0:(PathLengthW[i]-1))
  JumpRelI <- NULL
  for (j in 1:nN) {
    PathPosNaugJ <-
      StartPosNaug[j]+(0:(PathLengthN[j]-1))

    JumpRelIJ <- 1 * (rep(Waug[PathPosWaugI],
      length.out=PathLengthW[i]*PathLengthN[j])
      < rep.int(Naug[PathPosNaugJ], rep.int(PathLengthW[i],
      PathLengthN[j])))

    dim(JumpRelIJ) <-
      c(PathLengthW[i], PathLengthN[j])

    if (is.null(JumpRelI)) JumpRelI <-
      pmax(colSums(JumpRelIJ)-1, 0)
    else JumpRelI <-
      c(JumpRelI, pmax(colSums(JumpRelIJ)-1, 0))
  } ## next (j in 1:n)
  if (is.null(NoJumpsWiBeforeYjk)) NoJumpsWiBeforeYjk <-
    JumpRelI
  else NoJumpsWiBeforeYjk <-
    rbind(NoJumpsWiBeforeYjk, JumpRelI)
} ## next (i in 1:n)
rm(PathPosWaugI, PathPosNaugJ, JumpRelIJ, JumpRelI)
```

3BC. Die Matrix $Y_{iN_i(Y_{jk-})}$: Aus den Größen n_N , $K = \text{sum}(\text{PathLengthN})$, StartPosNaug und $\text{NoJumpsNiBeforeYjk}$ läßt sich ermitteln, wann ein Prozeß

N_i letztmalig vor Y_{jk} gesprungen ist, diese Größe entspricht $Y_{iN_i(Y_{jk-})}$. Diese Werte lassen sich wieder in einer Matrix anordnen:

$$\left(\begin{array}{c} \vdots \\ \cdots Y_{iN_i(Y_{jk-})} \cdots \\ \vdots \end{array} \right)_{j,k} \in [t_{\min}, t_{\max})^{(K+n_N) \times n_N}$$

Wir erinnern insbesondere an die Konvention aus Abschnitt 4.2.2, die $Y_{jk} = t_{\min}$ für $k \leq 0$ festlegt. Damit sind die Werte der Matrix auch für die ersten Sprünge in der Stichprobe definiert.

Der Spaltenindex i durchläuft dabei wiederum nur nichtleere Pfade ($K_i > 0$). Die Zeilen sind wie üblich durch die beiden Größen $j = 1, \dots, n$ (mit $K_j > 0$) und $k = 1, \dots, K_j$ indiziert.

Letztlich muß diese Matrix natürlich aus `Naug` erstellt werden. Es wird sich allerdings als günstig erweisen, anstelle der Werte $Y_{iN_i(Y_{jk-})}$ selbst ihre Positionen im Vektor `Naug` zu speichern:

```
PosLastJumpNiBeforeYjk <-
  matrix(rep(StartPosNaug, each=sum(PathLengthN)), ncol=nN)
  + NoJumpsNiBeforeYjk
```

Also gilt

$$\text{Naug}[\text{PosLastJumpNiBeforeYjk}] = (Y_{iN_i(Y_{jk-})})_{j,k}$$

3DE. Der Vektor $(Y_{jk} - X_{ih})$: Aus (B.2.2) geht hervor, daß wir gerade die Differenzen $Y_{jk} - X_{ih}$ für $j = 1, \dots, n$, $k = 1, \dots, K_j$, $i = 1, \dots, n$ und $h = 1, \dots, W_i(Y_{jk})$ benötigen, um die postulierte Auswirkung von W_i auf $\bar{\Lambda}_{\vartheta, n}(\cdot)$ zu quantifizieren. Wir stellen diese Terme nun als Vektor bereit; natürlich werden leere Pfade hier wiederum nicht berücksichtigt:

$$\Phi_{01} := (\cdots, Y_{jk} - X_{ih}, \cdots)_{j,k,i,h} \in \mathbb{R}^{\sum_{j,k,i} W_i(Y_{jk})}$$

```
Phi01 <- rep(Naug, times=colSums(NoJumpsWiBeforeYjk))
  - Waug[sequence(NoJumpsWiBeforeYjk[NoJumpsWiBeforeYjk>0])
  + rep(rep(StartPosWaug, times=sum(PathLengthN)),
  times=NoJumpsWiBeforeYjk)]
```

4B. Der Vektor $\bar{N}_n(Y_{jk})$: Nach Schritt 3BC wird die Matrix `NoJumpsNiBeforeYjk` hier nun letztmalig benötigt. Mit ihrer Hilfe können wir den Vektor

$$\text{AvgNatYjk} = (\cdots, \bar{N}_n(Y_{jk}), \cdots) \in \mathbb{R}^{K+n_N}$$

bereitstellen, den wir bereits in (B.2.1) betrachtet haben:

```
AvgNatYjk <- (rowSums(NoJumpsNiBeforeYjk)+1)/n
```

4C. Die Matrix $(Y_{jk} - Y_{iN_i(Y_{jk-})})$: Aus (B.2.2) ist ersichtlich, daß in unserem Modell der Intensität die Terme $Y_{jk} - Y_{iN_i(Y_{jk-})}$ benötigt werden, welche wir nun leicht aus n , `Naug`, `StartPosNaug` und `PosLastJumpNiBeforeYjk` bestimmen:

$$\Phi_{10} := \begin{pmatrix} \vdots \\ \cdots Y_{jk} - Y_{iN_i(Y_{jk-})} \cdots \\ \vdots \end{pmatrix}_{j,k} \in \mathbb{R}^{(K+n_N) \times n_N}$$

Da uns die Differenzen für $N_i(Y_{jk}) = 0$ nicht interessieren, markieren wir sie im R-Code durch ein negatives Vorzeichen. Negative Terme werden später in den Optimierungsschritten entsprechend verarbeitet:

```
Phi10 <- matrix((Naug - Naug[PosLastJumpNiBeforeYjk])
  * (-1)^(Naug[PosLastJumpNiBeforeYjk]==0), ncol=nN)
```

4DE. Der Vektor $(\sum_i W_i(Y_{jk}))$: Nach Schritt 3DE wird die Matrix `NoJumpsWiBeforeYjk` hier zum letzten Mal verwendet. Durch ihre Spaltensummen wird der Vektor

$$\text{NoJumpsWBeforeYjk} = \left(\cdots, \sum_{i=1}^n W_i(Y_{jk}), \cdots \right) \in \mathbb{N}^{K+n_N}$$

definiert, welcher in den Optimierungsschritten hilfreich sein wird:

```
NoJumpsWBeforeYjk <- colSums(NoJumpsWiBeforeYjk)
```

B.3.2 Bestimmung des Kleinst-Quadrate-Schätzers

Die generische R-Routine `nlm` zur nichtlinearen Optimierung benötigt einerseits die zu minimierende Funktion `f` und andererseits geeignete Startwerte `p` für die zu bestimmenden Parameterwerte. Wir verwenden als Zielfunktion:

```
abstand <- function(p) sum((mittlintensitaet(p)-AvgNatYjk)^2)
```

Diese Funktion greift auf die globale Variable `AvgNatYjk` aus Schritt 4B zurück. Außerdem wird die Funktion `mittlintensitaet` aufgerufen, die zu jedem Parameterwert ϑ die Werte $\bar{\Lambda}_{\vartheta,n}(Y_{jk})$, $j = 1, \dots, n$, $k = 1, \dots, K_j$, als K -dimensionalen Vektor liefert:

```
mittlintensitaet <- function(p) {
  ## the auxiliary variable Phi20 is stored globally
  ## to save memory space
  Phi20 <<- (exp(-p[5]*pmax(Phi10,0))*(p[5]*sin(p[2]*Naug+p[3])+
    p[2]*cos(p[2]*Naug+p[3]))-
    matrix(p[5]*sin(p[2]*Naug[PosLastJumpNiBeforeYjk]+p[3])+
    p[2]*cos(p[2]*Naug[PosLastJumpNiBeforeYjk]+p[3]),
    ncol=nN)) *
```

```

(Naug[PosLastJumpNiBeforeYjk]>0)

Intens <- p[4] * Naug +
  p[1]/p[2] * (cos(p[3]) - cos(p[2] * Naug + p[3])) +
  p[4]/(n*p[5]) * (rowSums(exp(-p[5]*pmax(Phi10,0))-1) +
  rowSums(matrix((unlist(tapply(
  (exp(-p[5]*pmax(Phi10,0))-1)[EntriesPathNi],
  EntriesPathNi[,2], cumsum), use.names=FALSE))
  [PosLastJumpNiBeforeYjk], ncol=nN))) +
  p[1]/(n*p[2]^2+n*p[5]^2) * (rowSums(Phi20) +
  rowSums(matrix((unlist(tapply(
  Phi20[EntriesPathNi], EntriesPathNi[,2], cumsum),
  use.names=FALSE)) [PosLastJumpNiBeforeYjk],
  ncol=nN)))

IntensCondPoisson <- p[6]/(n*p[7])*tapply(
  exp(p[7]*Phi01)-1, rep(1:sum(PathLengthN),
  times=NoJumpsWBeforeYjk), sum)

names(Intens) <- 1:sum(PathLengthN)
Intens[names(IntensCondPoisson)] <-
Intens[names(IntensCondPoisson)]+IntensCondPoisson
drop(Intens)
}

```

Wir wählen Startwerte für $\text{startparam} := (\alpha, \dots, \eta)$ gemäß der in Abschnitt 4.9 angestellten Überlegungen:

```

tmax <- max(Naug)
Ntmax <- (sum(PathLengthN)-nN)/n
Ntmaxqt <- sum(Naug>0 & Naug <= tmax/4)/n

lsparam[2] <- 2*pi/tmax
lsparam[3] <- -pi/2
## initial values of p[2] and p[3] ensure one full season
## between 0 and tmax with a maximum around tmax/2 (summer)

lsparam[4] <- Ntmax / tmax
## initial value of p[4] ensures precise match of Poisson
## Intensity part with average observed path at tmax

Ntmaxthreeqt <- sum(Naug>0 & Naug <= 3*tmax/4)/n
lsparam[1] <- lsparam[2]/2 * (Ntmaxthreeqt -
  lsparam[4]*tmax/2 - Ntmaxqt)
## initial value of p[1] ensures least squares between
## Poisson Intensity part and average observed path
## at 0.25 tmax and 0.75 tmax

## assume effect of previous purchase acts vanishes
## (1% remaining) after a week

```

```

lsparam[5] <- log(100)/7

## assume importance of advertising roughly equals number
## of purchase acts / number of ad contacts
InitAdEffect <- (sum(PathLengthN)-n)/(sum(PathLengthW)-n)

## rescale lsparam[1] and lsparam[4] to "make space" for ad effect
lsparam[1] <- lsparam[1]*(1-InitAdEffect)
lsparam[4] <- lsparam[4]*(1-InitAdEffect)

## determine decay rate of advertising contacts by observed
## median between contacts and next purchase act
MedYminX <- median(Phi01[EntriesPathNi])
lsparam[7] <- log(0.5)/MedYminX

## subsequently: determine jumpsize of ad contact so that area
## under all "ad contact-intensities" equals
## the assumed initial importance of ad contacts:
lsparam[6] <- -lsparam[7]*InitAdEffect^2

```

Nun kann der KQS für das Modell (4.8.3) mit der Anweisung

```
optimparam <- nlm(f=abstand, p=startparam)$estimate
```

bestimmt werden.

B.3.3 Bestimmung der standardisierenden Matrix

Um die Differentiation und Integration in (B.2.3) numerisch durchzuführen, verwenden wir die R-Routinen `fdHess` (aus dem package `nlme`) und `integrate`. Das Integrationsprogramm benötigt seinerseits eine Funktion, die den Integranden an allen Elementen ihres vektorwertigen Arguments berechnet. Wir können dafür nicht auf die bereits definierte Funktion `mittlintensitaet` zurückgreifen, da sie auf die Berechnung der Intensität an den Sprungstellen der Stichprobe zugeschnitten ist. Deshalb führen wir die neue Funktion

```
mittlintensvec <- function(xvec,p) sapply(xvec,mittlintenspktw,p)
```

ein, die einen Vektor `xvec` von beliebigen Punkten aus I verarbeiten kann. `mittlintensvec` ruft dazu die folgende Funktion auf, die ein skalares Argument `x` und den Parametervektor erwartet:

```

mittlintenspktw <- function(x,p) {

  LastJumpNiBeforeX <- Naug[unlist(tapply(Naug < x,
    rep(1:nN,times=PathLengthN),sum),
    use.names=FALSE) +
    c(0,cumsum(PathLengthN)[-nN])]

```

```

## the following auxiliary variable can be prepared globally
LastJumpNiBeforeCurrent <-
  Naug[1:(sum(PathLengthN))-(Naug>0)]

p[4] * x + p[1]/p[2] * (cos(p[3]) - cos(p[2] * x + p[3])) +
  p[4]/(n*p[5]) * sum((exp(-p[5] *
    (Naug-LastJumpNiBeforeCurrent)) - 1) *
    (Naug < x & LastJumpNiBeforeCurrent > 0)) +
  p[1]/(n*p[2]^2+n*p[5]^2) * sum((exp(-p[5] *
    (Naug-LastJumpNiBeforeCurrent))*(p[5]*sin(p[2]*Naug +
    p[3]) + p[2]*cos(p[2]*Naug + p[3])) -
    (p[5]*sin(p[2]*LastJumpNiBeforeCurrent + p[3]) +
    p[2]*cos(p[2]*LastJumpNiBeforeCurrent + p[3]))) *
    (Naug < x & LastJumpNiBeforeCurrent > 0)) +
  p[4]/(n*p[5]) * sum((exp(-p[5] *
    (x-LastJumpNiBeforeX)) - 1) * (LastJumpNiBeforeX > 0)) +
  p[1]/(n*p[2]^2+n*p[5]^2) * sum((exp(-p[5] *
    (x-LastJumpNiBeforeX))*(p[5]*sin(p[2]*x + p[3]) +
    p[2]*cos(p[2]*x + p[3])) -
    (p[5]*sin(p[2]*LastJumpNiBeforeX + p[3]) +
    p[2]*cos(p[2]*LastJumpNiBeforeX + p[3]))) *
    (LastJumpNiBeforeX > 0)) +
  p[6]/(n*p[7]) * sum((exp(p[7]*(x - Waug)) - 1) *
    (Waug < x & Waug > 0))
}

```

Im Integranden von (B.2.3) wird auch die differentielle Intensität benötigt, die wir entsprechend zur Verfügung stellen:

```

diffmittlintensvec <- function(xvec,p) {
  sapply(xvec,diffmittlintenspktw,p)
}

```

diffmittlintensvec kann Vektoren xvec verarbeiten, es verwendet für jedes Element x des Vektors die Funktion:

```

diffmittlintenspktw <- function(x,p) {

  LastJumpNiBeforeX <- Naug[unlist(tapply(Naug < x,
    rep(1:nN,times=PathLengthN),sum),
    use.names=FALSE)+c(0,cumsum(PathLengthN)[-nN])]

  sum((p[1]*sin(p[2]*x + p[3]) + p[4])/n * (1 - exp(-p[5]*
    (x - LastJumpNiBeforeX))*(LastJumpNiBeforeX > 0))) +
    p[6]/n * sum(exp(p[7]*(x - Waug)) * (Waug < x & Waug > 0))
}

```

Das Integral in (B.2.3) wird nun von der folgenden Funktion berechnet:

```

TheoDistance <- function(p) {

```

```

Integrand <- function(xvec,p) {
  0.5*(mittlintensvec(xvec,p)-
    mittlintensvec(xvec,optimparam))^2 *
  diffmittlintensvec(xvec,optimparam)
}
integrate(Integrand,0,364,p=p)$value
}

```

Schließlich bilden wir die zweite Ableitung von `TheoDistance` nach dem Parameter und legen die resultierende Matrix als Näherung von $\Phi_n(\vartheta_0)$ in `StdMatrix` ab:

```
StdMatrix <- fdHess(optimparam,TheoDistance)$Hessian
```

B.3.4 Bestimmung der i. i. d. Terme

Wir wenden uns der Berechnung der Ausdrücke in (B.2.8) für $i = 1, \dots, n$ zu. Dabei müssen wir nun auch die Pfade ohne Sprünge berücksichtigen. Wir verwenden dazu die n -dimensionalen Vektoren `StartPosNaugAll` und `StartPosWaugAll`, die wie in den Schritten 0ABCD und 0E des Abschnitts B.3.1 definiert sind, jedoch zusätzlich Null-Einträge für leere Pfade enthalten. Wir bestimmen aus diesen Größen:

```

StartPosNaugAll <- c( 0,cumsum( PathLengthNall)[-n])+1
StartPosWaugAll <- c( 0,cumsum( PathLengthWall)[-n])+1

```

Mit diesen Hilfsgrößen können wir nun die Intensitäten $\lambda_{\vartheta,i}(\cdot)$ berechnen; wir geben wieder eine Version für skalare Argumente und eine darauf basierende für vektorielle Argumente an:

```

diffintenspktw <- function(x,p,j) {
  LastJumpNjBeforeX <- max(Naug[StartPosNaugAll[j]+
    0:max((PathLengthNall[j]-1),0)]*(Naug[StartPosNaugAll[j]+
    0:max((PathLengthNall[j]-1),0)]<x))

  JumpsWj <- Waug[StartPosWaugAll[j]+
    0:max((PathLengthWall[j]-1),0)]

  (p[1]*sin(p[2]*x + p[3]) + p[4]) *
  (1 - exp(-p[5]*(x - LastJumpNjBeforeX))*
  (LastJumpNjBeforeX > 0)) +
  p[6] * sum(exp(p[7]*(x - JumpsWj)) *
  (JumpsWj < x & JumpsWj > 0))
}

diffintensvec <- function(xvec,p,j) {
  sapply(xvec,diffintenspktw,p,j)
}

```

Jetzt geben wir eine Funktion an, die den Ausdruck in eckigen Klammern aus (B.2.8) für den Parameter $p \in \Theta$ und den Index $j = 1, \dots, n$ auswertet. Zur numerischen Bestimmung des Integrals wird eine weitere Funktion aufgerufen, die wir im Anschluß definieren:

```
square.bracket.in.B.2.9 <- function(p,j) {

  ## we use AvgNAtYjk, but "re-"multiplied with n - is there
  ## a more convenient way to store this information?
  ## to save memory, we use global auxiliary variables here
  ## and assign with "<<-"; these quantities will also be used
  ## within the call of "integrand.in.B.2.9"

  mittlintenspktwAtTmaxP <<- mittlintenspktw(364,p)
  mittlintensAtJumps <<- mittlintensitaet(p)

  integral.with.respect.to.path <- 0
  if(PathLengthNall[j]>0) {
    integral.with.respect.to.path <- PathLengthNall[j]/(2*n) *
      mittlintenspktwAtTmaxP^2 - (2*n)^(-1) *
      sum(mittlintensAtJumps[StartPosNaugAll[j]+
        0:(PathLengthNall[j]-1)]^2) -
      sum(PathLengthN)*PathLengthNall[j] / (n^2) *
      mittlintenspktwAtTmaxP +
      n^(-2) * sum(mittlintensAtJumps[StartPosNaugAll[j]+
        0:(PathLengthNall[j]-1)] *
        n*AvgNAtYjk[StartPosNaugAll[j]+
        0:(PathLengthNall[j]-1)]) +
      n^(-2) * crossprod(
        NoJumpsNiBeforeYjk[,sum(PathLengthNall[1:j]>0)],
        mittlintensAtJumps)
  }

  integrate(integrand.in.B.2.9,0,364,
    subdivisions=300,p=p,j=j)$value +
    integral.with.respect.to.path
}

## next, we provide the integrand of the integral in (B.2.9) at a
## vector of arguments xvec:

integrand.in.B.2.9 <- function(xvec,p,j) {

  ## "mittlintenspktwAtTmaxP", "mittlintensAtJumps" have been
  ## set globally by "square.bracket.in.B.2.9"

  mittlintensvecAtX <- mittlintensvec(xvec,p)

  ## the power in the following statement is taken
  ## element-wise for the vectors
```

```

(0.5 * mittlintenspktwAtTmaxP^2 - 0.5 * mittlintensvecAtX^2 -
  sum(PathLengthN)/n * mittlintenspktwAtTmaxP +
  n^(-1) * rowSums(
    outer(mittlintensvecAtX,mittlintensAtJumps,max))) *
  diffintensvec(xvec,p,j)
}

```

Im Anschluß führen wir die Differentiation in (B.2.8) numerisch aus und erstellen die Matrix `iidTerms` mit den Spalten T_j :

```

iidTerms <- NULL
for (lauf in 1:n) {
  iidTerms <- cbind(iidTerms,fdHess(optimparam,
    square.bracket.in.B.2.9,j=lauf)$gradient)
}

```

B.3.5 Bestimmung der asymptotischen Kovarianzmatrix

Zunächst berechnen wir die Stichprobenvarianz der T_j gemäß (B.2.7):

```
VarIID <- var(t(iidTerms))
```

Dann erhalten wir mit der Näherung (B.2.6) die asymptotische Kovarianzmatrix des KQS:

```
VarLSE <- t(solve(StdMatrix)) %*% VarIID %*% solve(StdMatrix)
```

Literatur

- [1] Aalen, O. O.; Gjessing, H. K.: *Understanding the Shape of the Hazard Rate: A Process Point of View*. Statistical Science 16 (1), 2001.
- [2] Ailawadi, K. L.; Neslin, S. A.: *The Effect of Promotion on Consumption: Buying more and Consuming It Faster*. Journal of Marketing Research 35, 1998.
- [3] Akaike, H.: *A New Look at Statistical Model Identification*. IEEE Transactions on Automatic Control 19, 1974.
- [4] Allenby, G. M.; Lenk, P. J.: *Modeling Household Purchase Behavior With Logistic Normal Regression*. Journal of the American Statistical Association 89 (428), 1994.
- [5] Andersen, P. K.; Borgan, Ø; Gill, R. D.; Keiding, N.: *Statistical Models Based on Counting Processes*. New York, Berlin, Heidelberg: Springer 1993.

- [6] Aubry, C.; Dabye, A. S.: *Asymptotic normality of the minimum distance estimators for a Poisson process with a discontinuous intensity function*. Journal of Statistical Planning and Inference 99, pp. 3-23: Elsevier, 2001.
- [7] Bauer, H.: *Maß- und Integrationstheorie*. Berlin, New York: de Gruyter, 1992².
- [8] Bauer, H.: *Probability Theory*. Berlin, New York: de Gruyter, 1996.
- [9] Ben-Akiva, M.; Lerman, S. R.: *Discrete Choice Analysis*. Cambridge, MA: MIT Press 1985.
- [10] Bickel, P. J.; Wichura, M. J.: *Convergence Criteria for Multiparameter Stochastic Processes and Some Applications*. The Annals of Mathematical Statistics 42 (5), pp. 1656-1670, 1971.
- [11] Billingsley, P.: *Convergence of Probability Measures*. New York: John Wiley 1968.
- [12] Brémaud, P.: *Point Processes and Queues - Martingale Dynamics*. New York: Springer, 1981.
- [13] Bronnenberg, B. J.: *Advertising Frequency Decisions in a Discrete Markov Process Under a Budget Constraint*. Journal of Marketing Research 35, 1998.
- [14] Bucklin, R. E.; Lattin, J. M.: *A Two-State Model of Purchase Incidence and Brand Choice*. Marketing Science 10 (1), 1991.
- [15] Chatfield, C.; Ehrenberg, A. S. C.; Goodhardt, G. J.: *Progress on a Simplified Model of Stationary Purchasing Behavior*. Journal of the Royal Statistical Society A 79, 1966.
- [16] Chatfield, C.; Goodhardt, G. J.: *A Consumer Purchasing Model with Erlang Interpurchase Times*. Journal of the American Statistical Association 68, 1973.
- [17] Chintagunta, P. K.: *Investigating Purchase Incidence, Brand Choice And Purchase Quantity Decisions Of Households*. Marketing Science 12 (2), 1993.
- [18] Chintagunta, P. K.: *Inertia and Variety Seeking in a Model of Brand-Purchase Timing*. Marketing Science 17 (3), 1998.
- [19] Chintagunta, P. K.: *Variety Seeking, Purchase Timing and the Lightning Bolt Brand Choice Model*. Management Science 45 (4), 1999.
- [20] Chintagunta, P. K.; Haldar, S.: *Investigating Purchase Timing Behavior in Two Related Product Categories*. Journal of Marketing Research 35, 1998.
- [21] Chintagunta, P. K.; Jain, D. C.; Vilcassim, N. J.: *Investigating Heterogeneity in Brand Preferences in Logit Models for Panel Data*. Journal of Marketing Research 28, 1991.

- [22] Chintagunta, P. K.; Prasad, A. R.: *An Empirical Investigation of the Dynamic McFadden Model of Purchase Timing and Brand Choice: Implications for Market Structure*. Journal of Business and Economic Statistics 16 (1), 1998.
- [23] Deighton, J.; Henderson, C. M.; Neslin, S. A.: *The Effects of Advertising on Brand Switching and Repeat Purchasing*. Journal of Marketing Research 31, 1994.
- [24] Dorogovtsev, A. Ya.; Kukush, A. G.: *Asymptotic properties of a nonparametric intensity estimator of a nonhomogeneous Poisson process*. Cybernetics and Systems Analysis 32 (1), pp. 74-85: Plenum Publishing Corporation, 1996.
- [25] East, R.; Hammond, K.: *The Erosion of Repeat-Purchase Loyalty*. Marketing Letters 7, 1996.
- [26] Ehrenberg, A. S. C.: *The Pattern of Consumer Purchases*. Applied Statistics 8, 1959.
- [27] Ehrenberg, A. S. C.: *Repeat-Buying: Facts, Theory And Applications*. London: Charles Griffin & Company Ltd. and New York: Oxford University Press, 1988.
- [28] Ehrenberg, A. S. C.: *Comments on How Advertising Works*. Marketing and Research Today August 1992.
- [29] Ehrenberg, A. S. C.: *Repetitive Advertising and the Consumer*. Journal of Advertising Research November, December 2000.
- [30] Fader, P. S.; Hardie, B. G. S.: *Modeling the Evolution of Repeat Buying*. Available at www-marketing.wharton.upenn.edu/ideas/pdf/98-031.pdf
- [31] Fleming, T. R.; Harrington, D. P.: *Counting Processes and Survival Analysis*. New York: John Wiley 1991.
- [32] Forster, O.: *Analysis 2*. Braunschweig, Wiesbaden: Friedrich Vieweg, 1984⁵.
- [33] Franses, P. H.; Paap, R.: *Quantitative Models in Marketing Research*. Cambridge: University Press, 2001.
- [34] Gönül, F.; Srinivasan, K.: *Consumer Purchase Behavior in a Frequently Bought Product Category: Estimation Issues and Managerial Insights from a Hazard Function Model with Heterogeneity*. Journal of the American Statistical Association 88 (424), 1993.
- [35] Gönül, F.; Srinivasan, K.: *Modeling Multiple Sources of Heterogeneity in Multinomial Logit Models: Methodological and Managerial Issues*. Marketing Science 12 (3), 1993.
- [36] Gould, B. W.: *Consumer promotion and purchase timing: the case of cheese*. Applied Economics 29, 1997.
- [37] Gortsev, A. M., Klimov, I. S.: *Estimated dead period and intensity of a Poisson process*. Radio and Communications Technology 1, (2), pp. 10-16: Allerton Press, 1996.

- [38] Günther, M.; Vossebein, U.; Wildner, R.: *Marktforschung mit Panels*. Wiesbaden: Gabler, 1998.
- [39] Gupta, S.: *Impact of Sales Promotions on When, What and How Much to Buy*. Journal of Marketing Research 25, 1988.
- [40] Gupta, S.: *Stochastic Models of Interpurchase Time With Time-Dependent Covariates*. Journal of Marketing Research 28, 1991.
- [41] Heckman, J.; Singer, B.: *A Method for minimizing the Impact of Distributional Assumptions in Econometric Models for Duration Data*. Econometrica 52, 1984.
- [42] Helmets, R.; Zitikis, R.: *On estimation of Poisson intensity functions*. Annals of the Institute of Statistical Mathematics Tokyo 51, (2), pp. 265-280: 1999.
- [43] Hewitt, E.; Savage, L. J.: *Symmetric Measures on Cartesian Products*. Transactions of the American Mathematical Society 80, pp. 470-501, Menasha, Wisconsin: George Banta Publishing Company, 1955.
- [44] Hruschka, H.: *Marketing-Entscheidungen*. München: Vahlen 1996.
- [45] Ibragimov, I. A.; Has'minskii, R. Z.: *Statistical Estimation - Asymptotic Theory*. New York, Heidelberg, Berlin: Springer, 1981.
- [46] Jain, D. C.; Vilcassim, N. J.: *Estimating Household Purchase Rates for Consumer Non-Durable Goods*. Applied Stochastic Models and Data Analysis 10, pp. 15-26, 1994.
- [47] Jain, D. C.; Vilcassim, N. J.; Chintagunta, P. K.: *A Random-Coefficients Logit Brand-Choice Model Applied to Panel Data*. Journal of Business and Economic Statistics 12, (3), 1994.
- [48] Jones, J. M.: *A Composite Heterogeneous Model for Brand Choice Behavior*. Management Science 19, 1973.
- [49] Jones, J. P.: *Over-Promise and Under-Delivery*. Marketing and Research Today 19, 1991.
- [50] Koyck, L. M.: *Distributed Lags and Investment Analysis*. Amsterdam: North-Holland Publishing Company 1954.
- [51] Kutoyants, Yu. A.: *Statistical Inference for Spatial Poisson Processes*. New York, Berlin, Heidelberg: Springer, 1998.
- [52] Lancaster, T.: *The Econometric Analysis of Transition Data*. Cambridge: University Press 1990.
- [53] Leeflang, P. S. H.: *Mathematical Models in Marketing, a Survey, the Stage of Development, Some Extensions and Applications*. Leiden: H. E. Stenfert Kroese, 1974.
- [54] Leeflang, P. S. H.; Wittink, D. R.; Wedel, M.; Naert, P. A.: *Building Models for Marketing Decisions*. Dordrecht: Kluwer Academic Publishers 2000.

- [55] Little, J. D. C.: *Aggregate Advertising Models: The State of the Art*. Operations Research, 1979.
- [56] Lorenz, F.: *Lineare Algebra 1*. Mannheim, Leipzig, Wien, Zürich: BI-Wissenschaftsverlag, 1993³.
- [57] Maier, G.; Weiss, P.: *Modelle diskreter Entscheidungen - Theorie und Anwendung in den Sozial- und Wirtschaftswissenschaften*. Wien, New York: Springer 1989.
- [58] Massy, W. F.; Montgomery, D. B.; Morrison, D. G.: *Stochastic Models of Buying Behavior*. Cambridge, MA: M.I.T. Press 1970.
- [59] Mino, H.: *Parameter Estimation of the Intensity Process of Self-Exciting Point Processes Using the EM Algorithm*. IEEE Transactions on Instrumentation and Measurement 50 (3), pp. 658-664, 2001.
- [60] Morrison, D. G.; Schmittlein, D. C.: *Generalizing the NBD Model for Customer Purchases: What are the Implications and is it Worth the Effort?*. Journal of Business and Economic Statistics 6, 1988.
- [61] Narendranathan, W.; Stewart, M. B.: *How does the Benefit vary as Unemployment Spells lengthen?* Journal of Applied Econometrics 8, 1993.
- [62] Ovchinnikova, T. M.; Trifonov, A. P.: *The Instant of Change in the Unknown Intensity of a Poisson Flow: Detection and Estimation*. Automation and Remote Control 60 (1), pp. 52-60: Kluwer Academics/Plenum Publishers: 1999.
- [63] Pedryck, J. H.; Zufryden, F. S.: *Evaluating the Impact of Advertising Media Plans. A Model of Consumer Purchase Dynamics Using Single-Source Data*. Marketing Science 1991.
- [64] Pollard, D.: *The Minimum Distance Method of Testing*. Metrika 27, pp. 43-70, Wien: Physica-Verlag, 1980.
- [65] Poulsen, C. S.: *Mixed Markov and Latent Markov Modeling Applied to Brand Choice Behavior*. International Journal of Research in Marketing 7, 1990.
- [66] Rosenbach, M: *Werbung: 'Weil isch es mir wert bin'*. Hamburg: Der Spiegel 49, 2001.
- [67] Serfling, R. J.: *Approximation Theorems of Mathematical Statistics*. New York: John Wiley 1980.
- [68] Sichel, H. S.: *Repeat Buying and the Generalized Inverse Gaussian-Poisson Distribution*. Applied Statistics 31, 1982.
- [69] Sikkel, D.; Hoogendoorn, A. W.: *Models for Monthly Penetrations with Incomplete Panel Data*. Statistica Neerlandica 49, 1995.
- [70] Schladitz, K.: *Estimation of the intensity of stationary flat processes*. Advanced Applied Probability (SGSA) 32, pp. 114-139: Applied Probability Trust, 2000.

- [71] Snyder, D. L.; Miller, M. I.: *Random Point Processes in Time and Space*. New York: Springer, 1991².
- [72] Stute, W.: *Parameter Estimation In Smooth Empirical Processes*. Stochastic Processes and their Applications 22, pp. 223-244, North-Holland: Elsevier Science Publishers B. V., 1986.
- [73] Sueyoshi, G. T.: *A Class Of Binary Response Models For Grouped Duration Data*. Journal of Applied Econometrics 10, pp. 411-431, 1995.
- [74] Tapiero, C. S.: *The NBD repeat purchase process and M/G/∞ queues*. International Journal of Production Economics 63, pp. 141-145, 2000.
- [75] Vakratsas, D.: *Household cost effects on purchase timing decisions: do demographics matter?*. Journal of Consumer Marketing 15, (1), pp. 6-22: MCB University Press, 1998.
- [76] Vilcassim, N. J.; Jain, D. C.: *Modeling Purchase-Timing and Brand-Switching Behavior Incorporating Explanatory Variables and Unobserved Heterogeneity*. Journal of Marketing Research 28, pp. 29-41, 1991.
- [77] Wangler, A.: *Heterogenitätsprobleme in der Verlaufsdatenanalyse*. Frankfurt am Main: Peter Lang GmbH, Europäischer Verlag der Wissenschaften, 1997.
- [78] Wedel, M.; DeSarbo, W. S.; Bult, J. R.; Ramaswamy, V.: *A Latent Class Poisson Regression Model for Heterogeneous Count Data*. Journal of Applied Econometrics 8, 1993.
- [79] Wedel, M.; Kamakura, W. A.; DeSarbo, W. S.; Ter Hofstede, F.: *Implications for Asymmetry, Nonproportionality and Heterogeneity in Brand Switching from Piece-wise Exponential Mixture Hazard Models*. Journal of Marketing Research 32, 1995.
- [80] Young, M. R.; DeSarbo, W. S.; Morwitz, V. G.: *The Stochastic Modeling of Purchase Intentions and Behavior*. Management Science 4, (2), 1998.

Erklärung

Hiermit versichere ich, daß ich diese Arbeit selbständig verfaßt und nur die angegebenen Quellen und Hilfsmittel benutzt habe.

Frankfurt / Main, 4. April 2005