Modalwertschätzung in der nichtparametrischen Kurvenschätzung und Blockwise Bootstrap für den geschätzten empirischen Prozess

D i s s e r t a t i o n

zur Erlangung des akademischen Grades
doctor rerum naturalium (Dr. rer. nat.)

vorgelegt dem Rat der Fakultät für Mathematik und Informatik
der Friedrich-Schiller-Universität Jena

von Dipl.-Math. Barbara Wieczorek
geboren am 31.10.1980 **in** München

Gutachter

1. Prof. Dr. Michael H. Neumann (Jena)
2. Prof. Dr. Natalie Neumeyer (Hamburg)
3. Prof. Dr. Eckhard Liebscher (Merseburg)

Tag der öffentlichen Verteidigung: 28. Februar 2013

Zusammenfassung

Die folgende Arbeit befasst sich mit Fragestellungen der asymptotischen Statistik. Der erste Teil ist der Modalwertschätzung im Rahmen der nichtparametrischen Kurvenschätzung in verschiedenen Modellen gewidmet. Zunächst wird das Problem der Modalwertschätzung in Dekonvolutionsmodellen betrachtet, wo interessierende Zufallsvariablen nicht direkt beobachtet werden können, sondern mit einer additiven Störgröße versehen sind. Durch Verwendung von für diese Problematik üblichen Kernschätzern für die Kurven selbst werden Konvergenzraten für den Modalwertschätzer erzielt, der durch Maximierung der Kurvenschätzer definiert ist. Die erzielten Konvergenzraten im Modell der Dichteschätzung und der Errors-in-Variables-Regression sind von der Glattheit der mindestens als zweimal stetig differenzierbar angenommen Kurve sowie der Schlechtgestelltheit des Dekonvolutionsproblems abhängig und erweisen sich als asymptotisch optimal. Im Modell mit direkten Beobachtungen wird der Fall einer Kurve mit nichtdifferenzierbarer Modalstelle unter Vorliegen stark mischender Beobachtungen beleuchtet. Zur Schätzung des Modalwertes wird sowohl die Maximierung des Kurvenschätzers über dem Kontinuum sowie über einem Gitter herangezogen, beide Schätzer erweisen sich wiederum als asymptotisch optimal.
Der zweite Teil dieser Arbeit behandelt den empirischen Prozesses mit geschätztem Parameter, basierend auf schwach abhängigen Beobachtungen. Zahlreiche Teststatistiken beruhen auf dem geschätzten empirischen Prozess, jedoch ist deren Verteilung von unbekannten Parametern abhängig. Daher wird der Originalprozess durch einen Prozess imitiert, welcher auf Pseudo-Beobachtungen beruht, die durch ein geeignetes Resampling-Verfahren generiert wurden. Für den Prozess der zugrundeliegenden Beobachtungen wird das auf Doukhan und Louhichi basierende Abhängigkeitskonzept der ψ-Weak Dependence verwendet, welches allgemeiner ist als das Mixing-Konzept und etwa auch innovationsgesteuerte Prozesse mit diskret verteilten Innovationen umfasst. Es wird zunächst die

Verteilungskonvergenz des geschätzten empirischen Prozesses gegen einen zentrierten Gaußprozess nachgewiesen. Weiterhin wird für den Bootstrap-Prozess die Verteilungskonvergenz in Wahrscheinlichkeit gegen einen zentrierten Gaußprozess gezeigt und daraus für geeignete Teststatistiken die schwache Konsistenz des Bootstrap-Verfahrens in der Supremums-Metrik abgeleitet.

Inhaltsverzeichnis

1 Modalwertschätzung im Rahmen der nichtparametrischen Kurvenschätzung 5

1.1 Modalwertschätzung in Dekonvolutionsmodellen 6

1.1.1 Kernschätzer in nichtparametrischen Dekonvolutionsmodellen 8

1.1.2 Obere Schranken für die Konvergenzraten der Modalwertschätzer 12

1.1.3 Untere Schranken für das quadratische Risiko des Modalwertschätzers 23

1.1.4 Hilfsresultate 33

1.2 Regressionskurven im Fixed Design Modell mit nichtglatter Modalstelle 48

1.2.1 Kernschätzer im Fixed Design Modell 50

1.2.2 Obere Schranken für die Konvergenzrate 52

1.2.3 Untere Schranken für die Konvergenzrate 55

1.2.4 Beweise für die Resultate im Fixed Design Modell 56

1.2.5 Beweise der Hilfsresultate 61

2 Blockwise Bootstrap für den geschätzten empirischen Prozess bei schwach abhängigen Beobachtungen 67

2.1 Empirische Prozesse mit geschätztem Parameter und Goodness-of-Fit 67

2.2 Verteilungskonvergenz des geschätzten empirischen Prozesses . . . 73

2.2.1 Verteilungskonvergenz von Prozessen 73

2.2.2 Voraussetzungen und Grenzverhalten des geschätzten empirischen Prozesses 76

2.2.3 Beweis der Verteilungskonvergenz des Originalprozesses . . 79

2.2.4 Konvergenz des Originalprozesses unter fixierten Alternativen 109
2.3 Imitation des Originalprozesses durch einen Bootstrap-Prozess . . 110
2.3.1 Voraussetzungen und Konvergenz des Bootstrap-Prozesses 113
2.3.2 Beweis der Verteilungskonvergenz des Bootstrap-Prozesses 116
2.4 Anwendung auf Teststatistiken und Bootstrap-Konsistenz 162
2.5 Hilfsresultate 165
2.5.1 Abschätzung für Kovarianzen durch lipschitzstetige Glättungen 165
2.5.2 Hilfsresultate zur Konvergenz des Originalprozesses 169
2.5.3 Hilfsresultate zur Konvergenz des Bootstrap-Prozesses . . . 172

3 Ausblick **183**

Symbolverzeichnis **184**

Literaturverzeichnis **188**

1 Modalwertschätzung im Rahmen der nichtparametrischen Kurvenschätzung

Im Themengebiet der nichtparametrischen Schätzung von Wahrscheinlichkeitsdichten und Regressionskurven befasst sich die Modalwertschätzung mit der Fragestellung, an welcher Stelle die interessierende Kurve ihr Maximum erreicht. Im Fall von Dichtefunktionen bezeichnet der Modalwert also die Stelle, an welcher die meiste Wahrscheinlichkeitsmasse konzentriert ist. Die Schätzung des Modalwertes ist insbesondere bei Fragestellungen von Interesse, bei der die unbekannte Dichtefunktion der Beobachtungen als nichtsymmetrisch angenommen wird, während sich hierfür bei symmetrischen Dichten ebenfalls Erwartungswert und Median als zu schätzende Größen anbieten. Anwendung findet die Schätzung des Modalwertes einer Dichte etwa im Bereich der Methode der Molekularen Uhr, wie in Hedges und Shah [37] dargelegt. Ziel hierbei ist es, den Zeitpunkt in der Evolutionsgeschichte abzuschätzen, zu welchem die Aufspaltung zweier Arten stattgefunden hat. Realisiert wird dies, indem jeweils der Zeitpunkt der evolutionsgeschichtlichen Aufspaltung von bestimmten Proteinen anhand der Unterschiede in der DNA-Sequenz geschätzt wird. Der Modalwert der Verteilung dieser Zeitschätzungen wird nun als der Zeitpunkt der Aufspaltung beider Spezies angenommen. Indem die Zeitschätzungen für eine Vielzahl von Proteinen durchgeführt werden, kann der Modalwert anhand dieser Daten geschätzt werden.
In der Regressionskurvenschätzung spielt der Modalwert etwa zur Durchführung von Gruppenvergleichen eine Rolle. Ein klassisches Anwendungsbeispiel wurde von Müller [50] genannt. Hierbei fungieren die Modalwerte in der zweiten bzw.

ersten Ableitung von menschlichen Wachstumskurven in bestimmten Altersgruppen als Indikatoren für Wachstumsschübe. Gruppenvergleiche können hier beispielshalber zwischen Mädchen und Jungen angestellt werden.

Für den Fall, dass die den interessierenden Kurven zugrundeliegenden Daten direkt beobachtet werden können, liegen bereits zahlreiche Resultate zur Modalwertschätzung vor.
In Parzen [56] wurde erstmals die Konsistenz und die asymptotische Normalität des auf einem Kerndichteschätzer, nämlich dem Rosenblatt-Parzen-Schätzer, basierenden Modalwertschätzers bewiesen. Eddy [25] konnte etwas bessere Konvergenzraten dieses Modalwertschätzers nachweisen. Untere Schranken für das quadratische Risiko des Modalwertschätzers wurden erstmals in Hasminskii [36] untersucht. Hierbei wurde vorausgesetzt, dass die zugrundeliegende Dichte zweimal stetig differenzierbar ist. In Romano [67] wurden diese Resultate auf Funktionen höherer Glattheit erweitert und untere Schranken in Abhängigkeit der Glattheit der Kurve bewiesen. Ebenso wurde in Romano [67] gezeigt, dass diese Raten zugleich auch obere Schranken für die Konvergenzgeschwindigkeit des Modalwertschätzers sind.

In Müller [50] und Müller [51] wurden im Rahmen der Modalwertschätzung für Regressionsfunktionen für das quadratische Risiko des Modalwertschätzers dieselben Raten wie in Eddy [25] erzielt. Diese stimmen wiederum im Falle dreimal differenzierbarer Funktionen mit den in Romano [67] erzielten Resultaten überein.

1.1 Modalwertschätzung in Dekonvolutionsmodellen

Während in den oben erwähnten Arbeiten die der Dichteschätzung zugrundeliegenden Daten als direkt beobachtbar angenommen wurden, geht man in Dekonvolutionsmodellen davon aus, dass die interessierenden Daten nur indirekt beobachtet werden können. Sind im Falle der Dichteschätzung also X_i, $i = 1, ..., n$, die interessierende Zufallsvariablen, werden nicht diese beobachtet, sondern $Y_1, ..., Y_n$

mit

$$Y_i = X_i + \varepsilon_i.$$

Hierbei sind die X_i und ε_i unabhängig und verteilt wie X und ε. Eine übliche Methode zur Schätzung des Modalwertes der interessierenden Dichte f_X von X ist es, dass zunächst in geeigneter Weise die Schätzung der Dichte erfolgt und dann als Schätzer der Modalwert des Dichteschätzers herangezogen wird. Diese Methode wird auch in der folgenden Arbeit angewendet.
Zur Dichteschätzung im Dekonvolutionsmodell mithilfe von Kernschätzern existieren zahlreiche Resultate. Die gleichmäßige Konsistenz des auf Dekonvolutionskernen beruhenden Dichteschätzers wurde in Liu und Taylor [48] nachgewiesen. Stefanski [71] und Stefanski und Carrol [72] erzielten obere Schranken für die Konvergenzraten. Was die Optimalität der Konvergenzraten anbelangt, so konnte Fan [28] obere und entsprechende untere Schranken für das quadratische Risiko dieses Kurvenschätzers zeigen.
Analoge Resultate bestehen für die der Dichteschätzung mit indirekten Beobachtungen verwandte Fragestellung der Regression mit Fehlern in den Variablen (auch als Errors-in-Variables-Regression bezeichnet). Geschätzt werden soll hierbei der Modalwert der Regressionskurve $\mathbb{E}(Z|X = x)$, wobei wiederum die Regressorvariable X nicht beobachtbar ist, sondern in analoger Weise wie im oben beschriebenen Modell nur $Y = X + \varepsilon$.
Fan [28] konnte für die Regressionskurvenschätzung mit Fehlern in den Variablen ebenfalls obere und untere Schranken für das quadratische Risiko nachweisen.

Was nun die Modalwertschätzung in Dekonvolutinsmodellen anbelangt, so existiert eine entsprechende Arbeit, in welcher im Fall der Dichteschätzung Konvergenzraten des Modalwertschätzers erwähnt sind, siehe Rachdi und Sabre [63]. Allerdings sind die dort erzielten Raten offensichtlich abgeleitet von denjenigen Raten, welche für die gleichmäßige Konsistenz des Kurvenschätzers im Falle zweimal stetig differenzierbarer Funktionen nachgewiesen wurden. Die so erzielten Raten des Modalwertschätzers sind zum einen nicht optimal, und zum anderen hängen sie ebenso nicht von höherer Glattheit der zugrundeliegenden Dichte ab.

In der aktuellen Arbeit werden obere Schranken für die Konvergenzrate nachgewiesen, die eine Verbesserung und Verallgemeinerung der Resultate von Rachdi

und Sabre [63] darstellen. Die Konvergenzrate hängt hierbei von der Glattheit der zugrundeliegenden zu schätzenden Kurve ab, sowohl im Falle der Dichte- als auch der Regressionskurvenschätzung in Dekonvolutionsmodellen. Es wird gezeigt, dass unter Voraussetzung höherer Glattheit der zugrundeliegenden Kurve bessere Raten erzielt werden können.
Unlängst wurde in Meister [49] ebenfalls die Modalwertschätzung für Dichten im Dekonvolutionsmodell betrachtet. Es werden dort keinerlei Stetigkeits- oder Glattheitsanforderungen an die zugrundeliegende Kurve gestellt und Minimaxraten erzielt für eine Funktionenklasse, die somit allgemeiner ist als in sämtlichen anderen Arbeiten, die zur Modalwertschätzung vorliegen. Der dort verwendete Schätzer maximiert nicht einen Kurvenschätzer über ein gesamtes Intervall, sondern schätzt die Wahrscheinlichkeitsmasse einer endlichen Anzahl kleiner Intervalle. Die dort erzielten Raten sind für diese allgemeine Funktionenklasse optimal und stimmen überein mit den in Wieczorek [78] erzielten Raten für zweimal stetig differenzierbare Funktionen.
Optimale obere Schranken für das quadratische Risiko des Modalwertschätzers wurden bislang aus der asymptotischen Normalität des Modalwertschätzers abgeleitet. In der vorliegenden Arbeit werden obere Schranken erzielt, indem der Kontrast der Funktion an der Modalstelle analysiert wird, ohne auf die Raten der gleichmäßigen Konsistenz des Kurvenschätzers zurückgreifen zu müssen, was keine optimale Konvergenzrate des Modalwertschätzers liefern würde. Dies wurde bereits in Neumann [54] im Fall der Change-Point-Schätzung umgesetzt.
Ebenso wird gezeigt werden, dass die erzielten oberen Schranken optimal sind. Dies wird über die Analyse eindimensionaler Teilprobleme realisiert, wie es in Donoho und Liu [21] erstmalig vorgeschlagen wurde.
Die entsprechenden Resultate sind in Wieczorek [78] zusammengefasst.

1.1.1 Kernschätzer in nichtparametrischen Dekonvolutionsmodellen

Wie bereits im vorigen Abschnitt erläutert, wird die Schätzung des Modalwertes der jeweiligen Kurve über die Bestimmung des Modalwertes eines geeigneten Kernschätzers dieser Kurve erfolgen. Im folgenden werden nun zunächst zwei

Kernschätzer vorgestellt, welche in den oben beschriebenen Dekonvolutionsmodellen Verwendung finden.

Bei der Dichteschätzung ist der verwendete Schätzer vom Rosenblatt-Parzen-Typ, wobei hier anstelle der gewöhnlichen Kernfunktion ein Dekonvolutionskern verwendet wird. Im Regressionsmodell mit Fehlern in den Variablen wird ein Schätzer vom Nadaraya-Watson-Typ eingesetzt, ebenso mit einem Dekonvolutionskern anstelle des gewöhnlichen Kernes.

An die zu schätzenden Funktionen stellen wir in beiden Fällen die Anforderung, dass diese in der folgenden Klasse p-mal differenzierbarer Funktionen mit eindeutiger Modalstelle θ liegen, wobei $p \geq 2$ gilt:

$$\begin{aligned}\mathcal{F}_p := \{f(\cdot)\colon\, & f^{(p)}(\cdot) \text{ existiert, ist stetig und beschränkt,} \\ & \exists\theta\colon \forall\mu > 0\, \exists\delta > 0\colon\, f(\theta) - f(x) > \delta\ \forall x \text{ mit } |x-\theta| > \mu, \quad f''(\theta) < 0\}.\end{aligned}$$

Im Fall der Dichteschätzung sind nur Wahrscheinlichkeitsdichten in dieser Klasse zugelassen, d.h. in diesem Fall wird folgende Klasse von Funktionen betrachtet:

$$\widetilde{\mathcal{F}}_p := \mathcal{F}_p \cap \{f(\cdot) :\ f \text{ ist Wahrscheinlichkeitsdichte}\}.$$

Dichteschätzung im Dekonvolutionsmodell

Wir richten nun unser Augenmerk auf die Schätzung einer Wahrscheinlichkeitsdichte im Dekonvolutionsmodell. Ausgangspunkt sind unabhängige und identisch verteilte Beobachtungen $Y_1, ..., Y_n$, welche verteilt sind wie

$$Y = X + \varepsilon. \tag{1.1}$$

Die Zufallsvariable X besitze hierbei die Dichte f_X, und X und ε werden als unabhängig vorausgesetzt. Wir sind also an der Verteilung von X interessiert, können X aber nur indirekt über die Zufallsvariable Y beobachten, mit dem Fehler ε.
Die Verteilung von ε wird als bekannt vorausgesetzt. Ebenso wird für diese Verteilung gewöhnliche Glattheit vom Grad β angenommen. Dies beduetet, dass die charakteristische Funktion ϕ_ε der Verteilung von ε folgende Ungleichung erfüllt,

für positive Konstanten c_0 und c_1:

$$c_0(1+|t|)^{-\beta} \leq |\phi_\varepsilon(t)| \leq c_1(1+|t|)^{-\beta}, \forall t \in \mathbb{R}. \tag{1.2}$$

Zahlreiche bekannte Verteilungen erfüllen die Bedingung der gewöhnlichen Glattheit, darunter Gamma-, Laplace-, Hyperexponential- und Pearson-Verteilung. (siehe Ushakov [75], Appendix B).

Aufgrund der Unabhängigkeit von X und ε lässt sich die charakteristische Funktion von Y als das Produkt der charakteristischen Funktionen von X und ε darstellen:

$$\phi_Y(\cdot) = \phi_X(\cdot)\phi_\varepsilon(\cdot)\,.$$

Da aufgrund von Voraussetzung (1.2) ϕ_ε strikt positiv ist, kann die charakteristische Funktion als Quotient von ϕ_Y und ϕ_ε dargestellt werden. Ein möglicher Schätzer für ϕ_X ergibt sich daher folgendermaßen:

$$\hat{\phi}_X = \frac{\tilde{\phi}_Y}{\phi_\varepsilon},$$

wobei $\tilde{\phi}_Y$ ein geeigneter Schätzer für die charakteristische Funktion von Y ist. Als Schätzer für die Dichte f_X kann somit die inverse Transformation von $\hat{\phi}_X$ herangezogen werden. Dieser Schätzer wird im folgenden mit $\hat{f}_{X,n}$ bezeichnet. Um sicherzustellen, dass die inverse Transformation von $\hat{\phi}_X$ existiert, muss die Integrierbarkeit von $\frac{\tilde{\phi}_Y}{\phi_\varepsilon}$ sichergestellt werden, was durch entsprechende Wahl des Schätzers $\tilde{\phi}_Y$ erfolgt. Dies gilt, indem gesetzt wird

$$\tilde{\phi}_Y = \hat{\phi}_Y(t)\phi_K(b_n t),$$

wobei $\hat{\phi}_Y(t) = \frac{1}{n}\sum\limits_{j=1}^{n} e^{itY_j}$ die empirische charakteristische Funktion von Y bezeichnet. b_n ist eine Bandbreite, welche nichtzufällig gewählt wird, und $K(\cdot)$ bezeichnet einen Kern der Ordnung $p-1$.
Die charakteristische Funktion des Kerns erfülle folgende Bedingung, welche die Endlichkeit aller Momente des Dichteschätzers garantiert:

$$\int\limits_{\mathbb{R}} (|u|^{1+\beta}|\phi_K(u)|)^r du < \infty \quad \text{für } 1 \leq r \leq 2. \tag{1.3}$$

Der Dichteschätzer $\hat{f}_{X,n}$ an der Stelle t kann demnach geschrieben werden als

$$\begin{aligned}
\hat{f}_{X,n}(t) &= \frac{1}{2\pi}\int\limits_{\mathbb{R}} e^{-itu}\,\hat{\phi}_X(u)du \;=\; \frac{1}{2\pi}\int\limits_{\mathbb{R}} e^{-itu}\,\frac{\hat{\phi}_Y(u)\,\phi_K(b_n u)}{\phi_\varepsilon(u)}\,du \\
&= \frac{1}{nb_n}\sum_{j=1}^{n}\frac{1}{2\pi}\int\limits_{\mathbb{R}} e^{-i(\frac{t-Y_j}{b_n})u}\,\frac{\phi_K(u)}{\phi_\varepsilon(u/b_n)}\,du \\
&= \frac{1}{nb_n}\sum_{j=1}^{n} K^*_{b_n}\big(\frac{t-Y_j}{b_n}\big)\,, \qquad (1.4)
\end{aligned}$$

wobei

$$K^*_{b_n}(x) \;:=\; \frac{1}{2\pi}\int\limits_{\mathbb{R}} e^{-ixu}\,\frac{\phi_K(u)}{\phi_\varepsilon(u/b_n)}\,du.$$

Mit $K^*_{b_n}(\cdot)$ wird also der Dekonvolutionskern bezeichnet.
Optimale lokale und globale Raten wurden in Fan [28] und Fan [29] für derartige Kernschätzer nachgewiesen.
Auch wenn die Wahl der Bandbreite im hier untersuchten Fall nicht datengesteuert erfolgt, könnte in der Praxis eine Bestimmung der Bandbreite via Kreuzvalidierung erfolgen. Dieser Ansatz wurde bereits in Hesse [40] für Dichteschätzung in Dekonvolutionsmodellen verfolgt. Was Modalwertschätzung in Modellen mit direkten Beobachtungen anbelangt, so wurde diese Methode von Romano [67] in der Dichteschätzung und von Müller [50] in der Regressionsanalyse verwendet, wobei dieselben asymptotischen Resultate erzielt wurden wie im Fall nichtdatenabhängiger Bandbreiten.

Kurvenschätzung im Regressionsmodell mit Fehlern in den Variablen

Im Regressionsmodell mit Fehlern in den Variablen geht es darum, die Abhängigkeit einer Variable Z von einer Regressorvariable X zu beschreiben, wobei X nicht direkt beobachtet werden kann.
Gegeben sind also Tupel von Beobachtungen (Y_i, Z_i), $i = 1, ..., n$. Die Tupel (Y_i, Z_i) seien hierbei unabhängig und identisch verteilt. Die Z_i sind verteilt wie Z, während die Y_i dieselbe Verteilung besitzen wie $Y = X + \varepsilon$. ε und X sind wiederum unabhängig.
Der Zufallsvektor (X, Z) besitze eine zweidimensionale Dichte, welche mit $f_{X,Z}$

bezeichnet werde.
Die zu schätzende Regressionsfunktion ist nun gegeben durch

$$m(x) = \mathbb{E}(Z|X = x) = \int z f_{Z|X=x}(z)dz = \frac{\int z f_{X,Z}(x,z)dz}{\int f_{X,Z}(x,z)dz} =: \frac{r(x)}{f_X(x)}\,.$$

Für dieses Modell wurde in Fan und Truong [30] ein Schätzer vom Nadaraya-Watson-Typ vorgeschlagen. Ebenso wie bei der Dichteschätzung im Dekonvolutionsmodell wird hier ein Dekonvolutionskern zur Schätzung der Randdichte von X verwendet. Der Schätzer für die Regressionsfunktion ist dann gegeben durch

$$\hat{m}_n(x) = \frac{\frac{1}{nb_n}\sum\limits_{j=1}^{n} K^*_{b_n}\left(\frac{x-Y_j}{b_n}\right) Z_j}{\frac{1}{nb_n}\sum\limits_{j=1}^{n} K^*_{b_n}\left(\frac{x-Y_j}{b_n}\right)} = \frac{\hat{r}_n(x)}{\hat{f}_{X,n}(x)}. \tag{1.5}$$

Ebenso wie im vorherigen Abschnitt bezeichnet

$$\frac{1}{nb_n}\sum_{j=1}^{n} K^*_{b_n}\left(\frac{x-Y_j}{b_n}\right)$$

den Schätzer für f_X, wobei die charakteristische Funktion der Kernfunktion $K(\cdot)$ wiederum Gleichung (1.3) erfüllt.

Die Verteilung der Fehlervariablen ε wird wiederum als gewöhnlich glatt angenommen, wie durch Gleichung (1.2) beschrieben.
Optimale lokale und globale Raten für diesen Kernschätzer wurden ebenfalls in Fan und Truong [30] gezeigt.

1.1.2 Obere Schranken für die Konvergenzraten der Modalwertschätzer

Modalwertschätzung für Dichten im Dekonvolutionsmodell

Wir sind nun daran interessiert, den Modalwert der Dichtefunktion f_X zu schätzen, basierend auf der Dichteschätzung im Dekonvolutionsmodell, wie im Abschnitt 1.1.1 beschrieben. Der Modalwert der Dichte f_X sei folgendermaßen definiert

$$\theta = \operatorname{argmax} f_X(x).$$

Die Eindeutigkeit des Modalwertes wird dadurch sichergestellt, dass die Dichte f_X in der oben eingeführten Klasse $\widetilde{\mathcal{F}}_p$ liegt.

Des weiteren werden wir annehmen, dass der Modalwert in einem kompakten Intervall $[\tau_1, \tau_2]$ liegt, welches als bekannt vorausgesetzt wird.

Als Schätzer für den unbekannten Wert θ definieren wir

$$\hat{\theta}_n = \inf\{t|\, \hat{f}_{X,n}(t) = \max_{s\in[\tau_1,\tau_2]}\{\hat{f}_{X,n}(s)\}\}.$$

Die Messbarkeit dieses Schätzers wurde in Eddy [25] nachgewiesen.
Die Definition von $\hat{\theta}_n$ in obiger Weise ist die am meisten verbreitete Methode um den Modalwert einer Funktion zu schätzen. Dennoch bestehen einige andere Methoden zur Schätzung des Modalwertes. In Vieu [76] wurden Konvergenzraten eines Modalwertschätzers nachgewiesen, welcher auf der Schätzung der ersten Ableitung der zugrundeliegenden Kurve beruht. Dies wird sowohl für nicht datengesteuerte als auch datengesteuerte Bandbreiten untersucht. In Bickel und Fan [9] wurden Konvergenzraten über einen Maximum-Likelihood-Ansatz bestimmt, und Shoung und Zhang [70] verwendeten einen Kleinste-Quadrate-Schätzer. In Meister [49] erfolgt die Schätzung des Modalwertes über die Schätzung der Wahrscheinlichkeitsmasse endlich vieler Intervalle, wie bereits eingangs erwähnt.

Was nun obere Schranken für die Konvergenzrate des oben definierten Modalwertschätzers im Dekonvolutionsmodell anbelangt, wurde in Wieczorek [78] das folgende Resultat hinsichtlich des quadratischen Risikos von $\hat{\theta}_n$ bewiesen:

Theorem 1.1.1
Es seien $Y_1, ..., Y_n$ gemäß Modell (1.1)*unabhängig und identisch verteilt wie $Y = X + \varepsilon$, wobei die charakteristische Funktion von ε gewöhnlich glatt der Ordnung β sei, $\beta \geq 0$, siehe hierzu* (1.2).
Als Kurvenschätzer für $f_X \in \mathcal{F}_p$, $p \geq 2$, diene der Kernschätzer mit Dekonvolutionskern, wie in (1.4) *definiert. Hierbei sei K ein Kern der Ordnung $p-1$, für welchen* (1.3) *gelte. Die Bandbreite b_n sei von der Ordnung $n^{-\frac{1}{2\beta+2p+1}}$. Dann gilt für das quadratische Risiko der Modalwertschätzers $\hat{\theta}_n$, dass*

$$\mathbb{E}[(\hat{\theta}_n - \theta)^2] = O(n^{-\frac{2(p-1)}{2p+2\beta+1}}).$$

Im Fall $\beta = 0$, welches der Fall direkter Beobachtungen ist, stimmt diese Rate überein mit der in Romano [67] erzielten oberen Schranke.

Bislang wurden im Fall direkter Beobachtungen optimale obere Schranken für das quadratische Risiko über Methoden nachgewiesen, welche auf der Theorie empirischer Prozesse beruhen. Hierzu wurde stets angenommen, dass die zugrundeliegende Funktionen mindestens dreimal differenzierbar ist. Obgleich inRachdi und Sabre [63] asymptotische Normalität des Modalwertschätzers für Wahrscheinlichkeitsdichten im Dekonvolutionsmodell nachgewiesen wurde, ergibt deren Annahme $nb_n^7 \to 0$ nicht die optimale Rate im dort betrachteten Fall.
Eine weitere Möglichkeit zum Nachweis oberer Schranken für die Konvergenzrate ist es, die Abweichung des Schätzers $\hat{\theta}_n$ von θ abzuschätzen, indem der Kontrast der Funktion an der Modalstelle analysiert wird. Dieser kann beschrieben werden durch

$$f_X(\theta) - f_X(x) \geq d_1(\theta - x)^\kappa \text{ mit } d_1 > 0, \text{ für } x \text{ in einer Umgebung von } \theta.$$

Je kleiner die Konstante κ ist, desto steiler ist also das Abfallen der Kurve an der Modalstelle.
Dieser Ansatz wurde zuerst in Müller [50] verfolgt. Dadurch kann die Konvergenzrate des Modalwertschätzers nach oben abgeschätzt werden durch die Rate der globalen gleichmäßigen Konvergenz des verwendeten Kurvenschätzers.
In Rachdi und Sabre [63] wurde ebenfalls ein ähnlicher Ansatz verwendet, um Raten für das quadratische Riskio des Modalwertschätzers abzuleiten. Allerdings ergibt dies einen zusätzlichen log-Term als Faktor in der erzielten Konvergenzrate, ebenso kann kein Nutzen aus höherer Glattheit der zugrundeliegenden Kurve gezogen werden.

Um optimale Raten, die von der Glattheit der Kurve abhängen, zu erzielen, wird in dieser Arbeit eine ähnliche Idee verfolgt. Hierbei wird die κ-te Potenz der Abweichung des Modalwertschätzers folgendermaßen nach oben beschränkt, ebenso wie in Müller [50]:

$$(\hat{\theta}_n - \theta)^\kappa \leq O(\hat{f}_{X,n}(\hat{\theta}_n) - f_X(\hat{\theta}_n) - (\hat{f}_{X,n}(\theta) - f_X(\theta))).$$

Die rechte Seite kann wiederum nach oben abgeschätzt werden. Hierbei wird zwischen den Fällen unterschieden, in welchen $\hat{\theta}_n$ innerhalb bestimmter Intervalle

eines Gitters liegt, wobei die Feinheit des Gitters mit wachsendem n zunimmt. Aufgrund der Glattheit sowohl der Kurve selbst als auch des Kurvenschätzers kann von der Nähe von θ und den einzelnen Gitterpunkten profitiert werden. Auf diese Weise wird eine genauere Abschätzung für die obere Schranke erzielt und dadurch können letztendlich optimale Raten für das quadratische Risiko des Modalwertschätzers $\hat{\theta}_n$ abgeleitet werden.

Im hier betrachtete Fall, in welchem die unbekannte Kurve gemäß Annahme der Klasse $\widetilde{\mathcal{F}}_p$ entstammt, ist der Kontrast von der Ordnung $\kappa = 2$, welches durch eine Taylorentwicklung bei θ gezeigt wird. Hieraus ist auch ersichtlich, dass die getätigte Annahme $f_X''(\theta) < 0$ nicht zwingend erforderlich ist. Die Rate vergrößert sich lediglich zu

$$O(n^{-\frac{1}{\kappa-1}\frac{2(p-1)}{2p+2\beta+1}}),$$

falls der Kontrast an der Modalstelle von der Ordnung $\kappa \geq 2$ ist. Der Einfachheit und Lesbarkeit halber wird der obere Fall bewiesen, wenngleich sich der Beweis unmittelbar auf den Fall übertragen lässt, in welchem der Kontrast schwächer als quadratisch ist.
Für kleineres κ, welches den Verlust der Differenzierbarkeit an der Modalstelle bedeutet, kann hingegen die Rate $O(n^{-1/2\kappa+1})$ bewiesen werden, wobei hier derselbe Ansatz verwendet werden kann wie in Wieczorek und Ziegler [79].

Beweis von Theorem 1.1.1
Um obere Schranken für die Konvergenzraten des Modalwertschätzers zu beweisen betrachten wir ein Gitter $\widetilde{G}_n$ der Breite h_n auf dem Intervall $[\tau_1, \tau_2]$. Der Gitterpunkt, welcher k Schritte von θ entfernt ist, werde mit $\theta_{n,k}$ bezeichnet. Das Gitter sei also folgendermaßen definiert:

$$\widetilde{G}_n \quad = \quad \{\theta_{n,k} = \theta + k \cdot h_n \in [\tau_1, \tau_2] |\, k \in \mathbb{Z}\}.$$

Um das quadratische Risiko zu beschränken, werden dann obere Schranken angegeben für die Wahrscheinlichkeiten, dass der Modalwertschätzer $\hat{\theta}_n$ innerhalb des Intervalls $[\theta_{n,k-1}, \theta_{n,k}]$ liegt. Abkürzend werden diese Intervalle mit $T_{n,k}$ bzeichnet.

Aufgrund der Voraussetzung $f_X''(\theta) < 0$ ergibt sich die folgende Kontrasteigenschaft an der Modalstelle der Kurve f_X, wobei x in einer μ-Umgebung von θ liegt,

für ein $\mu > 0$:

$$f_X(\theta) - f_X(x) \geq d_1(\theta - x)^2,\ d_1 > 0. \tag{1.6}$$

Dies wird durch eine Taylorentwicklung bei θ ersichtlich:

$$\begin{aligned} f_X(x) &= f_X(\theta) + (\theta - x) \cdot f_X'(\theta) + \frac{1}{2} \cdot (\theta - x)^2 \cdot f_X''(\xi) \\ &= f_X(\theta) + \frac{1}{2} \cdot (\theta - x)^2 \cdot f_X''(\xi). \end{aligned}$$

Der zweite Summand in der ersten Zeile entspricht dem Restglied nach Lagrange und ξ bezeichnet somit eine Stelle zwischen θ und x. Aufgrund der Voraussetzung $f_X''(\theta) < 0$ und der Stetigkeit von $f_X''(\cdot)$ gilt also auch $f_X''(\xi) < 0$, falls die μ-Umgebung von θ hinreichend klein gewählt ist. An dieser Stelle ist auch ersichtlich, dass bei Vorhandensein höherer stetiger Ableitungen auch bei Verschwinden höherer Ableitungen eine entsprechende Kontrasteigenschaft per Taylorentwicklung gezeigt werden kann, wobei sich hier der Exponent bei (1.6) erhöht, abhängig von der ersten nichtverschwindenden Ableitung bei θ.
Insgesamt ergibt sich im obigen Fall also für den Kontrast innerhalb der ε-Umgebung Eigenschaft (1.6):

$$f_X(\theta) - f_X(x) = (-\frac{1}{2}) \cdot f_X''(\xi) \cdot (\theta - x)^2 \geq d_1 \cdot (\theta - x)^2,\ d_1 > 0.$$

Außerhalb der μ-Umgebung gilt aufgrund der Eindeutigkeit der Modalstelle, wie in $\mathcal{F}_p$ definiert:

$$f_X(\theta) - f_X(x) \geq \delta, \text{ für ein } \delta > 0.$$

Das Gitter $\widetilde{G}_n$ kann somit in zwei Teilmengen unterteilt werden, wobei die eine diejenigen Gitterpunkte enthält, welche innerhalb der μ-Umgebung liegen, und die anderen jene Punkte, die außerhalb der μ-Umgebung liegen. Die zugehörigen Indizes werden in den Mengen, G_n, G_n^1 und G_n^2 zusammengefasst, welche also folgendermaßen definiert seien

$$\begin{aligned} G_n &= \{k \in \mathbb{Z} : \theta_{n,k} \in [\tau_1, \tau_2]\}, \\ G_n^1 &= \{k \in \mathbb{Z} : \theta_{n,k} \in (\theta - \mu, \theta + \mu)\}, \\ G_n^2 &= \{k \in \mathbb{Z} : \theta_{n,k} \in [\tau_1, \tau_2] \setminus (\theta - \mu, \theta + \mu)\}. \end{aligned}$$

Das quadratische Risiko des Modalwertschätzers kann somit folgendermaßen nach oben abgeschätzt werden:

$$\mathbb{E}\big((\theta - \hat{\theta}_n)^2\big) \quad \leq \quad \sum_{k \in G_n} \big(\sup_{t \in T_{n,k}} (t - \theta)^2 \big) \cdot \mathbb{P}(\hat{\theta}_n \in T_{n,k}).$$

Das Supremum im obigen Ausdruck kann nun für die direkt an θ angrenzenden Intervalle durch h_n^2 nach oben abgeschätzt werden. Für die weiter entfernten Intervalle ergibt sich die Abschätzung $(|k|+1)^2$. Man erhält also weiter:

$$\begin{aligned}
\mathbb{E}\big((\theta - \hat{\theta}_n)^2\big) \quad \leq \quad & h_n^2 \cdot \mathbb{P}(\hat{\theta}_n \in T_{n,0} \cup T_{n,1}) \\
& + h_n^2 \cdot \sum_{\substack{k \in G_n^1 \\ k \neq 0,1}} (|k|+1)^2 \cdot \mathbb{P}(\hat{\theta}_n \in T_{n,k}) \\
& + h_n^2 \cdot \sum_{\substack{k \in G_n^2 \\ k \neq 0,1}} (|k|+1)^2 \cdot \mathbb{P}(\hat{\theta}_n \in T_{n,k}).
\end{aligned}$$

Um nun zu einer Abschätzung des quadratischen Risikos zu gelangen, bleibt es also, die Wahrscheinlichkeiten abzuschätzen, dass $\hat{\theta}_n$ im Intervall $T_{n,k}$ liegt. Falls es gelingt zu beweisen, dass $\mathbb{P}(\hat{\theta}_n \in T_{n,k})$ genügend klein ist, sodass beide letzteren Summen konvergieren, so ergibt sich h_n^2 als Konvergenzrate für das quadratische Risiko.
Um dies zu beweisen, nehmen wir o.B.d.A. im Folgenden an, dass $k > 1$ gilt. Analoge Überlegungen gelten für $k < 0$. Wir haben

$$\begin{aligned}
& \mathbb{P}(\hat{\theta}_n \in T_{n,k}) \leq \mathbb{P}\Big(\sup_{t \in T_{n,k}} \big(\hat{f}_{X,n}(t) - \hat{f}_{X,n}(\theta)\big) \geq 0 \Big) \\
\leq \quad & \mathbb{P}\Big(\sup_{t \in T_{n,k}} (\hat{f}_{X,n}(t) - \hat{f}_{X,n}(\theta) - (f_X(t) - f_X(\theta))) \geq \inf_{t \in T_{n,k}} (f_X(\theta) - f_X(t)) \Big).
\end{aligned}$$

Für den Ausdruck auf der rechten Seite innerhalb der Wahrscheinlichkeit bestehen zwei verschiedene Arten unterer Schranken, je nachdem ob $T_{n,k}$ innerhalb der Umgebung von θ liegt, in welcher die Ungleichung (1.6) gilt, oder außerhalb.

Für $k \in G_n^1$ erhalten wir für den letztgenannten Ausdruck aufgrund von (1.6)

$$\begin{aligned}
& \mathbb{P}\Big(\sup_{t \in T_{n,k}} (\hat{f}_{X,n}(t) - \hat{f}_{X,n}(\theta) - (f_X(t) - f_X(\theta))) \geq \inf_{t \in T_{n,k}} (f_X(\theta) - f_X(t)) \Big) \\
\leq \quad & \mathbb{P}\Big(\sup_{t \in T_{n,k}} \big(\hat{f}_{X,n}(t) - \hat{f}_{X,n}(\theta) - (f_X(t) - f_X(\theta))\big) \geq d_1 \cdot ((|k| - 1) \cdot h_n)^2 \Big) \\
\leq \quad & \mathbb{P}\Big(\sup_{t \in T_{n,k}} |\hat{f}_{X,n}(t) - \mathbb{E}\hat{f}_{X,n}(t) - \big(\hat{f}_{X,n}(\theta) - \mathbb{E}\hat{f}_{X,n}(\theta)\big)| \geq \tilde{d}_1 \cdot (k \cdot h_n)^2 \Big) \\
+ \quad & \mathbb{P}\Big(\sup_{t \in T_{n,k}} |\mathbb{E}\hat{f}_{X,n}(t) - f(t) - \big(\mathbb{E}\hat{f}_{X,n}(\theta) - f_X(\theta)\big)| \geq \tilde{d}_1 \cdot (k \cdot h_n)^2 \Big) \\
=: \quad & \mathbb{P}_{1,k} + \mathbb{P}_{2,k} \,.
\end{aligned}$$

Die Glattheit des Schätzers gewährleistet, dass die Wahrscheinlichkeit, dass das Supremum innerhalb eines Gitterintervalls über die betragsmäßige Abweichung des zentrierten Schätzers vom Wert des zentrierten Schätzers an der Modalstelle eine bestimmte Schranke überschreitet, dieselbe Ordnung besitzt wie die entsprechende Wahrscheinlichkeit, wenn die Abweichung nur an einem festen Punkt innerhalb des Gitterintervalls betrachtet wird. Dies wird gezeigt, indem die Chaining-Technik angewendet wird und eine Momentenungleichung für die Differenzen des zentrierten Kurvenschätzers an zwei festen Punkten ausgenutzt wird. Die Wahrscheinlichkeit $\mathbb{P}_{1,k}$ kann nun durch Anwendung der Markov-Ungleichung nach oben abgeschätzt werden. Mithilfe von Lemma 1.1.1 (iii) und Lemma 1.1.6 ergibt sich also, für $r = 4$,

$$\mathbb{P}_{1,k} = |k|^{-4} O\Big(\frac{(\frac{1}{b_n})^4 (n b_n^{2\beta+1})^{-2}}{h_n^4} \Big).$$

Um eine geeignete obere Schranke für $\mathbb{P}_{2,k}$ zu erhalten, ist zu zeigen, dass für genügend großes n der deterministische Ausdruck auf der linken Seite innerhalb der Wahrscheinlichkeit kleiner als die rechte Seite ist, womit $\mathbb{P}_{2,k}$ für hinreichend großes n verschwindet. Hierzu wird zunächst die Ordnung der linken Seite berechnet, in Abhängigkeit der noch festzulegenden Bandbreite. Mithilfe von Lemma 1.1.1 (i) erhält man hierfür:

$$
\begin{aligned}
& \sup_{t\in T_{n,k}} |\big(\mathbb{E}\hat{f}_{X,n}(t) - \mathbb{E}\hat{f}_{X,n}(\theta)\big) - \big(f_X(t) - f_X(\theta)\big)| \\
= & \sup_{t\in T_{n,k}} \Big| \int_0^{t-\theta} \big(\int_{\mathbb{R}} \frac{1}{b_n} K\big(\frac{\theta - z}{b_n}\big) f_X'(z+u)dz - f_X'(\theta+u)\big) du \Big| \\
\leq & \sup_{t\in T_{n,k}} b_n^{p-1} \frac{||f_X^{(p)}||_\infty}{(p-1)!} \int_0^{t-\theta} \int_{\mathbb{R}} \Big| K(z) z^{p-1} \Big| dz\, du = O(k \cdot h_n \cdot b_n^{p-1}). \qquad (1.7)
\end{aligned}
$$

Falls nun die Wahrscheinlichkeit betrachtet wird, dass der Schätzer $\hat{\theta}_n$ in einem Gitterintervall außerhalb der μ-Umgebung liegt, in welcher die Kontrasteigenschaft (1.6) erfüllt ist, so kann man sich für die Abschätzung an dieser Stelle die Eindeutigkeit der Modalstelle zunutze machen und man erhält für $k \in G_n^2$

$$
\begin{aligned}
& \mathbb{P}(\hat{\theta}_n \in T_{n,k}) \\
\leq \quad & \mathbb{P}\big(\sup_{t\in T_{n,k}} |\hat{f}_{X,n}(t) - \mathbb{E}\hat{f}_{X,n}(t) - \big(\hat{f}_{X,n}(\theta) - \mathbb{E}\hat{f}_{X,n}(\theta)\big)| \geq \frac{\delta}{2}\big) \\
+ \quad & \mathbb{P}\big(\sup_{t\in T_{n,k}} |\mathbb{E}\hat{f}_{X,n}(t) - f(t) - \big(\mathbb{E}\hat{f}_{X,n}(\theta) - f(\theta)\big)| \geq \frac{\delta}{2}\big) \\
=: \quad & \tilde{\mathbb{P}}_{1,k} + \tilde{\mathbb{P}}_{2,k}.
\end{aligned}
$$

In analoger Weise wie oben kann die Wahrscheinlichkeit $\tilde{\mathbb{P}}_{1,k}$ wiederum mithilfe von Lemma 1.1.6 beschränkt werden:

$$
\tilde{\mathbb{P}}_{1,k} = O\Big(\frac{(\frac{kh_n}{b_n})^4 (nb_n^{2\beta+1})^{-2}}{\delta^4} \Big).
$$

Die Bandbreite b_n und die Gitterbreite h_n werden nun folgendermaßen gewählt:

$$
h_n \asymp b_n^{p-1} \text{ sowie } b_n \asymp n^{-\frac{1}{2\beta+2p+1}}.
$$

Damit gilt für ein bestimmtes k_0, dass $\mathbb{P}_{2,k}$ und $\tilde{\mathbb{P}}_{2,k}$ verschwinden für $|k| > k_0$, was aufgrund von (1.7) ersichtlich ist und aufgrund der Tatsache, dass $k \leq \lfloor \frac{|\tau_2 - \tau_1|}{h_n} \rfloor$. Mit den obigen Überlegungen kann das quadratische Risiko demnach folgender-

maßen beschränkt werden:

$$\begin{aligned}
&\mathbb{E}((\theta-\hat{\theta}_n)^2)\\
\leq\ & h_n^2+\sum_{|k|\leq k_0}(kh_n)^2+\sum_{\substack{k\in G_n^1\\ k\neq 0,1}} h_n^2k^{-2}C_1\frac{b_n^{-4}(nb_n^{2\beta+1})^{-2}}{h_n^4}\\
&+\sum_{\substack{k\in G_n^2\\ k\neq 0,1}} h_n^2k^2C_2\frac{(\frac{kh_n}{b_n})^4(nb_n^{2\beta+1})^{-2}}{\delta^4},
\end{aligned}$$

und mit $|k|\leq|\tau_2-\tau_1|/h_n$ kann weiter nach oben abgeschätzt werden

$$\begin{aligned}
&\leq h_n^2\Big(2+\sum_{|k|\leq k_0}k^2+\sum_{k=1}^{\infty}k^{-2}C_1\frac{(nb_n^{2\beta+1})^{-2}}{b_n^4h_n^4}+\sum_{k=1}^{\lfloor\frac{|\tau_2-\tau_1|}{h_n}\rfloor}k^2C_2(\frac{kh_n}{b_n})^4\frac{(nb_n^{2\beta+1})^{-2}}{\delta^4}\Big)\\
&\leq h_n^2\Big(2+C_3+\big(\tilde{C}_1\sum_{k=1}^{\infty}k^{-2}\big)+\tilde{C}_2\Big).
\end{aligned}$$

Der letzte Schritt erfolgt aufgrund der Wahl von h_n und b_n. Daher ergibt sich insgesamt als obere Schranke für das quadratische Risiko

$$\mathbb{E}\big((\theta-\hat{\theta}_n)^2\big)=O(h_n^2)=O(n^{-\frac{2(p-1)}{2p+2\beta+1}}).$$

□

Modalwertschätzung für Regressionskurven im Modell mit Fehlern in den Variablen

Unter Verwendung derselben Techniken wie im Fall der Dichteschätzung kann ein analoges Resultat für die obere Schranke des quadratischen Risikos des Modalwertschätzers ebenso im Regressionsmodell mit Fehlern in den Variablen erzielt werden. Der zu schätzende Modalwert ist in diesem Fall der Modalwert θ der Regressionsfunktion

$$m(x)=\mathbb{E}(Z|X=x).$$

Wiederum wird der Modalwertschätzer durch Maximierung des Schätzers der Kurve $m(\cdot)$ definiert:

$$\hat{\theta}_n=\inf\{t|\,\hat{m}_n(t)=\max_{s\in[\tau_1,\tau_2]}\{\hat{m}_n(s)\}\}. \tag{1.8}$$

Unter einigen zusätzlichen Anforderungen an die Randdichte f_X und den bedingten Erwartungswert $\mathbb{E}(Z|X=x)$ kann folgendes Resultat hinsichtlich oberer Schranken für das quadratische Risiko von $\hat{\theta}_n$ erzielt werden (siehe Wieczorek [78]):

Theorem 1.1.2
Seien $(Y_1, Z_1), ..., (Y_n, Z_n)$ unabhängig und identisch verteilt wie (Y, Z), mit $Y = X + \varepsilon$, wobei X und ε unabhängig seien und die charakteristische Funktion von ε die Bedingung (1.2) erfülle.
Sei weiter $m(x) = \mathbb{E}(Z|X=x) \in \mathcal{F}_p$, $p \geq 2$. $f_X^{(p)}$ existiere, sei stetig und beschränkt und es gelte $f_X \geq c > 0$ in $[\tau_1, \tau_2]$.
Das bedingte Moment $\mathbb{E}(|Z|^{4\cdot q}|X=x)$ sei gleichmäßig beschränkt in x für ein $q > 1$.
Der Kurvenschätzer für m sei definiert wie in (1.5), K ein Kern der Ordnung $p-1$, welcher (1.3) erfülle. Die Bandbreite b_n sei von der Ordnung $n^{-\frac{1}{2\beta+2p+1}}$.
Dann gilt

$$\mathbb{E}(\hat{\theta}_n - \theta)^2 = O(n^{-\frac{2(p-1)}{2p+2\beta+1}}).$$

Auch in diesem Fall kann das Resultat erweitert werden für den Fall, dass an der Modalstelle Ableitungen höherer Ordnung verschwinden, ebenso wie für den Fall geringerer als der oben geforderten Glattheit der zugrundeliegenden Kurve an der Modalstelle.

Beweis von Theorem 1.1.2

Der Beweis der oberen Schranke für die Konvergenzrate des Modalwertschätzers im Errors-in-Variables-Regressionsmodell ist im Wesentlichen analog zu demjenigen für den Dichteschätzer. Einige Veränderungen müssen jedoch vorgenommen werden. Zunächst kann das quadratische Risiko in diesem Fall folgendermaßen abgeschätzt werden:

$$\begin{aligned} & \mathbb{E}\big((\theta - \hat{\theta}_n)^2\big) \\ \leq & \sum_{k \in G_n} \sup_{t \in T_{n,k}} (t-\theta)^2 \cdot \mathbb{P}(A_n) + \sum_{k \in G_n} \sup_{t \in T_{n,k}} (t-\theta)^2 \cdot \mathbb{P}(\{\hat{\theta}_n \in T_{n,k}\} \cap \bar{A}_n). \end{aligned} \tag{1.9}$$

Hierbei ist A_n folgendermaßen definiert:

$$A_n := \{\omega : \sup_{t\in[\tau_1,\tau_2]} |\hat{f}_{X,n}(t) - f(t)| > c/2\}.$$

$\bar{A}_n$ ist also Teilmenge desjenigen Ereignisses, dass der Schätzer der Randdichte f_X von Null weg beschränkt ist. Diese Annahme wird benötigt, um zu gewährleisten, dass die im weiteren Verlauf benötigten Momente entsprechend nach oben beschränkt werden können.
Gemäß Lemma 1.1.3 bestimmt der zweite Term bei (1.9) die Konvergenzordnung des quadratischen Risikos.
Analog zur Dichteschätzung wird nun die Wahrscheinlichkeit abgeschätzt, dass der Modalwertschätzer $\hat{\theta}_n$ im Intervall $T_{n,k}$ liegt. Hierbei erhält man zunächst folgende Abschätzung:

$$\begin{aligned}
& \mathbb{P}(\{\hat{\theta}_n \in T_{n,k}\} \cap \bar{A}_n) \\
\leq\ & \mathbb{P}\Big(\{\sup_{t\in T_{n,k}} (\hat{m}_n(t) - \hat{m}_n(\theta)) \geq 0\} \cap \bar{A}_n\Big) \\
\leq\ & \mathbb{P}\Big(\{\sup_{t\in T_{n,k}} (\hat{m}_n(t) - \hat{m}_n(\theta) - (m(t) - m(\theta))) \geq \inf_{t\in T_{n,k}} (m(\theta) - m(t))\} \cap \bar{A}_n\Big)
\end{aligned}$$

Für $k \in G_n^1$, $k > 1$, gilt weiter

$$\begin{aligned}
\leq\ & \mathbb{P}\Big(\{\sup_{t\in T_{n,k}} \big(\hat{m}_n(t) - \hat{m}_n(\theta) - (m(t) - m(\theta))\big) \geq \tilde{d}_1(kh_n)^2\} \cap \bar{A}_n\Big) \\
\leq\ & \mathbb{P}\Big(\{\sup_{t\in T_{n,k}} \Big|\frac{\hat{r}_n(t)}{\hat{f}_{X,n}(t)} - \frac{\mathbb{E}\hat{r}_n(t)}{\mathbb{E}\hat{f}_{X,n}(t)} - (\frac{\hat{r}_n(\theta)}{\hat{f}_{X,n}(\theta)} - \frac{\mathbb{E}\hat{r}_n(\theta)}{\mathbb{E}\hat{f}_{X,n}(\theta)})\Big| \geq \tilde{d}_1(kh_n)^2\} \cap \bar{A}_n\Big) \\
& + \mathbb{P}\Big(\{\sup_{t\in T_{n,k}} \Big|\frac{\mathbb{E}\hat{r}_n(t)}{\mathbb{E}\hat{f}_{X,n}(t)} - \frac{r(t)}{f_X(t)} - (\frac{\mathbb{E}\hat{r}_n(\theta)}{\mathbb{E}\hat{f}_{X,n}(\theta)} - \frac{r(\theta)}{f_X(\theta)})\Big| \geq \tilde{d}_1(kh_n)^2\} \cap \bar{A}_n\Big) \\
\leq\ & \mathbb{P}\Big(\{\sup_{t\in T_{n,k}} \Big|\frac{\hat{r}_n(t)}{\hat{f}_{X,n}(t)} - \frac{\mathbb{E}\hat{r}_n(t)}{\mathbb{E}\hat{f}_{X,n}(t)} - (\frac{\hat{r}_n(\theta)}{\hat{f}_{X,n}(\theta)} - \frac{\mathbb{E}\hat{r}_n(\theta)}{\mathbb{E}\hat{f}_{X,n}(\theta)})\Big| \geq \tilde{d}_1(kh_n)^2\} \cap \bar{A}_n\Big) \\
& + \mathbb{P}\Big(\{\sup_{t\in T_{n,k}} \Big|\frac{\mathbb{E}\hat{r}_n(t)}{\mathbb{E}\hat{f}_{X,n}(t)} - \frac{r(t)}{f_X(t)} - (\frac{\mathbb{E}\hat{r}_n(\theta)}{\mathbb{E}\hat{f}_{X,n}(\theta)} - \frac{r(\theta)}{f_X(\theta)})\Big| \geq \tilde{d}_1(kh_n)^2\} \cap \bar{A}_n\Big) \\
=:\ & \mathbb{P}_{1,k} + \mathbb{P}_{2,k}\,.
\end{aligned}$$

Die Wahrscheinlichkeit $\mathbb{P}_{1,k}$ wird nun in analoger Weise abgeschätzt wie im Beweis von Theorem 1.1.1. Um Lemma 1.1.6 anwenden zu können, wird hierfür wiederum eine Abschätzung für die vierten Momente des Terms auf der linken Seite innerhalb der Wahrscheinlichkeit benötigt. Diese ist in Lemma 1.1.4 bewiesen.

Unter Verwendung von Lemma 1.1.6 erhält man erhält also auch hier:

$$\mathbb{P}_{1,k} = |k|^{-4} O\Big(\frac{(\frac{1}{b_n})^4 (n b_n^{2\beta+1})^{-2}}{h_n^4}\Big).$$

Hinsichtlich des Terms in $\mathbb{P}_{2,k}$ kann mithilfe von Lemma 1.1.5 ebenso wie im Beweis von Theorem 1.1.1 die Ordnung $O(k \cdot h_n \cdot b_n^{p-1})$ gezeigt werden.
Ebenfalls gilt für $k \in G_n^2$

$$\begin{aligned}
& \mathbb{P}(\{\hat{\theta}_n \in T_{n,k}\} \cap \bar{A}_n) \\
\leq \quad & \mathbb{P}\Big(\{\sup_{t \in T_{n,k}} |\hat{m}_n(t) - \mathbb{E}\hat{m}_n(t) - \big(\hat{m}_n(\theta) - \mathbb{E}\hat{m}_n(\theta)\big)| \geq \frac{\delta}{2}\} \cap \bar{A}_n\Big) \\
+ \quad & \mathbb{P}\Big(\{\sup_{t \in T_{n,k}} \Big|\frac{\mathbb{E}\hat{r}_n(t)}{\mathbb{E}\hat{f}_{X,n}(t)} - \frac{r(t)}{f_X(t)} - \big(\frac{\mathbb{E}\hat{r}_n(\theta)}{\mathbb{E}\hat{f}_{X,n}(\theta)} - \frac{r(\theta)}{f_X(\theta)}\big)\Big| \geq \frac{\delta}{2}\} \cap \bar{A}_n\Big) \\
=: \quad & \tilde{\mathbb{P}}_{1,k} + \tilde{\mathbb{P}}_{2,k}.
\end{aligned}$$

Es gilt auch hier

$$\tilde{\mathbb{P}}_{1,k} = O\Big(\frac{(\frac{k h_n}{b_n})^4 (n b_n^{2\beta+1})^{-2}}{\delta^4}\Big).$$

Nun erfolgt die Abschätzung des quadratischen Risikos in analoger Weise wie im vorherigen Theorem und somit folgt die Aussage von Theorem 1.1.2. □

1.1.3 Untere Schranken für das quadratische Risiko des Modalwertschätzers

Während im vorherigen Abschnitt obere Schranken für die Raten des Modalwertschätzers in den betrachteten Dekonvolutionsmodellen erzielt wurden, wird in den folgenden beiden Theoremen gezeigt, dass diese Raten auch optimal sind, sowohl für die Dichteschätzung als auch für das Errors-in-Variables-Regressionsmodell.
Diese beiden Resultate werden bewiesen, indem sogenannte eindimensionale Teilprobleme betrachtet werden. Hierbei werden Dichten gewählt, sodass die Situationen, welche von diesen Dichten erzeugt werden, nicht vollständig unterschieden werden können, wobei jedoch die Differenz der Modalwerte der interessierenden Funktionen von derselben Ordnung ist wie die Konvergenzrate des quadratischen Risikos von $\hat{\theta}_n$, die in den Theoremen 1.1.1 und 1.1.2 erzielt wurde.

Untere Schranken für die Dichteschätzung

Um untere Schranken zunächst für die Modalwertschätzung im Fall der Dichteschätzung nachzuweisen, wird im Folgenden angenommen, dass die Verteilung der Fehler ε eine Dichte besitzt, bezeichnet mit f_ε. Dann kann, unter den bisher getätigten Annahmen, folgendes Theorem bewiesen werden (siehe Wieczorek [78]):

Theorem 1.1.3
Es gelten die Voraussetzungen von Theorem 1.1.1. Zusätzlich besitze die Verteilung von ε eine Dichte f_ε mit folgender Eigenschaft:
$\int_{\mathbb{R}\setminus[c,d]} |f_\varepsilon^{(\alpha)}(x)|dx \leq C$, $\alpha \geq 2\beta + p$, *für ein kompaktes Intervall* $[c,d]$ *und* $C \in \mathbb{R}$.
Dann gilt, für einen beliebigen Modalwertschätzer $\hat{M}_n$ *für* f_X, *dass*

$$\sup_{f \in \widetilde{\mathcal{F}}_p} \mathbb{E}_f(\hat{M}_n - \theta)^2 \geq d \cdot n^{-\frac{2(p-1)}{2p+2\beta+1}} \qquad \text{für ein } d > 0.$$

$\mathbb{E}_f$ bezeichnet den Erwartungswert, wobei die Beobachtungen die Dichte f besitzen.
An dieser Stelle sollte darauf hingewiesen werden, dass die Bedingung an die Fehlerdichte keine Einschränkung darstellt, was die typischen Fälle gewöhnlich glatter Verteilungen anbelangt. Die Bedingung ist etwa erfüllt für die Gamma- und Hyperexponentialverteilung, für beliebig großes α, falls $0 \in [c,d]$, und für die Laplace-Verteilung, falls $\mu \in [c,d]$, wobei μ den Erwartungswert der Verteilung bezeichnet (siehe hierzu Ushakov [75], Appendix B).
Die beiden konkurrierenden Situationen werden in ähnlicher Weise konstruiert wie in Romano [67]. Hierbei wird eine fixierte Dichtefunktion $f_{X,1}$ betrachtet. Die Dichten $f_{X,2,n}$ werden erzeugt, indem eine Störfunktion zu $f_{X,1}$ addiert wird, in Abhängigkeit von n. Um die Vorgehensweise möglichst einfach zu gestalten, wird angenommen, dass $f_{X,1}$ sich in einer kompakten Umgebung der Modalstelle wie eine quadratische Parabel verhält. Die Störfunktion ist so gewählt, dass sich, abhängig von n, die Modalwerte der Parabel um eine Distanz verschieben, welche sich asymptotisch verhält wie $n^{-(p-1)/(2p+2\beta+1)}$. Dies entspricht der Ordnung der oberen Schranke für die Konvergenzrate des Modalwertschätzers.
Dann, der Vorgehensweise wie im Beweis von Lemma 4.1 in Neumann [54] folgend, wird gezeigt, dass der quadrierte Hellinger-Abstand der beiden Dichten von der Ordnung $O(n^{-1})$ ist. Dies bedeutet, dass die beiden Situationen asymptotisch

nicht unterscheidbar sind, was eine untere Schranke der Ordnung $n^{-(p-1)/(2p+2\beta+1)}$ für das quadratische Risiko ergibt.

Beweis von Theorem 1.1.3

Wie bereits geschildert, wird zum Beweis der unteren Schranken dieselbe Herangehensweise wie im Beweis von Lemma 4.1 in Neumann [54] gewählt. Hierzu wird eine Teilmenge von Dichten in $\widetilde{\mathcal{F}}_p$ betrachtet. Diese besteht aus einer Familie $f_{X,1,n}$ von Dichten sowie einer weiteren Familie $f_{X,2,n}$. Die jeweiligen Modalwerte $M(f_{X,1,n})$ und $M(f_{X,2,n})$ werden mit $\theta_{1,n}$ und $\theta_{2,n}$ bezeichnet.
Die beiden Familien werden so gewählt, dass zum einen der quadrierte Hellinger-Abstand klein ist. In unserem Fall soll gelten:

$$H^2(f_{X,1,n}, f_{X,2,n}) = O(n^{-1}), \tag{1.10}$$

wobei

$$H^2(f_{X,1,n}, f_{X,2,n}) = \frac{1}{2} \int\limits_{\mathbb{R}} \left(\sqrt{f_{X,1,n}(x)} - \sqrt{f_{X,2,n}(x)}\right)^2 \lambda(dx).$$

Zum anderen sollen die Familien so konstruiert sein, dass für die Modalwerte gilt:

$$|\theta_{1,n} - \theta_{2,n}| \geq \Delta_n, \tag{1.11}$$

wobei gelten soll, dass Δ_n ein von n abhängiger Ausdruck ist, der sich asymptotisch wie die in Theorem 1.1.1 erzielte Konvergenzordnung verhält, also $\Delta_n \asymp n^{-(p-1)/(2p+2\beta+1)}$.
In der durch $f_{X,1,n}$ generierten Situation besitzt der Vektor der Beobachtungen $(Y_1, ..., Y_n)$ eine Verteilung mit der n-dimensionalen Dichte $(f_{X,1,n} * f_\varepsilon)^{[n]}$. Aufgrund der Unabhängigkeit der Beobachtung handelt es sich hierbei um die Produktdichte aus $f_{X,1,n} * f_\varepsilon$:

$$(f_{X,1,n} * f_\varepsilon)^{[n]}(y_1, ..., y_n) = \prod_{i=1}^{n} (f_{X,1,n} * f_\varepsilon)(y_i).$$

In der anderen Situation ist die Dichte $(f_{X,2,n} * f_\varepsilon)^{[n]}$, wieder mit

$$(f_{X,2,n} * f_\varepsilon)^{[n]}(y_1, ..., y_n) = \prod_{i=1}^{n} (f_{X,2,n} * f_\varepsilon)(y_i).$$

Unter Berücksichtigung der Tatsache, dass die Beobachtungen $Y_1, ..., Y_n$ unabhängig sind ergibt sich somit für die Hellinger-Affinität $\rho(f_1, f_2) = \int \sqrt{f_1}\sqrt{f_2}$ der beiden n-dimensionalen Dichten, unter Voraussetzung von (1.10):

$$\begin{aligned}
& \rho\Big((f_{X,1,n} * f_\varepsilon)^{[n]}, (f_{X,2,n} * f_\varepsilon)^{[n]}\Big) \\
= \; & \rho^n(f_{X,1,n} * f_\varepsilon, f_{X,2,n} * f_\varepsilon) \\
= \; & \Big(1 - \frac{1}{2} H^2(f_{X,1,n} * f_\varepsilon, f_{X,2,n} * f_\varepsilon)\Big)^n \\
\geq \; & (1 - C'/n)^n \geq C > 0 \\
& (\text{wegen } (1 - C'/n)^n \overset{n\to\infty}{\to} e^{-C'}).
\end{aligned}$$

Die Hellinger-Affinität ist also von Null weg beschränkt. Betrachten wir nun einen beliebigen Modalwertschätzer $\hat{M}(Y_1, ..., Y_n)$, so gilt unter Verwendung der Definition der Hellinger-Affinität und wegen $|\theta_{1,n} - \theta_{2,n}|/\Delta_n \geq 1$, dass

$$\begin{aligned}
& \rho\Big((f_{X,1,n} * f_\varepsilon)^{[n]}, (f_{X,2,n} * f_\varepsilon)^{[n]}\Big) \\
\leq \; & \int \frac{|\theta_{1,n} - \theta_{2,n}|}{\Delta_n} \sqrt{(f_{X,1,n} * f_\varepsilon)^{[n]}(\underline{y})} \sqrt{(f_{X,2,n} * f_\varepsilon)^{[n]}(\underline{y})} \; \lambda^{[n]}(d\underline{y}) \\
\leq \; & \int \frac{|\hat{M}(\underline{y}) - \theta_{1,n}|}{\Delta_n} \sqrt{(f_{X,1,n} * f_\varepsilon)^{[n]}(\underline{y})} \sqrt{(f_{X,2,n} * f_\varepsilon)^{[n]}(\underline{y})} \; \lambda^{[n]}(d\underline{y}) \\
& + \int \frac{|\hat{M}(\underline{y}) - \theta_{2,n}|}{\Delta_n} \sqrt{(f_{X,1,n} * f_\varepsilon)^{[n]}(\underline{y})} \sqrt{(f_{X,2,n} * f_\varepsilon)^{[n]}(\underline{y})} \; \lambda^{[n]}(d\underline{y}) \\
= \; & \int \frac{|\hat{M}(\underline{y}) - \theta_{1,n}|}{\Delta_n} \sqrt{f^{[n]}_{Y,1,n}(\underline{y}) \wedge f^{[n]}_{Y,2,n}(\underline{y})} \sqrt{f^{[n]}_{Y,1,n}(\underline{y}) \vee f^{[n]}_{Y,2,n}(\underline{y})} \; \lambda^{[n]}(d\underline{y}) \\
& + \int \frac{|\hat{M}(\underline{y}) - \theta_{2,n}|}{\Delta_n} \sqrt{f^{[n]}_{Y,1,n}(\underline{y}) \wedge f^{[n]}_{Y,2,n}(\underline{y})} \sqrt{f^{[n]}_{Y,1,n}(\underline{y}) \vee f^{[n]}_{Y,2,n}(\underline{y})} \; \lambda^{[n]}(d\underline{y}).
\end{aligned}$$

Anwendung der Cauchy-Schwarz-Ungleichung und Abschätzung des Maximums durch die Summe ergibt weiterhin

$$\begin{aligned}
\leq \; & \sqrt{\int \Big(\frac{|\hat{M}(\underline{y}) - \theta_{1,n}|}{\Delta_n}\Big)^2 \; \big(f^{[n]}_{Y,1,n}(\underline{y}) \wedge f^{[n]}_{Y,2,n}(\underline{y})\big) \; \lambda^{[n]}(d\underline{y})} \\
& \cdot \sqrt{\int \big(f^{[n]}_{Y,1,n}(\underline{y}) + f^{[n]}_{Y,2,n}(\underline{y})\big) \; \lambda^{[n]}(d\underline{y})}
\end{aligned}$$

$$
\begin{aligned}
&+ \sqrt{\int \Big(\frac{|\hat{M}(\underline{y}) - \theta_{2,n}|}{\Delta_n}\Big)^2 \, \big(f^{[n]}_{Y,1,n}(\underline{y}) \wedge f^{[n]}_{Y,2,n}(\underline{y})\big) \, \lambda^{[n]}(d\underline{y})} \\
&\qquad \cdot \sqrt{\int \big(f^{[n]}_{Y,1,n}(\underline{y}) + f^{[n]}_{Y,2,n}(\underline{y})\big) \, \lambda^{[n]}(d\underline{y})} \\
\leq \;& 2 \sqrt{\mathbb{E}_{f_{X,1,n}} \Big(\frac{\hat{M}(\underline{y}) - \theta_{1,n}}{\Delta_n}\Big)^2} + 2 \sqrt{\mathbb{E}_{f_{X,2,n}} \Big(\frac{\hat{M}(\underline{y}) - \theta_{2,n}}{\Delta_n}\Big)^2} \\
\leq \;& 4 \max_{f_X \in \{f_{X,1,n}, f_{X,2,n}\}} \sqrt{\mathbb{E}_{f_X} \Big(\frac{\hat{M}(\underline{y}) - \theta_X}{\Delta_n}\Big)^2} \, .
\end{aligned}
$$

Da der letzte Ausdruck eine obere Schranke für die Hellinger-Affinität der Dichten der beiden konkurrierenden Situationen darstellt, die Hellinger-Affinität aber nach unten durch eine positive Konstante beschränkt ist, ergibt sich weiter

$$
\inf_{\hat{M}} \max_{f_X \in \widetilde{\mathcal{F}}_p} \mathbb{E}_{f_X} (\hat{M}(Y_1, ..., Y_n) - \theta_{X,n})^2 \geq \frac{1}{4} \, C \cdot \Delta_n^2,
$$

da der Schätzer für den Modalwert beliebig gewählt war.

Nun gilt es also, die beiden Familien konkurrierender Dichten $f_{X,1,n} \in \widetilde{\mathcal{F}}_p$ und $f_{X,2,n} \in \widetilde{\mathcal{F}}_p$ so zu konstruieren, dass die Bedingungen (1.10) und (1.11) erfüllt sind.
Wie bereits angedeutet, ist hierbei $f_{X,1,n} \equiv f_{X,1}$ eine fixierte Dichte. Die Funktionen $f_{X,2,n}$ werden durch Addition einer von n abhängigen Störfunktion gewonnen. Für die konkurrierenden Dichten gelten im Detail folgende Annahmen:

(i) $\quad f_{X,1}(x) > 0 \; \forall x$, mit $M(f_{X,1}) =: \theta_1 = 0$ und
$\quad f_{X,1}(x) = f(0) - x^2$ für $|x| \leq \tau_3$.

(ii) $\quad f_{X,2,n}(x) = f_{X,1}(x) + g_n(x)$, $M(f_{X,2,n}) =: \theta_{2,n}$,
$\quad$ wobei $g_n(x) = b_n^p \, g(\dfrac{x - \theta_1}{b_n})$,
$\quad$ und $g(\cdot)$ beschränkt mit $g(x) = -(2x + b_n^{p-2})$ für $|x| \leq \tau_4$, $\tau_4 \leq \tau_3$.
$\quad g^{(p)}(\cdot)$ existiert und ist stetig, $g(x) \equiv 0$ for $|x| \geq \tau_5$, $\tau_4 < \tau_5$.
$\quad \displaystyle\int_{\mathrm{supp}\,(g_n)} g_n(x) x^j dx = 0$ für $j = 0, ..., \alpha - 1$.

(Die Abhängigkeit von n der Funktion g wird in der verwendeten Schreibweise nicht explizit hervorgehoben.)

Die Annahmen für $f_{X,1}$ und $f_{X,2,n}$ stellen sicher, dass sowohl $f_{X,1}$ als auch $f_{X,2,n}$ sich in einer Umgebung der jeweiligen Modalstelle wie quadratische Parabeln verhalten. Es gilt in einer Umgebung von $0 = \theta_1$:

$$f_{X,2,n}(x) = f_{X,1}(0) - (x + b_n^{p-1})^2 \,.$$

Da $g(\cdot)$ beschränkt ist, gilt für genügend großes n auch $f_{X,2,n} \in \widetilde{\mathcal{F}}_p$ und es gilt für den Modalwert $\theta_{n,2} = -b_n^{p-1}$, woraus Eigenschaft (1.11) folgt.

Nun ist also nachzuweisen, dass die konstruierten Familien von Dichten die Anforderung (1.10) erfüllt. Es gilt zum einen

$$\begin{aligned} & H^2(f_{X,1} * f_\varepsilon, f_{X,2,n} * f_\varepsilon) \\ = & \int \left(\sqrt{(f_{X,1} * f_\varepsilon)(x)} - \sqrt{((f_{X,1} + g_n) * f_\varepsilon)(x)}\right)^2 dx \\ \leq & \int_{c-\eta}^{d+\eta} \frac{((g_n * f_\varepsilon)(x))^2}{(f_{X,1} * f_\varepsilon)(x)} dx + \int_{\mathbb{R}\setminus[c-\eta,d+\eta]} |(g_n * f_\varepsilon)(x)| dx =: I_1 + I_2 \,. \end{aligned}$$

Es gilt $f_{X,1} > 0$ für alle x, daher ist $f_{X,1} * f_\varepsilon$ von Null weg beschränkt auf dem kompakten Intervall $[c - \eta, d + \eta]$. Im Integranden von I_1 kann also der Nenner nach unten durch eine positive Konstante abgeschätzt werden und man erhält zunächst

$$I_1 \;\leq\; C \int_{\mathbb{R}} ((g_n * f_\varepsilon)(x))^2 \, dx$$

Die Gleichung von Plancherel liefert weiterhin

$$\begin{aligned} = & \; C \int_{\mathbb{R}} |\phi_{g_n}(\omega)\phi_\varepsilon(\omega)|^2 \, d\omega \\ = & \; C \cdot b_n^{2p+1} \int_{\mathbb{R}} |e^{i\frac{u}{b_n}\theta_1} \cdot \phi_g(u) \cdot \phi_\varepsilon(\frac{u}{b_n})|^2 \, du \end{aligned}$$

und aufgrund der gewöhnlichen Glattheit der Fehlerverteilung

$$\leq \qquad \tilde{C} \cdot b_n^{2p+2\beta+1} \int_{\mathbb{R}} (|\phi_g(u)||u|^{-\beta})^2 \, du$$

$$
\begin{aligned}
&\leq \quad \tilde{C} \cdot b_n^{2p+2\beta+1} \Big(\int\limits_{|u|<1} (|\phi_g(u) u^{(-\lceil\beta\rceil)}|^2 \, du + \int\limits_{|u|\geq 1} |\phi_g(u) u^{-\lfloor\beta\rfloor}|^2 \, du \Big) \\
&\leq \quad \tilde{C} \cdot b_n^{2p+2\beta+1} \Big(\int\limits_{\mathbb{R}} |(\phi_{g^{(-\lceil\beta\rceil)}}(u))|^2 \, du + \int\limits_{\mathbb{R}} |\phi_{g^{(-\lfloor\beta\rfloor)}}(u)|^2 \, du \Big) \\
&\overset{Plancherel}{=} \quad \tilde{C} \cdot b_n^{2p+2\beta+1} \Big(\int\limits_{\text{supp (g)}} |g^{(-\lceil\beta\rceil)}(u)|^2 \, du + \int\limits_{\text{supp (g)}} |g^{(-\lfloor\beta\rfloor)}(u)|^2 \, du \Big) \\
&= \quad O(b_n^{2p+2\beta+1}) = O(n^{-1}) \text{ für } b_n = O(n^{-\frac{1}{2p+2\beta+1}}) \, .
\end{aligned}
$$

Die beiden zuletzt auftretenden Integrale sind beschränkt aufgrund der Beschränktheit der Funktion g.

Hinsichtlich des Integranden von I_2 gilt, aufgrund der Differenzierbarkeit der Fehlerdichte f_ε, unter Anwendung von Voraussetzung (ii),

$$
\begin{aligned}
& |(g_n * f_\varepsilon)(x)| \\
= \; & \Big| \int\limits_{\text{supp (g}_\text{n})} g_n(y) f_\varepsilon(x-y) \, dy \Big| \\
= \; & | \int\limits_{\text{supp (g}_\text{n})} g_n(y) \Big(\sum_{j=0}^{\alpha-1} \frac{(-y)^j}{j!} f^{(j)}(x) + \int\limits_x^{x-y} \frac{f_\varepsilon^{(\alpha)}(u)}{(\alpha-1)!} (x-u)^{(\alpha-1)} du \Big) \, dy| \\
\leq \; & C \cdot b_n^{\alpha+p} \int\limits_{-\tau_5}^{\tau_5} \big|g(u)\big| du \int\limits_{\text{supp (g}_\text{n})} |f_\varepsilon^{(\alpha)}(x-z)| \, dz \, .
\end{aligned}
$$

Also gilt für I_2 folgende Abschätzung

$$
\begin{aligned}
I_2 \; \leq \; & C(\tau_5 b_n)^{\alpha+p} \int\limits_{\mathbb{R}\setminus[c-\eta,d+\eta]} \int\limits_{\text{supp (g}_\text{n})} |f_\varepsilon^{(\alpha)}(x-z)| dz \, dx \\
= \; & \tilde{C} b_n^{\alpha+p} \Big(\int\limits_{\text{supp (g}_\text{n})} \int\limits_{-\infty}^{c-\eta-z} |f_\varepsilon^{(\alpha)}(u)| du + \int\limits_{d+\eta-z}^{\infty} |f_\varepsilon^{(\alpha)}(u)| du \, dz \Big) \\
\leq \; & \tilde{C} b_n^{\alpha+p} \Big(\int\limits_{\text{supp (g}_\text{n})} \int\limits_{-\infty}^{c-\eta+\tau_5 b_n} |f_\varepsilon^{(\alpha)}(u)| du + \int\limits_{d+\eta-\tau_5 b_n}^{\infty} |f_\varepsilon^{(\alpha)}(u)| du \, dz \Big) \\
\leq \; & O(b_n^{\alpha+p+1}) \, ,
\end{aligned}
$$

falls c und d so gewählt werden, dass die geforderte Bedingung an f_ε erfüllt ist. Die Ordnung dieses Terms ist nun kleiner oder gleich derjenigen von I_1, falls

$$2\beta + 2p + 1 \leq \alpha + p + 1 \Leftrightarrow 2\beta + p \leq \alpha.$$

Somit gilt für $b_n \asymp n^{-\frac{1}{n^{2p+2\beta+1}}}$, dass

$$\inf_{\hat{M}} \max_{f_X \in \{f_{X,1}, f_{X,2,n}\}} \sqrt{\mathbb{E}_{f_X}(\hat{M}(Y_1, ..., Y_n) - \theta_X)^2} \geq \frac{1}{4} C\, b_n^{2(p-1)} \asymp n^{-\frac{2(p-1)}{2p+2\beta+1}}. \qquad \square$$

Untere Schranken im Regressionsmodell mit Fehlern in den Variablen

Analog zu den unteren Schranken, die im vorhergehenden Abschnitt erzielt wurden, können entsprechende Raten auch im Regressionsmodell mit Fehlern in den Variablen erzielt werden. Die Voraussetzungen an die Fehlerdichte, welche bereits für den Nachweis unterer Schranken im Fall der Dichteschätzung getätigt wurden, werden auch hier benötigt. Insgesamt gilt folgendes Theorem (siehe Wieczorek [78])

Theorem 1.1.4
Es gelten die Voraussetzungen von Theorem 1.1.2 und $\int_{\mathbb{R}\setminus[c,d]} |f_\varepsilon^{(\alpha)}(x)|dx \leq C$, $\alpha \geq 2\beta + p$, *für ein kompaktes Intervall* $[c, d]$ *und* $C \in \mathbb{R}$. *Dann gilt für einen beliebigen Modalwertschätzer* $\hat{M}_n$ *für* $m(x) = \mathbb{E}(Z|X = x)$, *dass*

$$\sup_{m \in \mathcal{F}_p} \mathbb{E}_m(\hat{M}_n - \theta)^2 \geq d \cdot n^{-\frac{2(p-1)}{2p+2\beta+1}} \qquad \textit{für ein } d > 0.$$

Ähnlich wie im Beweis des vorherigen Theorems werden zwei Familien zweidimensionaler Dichten erzeugt. Hierbei besteht eine aus einer fixierten Dichte, $f_{X,Z,1}(x, z)$ und einer weiteren, $f_{X,Z,2,n}$, für $n \in \mathbb{N}$. Letztere wird erzeugt, indem eine Störfunktion zu $f_{X,Z,1}$ addiert wird, so dass einerseits die Randdichten bezüglich X identisch sind, andererseits die bedingten Erwartungswerte von Z unter der Bedingung $X = x$ verschiedene Modalwerte besitzen, aber immer noch ähnlich genug sind, dass der Hellinger-Abstand der beiden Dichten von der Ordnung $O(n^{-1})$ ist.
Wiederum setzen wir voraus, dass der bedingte Erwartungswert sich in einer Umgebung des Modalwertes wie eine Parabel verhält.

Beweis von Theorem 1.1.4
Im Modell mit Fehlern in den Variablen besitzt die Verteilung von (X, Z) unter

den Voraussetzungen von Theorem 1.1.2 eine zweidimensionale Dichten. Um konkurrierende Situationen zu erzeugen, werden wiederum zwei Familien von zweidimensionalen betrachtet, bezeichnet mit $f_{X,Z,1}(x,z)$ und $f_{X,Z,2,n}(x,z)$. Die zugehörigen Regressionsfunktionen werden mit m_1 und $m_{2,n}$ bezeichnet.
Es werden folgende Voraussetzungen an $f_{X,Z,1}(x,z)$ und $f_{X,Z,2,n}(x,z)$ gestellt

(i') $f_{X,Z,1}(x,z)$ ist stetig und $f_{X,Z,1}(x,z) > 0 \quad \forall x, z$;

$$\int z f_{X,Z,1}(x,z)dz =: r_1(x) \in \mathcal{F}_p,$$

$$\int z f_{X,Z,1}(x,z)dz \equiv 0,\ x \notin [-1,1],$$

$$M\big(m_1\big) =: \theta_1 = 0,$$

und $m_1(x) = m_1(0) - x^2$ für $x \in [-\tau_3, \tau_3]$, $1 > \tau_3 > 0$

$$\int f(x,z)dz = f_X(x) \equiv \text{const} =: c,\ x \in [-1,1]\,,$$

(ii') $f_{X,Z,2,n}(x,z) = f_{X,Z,1}(x,z) + g_n(x,z) = f_{X,Z,1}(x,z) + h(z) \cdot b_n^p g(x/b_n)$,

wobei

h beschränkt, $\operatorname{supp}(h) \subseteq [-1,1]$, $\int z h(z)dz = 1$, $\int h(z)dz = 0$.

g wird so gewählt wie im vorherigen Abschnitt.

Dann gilt mit

$$m_1(x) = \mathbb{E}_{f_{X,Z,1}}(Z|X=x) = \frac{\int z f_{X,Z,1}(x,z)dz}{\int f_{X,Z,1}(x,z)dz}\,,$$

dass

$$\begin{aligned} m_{2,n}(x) &= \mathbb{E}_{f_{X,Z,2,n}}(Z|X=x) = \frac{\int z f_{X,Z,2,n}(x,z)dz}{\int f_{X,Z,2,n}(x,z)dz} \\ &= \frac{\int z f_{X,Z,1}(x,z)dz}{f_X(x)} + \frac{b_n^p g(x/b_n)}{f_X(x)} \\ &= m_1(x) + \frac{1}{c} b_n^p g(x/b_n)\,. \end{aligned}$$

Es sei bemerkt, dass m_1 und $m_{2,n}$ in $\mathcal{F}_p$ liegen, denn die Randdichte f_X ist konstant in einer Umgebung des Modalwertes. Die Beobachtungen (Y_i, Z_i) besitzen

eine Verteilung mit Dichte

$$(f_{X,Z,1} * f_\varepsilon)(y,z) \;:=\; \int\limits_{\mathbb{R}} f_{X,Z,1}(y-\varepsilon,z) f_\varepsilon(\varepsilon)\, d\varepsilon$$

bzw.

$$(f_{X,Z,2,n} * f_\varepsilon)(y,z) \;:=\; \int\limits_{\mathbb{R}} f_{X,Z,2,n}(y-\varepsilon,z) f_\varepsilon(\varepsilon)\, d\varepsilon.$$

Da $f_{X,Z,1}(x,z)$ positiv auf $\mathbb{R}^2$ ist, ist $(f_{X,Z,1} * f_\varepsilon)$ auf dem kompakten Intervall $[c_1-\eta, d_1+\eta]$ nach unten beschränkt durch eine Konstante $1/C$, $C>0$, wobei $[c_2,d_2]$ so gewählt ist, dass $c_2 \le -1$, $d_2 \ge 1$.
Für den Hellinger-Abstand der beiden zweidimensionalen Dichten gilt

$$\begin{aligned}
& H^2(f_{X,Z,1} * f_\varepsilon, f_{X,Z,2,n} * f_\varepsilon) \\
\le\; & \int\limits_{c_1-\eta}^{d_1+\eta} \int\limits_{c_2}^{d_2} \frac{\big(h(z) g_n(x) * f_\varepsilon(x)\big)^2}{(f_{X,Z,1}(x,z) * f_\varepsilon(x))}\, dz dx \\
& + \int\limits_{\mathbb{R}^2 \setminus [c_1-\eta, d_1+\eta]\times[c_2,d_2]} h(z) g_n(x) * f_\varepsilon(x) d(x,z) \\
=:\; & I_1 + I_2\,.
\end{aligned}$$

Für I_1 erhält man folgende Abschätzung, wobei das Vorgehen analog zum Fall der Dichteschätzung ist:

$$\begin{aligned}
I_1 &\le C \int\limits_{c_2}^{d_2} \int\limits_{c_1-\eta}^{d_1+\eta} h(z) g_n(x) * f_\varepsilon(x) dx dz \\
&\le ||h||_\infty (d_2 - c_2) \cdot C \int\limits_{c_1-\eta}^{d_1+\eta} g_n(x) * f_\varepsilon(x) dx \\
&= O(b_n^{2p+2\beta+1}), \text{ für } b_n = O(n^{-\frac{1}{2p+2\beta+1}}).
\end{aligned}$$

Für I_2 gilt

$$I_2 \;=\; \int\limits_{\mathbb{R}^2 \setminus [c_1-\eta, d_1+\eta]\times[c_2,d_2]} h(z) g_n(x) * f_\varepsilon(x) d(x,z)$$

$$= \int\limits_{\mathbb{R}\setminus[c_1-\eta,d_1+\eta]} \int\limits_{\mathbb{R}} h(z) g_n(x) * f_\varepsilon(x)\, dz dx$$
$$+ \int\limits_{[c_1-\eta,d_1+\eta]} \int\limits_{\mathbb{R}\setminus[c_2,d_2]} h(z) g_n(x) * f_\varepsilon(x)\, dz dx\,.$$

Aufgrund der Wahl von c_2 und d_2 verschwindet das zweite Integral und mit

$$|h(z) g_n(x) * f_\varepsilon(x)| \;\leq\; |h(z)|\, b_n^{a+p} \frac{1}{(\alpha-1)!} \int\limits_{\mathrm{supp}\,(\mathrm{g_n})} |f_\varepsilon^{(\alpha)}(x-u)|\, du$$

folgt das analoge Resultat wie im Fall der Dichteschätzung.

1.1.4 Hilfsresultate

Um die in den Beweisen der Theoreme 1.1.1 und 1.1.2 auftretenden Wahrscheinlichkeiten nach oben beschränken zu können, spielen Momentenungleichungen für die Kurvenschätzer eine bedeutende Rolle, sowohl für den Dichteschätzer als auch für den Regressionskurvenschätzer.

Zunächst werden einige Resultate für Momente des Dichteschätzers im Dekonvolutionsmodell angegeben.

Lemma 1.1.1 *Falls die Kernfunktion der Bedingung* (1.3) *genügt, dann gilt für jedes kompakte Intervall I, für $r \geq 2$*

$$\begin{aligned}
(i) \quad & \mathbb{E}(\hat{f}_{X,n}(t)) = \frac{1}{b_n} \int K\big(\frac{t-u}{b_n}\big) f_X(u) du \\
(ii) \quad & \mathbb{E}\Big(\big|\hat{f}_{X,n}(t) - \mathbb{E}\hat{f}_{X,n}(t)\big|^r\Big) = O((n b_n^{2\beta+1})^{-r/2}) \\
(iii) \quad & \mathbb{E}\Big(\big|\hat{f}_{X,n}(t_1) - \mathbb{E}\hat{f}_{X,n}(t_1) - (\hat{f}_{X,n}(t_2) - \mathbb{E}\hat{f}_{X,n}(t_2))\big|^r\Big) \\
& \quad = O(\big|\frac{t_1-t_2}{b_n}\big|^r (n b_n^{2\beta+1})^{-r/2})
\end{aligned}$$

gleichmäßig für t, t_1, $t_2 \in I$.

Beweis:

zu (i):

$$\mathbb{E}(\hat{f}_{X,n}(t)) \;=\; \mathbb{E}\Big(\frac{1}{n b_n} \sum_{j=1}^{n} K_{b_n}^*\big(\frac{t-Y_j}{b_n}\big)\Big)$$

$$
\begin{aligned}
&= \mathbb{E}\Big(\frac{1}{nb_n}\sum_{j=1}^{n}\frac{1}{2\pi}\int\limits_{\mathbb{R}} e^{-i(\frac{t-Y_j}{b_n})u}\,\frac{\phi_K(u)}{\phi_\varepsilon(\frac{u}{b_n})}\,du\Big)\\
&= \mathbb{E}\Big(\frac{1}{2\pi}\cdot\frac{1}{b_n}\int\limits_{\mathbb{R}} e^{-i(\frac{t-Y}{b_n})u}\,\frac{\phi_K(u)}{\phi_\varepsilon(\frac{u}{b_n})}\,du\Big)\\
&= \mathbb{E}\Big(\frac{1}{2\pi}\int\limits_{\mathbb{R}} e^{-i(t-Y)u}\,\frac{\phi_K(ub_n)}{\phi_\varepsilon(u)}\,du\Big)\\
&= \mathbb{E}\Big(\frac{1}{2\pi}\int\limits_{\mathbb{R}} e^{-i(t-(X+\varepsilon))u}\,\frac{\phi_K(ub_n)}{\phi_\varepsilon(u)}\,du\Big).
\end{aligned}
$$

Aufgrund der Unabhängigkeit von X und ε und unter Ausnutzung der Definition der charakteristischen Funktion entspricht dies

$$
\begin{aligned}
&\frac{1}{2\pi}\int\limits_{\mathbb{R}} e^{-itu}\cdot\phi_X(u)\cdot\phi_\varepsilon(u)\cdot\frac{\phi_K(ub_n)}{\phi_\varepsilon(u)}\,du\\
=\;&\frac{1}{2\pi}\int\limits_{\mathbb{R}} e^{-itu}\cdot\phi_X(u)\cdot\phi_K(ub_n)\,du\\
=\;&\frac{1}{2\pi}\int\limits_{\mathbb{R}} e^{-itu}\int\limits_{\mathbb{R}} e^{iuy}\,f_X(y)\,dy\int\limits_{\mathbb{R}} e^{iux}\frac{1}{b_n}\,K(\frac{x}{b_n})\,dx\,du\\
=\;&\frac{1}{2\pi}\int\limits_{\mathbb{R}} e^{-itu}\int\limits_{\mathbb{R}}\int\limits_{\mathbb{R}} e^{iuy}f_X(y)\,e^{iu(x-y)}\,\frac{1}{b_n}\,K(\frac{x-y}{b_n})\,dy\,dx\,du\\
=\;&\frac{1}{2\pi}\int\limits_{\mathbb{R}} e^{-itu}\int\limits_{\mathbb{R}} e^{iux}\int\limits_{\mathbb{R}} f_X(y)\,\frac{1}{b_n}\,K(\frac{x-y}{b_n})\,dy\,dx\,du\\
=\;&\frac{1}{b_n}\int\limits_{\mathbb{R}} K(\frac{t-u}{b_n})f_X(u)\,du.
\end{aligned}
$$

<u>zu (iii)</u>:
Zunächst kann geschrieben werden

$$
\begin{aligned}
&\mathbb{E}\Big(\big|\hat{f}_{X,n}(t_1)-\mathbb{E}\hat{f}_{X,n}(t_1)-(\hat{f}_{X,n}(t_2)-\mathbb{E}\hat{f}_{X,n}(t_2))\big|^r\Big)\\
=\;&\mathbb{E}\Big(\big|\sum_{j=1}^{n}\frac{1}{nb_n}\big(K^*_{b_n}(\frac{t_1-Y_j}{b_n})-K^*_{b_n}(\frac{t_2-Y_j}{b_n})\big)\\
&\quad-\mathbb{E}\Big(\frac{1}{nb_n}\big(K^*_{b_n}(\frac{t_1-Y_j}{b_n})-K^*_{b_n}(\frac{t_2-Y_j}{b_n})\big)\Big)\big|^r\Big)\\
=:\;&\mathbb{E}\Big(\big|\sum_{j=1}^{n}\big(Z_{j,n,\Delta t}-\mathbb{E}Z_{j,n,\Delta t}\big)\big|^r\Big).
\end{aligned}
$$

Hieraus ist ersichtlich, dass die Momente einer Summe unabhängiger zentrierter Zufallsgrößen vorliegen. Damit kann die Rosenthal-Ungleichung (siehe Braverman [12], S. 21) angewendet werden und wir erhalten

$$\leq \quad C\Big\{\sum_{j=1}^{n}\mathbb{E}\big(\big|Z_{j,n,\Delta t}-\mathbb{E}Z_{j,n,\Delta t}\big|^r\big)+\Big(\sum_{j=1}^{n}\mathbb{E}\big(|Z_{j,n,\Delta t}-\mathbb{E}Z_{j,n,\Delta t}|\big)^2\Big)^{r/2}\Big\}$$

und unter Verwendung der C_r-Ungleichung (siehe Shorack und Wellner [69], S. 843)

$$\leq \quad C\Big\{C_r\sum_{j=1}^{n}\mathbb{E}\big(\big|Z_{j,n,\Delta t}\big|^r\big)+\Big(C_2\sum_{j=1}^{n}\mathbb{E}\big(|Z_{j,n,\Delta t}|\big)^2\Big)^{r/2}\Big\}. \tag{1.12}$$

Nun gilt es, obere Schranken für $\mathbb{E}(|Z_{j,n,\Delta t}|^r)$ zu berechnen:
Hierfür sehen wir zunächst, dass $\frac{1}{r}+\frac{r-1}{r}=1$ und $1<\frac{r}{r-1}\leq 2$. In diesem Fall, ergibt das Riesz-Thorin-Theorem (siehe Werner [77]) folgendes, wobei mit $||\cdot||_r$ die L^r-Norm bezeichnet wird

$$\big(||\phi_f||_r^r\big)^{1/r}\leq\big(||f||_{\frac{r}{r-1}}^{\frac{r}{r-1}}\big)^{\frac{r-1}{r}} \quad\Leftrightarrow\quad ||\phi_f||_r^r\leq(||f||_{\frac{r}{r-1}}^{\frac{r}{r-1}})^{r-1} \tag{1.13}$$

Also gilt

$$\begin{aligned}
&\mathbb{E}\big(\big|Z_{j,n,\Delta t}\big|^r\big)\\
=\;&\mathbb{E}\Big(\Big|\frac{1}{nb_n}\cdot\frac{1}{2\pi}\int_{\mathbb{R}}\big[e^{-i\frac{t_1-Y_j}{b_n}u}-e^{-i\frac{t_2-Y_j}{b_n}u}\big]\frac{\phi_K(u)}{\phi_\varepsilon(u/b_n)}\,du\Big|^r\Big)\\
=\;&\frac{1}{n^r}\frac{1}{(2\pi)^r}\mathbb{E}\Big|\int_{\mathbb{R}}e^{iY_ju}\big[e^{-it_1u}-e^{-it_2u}\big]\frac{\phi_K(b_nu)}{\phi_\varepsilon(u)}du\Big|^r\\
=\;&\frac{1}{n^r}\frac{1}{(2\pi)^r}\mathbb{E}\int_{\mathbb{R}}\Big|\int_{\mathbb{R}}e^{ixu}e^{i\varepsilon_ju}\big[e^{-it_1u}-e^{-it_2u}\big]\frac{\phi_K(b_nu)}{\phi_\varepsilon(u)}du\Big|^r f_X(x)dx\\
\leq\;&\frac{||f_X||_\infty}{n^r}\frac{1}{(2\pi)^r}\mathbb{E}\int_{\mathbb{R}}\Big|\int_{\mathbb{R}}e^{ixu}e^{i\varepsilon_ju}\big[e^{-it_1u}-e^{-it_2u}\big]\frac{\phi_K(b_nu)}{\phi_\varepsilon(u)}du\Big|^r dx
\end{aligned}$$

und unter Anwendung von (1.13)

$$\leq\quad\frac{||f_X||_\infty}{n^r}\frac{1}{(2\pi)^r}\mathbb{E}\Big(\int_{\mathbb{R}}\Big(\big|e^{i\varepsilon_ju}\big|\underbrace{\big|e^{-it_1u}-e^{-it_2u}\big|}_{\leq|t_1-t_2|\cdot|u|}\big|\frac{\phi_K(b_nu)}{\phi_\varepsilon(u)}\big|\Big)^{\frac{r}{r-1}}du\Big)^{r-1}.$$

Im Fall von Aussage (ii) des Lemmas wird der Ausdruck $\left|e^{-it_1u} - e^{-it_2u}\right|$ ersetzt durch $\left|e^{-itu}\right|$ und kann daher durch 1 nach oben beschränkt werden. Weiter kann letzterer Ausdruck folgendermaßen abgeschätzt werden:

$$\begin{aligned}
&\leq \frac{||f_X||_\infty}{n^r}\frac{1}{(2\pi)^r}\mathbb{E}\Big(\int\limits_{\mathbb{R}}\Big(\left|e^{i\varepsilon_j u}\right|\cdot|t_1-t_2|\cdot|u|\cdot\Big|\frac{\phi_K(b_nu)}{\phi_\varepsilon(u)}\Big|\Big)^{\frac{r}{r-1}}du\Big)^{r-1}\\
&\leq \frac{1}{n}\frac{1}{(nb_n)^{r-1}}\frac{1}{(2\pi)^r}\Big(\Big|\frac{t_1-t_2}{b_n}\Big|^{\frac{r}{r-1}}O\Big(\frac{1}{b_n^\beta}\Big)^{\frac{r}{r-1}}\int\limits_{\mathbb{R}}(|\phi_K(u)||u|^{1+\beta})^{\frac{r}{r-1}}du\Big)^{r-1}\\
&= \frac{1}{n}O\Big((nb_n^{\frac{r}{r-1}\beta+1})^{-(r-1)}\Big)\Big|\frac{t_1-t_2}{b_n}\Big|^r,
\end{aligned}$$

was gleichmäßig für alle t_1, t_2 in I gilt.
Somit ist (1.12) von der Ordnung

$$O\Big((nb_n^{\frac{r}{r-1}\beta+1})^{-(r-1)}\Big|\frac{t_1-t_2}{b_n}\Big|^r\Big) + O\Big(\big((nb_n^{2\beta+1})\big)^{-r/2}\Big|\frac{t_1-t_2}{b_n}\Big|^r\Big).$$

Da $\frac{r}{r-1} \leq 2$ falls $r \geq 1$, gilt weiterhin $nb_n^{\frac{r}{r-1}\beta+1} \geq nb_n^{2\beta+1}$ und mit $r-1 \geq r/2$ folgt für $r \geq 2$

$$(nb_n^{\frac{r}{r-1}\beta+1})^{-(r-1)} \leq (nb_n^{2\beta+1})\big)^{-r/2}.$$

Also folgt (iii). Hierbei gilt, dass aufgrund von (1.3) alle Momente für $r \geq 2$ existieren. □

Auch im Fall der Schätzung des Modalwertes im Errors-in-Variables-Regressionsmodell werden Resultate hinsichtlich der Momente von Nenner und Zähler des Kurvenschätzers benötigt.
Bezüglich der Momente des Zählers liefert das nächste Lemma die benötigten Aussagen:

Lemma 1.1.2 *Falls* (1.3) *erfüllt ist und das bedingte Moment der Ordnung* r, $\mathbb{E}(|Z|^r|X=x)$, *existiert und gleichmäßig beschränkt ist, dann gilt für jedes kompakte Intervall* I, *für* $r \geq 2$

$$(i) \qquad \mathbb{E}(\hat{r}_n(t)) = \frac{1}{b_n} \int_{\mathbb{R}} K\big(\frac{t-x}{b_n}\big) f_X(x) m(x) dx$$

$$(ii) \qquad \mathbb{E}\Big(\big|\hat{r}_n(t_1) - \mathbb{E}\hat{r}_n(t_1)\big|^r\Big) = O((nb_n^{2\beta+1})^{-r/2})$$

$$(iii) \qquad \mathbb{E}\Big(\big|\hat{r}_n(t_1) - \mathbb{E}\hat{r}_n(t_1) - (\hat{r}_n(t_2) - \mathbb{E}\hat{r}_n(t_2))\big|^r\Big)$$
$$= O(\big|\frac{t_1-t_2}{b_n}\big|^r (nb_n^{2\beta+1})^{-r/2})$$

gleichmäßig für t, t_1, $t_2 \in I$.

Beweis:

zu (i)

Diese wohlbekannte Aussage kann ebenso in der einschlägigen Literatur, etwa bei Fan und Truong [30] nachgelesen werden.

Zunächst gilt für die gemeinsame Dichte $f_{Y,Z}(y,z)$ des Zufallsvektors (Y,Z) folgendes, wobei F_ε die Verteilungsfunktion der Fehlervariable ε bezeichnet:

$$f_{Y,Z}(y,z) \quad = \quad \int_{\mathbb{R}} f_{X,Z}(y-u,z)\, dF_\varepsilon(u)$$

Damit gilt für den Erwartungswert des Zählers:

$$\begin{aligned}
& \mathbb{E}(\hat{r}_n(t)) \\
= \; & \mathbb{E}\Big(\frac{1}{nb_n}\sum_{j=1}^n K^*_{b_n}\big(\frac{t-Y_j}{b_n}\big) Z_j\Big) \\
= \; & \frac{1}{b_n} \cdot \mathbb{E}\Big(K^*_{b_n}\big(\frac{t-Y}{b_n}\big) Z\Big) \\
= \; & \frac{1}{2\pi} \cdot \mathbb{E}\Big(Z \cdot \int_{\mathbb{R}} e^{-i(t-Y)u}\, \frac{\phi_K(ub_n)}{\phi_\varepsilon(u)}\, du\Big) \\
= \; & \frac{1}{2\pi} \cdot \int_{\mathbb{R}} \int_{\mathbb{R}} z \Big(\int_{\mathbb{R}} e^{-i(t-y)u}\, \frac{\phi_K(ub_n)}{\phi_\varepsilon(u)}\, du\Big)\Big(\int_{\mathbb{R}} f_{X,Z}(y-\varepsilon,z)\, dF_\varepsilon(\varepsilon)\Big)\, dydz \\
= \; & \frac{1}{2\pi} \cdot \int_{\mathbb{R}} z \Big(\int_{\mathbb{R}} e^{-i(t-y)u}\, \frac{\phi_K(ub_n)}{\phi_\varepsilon(u)} \int_{\mathbb{R}} e^{iuy} \int_{\mathbb{R}} f_{X,Z}(y-\varepsilon,z)\, dF_\varepsilon(\varepsilon)\, dydu\Big) dz
\end{aligned}$$

$$
\begin{aligned}
&= \frac{1}{2\pi} \cdot \int_{\mathbb{R}} z \Big(\int_{\mathbb{R}} e^{-itu} \frac{\phi_K(ub_n)}{\phi_\varepsilon(u)} \int_{\mathbb{R}} e^{iux} f_{X,Z}(x,z)\, dx \int_{\mathbb{R}} e^{iu\varepsilon}\, dF_\varepsilon(\varepsilon)\, du \Big) dz \\
&= \frac{1}{2\pi} \cdot \int_{\mathbb{R}} z \Big(\int_{\mathbb{R}} e^{-itu} \frac{\phi_K(ub_n)}{\phi_\varepsilon(u)} \phi_\varepsilon(u) \int_{\mathbb{R}} e^{iux} f_{X,Z}(x,z)\, dxdu \Big)\, dz \\
&= \frac{1}{2\pi} \cdot \int_{\mathbb{R}} z \Big(\int_{\mathbb{R}} e^{-itu} \phi_K(ub_n) \int_{\mathbb{R}} e^{iux} f_{X,Z}(x,z)\, dxdu \Big)\, dz \\
&= \frac{1}{2\pi} \cdot \int_{\mathbb{R}} z \Big(\int_{\mathbb{R}} \Big(\int_{\mathbb{R}} e^{-i(\frac{t-x}{b_n})tu} \phi_K(u)\, du \Big) f_{X,Z}(x,z)\, dx \Big)\, dz \\
&= \int_{\mathbb{R}} z \Big(\int_{\mathbb{R}} K(\frac{t-x}{b_n})\, f_{X,Z}(x,z)\, dx \Big)\, dz \\
&= \int_{\mathbb{R}} K(\frac{t-x}{b_n}) \int_{\mathbb{R}} z \cdot f_{X,Z}(x,z)\, dz\, dx \\
&= \int_{\mathbb{R}} K(\frac{t-x}{b_n}) \int_{\mathbb{R}} z \cdot f_{Z|X=x}(z)\, dz\, f_X(x)\, dx \\
&= \int_{\mathbb{R}} K(\frac{t-x}{b_n})\, f_X(x)\, m(x)\, dx.
\end{aligned}
$$

<u>zu (iii):</u>
Zunächst gilt:

$$
\begin{aligned}
& \mathbb{E}\Big(\big| \hat{r}_n(t_1) - \mathbb{E}\hat{r}_n(t_1) - (\hat{r}_n(t_2) - \mathbb{E}\hat{r}_n(t_2)) \big|^r \Big) \\
= \; & \mathbb{E}\Big(\big| \hat{r}_n(t_1) - \hat{r}_n(t_2) - (\mathbb{E}\hat{r}_n(t_1) - \mathbb{E}\hat{r}_n(t_2)) \big|^r \Big) \\
= \; & \mathbb{E}\Big(\Big| \sum_{j=1}^n Z_j \cdot \frac{1}{nb_n} \big(K^*_{b_n}(\frac{Y_j - t_1}{b_n}) - K^*_{b_n}(\frac{Y_j - t_2}{b_n}) \big) \\
& \qquad - \mathbb{E}\Big(Z_j \cdot \frac{1}{nb_n} \big(K^*_{b_n}(\frac{Y_j - t_1}{b_n}) - K^*_{b_n}(\frac{Y_j - t_2}{b_n}) \big) \Big) \Big|^r \Big) \\
=: \; & \mathbb{E}\Big(\Big| \sum_{j=1}^n \big(U_{j,n,\Delta t} - \mathbb{E}U_{j,n,\Delta t} \big) \Big|^r \Big).
\end{aligned}
$$

In Analogie zum vorherigen Lemma, Aussage (iii), erfolgt nun die Abschätzung des r-ten absoluten Momentes von $U_{j,n,\Delta t}$:

$$
\begin{aligned}
& \mathbb{E}(|U_{j,n,\Delta t}|^r) \\
= \; & \mathbb{E}\Big(\frac{1}{(nb_n)^r} \big| K^*_{b_n}(\frac{t_1 - Y_j}{b_n}) - K^*_{b_n}(\frac{t_2 - Y_j}{b_n}) \big|^r |Z_j|^r \Big)
\end{aligned}
$$

$$
\begin{aligned}
&= \frac{1}{(nb_n)^r} \int\limits_{\mathbb{R}} \int\limits_{\mathbb{R}} \big|K^*_{b_n}\big(\frac{t_1-y}{b_n}\big) - K^*_{b_n}\big(\frac{t_2-y}{b_n}\big)\big|^r |z|^r f_{Y,Z}(y,z) dy\, dz \\
&= \frac{1}{(nb_n)^r} \int\limits_{\mathbb{R}} \int\limits_{\mathbb{R}} \big|K^*_{b_n}\big(\frac{t_1-y}{b_n}\big) - K^*_{b_n}\big(\frac{t_2-y}{b_n}\big)\big|^r |z|^r \int\limits_{\mathbb{R}} f_{X,Z}(y-\varepsilon,z) dF_\varepsilon(\varepsilon)\, dy\, dz \\
&= \frac{1}{n^r}\frac{1}{(2\pi)^r} \int\limits_{\mathbb{R}} \int\limits_{\mathbb{R}} \Big| \int\limits_{\mathbb{R}} e^{ixu} e^{i\varepsilon u}[e^{-it_1u} - e^{-it_2u}] \frac{\phi_K(b_n u)}{\phi_\varepsilon(u)} du \Big|^r \cdot \\
&\qquad \cdot \int\limits_{\mathbb{R}} |z|^r f_{X,Z}(x,z)\, dF_\varepsilon(\varepsilon)\, dx\, dz \\
&= \frac{1}{n^r}\frac{1}{(2\pi)^r} \int\limits_{\mathbb{R}} \int\limits_{\mathbb{R}} \Big| \int\limits_{\mathbb{R}} e^{ixu} e^{i\varepsilon u}[e^{-it_1u} - e^{-it_2u}] \frac{\phi_K(b_n u)}{\phi_\varepsilon(u)} du \Big|^r \cdot \\
&\qquad \cdot \int\limits_{\mathbb{R}} |z|^r f_{Z|X=x}(z)\, dz \cdot f_X(x)\, dF_\varepsilon(\varepsilon)\, dx \\
&= \frac{1}{n^r}\frac{1}{(2\pi)^r} \int\limits_{\mathbb{R}} \int\limits_{\mathbb{R}} \Big| \int\limits_{\mathbb{R}} e^{ixu} e^{i\varepsilon u}[e^{-it_1u} - e^{-it_2u}] \frac{\phi_K(b_n u)}{\phi_\varepsilon(u)} du \Big|^r \\
&\qquad \cdot \mathbb{E}(|Z|^r | X = x) \cdot f_X(x)\, dF_\varepsilon(\varepsilon)\, dx \\
&\leq \sup_{x\in[\tau_1,\tau_2]} \mathbb{E}(|Z|^r | X = x) \cdot \frac{||f_X||_\infty}{n^r \cdot (2\pi)^r} \cdot \\
&\qquad \cdot \int\limits_{\mathbb{R}} \int\limits_{\mathbb{R}} \Big| \int\limits_{\mathbb{R}} e^{ixu} e^{i\varepsilon u}[e^{-it_1u} - e^{-it_2u}] \frac{\phi_K(b_n u)}{\phi_\varepsilon(u)} du \Big|^r dF_\varepsilon(\varepsilon)\, dx
\end{aligned}
$$

An dieser Stelle kann, analog zur entsprechenden Abschätzung im Falle des Dichteschätzers, Ungleichung (1.13) angewendet werden und es ergibt sich ebenfalls

$$
\begin{aligned}
&\leq \sup_{x\in[\tau_1,\tau_2]} \mathbb{E}(|Z|^r | X = x) \cdot ||f_X||_\infty \cdot \\
&\qquad \cdot O\Big(\frac{1}{n(nb_n)^{\beta+r-1}}\Big) \cdot \frac{|t_1-t_2|^r}{b_n^r} \cdot \Big(\int \big(|u|^{\beta+1} \big|\phi_K(u)\big|\big)^{\frac{r}{r-1}} du \Big)^{r-1} \\
&= \frac{1}{n} \cdot O\Big((nb_n^{\frac{r}{r-1}\beta+1})^{r-1} \Big) \cdot \Big|\frac{t_1-t_2}{b_n}\Big|^r = \frac{1}{n} \cdot O\Big((nb_n^{2\beta+1})^{\frac{r}{2}} \Big) \cdot \Big|\frac{t_1-t_2}{b_n}\Big|^r
\end{aligned}
$$

für $r \geq 2$. Somit folgt (iii). □

Um Raten für die Momente des Nadaraya-Watson-Typ-Schätzers (1.5) zu erzielen, muss gewährleistet sein, dass die Wahrscheinlichkeit, dass der Nenner des Schätzers kleiner als eine Konstante ist, genügend klein ist, so dass dieses Ereignis die Konvergenzrate nicht beeinträchtigt. Dieses Resultat, was ein Standardresultat für Kerndichteschätzer darstellt, wird im nächsten Lemma angegeben.

Lemma 1.1.3 *Mit $b_n \asymp n^{-\frac{1}{2p+2\beta+1}}$ gilt unter den Voraussetzungen von Theorem 1.1.1, dass*

$$\forall B > 0 : \mathbb{P}(\sup_{t\in[\tau_1,\tau_2]} |\hat{f}_{X,n}(t) - f_X(t)| \geq B) = O(n^{-d})$$

für beliebig gewähltes $d > 0$.

Beweis:

Wie in der Literatur üblich (siehe zum Beispiel Müller und Stadtmüller [52]), wird dies gezeigt, indem ein Gitter T_n auf dem Intervall $[\tau_1, \tau_2]$ betrachtet wird. Die Feinheit dieses Gitters sei von der Ordnung n^{-3}. Die Abbildung $a_n : [0,1] \to T_n$ weist jedem Punkt in $[\tau_1, \tau_2]$ den nächstgelegenen Gitterpunkt in T_n zu.
So können wir schreiben, für $B > 0$:

$$\begin{aligned}
& \mathbb{P}(\sup_{t\in[\tau_1,\tau_2]} |\hat{f}_{X,n}(t) - f_X(t)| \geq B) \\
\leq\ & \mathbb{P}(\sup_{t\in[\tau_1,\tau_2]} |\hat{f}_{X,n}(t) - \hat{f}_{X,n}(a_n(t))| \geq B/4) \\
& + \mathbb{P}(\max_{t\in T_n} |\hat{f}_{X,n}(a_n(t)) - \mathbb{E}\hat{f}_{X,n}(a_n(t))| \geq B/4) \\
& + \mathbb{P}(\sup_{t\in[\tau_1,\tau_2]} |\mathbb{E}\hat{f}_{X,n}(a_n(t)) - \mathbb{E}\hat{f}_{X,n}(t)| \geq B/4) \\
& + \mathbb{P}(\sup_{t\in[\tau_1,\tau_2]} |\mathbb{E}\hat{f}_{X,n}(t) - f_X(t)| \geq B/4) \\
=:\ & \mathbb{P}_1 + \mathbb{P}_2 + \mathbb{P}_3 + \mathbb{P}_4
\end{aligned}$$

Was $\mathbb{P}_1$ and $\mathbb{P}_3$ anbelangt, so stellt die Lipschitz-Stetigkeit des Schätzers sicher, dass diese beiden Wahrscheinlichkeiten klein sind, denn wir haben, unter Durchführung ähnlicher Rechnungen wie in Lemma 1.1.1 (ii):

$$\begin{aligned}
& |\hat{f}_{X,n}(t) - \hat{f}_{X,n}(a_n(t))| \\
=\ & \frac{1}{nb_n} \Big| \sum_{j=1}^{n} K^*_{b_n}\big(\frac{t - Y_j}{b_n}\big) - K^*_{b_n}\big(\frac{a_n(t) - Y_j}{b_n}\big) \Big| \\
\leq\ & \frac{1}{b_n^{1+\beta}} L_K\, |t - a_n(t)|.
\end{aligned}$$

Des weiteren gilt

$$\frac{1}{b_n^{1+\beta}} L_K |t - a_n(t)| < B/4$$

für $b_n \asymp n^{-\frac{1}{2p+2\beta+1}}$ und $|t - a_n(t)| \leq n^{-3}$, für hinreichend großes n, was ebenso für den Erwartungswert gilt, und somit gilt $\mathbb{P}_1 = 0$ und $\mathbb{P}_3 = 0$.
Für $\mathbb{P}_2$ gilt, gemäß Lemma 1.1.1 (ii)

$$\begin{aligned}
\mathbb{P}_2 &\leq \sum_{t \in a_n} \mathbb{P}(|\hat{f}_{X,n}(a_n(t)) - \mathbb{E}\hat{f}_{X,n}(a_n(t))| \geq B/4) \\
&= O(n^3) \cdot (B/4)^r \cdot (nb_n^{2\beta+1})^{-r/2} \\
&= (B/4)^r \cdot O(n^3 n^{-\frac{rp}{2p+2\beta+1}}) = O(n^{-d}), \qquad \text{für } r \text{ hinreichend groß.}
\end{aligned}$$

Der Bias besitzt die gleichmäßige Ordnung b_n^{p-1}, und somit gilt $\mathbb{P}_4 = 0$ für hinreichend großes n, wegen

$$b_n^p \asymp (nb_n^{2\beta+1})^{-1/2}.$$

□

Im nächsten Lemma werden die absoluten Momente des stochastischen Anteils der Differenz des Nadaraya-Watson-Typ-Kurvenschätzers (siehe (1.5)) an zwei verschiedenen Punkten betrachtet. Wie in Theorem 1.1.2 wird hier angenommen, dass für die Dichte f_X gilt, dass $f_X \geq c > 0$ im Intervall $[\tau_1, \tau_2]$. Wie im Beweis von Theorem 1.1.2 definiert, betrachten wir das Ereignis $\bar{A}_n$, dass der Schätzer für f_X innerhalb einer $c/2$-Umgebung der tatsächlichen Kurve f_X liegt. Wie im zuletzt angegebenen Lemma gesehen, besitzt das entsprechende Gegenereignis eine Wahrscheinlichkeit der Ordnung $O(n^{-d})$, für beliebiges $d > 0$. Falls ω in $\bar{A}_n$ liegt, bedeutet dies ebenfalls, dass der Kurvenschätzer $\hat{f}_{X,n}$ nach unten von Null weg beschränkt ist und nach oben beschränkt ist durch $\max_{t \in [\tau_1, \tau_2]} f_X(t) + c/2$.

Lemma 1.1.4 *Falls $\mathbb{E}(|Z|^{r \cdot q}|X = x)$ existiert und gleichmäßig beschränkt ist für $q > 1$, dann gilt*

$$\begin{aligned}
&\mathbb{E}\Big(\Big|\frac{\hat{r}_n(t_1)}{\hat{f}_{X,n}(t_1)} - \frac{\hat{r}_n(t_2)}{\hat{f}_{X,n}(t_2)} - \Big(\frac{\mathbb{E}(\hat{r}_n(t_1))}{\mathbb{E}(\hat{f}_{X,n}(t_1))} - \frac{\mathbb{E}\hat{r}_n(t_2)}{\mathbb{E}\hat{f}_{X,n}(t_2)}\Big)\Big|^r \cdot \mathbb{1}\{\bar{A}_n\}\Big) \\
= \; & O\Big(\Big|\frac{t_1 - t_2}{b_n}\Big|^r \Big(\frac{1}{nb_n^{2\beta+1}}\Big)^{r/2}\Big).
\end{aligned}$$

Beweis:

Aufgrund der Tatsache, dass der Kurvenschätzer im Intervall $[\tau_1, \tau_2]$ für $\omega \in \bar{A}_n$ nach unten von Null weg beschränkt ist, kann der Ausgangsterm zunächst

folgendermaßen abgeschätzt werden (der Faktor $\mathbb{1}\{\bar{A}_n\}$ wird im folgenden der Lesbarkeit halber unterdrückt)

$$\begin{aligned}
&\mathbb{E}\Big|\frac{\hat{r}_n(t_1)}{\hat{f}_{X,n}(t_1)} - \frac{\hat{r}_n(t_2)}{\hat{f}_{X,n}(t_2)} - \Big(\frac{\mathbb{E}(\hat{r}_n(t_1))}{\mathbb{E}(\hat{f}_{X,n}(t_1))} - \frac{\mathbb{E}\hat{r}_n(t_2)}{\mathbb{E}\hat{f}_{X,n}(t_2)}\Big)\Big|^r \\
&\leq \frac{1}{C}\cdot\mathbb{E}\big|\hat{r}_n(t_1)\hat{f}_{X,n}(t_2)\mathbb{E}\hat{f}_{X,n}(t_1)\mathbb{E}\hat{f}_{X,n}(t_2) - \hat{r}_n(t_2)\hat{f}_{X,n}(t_1)\mathbb{E}\hat{f}_{X,n}(t_1)\mathbb{E}\hat{f}_{X,n}(t_2) \\
&\qquad -\hat{f}_{X,n}(t_1)\hat{f}_{X,n}(t_2)\mathbb{E}\hat{r}_n(t_1)\mathbb{E}\hat{f}_{X,n}(t_2) + \hat{f}_{X,n}(t_1)\hat{f}_{X,n}(t_2)\mathbb{E}\hat{r}_n(t_2)\mathbb{E}\hat{f}_{X,n}(t_1)\big|^r \\
&=: \frac{1}{C}\cdot\mathbb{E}\big(|T_1|^r\big);
\end{aligned}$$

Des weiteren gilt, unter Durchführung geeigneter Nulladditionen:

$$\begin{aligned}
T_1 \;=\;& \hat{r}_n(t_1)\hat{f}_{X,n}(t_2)\mathbb{E}\hat{f}_{X,n}(t_1)\mathbb{E}\hat{f}_{X,n}(t_2) - \hat{r}_n(t_2)\hat{f}_{X,n}(t_2)\mathbb{E}\hat{f}_{X,n}(t_1)\mathbb{E}\hat{f}_{X,n}(t_2) \\
&+ \hat{r}_n(t_2)\hat{f}_{X,n}(t_2)\mathbb{E}\hat{f}_{X,n}(t_1)\mathbb{E}\hat{f}_{X,n}(t_2) - \hat{r}_n(t_2)\hat{f}_{X,n}(t_1)\mathbb{E}\hat{f}_{X,n}(t_1)\mathbb{E}\hat{f}_{X,n}(t_2) \\
&- \hat{f}_{X,n}(t_1)\hat{f}_{X,n}(t_2)\mathbb{E}\hat{r}_n(t_1)\mathbb{E}\hat{f}_{X,n}(t_2) + \hat{f}_{X,n}(t_1)\hat{f}_{X,n}(t_2)\mathbb{E}\hat{r}_n(t_2)\mathbb{E}\hat{f}_{X,n}(t_2) \\
&- \hat{f}_{X,n}(t_1)\hat{f}_{X,n}(t_2)\mathbb{E}\hat{r}_n(t_2)\mathbb{E}\hat{f}_{X,n}(t_2) + \hat{f}_{X,n}(t_1)\hat{f}_{X,n}(t_2)\mathbb{E}\hat{r}_n(t_2)\mathbb{E}\hat{f}_{X,n}(t_1)
\end{aligned}$$

Ausklammern ergibt weiterhin

$$\begin{aligned}
=\;& [\hat{r}_n(t_1) - \hat{r}_n(t_2)]\cdot[\hat{f}_{X,n}(t_2)\mathbb{E}\hat{f}_{X,n}(t_1)\mathbb{E}\hat{f}_{X,n}(t_2)] \\
&- [\mathbb{E}\hat{r}_n(t_1) - \mathbb{E}\hat{r}_n(t_2)]\cdot[\hat{f}_{X,n}(t_1)\hat{f}_{X,n}(t_2)\mathbb{E}\hat{f}_{X,n}(t_2)] \\
&+ [\hat{f}_{X,n}(t_2) - \hat{f}_{X,n}(t_1)]\cdot[\hat{r}_n(t_2)\mathbb{E}\hat{f}_{X,n}(t_1)\mathbb{E}\hat{f}_{X,n}(t_2)] \\
&- [\mathbb{E}\hat{f}_{X,n}(t_2) - \mathbb{E}\hat{f}_{X,n}(t_1)]\cdot[\hat{f}_{X,n}(t_1)\hat{f}_{X,n}(t_2)\mathbb{E}\hat{r}_n(t_2)].
\end{aligned}$$

Weitere Nulladditionen führen zu

$$\begin{aligned}
=\;& [\hat{r}_n(t_1) - \hat{r}_n(t_2)]\cdot[\hat{f}_{X,n}(t_2)\mathbb{E}\hat{f}_{X,n}(t_1)\mathbb{E}\hat{f}_{X,n}(t_2)] \\
-\;& [\mathbb{E}\hat{r}_n(t_1) - \mathbb{E}\hat{r}_n(t_2)]\cdot[\hat{f}_{X,n}(t_2)\mathbb{E}\hat{f}_{X,n}(t_1)\mathbb{E}\hat{f}_{X,n}(t_2)] \\
+\;& [\mathbb{E}\hat{r}_n(t_1) - \mathbb{E}\hat{r}_n(t_2)]\cdot[\hat{f}_{X,n}(t_2)\mathbb{E}\hat{f}_{X,n}(t_1)\mathbb{E}\hat{f}_{X,n}(t_2)] \\
-\;& [\mathbb{E}\hat{r}_n(t_1) - \mathbb{E}\hat{r}_n(t_2)]\cdot[\hat{f}_{X,n}(t_2)\hat{f}_{X,n}(t_1)\mathbb{E}\hat{f}_{X,n}(t_2)] \\
+\;& [\hat{f}_{X,n}(t_2) - \hat{f}_{X,n}(t_1)]\cdot[\hat{r}_n(t_2)\mathbb{E}\hat{f}_{X,n}(t_1)\mathbb{E}\hat{f}_{X,n}(t_2)] \\
-\;& [\mathbb{E}\hat{f}_{X,n}(t_2) - \mathbb{E}\hat{f}_{X,n}(t_1)]\cdot[\hat{r}_n(t_2)\mathbb{E}\hat{f}_{X,n}(t_1)\mathbb{E}\hat{f}_{X,n}(t_2)] \\
+\;& [\mathbb{E}\hat{f}_{X,n}(t_2) - \mathbb{E}\hat{f}_{X,n}(t_1)]\cdot[\hat{r}_n(t_2)\mathbb{E}\hat{f}_{X,n}(t_1)\mathbb{E}\hat{f}_{X,n}(t_2)] \\
-\;& [\mathbb{E}\hat{f}_{X,n}(t_2) - \mathbb{E}\hat{f}_{X,n}(t_1)]\cdot[\mathbb{E}\hat{r}_n(t_2)\mathbb{E}\hat{f}_{X,n}(t_1)\mathbb{E}\hat{f}_{X,n}(t_2)]
\end{aligned}$$

$$
\begin{aligned}
&+ \quad [\mathbb{E}\hat{f}_{X,n}(t_2) - \mathbb{E}\hat{f}_{X,n}(t_1)] \cdot [\mathbb{E}\hat{r}_n(t_2)\mathbb{E}\hat{f}_{X,n}(t_1)\mathbb{E}\hat{f}_{X,n}(t_2)] \\
&- \quad [\mathbb{E}\hat{f}_{X,n}(t_2) - \mathbb{E}\hat{f}_{X,n}(t_1)] \cdot [\mathbb{E}\hat{r}_n(t_2)\hat{f}_{X,n}(t_1)\mathbb{E}\hat{f}_{X,n}(t_2)] \\
&+ \quad [\mathbb{E}\hat{f}_{X,n}(t_2) - \mathbb{E}\hat{f}_{X,n}(t_1)] \cdot [\mathbb{E}\hat{r}_n(t_2)\hat{f}_{X,n}(t_1)\mathbb{E}\hat{f}_{X,n}(t_2)] \\
&- \quad [\mathbb{E}\hat{f}_{X,n}(t_2) - \mathbb{E}\hat{f}_{X,n}(t_1)] \cdot [\mathbb{E}\hat{r}_n(t_2)\hat{f}_{X,n}(t_1)\hat{f}_{X,n}(t_2)].
\end{aligned}
$$

Ausklammern vereinfacht den Term T_1 schließlich zu

$$
\begin{aligned}
T_1 \;=\; & \Big([\hat{r}_n(t_1) - \hat{r}_n(t_2)] - [\mathbb{E}\hat{r}_n(t_1) - \mathbb{E}\hat{r}_n(t_2)]\Big) \\
& \cdot \Big(\hat{f}_{X,n}(t_2)\mathbb{E}\hat{f}_{X,n}(t_1)\mathbb{E}\hat{f}_{X,n}(t_2)\Big) \\
+ \; & \Big(\mathbb{E}\hat{f}_{X,n}(t_1) - \hat{f}_{X,n}(t_1)\Big) \cdot \Big(\mathbb{E}\hat{r}_n(t_1) - \mathbb{E}\hat{r}_n(t_2)\Big) \cdot \Big(\hat{f}_{X,n}(t_2)\mathbb{E}\hat{f}_{X,n}(t_2)\Big)
\end{aligned}
$$

$$
\begin{aligned}
+ \; & \Big([\hat{f}_{X,n}(t_2) - \hat{f}_{X,n}(t_1)] - [\mathbb{E}\hat{f}_{X,n}(t_2) - \mathbb{E}\hat{f}_{X,n}(t_1)]\Big) \\
& \cdot \Big(\hat{r}_n(t_2)\mathbb{E}\hat{f}_{X,n}(t_1)\mathbb{E}\hat{f}_{X,n}(t_2)\Big) \\
+ \; & \Big(\mathbb{E}\hat{f}_{X,n}(t_2) - \mathbb{E}\hat{f}_{X,n}(t_1)\Big) \cdot \Big(\hat{r}_n(t_2) - \mathbb{E}\hat{r}_n(t_2)\Big) \cdot \Big(\mathbb{E}\hat{f}_{X,n}(t_1)\mathbb{E}\hat{f}_{X,n}(t_2)\Big) \\
+ \; & \Big(\mathbb{E}\hat{f}_{X,n}(t_2) - \mathbb{E}\hat{f}_{X,n}(t_1)\Big) \cdot \Big(\mathbb{E}\hat{f}_{X,n}(t_1) - \hat{f}_{X,n}(t_1)\Big) \cdot \Big(\mathbb{E}\hat{r}_n(t_1)\mathbb{E}\hat{f}_{X,n}(t_2)\Big) \\
+ \; & \Big(\mathbb{E}\hat{f}_{X,n}(t_2) - \mathbb{E}\hat{f}_{X,n}(t_1)\Big) \cdot \Big(\mathbb{E}\hat{f}_{X,n}(t_2) - \hat{f}_{X,n}(t_2)\Big) \cdot \Big(\mathbb{E}\hat{r}_n(t_2)\hat{f}_{X,n}(t_1)\Big).
\end{aligned}
$$

T_1 kann also als die Summe von Produkten von je zwei bzw. drei Zufallsvariablen dargestellt werden. Unter Anwendung der C_r-Ungleichung (Shorack und Wellner [69], S. 843) kann das r-te Moment von T_1 nun zunächst abgeschätzt werden durch die Summe der r-ten Momente der jeweiligen Produkte, multipliziert mit einer Konstanten. Auf die einzelnen Summanden kann nun jeweils die Hölder-Ungleichung angewendet werden, und man erhält so folgende Abschätzung:

$$
\begin{aligned}
& \mathbb{E}(|T_1|^r) \\
\leq \; & C_r \cdot \Big\{ \Big(\mathbb{E}\Big|[\hat{r}_n(t_1) - \hat{r}_n(t_2)] - [\mathbb{E}\hat{r}_n(t_1) - \mathbb{E}\hat{r}_n(t_2)]\Big|^{rq_1}\Big)^{1/q_1} \\
& \qquad \cdot \Big(\mathbb{E}\Big|\hat{f}_{X,n}(t_2)\mathbb{E}\hat{f}_{X,n}(t_1)\mathbb{E}\hat{f}_{X,n}(t_2)\Big|^{rq_2}\Big)^{1/q_2} \\
& + \Big(\mathbb{E}\Big|\mathbb{E}\hat{r}_n(t_1) - \mathbb{E}\hat{r}_n(t_2)\Big|^{rq_1}\Big)^{1/q_1} \\
& \qquad \cdot \Big(\mathbb{E}\Big|\mathbb{E}\hat{f}_{X,n}(t_1) - \hat{f}_{X,n}(t_1)\Big|^{rq_3}\mathbb{E}\Big|\hat{f}_{X,n}(t_2)\mathbb{E}\hat{f}_{X,n}(t_2)\Big|^{rq_3}\Big)^{1/q_3} \\
& + \Big(\mathbb{E}\Big|\hat{r}_n(t_2)\mathbb{E}\hat{f}_{X,n}(t_1)\mathbb{E}\hat{f}_{X,n}(t_2)\Big|^{rq_1}\Big)^{1/q_1}
\end{aligned}
$$

$$
\begin{aligned}
&\cdot \Big(\mathbb{E}\Big|\hat{f}_{X,n}(t_2) - \hat{f}_{X,n}(t_1) - [\mathbb{E}\hat{f}_{X,n}(t_2) - \mathbb{E}\hat{f}_{X,n}(t_1)]\Big|^{rq_2}\Big)^{1/q_2} \\
+ \Big(&\mathbb{E}\Big|\hat{r}_n(t_2) - \mathbb{E}\hat{r}_n(t_2)\Big|^{rq_1}\Big)^{1/q_1} \\
&\cdot \Big(\mathbb{E}\Big|\mathbb{E}\hat{f}_{X,n}(t_2) - \mathbb{E}\hat{f}_{X,n}(t_1)\Big|^{rq_3}\Big|\mathbb{E}\hat{f}_{X,n}(t_1)\mathbb{E}\hat{f}_{X,n}(t_2)\Big|^{rq_3}\Big)^{1/q_3} \\
+ \Big(&\Big|\mathbb{E}\hat{r}_n(t_2)\mathbb{E}\hat{f}_{X,n}(t_2)\Big|^{rq_1}\Big)^{1/q_1} \\
&\cdot \Big(\mathbb{E}\Big|\mathbb{E}\hat{f}_{X,n}(t_2) - \mathbb{E}\hat{f}_{X,n}(t_1)\Big|^{rq_3}\mathbb{E}\Big|\mathbb{E}\hat{f}_{X,n}(t_1) - \hat{f}_{X,n}(t_1)\Big|^{rq_3}\Big)^{1/q_3} \\
+ \Big(&\mathbb{E}\Big|\hat{f}_{X,n}(t_1)\mathbb{E}\hat{r}_n(t_2)\Big|^{rq_1}\Big)^{1/q_1} \\
&\cdot \Big(\mathbb{E}\Big|\mathbb{E}\hat{f}_{X,n}(t_2) - \mathbb{E}\hat{f}_{X,n}(t_1)\Big|^{rq_3}\mathbb{E}\Big|\mathbb{E}\hat{f}_{X,n}(t_2) - \hat{f}_{X,n}(t_2)\Big|^{rq_3}\Big)^{1/q_3}\Big\}.
\end{aligned}
$$

Für die Konstanten q_1, q_2 und q_3 gilt hierbei $q_2 = q_1/(q_1 - 1)$, sowie $q_3 = 2q_1/(q_1 - 1)$, während q_1 beliebig nahe an 1 gewählt werden kann, da für den Dichteschätzer jede Momentenbedingung aufgrund von (1.3) erfüllt ist. Dadurch wird gewährleistet, dass für die bedingten Momente von Z unter der Bedingung $X = x$ keine Momentenbedingung benötigt wird, welche über diejenige hinausgeht, die in Theorem 1.1.2 gefordert ist. Der erste und dritte Summand bestimmen die Konvergenzrate, da die anderen Terme höchstens von derselben Ordnung sind. Mit Lemma 1.1.1 und Lemma 1.1.2 folgt das gewünschte Resultat. □

Für den im Beweis von Theorem 1.1.2 auftretenden Biasterm ist ebenfalls zu zeigen, von welcher Ordnung dieser ist, um zu gewährleisten, dass ab hinreichend großem n die entsprechende Wahrscheinlichkeit $\mathbb{P}_{2,k}$ Null wird. Unter den Voraussetzungen von Theorem 1.1.2 wird hierfür dieselbe Ordnung erzielt wie für den entsprechenden Term in Theorem 1.1.1. Diese Aussage gibt das folgende Lemma, wobei die Vorgehensweise des Beweises ähnlich ist wie im Lemma zuvor und die Rechnungen daher verkürzt dargestellt werden.

Lemma 1.1.5 *Es gelten die Bedingungen von Theorem 1.1.2. Dann gilt:*

$$\sup_{t \in T_{n,k}} \left| \frac{\mathbb{E}\hat{r}_n(t)}{\mathbb{E}\hat{f}_{X,n}(t)} - \frac{\mathbb{E}\hat{r}_n(\theta)}{\mathbb{E}\hat{f}_{X,n}(\theta)} - \left(\frac{r(t)}{f_X(t)} - \frac{r(\theta)}{f_X(\theta)}\right) \right| = O(k \cdot h_n \cdot b_n).$$

Beweis:

Es gilt, unter Durchführung analoger Überlegungen wie im vorherigen Lemma

$$
\begin{aligned}
& \Big| \frac{\mathbb{E}\hat{r}_n(t)}{\mathbb{E}\hat{f}_{X,n}(t)} - \frac{\mathbb{E}\hat{r}_n(\theta)}{\mathbb{E}\hat{f}_{X,n}(\theta)} - \Big(\frac{r(t)}{f_X(t)} - \frac{r(\theta)}{f_X(\theta)} \Big) \Big| \\
\leq \; & \frac{1}{C} \cdot |T_2|
\end{aligned}
$$

mit

$$
\begin{aligned}
T_2 \; := \; & \Big([\mathbb{E}\hat{r}_n(t) - \mathbb{E}\hat{r}_n(\theta)] - [r(t) - r(\theta)] \Big) \\
& \cdot \Big(\mathbb{E}\hat{f}_{X,n}(\theta) f_X(t) f_X(\theta) \Big) \\
+ \; & \Big(f_X(t) - \mathbb{E}\hat{f}_{X,n}(t) \Big) \cdot \Big(r(t) - r(\theta) \Big) \cdot \Big(\mathbb{E}\hat{f}_{X,n}(\theta) f_X(\theta) \Big) \\
+ \; & \Big([\mathbb{E}\hat{f}_{X,n}(\theta) - \mathbb{E}\hat{f}_{X,n}(t)] - [f_X(\theta) - f_X(t)] \Big) \\
& \cdot \Big(\mathbb{E}\hat{r}_n(\theta) f_X(t) f_X(\theta) \Big) \\
+ \; & \Big(f_X(\theta) - f_X(t) \Big) \cdot \Big(\mathbb{E}\hat{r}_n(\theta) - r(\theta) \Big) \cdot \Big(f_X(t) f_X(\theta) \Big) \\
+ \; & \Big(f_X(\theta) - f_X(t) \Big) \cdot \Big(f_X(t) - \mathbb{E}\hat{f}_{X,n}(t) \Big) \cdot \Big(r(t) f_X(\theta) \Big) \\
+ \; & \Big(f_X(\theta) - f_X(t) \Big) \cdot \Big(f_X(\theta) - \mathbb{E}\hat{f}_{X,n}(\theta) \Big) \cdot \Big(r(\theta) \mathbb{E}\hat{f}_{X,n}(t) \Big).
\end{aligned}
$$

Bestimmend für die Ordnung sind hier wiederum der erste und dritte Summand des Terms T_2. Für den Biasterm des dritten Summanden gilt gemäß der Argumentation im Beweis von Lemma 1.1.1 die Ordnung $O(k \cdot h_n \cdot b_n)$. Für den Bias-Ausdruck im ersten Summanden kann in ähnlicher Weise wie für den Dichteschätzer folgende Abschätzung getroffen werden:

$$
\begin{aligned}
& \sup_{t \in T_{n,k}} \big| [\mathbb{E}\hat{r}_n(t) - \mathbb{E}\hat{r}_n(\theta)] - [r(t) - r(\theta)] \big| \\
= \; & \sup_{t \in T_{n,k}} \Big| \int_{\mathbb{R}} \frac{1}{b_n} K\big(\frac{t-x}{b_n}\big) f_X(x) m(x) dx - \int_{\mathbb{R}} \frac{1}{b_n} K\big(\frac{\theta - x}{b_n}\big) f_X(x) m(x) dx \Big| \\
& \quad - \big(f_X(t) m(t) - f_X(\theta) m(\theta) \big) \\
= \; & \sup_{t \in T_{n,k}} \Big| \int_{\mathbb{R}} \frac{1}{b_n} K\big(\frac{\theta - z}{b_n}\big) (f_X(z + t - \theta) m(z + t - \theta) - f_X(z) m(z)) dz \\
& \quad - (f_X(t) m(t) - f_X(\theta) m(\theta)) \Big| \\
= \; & \sup_{t \in T_{n,k}} \Big| \int_0^{t-\theta} \Big(\int_{\mathbb{R}} \frac{1}{b_n} K\big(\frac{\theta - z}{b_n}\big) (f_X \cdot m)'(z + u) dz - (f_X \cdot m)'(\theta + u) \Big) du \Big|
\end{aligned}
$$

$$
\begin{aligned}
&= \sup_{t\in T_{n,k}} \Big| \int_0^{t-\theta} \Big(\int_{\mathbb{R}} K(z)\big((f_X\cdot m)'(\theta+u-zb_n) - (f_X\cdot m)'(\theta+u)\big)\,dz\Big)\,du\Big| \\
&= \sup_{t\in T_{n,k}} \Big| \int_0^{t-\theta} \Big(\int_{\mathbb{R}} K(z)\Big[\Big(\sum_{k=2}^{p-1} \frac{(f_X\cdot m)^{(k)}(\theta+u)}{(k-1)!}\,(zb_n)^{k-1}\Big) \\
&\qquad\qquad - \frac{(f_X\cdot m)^{(p)}(\theta+u+\tilde{z})}{(p-1)!}\,(zb_n)^{p-1}\Big]\,dz\Big)\,du\Big| \\
&\le \sup_{t\in T_{n,k}} b_n^{p-1}\cdot \frac{||(f_X\cdot m)^{(p)}||_\infty}{(p-1)!} \int_0^{t-\theta}\int_{\mathbb{R}} |K(z)||z^{p-1}|dz\,du \\
&= O(k\cdot h_n\cdot b_n^{p-1}).
\end{aligned}
$$

Die Biasterme im zweiten, vierten, fünften und sechsten Summanden sind jeweils von der Ordnung $O(b_n^p)$, denn

$$
\begin{aligned}
&\sup_{t\in T_{n,k}} \big|\mathbb{E}\hat{r}_n(t) - r(t)\big| \\
&= \sup_{t\in T_{n,k}} \Big| \int_{\mathbb{R}} K\big(\frac{t-x}{b_n}\big)\big(f_X(x)m(x) - f_X(t)m(t)\big)dx\Big| \\
&= \sup_{t\in T_{n,k}} \Big| \int_{\mathbb{R}} K(z)\big[(f_X\cdot m)(t-b_nz) - (f_X\cdot m)(t)\big]\,dz\Big| \\
&= \sup_{t\in T_{n,k}} \Big| \int_{\mathbb{R}} K(z)\Big[\Big(\sum_{k=1}^{p-1} \frac{(f_X\cdot m)^{(k)}(t)}{k!}\cdot (zb_n)^k\Big) + \frac{(f_X\cdot m)^{(p)}(\tilde{t})}{p!}\cdot (zb_n)^p\Big]dz\Big| \\
&\le b_n^p\cdot \frac{||(f_X\cdot m)^{(p)}||_\infty}{p!}\cdot \int_{\mathbb{R}} |K(z)|\cdot|z^p|\,dz \\
&= O(b_n^p).
\end{aligned}
$$

Analog wird der Bias von $\hat{f}_{X,n}(\cdot)$ berechnet.

Aufgrund der Eigenschaften der Funktionen $f_X(\cdot)$ bzw. $m(\cdot)$ ist

$$
\sup_{t\in T_{n,k}} |f_X(\theta) - f_X(t)| \text{ bzw. } \sup_{t\in T_{n,k}} |r(t) - r(\theta)|
$$

jeweils von der Ordnung $O(k\cdot h_n)$.

Des weiteren sind $f_X(\cdot)$ und $r(\cdot)$ im Intervall $[\tau_1, \tau_2]$ beschränkt, und dies gilt ebenso für $\mathbb{E}\hat{f}_{X,n}(\cdot)$ bzw. $\mathbb{E}\hat{r}_n(\cdot)$. Damit folgt also insgesamt die gewünschte Ordnung und somit die Aussage des Lemmas. $\square$

Das folgende Resultat wird benötigt, um die Wahrscheinlichkeit zu beschränken, dass der Modalwertschätzer $\hat{\theta}_n$ im Intervall $[\theta_{n,k-1}, \theta_{n,k}]$ liegt.

Lemma 1.1.6 *Falls für den stochastischen Prozess $(Z_n(t))_{t\in\mathbb{R}}$ für t_1, t_2 in $[\tau_1, \tau_2]$ gilt, dass,*

$$\mathbb{E}(|Z_n(t_1) - Z_n(t_2)|^r) = O\big(\big|\frac{t_1 - t_2}{b_n}\big|^r (nb_n^{2\beta+1})^{-r/2}\big)$$

und falls die Pfade von $(Z_n(t))_{t\in\mathbb{R}}$ stetig sind, dann folgt für $k \neq 0$

$$\mathbb{P}\big(\sup_{t\in[\theta_{n,k-1},\theta_{n,k}]} |Z_n(t_1) - Z_n(\theta)| \geq c\cdot(kh_n)^2\big) = |k|^{-r} O\big(\frac{(\frac{1}{b_n})^r (nb_n^{2\beta+1})^{-r/2}}{h_n^r}\big)\,.$$

Beweis:

Es wird an dieser Stelle angenommen, dass k positiv ist, analoge Überlegungen gelten für negatives k. Für festes $l \in \mathbb{N}$ werden die Mengen $D_{n,k,l}$ und $D_{n,k}$ folgendermaßen definiert

$$D_{n,k,l} := \{\theta_{n,k-1} + j\cdot\frac{h_n}{2^l}|\,, j = 0, ..., 2^l\} \quad \text{und } D_{n,k} := \bigcup_{j\in\mathbb{N}} D_{n,k,l}.$$

$D_{n,k,l}$ bezeichnet ein Gitter der Breite $2^{-l}h_n$ auf dem Intervall $T_{n,k} := [\theta_{n,k-1}, \theta_{n,k}]$, während $D_{n,k}$ eine dichte Menge innerhalb von $T_{n,k}$ darstellt.

Definiere nun die Abbildung

$$t_l(\cdot):\ D_{n,k,l+1} \to D_{n,k,l} \quad \text{durch} \quad t_l(t) = \min\{s \in D_{n,k,l} : s = \operatorname{argmin}|s-t|\}.$$

$t_l(t)$ bezeichnet den Gitterpunkt in $D_{n,k,l}$, welcher am nächsten an $t \in D_{n,k,l+1}$ gelegen ist.

Wähle nun η_l, so dass $\eta_l := \tilde{c}(kh_n)^2 2^{-\gamma l}$ mit $\tilde{c} \leq c(1-2^{-\gamma})$ und $\gamma < 1 - \frac{1}{r}$.

Dann gilt

$$\begin{aligned} & \mathbb{P}(\sup_{t\in T_{n,k}} |Z_n(t) - Z(\theta)| \geq c\cdot(kh_n)^2) \\ \leq\ & \mathbb{P}(\sup_{t\in T_{n,k}} |Z_n(t) - Z(\theta)| \geq \sum_{j=0}^{\infty}\eta_j) \text{ mit } \sum_{j=0}^{\infty}\eta_j \leq c\cdot(kh_n)^2 \end{aligned}$$

$$
\begin{aligned}
&= \mathbb{P}(\sup_{t\in T_{n,k}\cap D_{n,k}} |Z_n(t) - Z(\theta)| \geq \sum_{j=0}^{\infty} \eta_j) \text{ (aufgrund der Stetigkeit von } Z_n(t)) \\
&\leq \sum_{t\in D_{n,k,0}} \mathbb{P}(|Z_n(t) - Z_n(\theta)| \geq \eta_0) \\
&\quad + \sum_{l=1}^{\infty} \sum_{t\in D_{n,k,l}} \mathbb{P}(|Z_n(t) - Z_n(t_{l-1}(t))| \geq \eta_l).
\end{aligned}
$$

Dies kann mithilfe der Markov-Ungleichung, unter Verwendung der Voraussetzungen des Lemmas wiederum abgeschätzt werden durch

$$
\begin{aligned}
&\leq |k|^{-r}\, O\Big(\frac{(\frac{1}{b_n})^r (nb_n^{2\beta+1})^{-r/2}}{h_n^r}\Big) \sum_{l=0}^{\infty} 2^{l(r(\gamma-1)+1)} \\
&= |k|^{-r}\, O\Big(\frac{(\frac{1}{b_n})^r (nb_n^{2\beta+1})^{-r/2}}{h_n^r}\Big).
\end{aligned}
$$

□

1.2 Regressionskurven im Fixed Design Modell mit nichtglatter Modalstelle

Während im vorherigen Abschnitt Modalwertschätzung unter der Voraussetzung, dass die interessierende Kurve differenzierbar ist und die getätigten Beobachtungen indirekt sind, betrachtet wurde, werden nun andere Annahmen an das zugrundeliegende Modell gestellt. Im folgenden wird ein Regressionsmodell mit fixiertem Design beleuchtet, wobei jedoch die Regressionskurve, deren Modalwert es zu schätzen gilt, an der Modalstelle nichtdifferenzierbar ist.

Während in den im Abschnitt 1 erwähnten Arbeiten zur Modalwertschätzung jeweils Differenzierbarkeit an der Modalstelle der zugrundeliegenden Funktionen vorausgesetzt wird, so wurden in der Literatur auch obere Schranken für die Modalwertschätzung nichtglatter Kurven erzielt. In der bereits im Zusammenhang mit der Modalwertschätzung in Dekonvolutionsmodellen erwähnten Arbeit von Meister [49] ist für die Dichteschätzung auch der Fall nichtstetiger und nichtglatter Funktionen enthalten und die erzielten Raten sind in der dort betrachteten allgemeinen Funktionenklasse optimal. In Abraham et al. [1] und Abraham et al. [2] wird die Schätzung des Modalwertes nichtglatter Wahrscheinlichkeitsdichten

betrachtet und es werden Raten erzielt, die im Fall nichtdifferenzierbarer Funktionen von steilerem Abfallen der Kurve an der Modalstelle profitiern. Die Schätzung des Modalwertes erfolgt hierbei, indem der entsprechende Kurvenschätzer nur über die endlich vielen Beobachtungen maximiert wird, die erzielten Raten sind jedoch nicht optimal. In Herrmann und Ziegler [39] wurden etwas bessere Raten unter Verwendung des gewöhnlichen Argmax-Schätzers bewiesen, jedoch tritt auch hier ein zusätzlicher logarithmischer Faktor in der Konvergenzrate auf.
Im folgenden wird nun die Modalwertschätzung für Regressionskurven im Fixed Design Modell betrachtet. Mit einem Ansatz wie in Neumann [54], welcher auch in Wieczorek [78] im Fall der Dekonvolutionsmodelle verwendet wurde, wurden in hier sowohl obere Schranken der Konvergenzraten der verwendeten Schätzer nachgewiesen als auch deren Optimalität gezeigt. Als Schätzer für den Modalwert wird zum einen der gängige Argmax-Schätzer verwendet, zum anderen ein Schätzer, der aus der Maximierung des jeweiligen Kurvenschätzers über Punkte eines endlichen Gitters resultiert.
Die erzielten Resultate wurden in Kooperation mit Dr. Klaus Ziegler erarbeitet und sind in Wieczorek und Ziegler [79] nachzulesen. Die Autorin hat insbesondere zum Beweis der oberen Schranken des Modalwertschätzers entscheidende Beiträge geleistet als auch die Betrachtung unterer Schranken angeregt und ausgearbeitet.

Modellannahmen

Das Regressionsmodell in Wieczorek und Ziegler [79] wird zunächst durch folgende Gleichung beschrieben:

$$Y_{i,n} = m(x_{i,n}) + \varepsilon_{i,n}, \quad i = 1, ..., n. \tag{1.14}$$

Für beliebiges n bilden dabei die $\varepsilon_{i,n}$ eine Folge α-mischender Zufallsgrößen.
Die Designpunkte $x_{i,n} \in [0,1]$ werden als gleichmäßig verteilt angenommen, in dem Sinne, dass gilt

$$|x_{i,n} - x_{i-1,n}| = O(n^{-1}).$$

Die Regressionsfunktion $m(\cdot)$ sei auf $[0,1]$ hölderstetig der Ordnung $\rho \leq 1$, es gelte also für eine Konstante $c > 0$, dass für alle x, $y \in [0,1]$

$$|m(x) - m(y)| \leq c \cdot |x - y|^{\rho}.$$

Die Modalstelle sei wiederum eindeutig, und es gelte

$$|m(\theta) - m(x)| > \delta \text{ für } |\theta - x| > \mu.$$

Die Konvergenzgeschwindigkeit des Modalwertschätzers wird beeinflusst durch das Verhalten der Regressionsfunktion an der Modalstelle θ, wie in Wieczorek und Ziegler [79] gezeigt. Dieses Verhalten werde beschrieben durch

$$m(\theta) - m(x) \geq c_1|\theta - x|^\rho. \tag{1.15}$$

Je kleiner der Exponent ρ ist, desto stärker fällt die Kurve also an der Modalstelle θ. Die Klasse von Funktionen, welche die soeben genannten Voraussetzungen erfüllen, werde mit $\mathcal{F}_\rho$ bezeichnet.

1.2.1 Kernschätzer im Fixed Design Modell

Da der verwendete Modalwertschätzer darauf beruht, einen geeigneten Kurvenschätzer zu maximieren, wird zunächst der verwendete Schätzer für die unbekannte Regressionskurve $m(\cdot)$ vorgestellt. In Wieczorek und Ziegler [79] wird hierfür ein Kurvenschätzer in allgemeiner Form vorgeschlagen:

$$\hat{m}_n(t) = \sum_{i=1}^{n} Y_{i,n} w_{i,n}(t), \tag{1.16}$$

wobei $w_{i,n}(\cdot)$ eine Gewichtsfunktion darstellt, die folgende Bedingungen erfüllt:

$$\sum_{i=1}^{n} (w_{i,n}(t))^r = O\big(\frac{1}{(nb_n)^{r-1}}\big) \text{ für } r > 3/\rho. \tag{1.17}$$

Für den Bias gelte

$$|\mathbb{E}[\hat{m}_n(t)] - m(t)| = O(b_n^\rho)|, \tag{1.18}$$

gleichmäßig in t, für eine Folge $b_n := b(n)$, wobei gilt $b_n \overset{n\to\infty}{\to} 0$ und $nb_n \overset{n\to\infty}{\to} \infty$. Die gängigen Kernschätzer erfüllen diese Bedingungen, wobei b_n hier die Bandbreite ist, so zum Beispiel der Gasser-Müller-Schätzer

$$\hat{m}_n^{GM}(t) = \sum_{i=1}^{n} Y_{i,n} \frac{1}{b_n} \int_{s_{i-1,n}}^{s_{i,n}} K(\frac{x-t}{b_n}) dx, \tag{1.19}$$

der Priestley-Chao-Benedetti-Schätzer

$$\hat{m}_n^{PC}(t) = \sum_{i=1}^{n} Y_{i,n}(x_{i,n} - x_{i-1,n})\frac{1}{b_n}K(\frac{x_{i,n}-t}{b_n}) \tag{1.20}$$

und der Nadaraya-Watson-Schätzer

$$\hat{m}_n^{NW}(t) = \frac{\sum_{i=1}^{n} Y_{i,n}K(\frac{x_{i,n}-t}{b_n})}{\sum_{i=1}^{n} K(\frac{x_{i,n}-t}{b_n})}\,. \tag{1.21}$$

Die im Gasser-Müller-Schätzer auftretenden Punkte $s_{i,n}$ liegen zwischen $x_{i-1,n}$ und $x_{i,n}$.
Die Kernfunktion $K(\cdot)$ sei lipschitzstetig und besitze kompakten Träger.
Es werden nun zwei Modalwertschätzer folgendermaßen definiert:

$$\hat{\theta}_n^c = \min\{t | t = \operatorname{argmax}_{s\in[0,1]} \hat{f}_n(s)\} \tag{1.22}$$

und

$$\hat{\theta}_n^g = \min\{t | t = \operatorname{argmax}_{s\in\widetilde{G}_n} \hat{f}_n(s)\}. \tag{1.23}$$

Es sei hierbei $\widetilde{G}_n = \{k\cdot h_n,\ k \in \mathbb{Z}\}\cap[0,1]$ ein äquidistantes Gitter der Schrittweite h_n auf $[0,1]$.
Im folgenden wird gezeigt werden, dass die Konvergenzrate beider Schätzer optimal sind, falls Schrittweite und Bandbreite von derselben Ordnung gewählt werden, nämlich:

$$h_n \asymp b_n \asymp n^{-\frac{1}{2\rho+1}}.$$

Die Schätzung des Modalwertes im Falle von $\hat{\theta}_n^g$ erfolgt also, indem der entsprechende Kurvenschätzer über endlich viele Punkte maximiert wird, während bei $\hat{\theta}_n^c$ die Maximierung über dem Kontinuum erfolgt.

Hieraus ist ersichtlich, dass bei Verwendung des Schätzers $\hat{\theta}_n^g$ über weniger als n Punkte maximiert wird, da gilt $nh_n \to \infty$, während in Abraham et al. [1] die Maximierung über n Punkte erfolgt.

1.2.2 Obere Schranken für die Konvergenzrate

Um obere Schranken für die Konvergenzrate der beiden Modalwertschätzer $\hat{\theta}_n^c$ und $\hat{\theta}_n^g$ zu zeigen werden zunächst die Anforderungen an die Fehlervariablen genauer spezifiziert. Die Folge $\varepsilon_{i,n}$ sei definiert auf einem Wahrscheinlichkeitsraum $(\Omega_n, \mathcal{B}_n, \mathbb{P}_n)$. Für das jeweilige Wahrscheinlichkeitsmaß werde im Folgenden $\mathbb{P}$ geschrieben.
Die $\varepsilon_{i,n}$ bilden für festes n eine α-mischende Folge. Der Mischungskoeffizient sei hierbei definiert durch

$$\alpha_n(j) = \sup_{m \in \{1,\dots,n\}} \sup\{|\mathbb{P}(A \cap B) - \mathbb{P}(A)\mathbb{P}(B)| : A \in \mathcal{F}_1^m,\, B \in \mathcal{F}_{m+j}^n\}. \quad (1.24)$$

$\mathcal{F}_m^k$ (wiederum abhängig von n) bezeichne hierbei die von $(\varepsilon_{i,n} : m \leq i \leq k)$ erzeugte σ-Algebra.
Die Mischungskoeffizienten $\alpha_n(j)$ fallen hierbei polynomiell, gleichmäßig in n:

$$\alpha_n(j) \leq a j^{-\nu}, \text{ für } a > 0 \text{ und } \nu > 0\,.$$

Um Raten für das quadratische Risiko des jeweiligen Modalwertschätzers zu erhalten, wird ähnlich zur Vorgehensweise in Wieczorek [78] die Wahrscheinlichkeit begrenzt, dass der Modalwertschätzer auf dem Gitter $\widetilde{G}_n$ k Schritte von der wahren Modalstelle θ entfernt ist. Diese Wahrscheinlichkeit kann wiederum beschränkt werden aufgrund der Kontrasteigenschaft der Kurve m an der Modalstelle.
Während zur Behandlung des stochastischen Anteils in Wieczorek [78] eine Rosenthal-Ungleichung für unabhängige Zufallsvariable verwendet wurde, wird im Fall mischender Fehlervariablen eine Momentenungleichung vom Rosenthal-Typ herangezogen, welche in Yang [80] bewiesen wurde. Der α-Mischungskoeffizient ist hierbei ähnlich definiert wie in (1.24), wobei der zugrundeliegende Wahrscheinlichkeitsraum und somit auch der Mischungskoeffizient nicht von n abhängig ist.

Lemma 1.2.1 *Für $r > 2$, und $\delta > 0$ sei $\{\xi_i,\, i \geq 1\}$ eine α-mischende Folge von Zufallsvariablen mit $\mathbb{E}\xi_i = 0$ und $\mathbb{E}|\xi_i|^{r+\delta} < \infty$. Es gelte weiter*

$$\nu > r(r+\delta)/(2\delta)$$

und $\alpha(j) \leq aj^{-\nu}$ für ein $a > 0$. Dann existiert für beliebiges $\mu > 0$ eine positive Konstante $C = C(\mu, r, \delta, \nu, a) < \infty$, so dass

$$\mathbb{E}\big(\max_{1\leq j\leq n} |\sum_{i=1}^{j} \xi_i|^r\big) \leq C\Big\{n^\mu \sum_{i=1}^{n} \mathbb{E}|\xi_i|^r + \Big(\sum_{i=1}^{n} (\mathbb{E}|\xi_i|^{r+\delta})^{\frac{2}{r+\delta}}\Big)^{r/2}\Big\}.$$

Das Resultat überträgt sich unmittelbar auf Dreiecksschemata $\xi_{i,n}$, falls die Voraussetzungen des Lemmas gleichmäßig in n erfüllt sind, wie in (1.24) gefordert. Die Momentenungleichung gilt für jedes feste n für Folgen $\xi_{j,n}$. Die Konstante C gilt gleichmäßig in n, falls die Voraussetzungen gleichmäßig erfüllt sind.
Die Voraussetzungen von Lemma 1.2.1 sind beispielsweise für stationäre ARMA-Prozesse erfüllt, wenn die Innovationen Momente der Ordnung $r + \delta$ besitzen. Der Mischungskoeffizient fällt in diesem Fall exponentiell, ebenso wie in vielen anderen gängigen Modellen.
Die Anzahl der Momente, welche benötigt werden, hängt ab von der Gestalt der Modalstelle der Regressionskurve m, wie sich später zeigen wird.
Indem Lemma 1.2.1 auf die r-ten Momente des stochastischen Anteils der Abweichung des Kurvenschätzers von der tatsächlichen Kurve angewendet wird, erhält man folgendes Hilfsresultat, in Analogie zu den Betrachtungen im Dekonvolutionsmodell:

Lemma 1.2.2 *Es sei $\hat{m}_n$ ein Kurvenschätzer im Regressionsmodell* (1.14). *Falls die Gewichtsfunktionen die Bedingung* (1.17) *erfüllen und falls $\mathbb{E}(\varepsilon_{i,n}) = 0$ und $\mathbb{E}(|\varepsilon_{i,n}|^{r+\delta}) \leq M_{r+\delta} < \infty$, für $r > 2$, $\delta > 0$, gleichmäßig für alle i, n, ebenso wie $n^{-(1-\tilde{\mu})} = O(b_n)$, für ein $\tilde{\mu} > 0$, dann gilt*

$$\mathbb{E}(|\hat{m}_n(t) - \mathbb{E}\hat{m}_n(t)|^r) = O(nb_n)^{-r/2} \text{ gleichmäßig für alle } t.$$

Ähnlich wie in Wieczorek [78] werden mithilfe von Lemma 1.2.2 zunächst obere Schranken für den Modalwertschätzer $\hat{\theta}_n^g$ nachgewiesen, indem mithilfe von Lemma 1.2.2 jeweils die Wahrscheinlichkeiten beschränkt werden können, dass der Schätzer eine bestimmte Anzahl von Gitterpunkten vom tatsächlichen Modalwert entfernt ist. Man erhält also zunächst folgendes Resultat (siehe Wieczorek und Ziegler [79]):

Theorem 1.2.1 *Sei $\hat{m}_n(t) = \sum_{j=1}^{n} Y_{i,n} w_{i,n}(t)$ ein Kernschätzer im Regressionsmodell* (1.14)*. Für die Gewichtsfunktionen gelten die Voraussetzungen* (1.17) *und* (1.18)*. Die Regressionsfunktion m liege in der Klasse $\mathcal{F}_\rho$. Für die Fehlervariablen $\varepsilon_{i,n}$ gelten die Voraussetzungen von Lemma 1.2.2, mit $r > 3/\rho$. Dann gilt, mit $h_n \asymp b_n \asymp n^{-1/(2\rho+1)}$, dass*

$$\mathbb{E}(\hat{\theta}_n - \theta)^2 = O(n^{-\frac{2}{2\rho+1}}) \ .$$

Um die im soeben genannten Theorem erzielte Rate auch für den Argmax-Schätzer nachweisen zu können, welcher über das gesamte Intervall $[0,1]$ maximiert, müssen die Gewichtsfunktionen zusätzlich einer lipschitzähnlichen Bedingung genügen, ähnlich wie die Bedingung an die Momente des Kurvenschätzers in 1.1.1 und 1.1.2. Damit gilt folgendes (siehe Wieczorek und Ziegler [79])

Theorem 1.2.2 *Es gelten die Voraussetzungen des vorangegangenen Theorems sowie*

$$\sum_{i=1}^{n} |w_{i,n}(t_1) - w_{i,n}(t_2)|^r = O\Big(\Big(\frac{|t_1 - t_2|}{b_n}\Big)^r \frac{1}{(nb_n)^{r-1}}\Big) \ . \tag{1.25}$$

Dann folgt

$$\mathbb{E}(\hat{\theta}_n^c - \theta)^2 = O(n^{-\frac{2}{2\rho+1}}) \ .$$

Somit sind obere Schranken für die Konvergenzrate der beiden Modalwertschätzer (1.22) und (1.23) nachgewiesen. Nichtsdestoweniger hängt die Wahl der Bandbreite b_n vom unbekannten Parameter ρ ab. Was die datengesteuerte Wahl der Bandbreite anbelangt, so liegen hierzu Resultate vor für das Problem der Schätzung der Kurve im Fixed-Design-Regressionsmodell mit α-mischenden Fehlern,

so in Chu und Marron [17] und Hall et al. [34], wo die Methode der Leave-k-out-Kreuzvalidierung angewendet wurde. Es bliebe zu untersuchen, ob diese Methode auch im Fall der Modalwertschätzung angewendet werden kann.

1.2.3 Untere Schranken für die Konvergenzrate

In diesem Abschnitt wird gezeigt, dass die erzielten oberen Schranken für die Konvergenzraten der Modalwertschätzer $\hat{\theta}_n^g$ und $\hat{\theta}_n^c$ asymptotisch optimal sind. Dies wird ebenso wie im Fall der Dekonvolutionsmodelle gezeigt, indem eindimensionale Teilprobleme innerhalb der Funktionenklasse $\mathcal{F}_\rho$ betrachtet werden, wie in Donoho und Liu [21]. Für die hier betrachtete Fragestellung nehmen wir nun an, dass die Fehlervariablen unabhängig und standardnormalverteilt sind. Die hierfür in Wieczorek und Ziegler [79] im folgenden Theorem nachgewiesenen unteren Schranken gelten dann ebenso im allgemeineren Modell (1.14).

Theorem 1.2.3 *Gegeben sei das Regressionsmodell* (1.14)*, wobei die Fehlervariablen zeilenweise unabhängig und standardnormalverteilt seien. Es bezeichne $\hat{M}_n$ einen beliebigen Schätzer für den Modalwert θ der Regressionsfunktion, basierend auf den Beobachtungen $Y_1, ..., Y_n$. Dann gilt*

$$\sup_{m \in \mathcal{F}_\rho} \mathbb{E}_m\big((\hat{M}_n - \theta)^2\big) \geq C n^{-\frac{2}{2\rho+1}} .$$

Dieses Resultat verdeutlicht, dass auch in dieser Fragestellung für mischende Fehlervariablen asymptotisch dieselben optimalen Raten erzielt werden können wie unter Unabhängigkeit.

Um nun sicherzustellen, dass die oben genannten Resultate für die unter (1.19), (1.20) und (1.21) vorgestellten Kernschätzer gültig sind, wird mit dem folgenden Lemma gezeigt, unter welchen Bedingungen die hierfür notwendigen Voraussetzungen (1.17) und (1.18) erfüllt sind. Für den Schätzer $\hat{\theta}_n^c$ ist hierbei zusätzlich noch (1.25) zu zeigen.

Lemma 1.2.3 *Für die Designpunkte und die Kernfunktion gelte*

$$\begin{aligned}
&\max|x_{i,n} - x_{i-1,n}| = O(n^{-1}),\\
&s_{i+1,n} = \frac{x_{i+1,n} - x_{i,n}}{2},\ i = 1, ..., n-1,\ s_{0,n} = 0,\ s_{n,n} = 1,\\
&K(\cdot) \text{ sei stetig differenzierbar, } \operatorname{supp}(\mathrm{K}) \subseteq [-1, 1]\,.
\end{aligned}$$

Dann sind für den Gasser-Müller-Schätzer und den Priestley-Chao-Benedetti-Schätzer die Voraussetzungen (1.17), (1.18) *und* (1.25) *erfüllt.*
Falls zusätzlich eine lipschitzstetige Funktion g existiert, welche auf $[0, 1]$ *positiv ist, so dass*

$$\int_0^{x_{i,n}} g(u)du = \frac{i - \frac{1}{2}}{n} \text{ und } \int_0^{s_{i,n}} g(u)du = \frac{i}{n}$$

dann gelten (1.17), (1.18) *und* (1.25) *auch für den Nadaraya-Watson-Schätzer.*

1.2.4 Beweise für die Resultate im Fixed Design Modell

Beweis von Theorem 1.2.3

Die Beweisführung nutzt eine ähnliche Argumentation, wie sie im Beweis von Theorem 1.1.1 erfolgte. Während dort jedoch die Punkte des im Beweis verwendeten Gitters von θ abhängen, wird hier ein von θ unabhängiges Gitter auf $[0, 1]$ definiert, welches ebenfalls mit $\widetilde{G}_n$ bezeichnet wird, siehe Definition von (1.23):

$$\widetilde{G}_n = \{k \cdot h_n,\ k \in \mathbb{Z}\} \cap [0, 1]\}.$$

Zur Bezeichnung der Gitterpunkte wird weiter definiert:

$$\theta_{n,0} := \min\{t | t = \operatorname{argmin}_{s \in \widetilde{G}_n} |s - \theta|\} \text{ sowie } \theta_{n,k} := \theta_{n,0} + k \cdot h_n\,.$$

Des weiteren gelte $I := [0, 1]$ und $I := I_1 \cup I_2$, wobei I_1 eine kompakte Umgebung von θ bezeichne, in welcher die Kontrastbedingung (1.15) erfüllt ist. Weiter wird definiert

$$\begin{aligned}
&G_n := \{k \in \mathbb{Z} :\ \theta_{n,k} \in I\},\ G_n^1 := \{k \in \mathbb{Z} :\ \theta_{n,k} \in I_1\}\\
&G_n^2 := G_n \setminus G_n^1 = \{k \in \mathbb{Z} :\ \theta_{n,k} \in I_2\}\,.
\end{aligned}$$

Dann gilt

$$\begin{aligned}
\mathbb{E}[(\theta - \hat{\theta}_n^g)^2] &= \sum_{k \in G_n} (\theta - \hat{\theta}_n^g)^2 \cdot \mathbb{P}(\hat{\theta}_n^g = \theta_{n,k}) \\
&\leq 6h_n^2 + \sum_{\substack{k \in G_n \\ k \neq -1,0,1}} (\theta - \theta_{n,k})^2 \cdot \mathbb{P}(\hat{\theta}_n^g = \theta_{n,k}) \\
&\leq 6h_n^2 + \sum_{\substack{k \in G_n \\ k \neq -1,0,1}} (|k|+1)^2 \cdot \mathbb{P}(\hat{\theta}_n^g = \theta_{n,k}).
\end{aligned}$$

Weiter gilt nun für $k \in G_n^1$:

$$\begin{aligned}
\mathbb{P}(\hat{\theta}_n^g = \theta_{n,k}) &\leq \mathbb{P}(|\hat{m}_n(\theta_{n,k}) - \mathbb{E}[\hat{m}_n(\theta_{n,k})]|) \\
&\leq \mathbb{P}(|\hat{m}_n(\theta_{n,k}) - \mathbb{E}[\hat{m}_n(\theta_{n,k})]| \geq \tilde{c}_1 |kh_n|^\rho) \\
&\quad + \mathbb{P}(|\mathbb{E}[\hat{m}_n(\theta_{n,k})] - m(\theta_{n,k})| \geq \tilde{c}_1 |kh_n|^\rho) \\
&\quad + \mathbb{P}(|\mathbb{E}[\hat{m}_n(\theta)] - \hat{m}_n(\theta)| \geq \tilde{c}_1 |kh_n|^\rho) \\
&\quad + \mathbb{P}(|m(\theta) - \mathbb{E}[\hat{m}_n(\theta)]| \geq \tilde{c}_1 |kh_n|^\rho) \\
&\leq 2 \sup_{t \in [0,1]} \mathbb{P}(|\hat{m}_n(t) - \mathbb{E}\hat{m}_n(t)| \geq \tilde{c}_1 |kh_n|^\rho) \\
&\quad + 2 \sup_{t \in [0,1]} \mathbb{P}(|m(t) - \mathbb{E}[\hat{m}_n(t)]| \geq \tilde{c}_1 |kh_n|^\rho) \\
&=: 2 \cdot (\mathbb{P}_{1,k} + \mathbb{P}_{2,k}).
\end{aligned}$$

Wie im Beweis von Theorem 1.1.1 wird $\mathbb{P}_{1,k}$ mithilfe der Markov-Ungleichung abgeschätzt und mit Lemma 1.2.1 und aufgrund von $r > 3/\rho > 2$ gilt

$$\mathbb{P}_{1,k} \leq \sup_{t \in [0,1]} \frac{\mathbb{E}(|\sum_{i=1}^n \varepsilon_{i,n} w_{i,n}(t)|^r)}{(\tilde{c}_1 |kh_n|)^{r \cdot \rho}} = |k|^{-r\rho} \cdot O(1),$$

was gleichmäßig in t gilt, falls $h_n \asymp b_n \asymp n^{-1/(2\rho+1)}$.
Der in $\mathbb{P}_{2,k}$ auftretende Biasterm ist von der Ordnung $O(b_n^\rho)$ und daher ergibt sich

$$\mathbb{P}_{2,k} \leq \mathbb{P}\big(O(b_n^\rho \cdot h_n^{-\rho}) \geq |k|^\rho\big).$$

Für $k \in G_n^2$ ergibt sich

$$\mathbb{P}(\hat{\theta}_n^g = \theta_{n,k}) = O((nb_n)^{-r/2} \cdot \delta^{-r}).$$

$|k|$ ist beschränkt durch $1/h_n$, und somit ergibt sich insgesamt

$$\begin{aligned}
\mathbb{E}\big[(\theta-\hat{\theta}_n)^2\big] &\leq 2h_n^2\Big(3+\sum_{|k|\leq k_0} k^2 + 2\sum_{k\in G_n^1} k^{2-r\rho}C_1 + 2\sum_{k=1}^{\lfloor \frac{1}{h_n}\rfloor} k^2 C_2 h_n^r \frac{(nb_n)^{-r/2}}{\delta^r h_n^r}\Big) \\
&\leq 2h_n^2\Big(3+C_3+\big(\sum_{k\in G_n^1} k^{2-r\rho}C_1\big) + h_n^{r-3}C_2\frac{(nb_n)^{-r/2}}{\delta^r h_n^r}\Big) \\
&= 2h_n^2\Big(3+C_3+\big(\sum_{k\in G_n^1} k^{2-r\rho}\tilde{C}_1\big) + h_n^{r-3}\tilde{C}_2\Big) \\
&= O(n^{-2/(2\rho+1)}),
\end{aligned}$$

mit $h_n \asymp b_n$ und $b_n \asymp n^{-\frac{1}{2\rho+1}}$.

Beweis von Theorem 1.2.2

Das quadratische Risiko kann hier beschränkt werden durch

$$\mathbb{E}\big((\theta-\hat{\theta}_n^c)^2\big) \leq 6\cdot h_n^2 + 2\sum_{\substack{k\in G_n\\ k\neq -1,0,1}} (k\cdot h_n)^2 \cdot \mathbb{P}(\hat{\theta}_n^c \in [\theta_{n,k-1},\theta_{n,k}])\,.$$

Für k in G_n^1 kann die Wahrscheinlichkeit beschränkt werden durch

$$\begin{aligned}
&\mathbb{P}\big(\sup_{t\in[\theta_{n,k-1},\theta_{n,k}]} |\hat{m}_n(t)-\mathbb{E}\hat{m}_n(t)-(\hat{m}_n(\theta)-\mathbb{E}\hat{m}_n(\theta))| > c\cdot|k|^\rho h_n^\rho\big) \\
+\ &\mathbb{P}\big(\sup_{t\in[\theta_{n,k-1},\theta_{n,k}]} |m(t)-\mathbb{E}\hat{m}_n(t)-(m(\theta)-\mathbb{E}\hat{m}_n(\theta))| > c\cdot|k|^\rho h_n^\rho\big)\,.
\end{aligned}$$

Aufgrund der gleichmäßig gültigen Hölder-Bedingung für die Kurve m ist der Bias von derselben Ordnung wie im Beweis von Theorem 1.2.1.
Was den stochastischen Teil anbelangt, gilt folgendes Lemma, was wie im Dekonvolutionsmodell mittels Chaining gezeigt wird, unter Ausnutzung der Bedingungen an die Glattheit des Schätzers und der Kurve selbst.

Lemma 1.2.4 *Unter den Voraussetzungen von Theorem 1.2.1 gilt, für $k \neq 0,1$*

$$\begin{aligned}
&\mathbb{P}\big(\sup_{t\in[\theta_{n,k-1},\theta_{n,k}]} |\hat{m}_n(t)-\mathbb{E}\hat{m}_n(t)-(\hat{m}_n(\theta)-\mathbb{E}\hat{m}_n(\theta))| > c\cdot|k|^\rho h_n^\rho\big) \\
=\ &|k|^{-r\rho} O\big(\frac{(nb_n)^{-r/2}}{h_n^{\rho r}}\big)\,.
\end{aligned}$$

Ebenso wie im Beweis von Lemma 1.2.4 wird gezeigt

$$P\big(\sup_{t\in[\theta_{n,k-1},\theta_{n,k}]}|\hat{m}_n(t)-\mathbb{E}\hat{m}_n(t)-(\hat{m}_n(\theta)-\mathbb{E}\hat{m}_n(\theta))|>\frac{1}{4}\delta\big)=O\big(\frac{(nb_n)^{-r/2}}{\delta^r}\big)$$

für k in G_n^2.
Die restlichen Überlegungen erfolgen analog zum Beweis von Lemma 1.2.1. □

Beweis von Theorem 1.2.3

Es wird nun Modell (1.14) betrachtet, für den Spezialfall standardnormalverteilter, unabhängiger Fehlervariablen. Die unteren Schranken, die in diesem Fall erzielt werden, gelten dann ebenso im allgemeinen Modell (1.14).
Ähnlich zu Romano [67], wo untere Schranken für Dichten, welche an der Modalstelle differenzierbar sind, nachgewiesen wurden, wird das eindimensionale Teilproblem folgendermaßen konstruiert: Als Ausgangsfunktion diene folgende:

$$m_1(x)=-|x|^\rho \text{ mit der Modalstelle } \theta_1=0$$

und weiter

$$m_{2,n}(x)=m_1(x)+\gamma_n(x)=m_1(x)+b_n^\rho\cdot\gamma(\frac{x}{b_n}).$$

γ sei hierbei hölderstetig der Ordnung ρ und besitze kompakten Träger.
Darüber hinaus gelte $\int\gamma(x)dx=0$ und

$$\gamma(x)=|x|^\rho-|x+1|^\rho$$

in einer Umgebung von 0.
In einer Umgebung von 0 gilt demzufolge

$$m_{2,n}(x)=m_1(0)-|x+b_n|^\rho,$$

also wird der Modalwert $\theta_{n,2}$ von $m_{2,n}$ von Null weg verschoben um $-b_n$. Es gilt also für die Modalwerte von m_1 and $m_{2,n}$:

$$|\theta_1-\theta_{2,n}|\geq b_n\,. \tag{1.26}$$

Die Beobachtungsvektoren $(Y_{1,n},...,Y_{n,n})$ besitzen in beiden Fällen die Dichten

$$h_1^{Y_{1,n},...,Y_{n,n}}(\underline{y}):=h_1^{Y_{1,n},...,Y_{n,n}}(y_1,...,y_n)=\prod_{i=1}^n\frac{1}{\sqrt{2\pi}}\exp\big(-\frac{1}{2}(y_i-m_1(x_{i,n}))^2\big)$$

bzw.

$$h_{2,n}^{Y_{1,n},\dots,Y_{n,n}}(\underline{y}) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi}} \exp\big(-\frac{1}{2}(y_i - m_1(x_{i,n}) - \gamma_n(x_{i,n}))^2\big).$$

Für gegebene Regressionsfunktion m sind also die Beobachtungen Y_i verteilt wie die Fehlervariablen, wobei der Erwartungswert um den Wert $m(x_{i,n})$ verschoben ist. Es werde mit $\mathbb{E}_m$ der Erwartungswert bezeichnet, welcher dem Modell entspricht, in welchem m die zugrundeliegende Regressionsfunktion ist, und mit θ_m der entsprechende Modalwert.

Es bezeichne weiterhin $\hat{M}_n := \hat{M}(Y_{1,n}, ..., Y_{n,n})$ einen beliebigen Schätzer für den Modalwert θ, welcher auf den Beobachtungen $Y_{1,n}, ..., Y_{n,n}$ basiert. Dann besteht der folgende Zusammenhang zwischen Hellinger-Affinität der Dichten, welche die oben beschriebenen Situationen generieren, und dem quadratischen Risiko von $\hat{M}_n$:

Indem wiederum $\lambda^{[n]}$ das n-dimensionale Lebesgue-Maß bezeichne, gilt unter Beachtung von (1.26)

$$\begin{aligned}
& \rho(h_1^{Y_{1,n},\dots,Y_{n,n}}, h_{2,n}^{Y_{1,n},\dots,Y_{n,n}}) \\
= & \int \sqrt{h_1^{Y_{1,n},\dots,Y_{n,n}}(\underline{y})}\sqrt{h_{2,n}^{Y_{1,n},\dots,Y_{n,n}}(\underline{y})}\lambda^{[n]}(d\underline{y}) \\
\leq & \int \frac{|\hat{M}_n - \theta_1|}{b_n}\sqrt{h_1^{Y_{1,n},\dots,Y_{n,n}}(\underline{y}) \wedge h_{2,n}^{Y_{1,n},\dots,Y_{n,n}}(\underline{y})} \\
& \qquad \cdot\sqrt{h_1^{Y_{1,n},\dots,Y_{n,n}}(\underline{y}) \vee h_{2,n}^{Y_{1,n},\dots,Y_{n,n}}(\underline{y})}\lambda^{[n]}(d\underline{y}) \\
& + \int \frac{|\hat{M}_n - \theta_{2,n}|}{b_n}\sqrt{h_1^{Y_{1,n},\dots,Y_{n,n}}(\underline{y}) \wedge h_2^{Y_1,\dots,Y_n}(\underline{y})} \\
& \qquad \cdot\sqrt{h_1^{Y_{1,n},\dots,Y_{n,n}}(\underline{y}) \vee h_2^{Y_{1,n},\dots,Y_{n,n}}(\underline{y})}\lambda^{[n]}(d\underline{y}) \\
& \text{und mithilfe der Cauchy-Schwarz-Ungleichung} \\
\leq & \ 4 \max_{m\in\{m_1,m_{2,n}\}} \sqrt{\mathbb{E}_m\big(\frac{\hat{M}_n - \theta_m}{b_n}\big)^2} \leq 4 \sup_{m\in\mathcal{F}_\rho} \sqrt{\mathbb{E}_m\big(\frac{\hat{M}_n - \theta_m}{b_n}\big)^2}.
\end{aligned}$$

Es ist also wiederum nachzuweisen, dass die Hellinger-Affinität der beiden Dichten von Null weg beschränkt ist. Dann ergibt sich die gewünschte untere Schranke, falls $b_n \asymp n^{-1/(2\rho+1)}$ gewählt werden kann.

Die Hellinger-Affinität der beiden Dichten der jeweiligen Beobachtungsvektoren stellt sich nun folgendermaßen dar, wobei mit $H^2(\cdot,\cdot)$ den quadrierten Hellinger-Abstand bezeichnet und J_{θ_1} die Anzahl von Indizes, für welche $g(x_{i,n}/b_n)$ nicht verschwindet:

$$
\begin{aligned}
& \rho(h_1^{Y_{1,n},\dots,Y_{n,n}}, h_{2,n}^{Y_{1,n},\dots,Y_{n,n}}) \\
= & \int \sqrt{h_1^{Y_{1,n},\dots,Y_{n,n}}(\underline{y})}\sqrt{h_{2,n}^{Y_{1,n},\dots,Y_{n,n}}(\underline{y})}\lambda^{[n]}(d\underline{y}) \\
= & \prod_{i\in J_{\theta_1}} \int \sqrt{h_1^{Y_{i,n}}(y_i)}\sqrt{h_{2,n}^{Y_{i,n}}(y_i)}\lambda(dy_i) = \prod_{i\in J_{\theta_1}} (1-\frac{1}{2}H^2(h_1^{Y_{i,n}}, h_{2,n}^{Y_{i,n}}))\,.
\end{aligned}
$$

Die Anzahl der Faktoren im letztgenannten Produkt ist von der Ordnung $O(nb_n)$. Der Ausdruck ist also von Null weg beschränkt, falls der quadrierte Hellinger-Abstand von der Ordnung $O(1/(nb_n))$ ist.
Für den quadrierten Hellinger-Abstand gilt

$$
\begin{aligned}
& H^2(h_1^{Y_{i,n}},\, h_{2,n}^{Y_{i,n}}) \\
= & \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} \Big(\exp\big(-\frac{1}{4}(y-m_1(x_{i,n})-\gamma_n(x_{i,n}))^2\big) - \exp\big(-\frac{1}{4}(y-m_1(x_{i,n}))^2\big)\Big)^2 dy \\
= & \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} \big(\frac{1}{4}|\gamma_n(x_{i,n})|\big)^2\big(\frac{1}{2}(y^*-m_1(x_{i,n}))\big)^2\Big(\exp\big(-\frac{1}{4}(y^*-m_1(x_{i,n}))^2\big)\Big)^2 dy \\
& \qquad (y^* \text{ zwischen } y \text{ und } y-\gamma_n(x_{i,n})) \\
\leq & \; b_n^{2\rho}\cdot C = O(\frac{1}{nb_n}),
\end{aligned}
$$

falls $b_n \asymp n^{-\frac{1}{2\rho+1}}$, wobei die Konstante C gleichmäßig in i und n gilt. Damit ist die Aussage des Theorems gezeigt. □

1.2.5 Beweise der Hilfsresultate

Beweis von Lemma 1.2.2

Mit den Voraussetzungen an das Regressionsmodell (1.14) gilt

$$
\mathbb{E}(|\hat{m}_n(t)-\mathbb{E}\hat{m}_n(t)|^r) = \mathbb{E}(|\sum_{i=1}^n \varepsilon_{i,n} w_{i,n}(t)|^r)
$$

Bezeichne C_r eine gleichmäßig gültige, endliche obere Schranke für $\mathbb{E}(|\varepsilon_{i,n}|^r)$. Dann gilt mit Lemma 1.2.1

$$
\leq \; C\Big\{ n^\mu \sum_{i=1}^n \mathbb{E}(|\varepsilon_{i,n}|^r)|w_{i,n}(t)|^r + \Big(\sum_{i=1}^n (\mathbb{E}(|\varepsilon_{i,n} w_{i,n}(t)|^{r+\delta})^{\frac{2}{r+\delta}}\Big)^{r/2}\Big\}
$$

$$\begin{aligned}
&\leq C\Big\{n^{\mu}C_r\sum_{i=1}^{n}|w_{i,n}(t)|^r+\Big(\big(C_{r+\delta}\big)^{\frac{2}{r+\delta}}\sum_{i=1}^{n}|w_{i,n}(t)|^2\Big)^{r/2}\Big\}\\
&\leq O(\frac{1}{(nb_n)^{r-1}}n^{\mu})+O(\frac{1}{(nb_n)^{r/2}})=O(\frac{1}{(nb_n)^{r/2}}),
\end{aligned}$$

wegen $r>2$ und aufgrund der für b_n gültigen Annahme, da μ beliebig klein gewählt werden kann (wobei die Konstante C von μ abhängig ist). □

Beweis von Lemma 1.2.3

Aufgrund der Beschränktheit der Kernfunktion K und der Wahl der Designpunkte gilt für den Gasser-Müller-Schätzer (1.19)

$$w_{i,n}(t)=\frac{1}{b_n}\int\limits_{s_{i-1,n}}^{s_{i,n}}K\big(\frac{u-t}{b_n}\big)\,du=O((nb_n)^{-1}),$$

was gleichmäßig für alle t gilt. Mit (1.26) und aufgrund der Tatsache, dass K kompakten Träger besitzt, verschwindet $w_{i,n}(\cdot)$ nur für $i\in J_t$ nicht , wobei $|J_t|=O(nb_n)$. Daher gilt

$$\sum_{i=1}^{n}\big(\frac{1}{b_n}\int\limits_{s_{i-1,n}}^{s_{i,n}}K\big(\frac{u-t}{b_n}\big)\,du\big)^r=\sum_{j\in J_t}\big(\frac{1}{b_n}\int\limits_{s_{i-1,n}}^{s_{i,n}}K\big(\frac{u-t}{b_n}\big)\,du\big)^r=O(\frac{1}{(nb_n)^{r-1}})$$

gleichmäßig in t.
Analog gilt für den Priestley-Chao-Schätzer (1.20):

$$\begin{aligned}
&\sum_{i=1}^{n}((x_{i,n}-x_{i-1,n})\cdot\frac{1}{b_n}K\big(\frac{x_{i,n}-t}{b_n}\big))^r\\
=&\sum_{i\in J_t}((x_{i,n}-x_{i-1,n})\cdot\frac{1}{b_n}K\big(\frac{x_{i,n}-t}{b_n}\big))^r\\
=&\ O(\frac{1}{(nb_n)^{r-1}})\,.
\end{aligned}$$

Die Bedingung (1.25) folgt aufgrund der gleichmäßigen Lipschitzstetigkeit von K.

Für den Nadaraya-Watson-Schätzer (1.21) gilt

$$\Big|\frac{1}{nb_n}\sum_{i=1}^{n}\big(K((\frac{t-x_{i,n}}{b_n})\big)^r-\int\limits_0^1\frac{1}{b_n}\big(K\big(\frac{t-z}{b_n}\big)\big)^r g(t)dz\Big|$$

$$
\begin{aligned}
&\leq \sum_{i=1}^{n} \int_{s_{i-1,n}}^{s_{i,n}} \frac{1}{b_n} \Big| \big(K(\big(\frac{t-x_{i,n}}{b_n}\big)\big)^r g(\xi_{i,n}) - \big(K\big(\frac{t-z}{b_n}\big)\big)^r g(\xi_{i,n}) \Big| \\
&\qquad + \Big| \big(K\big(\frac{t-z}{b_n}\big)\big)^r g(\xi_{i,n}) - \big(K\big(\frac{t-z}{b_n}\big)\big)^r g(z) \Big| dz \\
&= \sum_{i \in J_t} \big(O(\frac{1}{(nb_n)^2}) + O(\frac{1}{n^2 b_n})\big) = O(\frac{1}{nb_n})\,.
\end{aligned}
$$

Weiter gilt

$$
\begin{aligned}
&\big(\frac{1}{nb_n} \sum_{i=1}^{n} K\big(\frac{t-x_{i,n}}{b_n}\big)\big)^r = \big(\int_0^1 \frac{1}{b_n} K\big(\frac{t-z}{b_n}\big) g(z) dz + O(\frac{1}{nb_n})\big)^r \\
= \; &\big(\int_0^1 \frac{1}{b_n} K\big(\frac{t-z}{b_n}\big) g(z) dz\big)^r + O(\frac{1}{nb_n})
\end{aligned}
$$

und somit erhält man

$$
\begin{aligned}
\sum_{i=1}^{n} (w_{i,n}(t))^r &= \frac{(\frac{1}{nb_n})^r \sum_{i=1}^{n} \big(K\big(\frac{t-x_{i,n}}{b_n}\big)\big)^r}{\big(\frac{1}{nb_n} \sum_{j=1}^{n} K\big(\frac{t-x_{j,n}}{b_n}\big)\big)^r} \\
&= \frac{1}{(nb_n)^{r-1}} \cdot \frac{\int_0^1 \frac{1}{b_n} \big(K\big(\frac{t-z}{b_n}\big)\big)^r g(z) dz + O(\frac{1}{nb_n})}{\big(\int_0^1 \frac{1}{b_n} K\big(\frac{t-z}{b_n}\big) g(z) dz\big)^r + O(\frac{1}{nb_n})} = O(\frac{1}{(nb_n)^{r-1}})
\end{aligned}
$$

gleichmäßig für $t \in [0,1]$.
Was (1.25) anbelangt, so wird die Aussage nachgewiesen, indem der Ausdruck auf der linken Seite von (1.25) auf einen gemeinsamen Hauptnenner gebracht wird. Die einzelnen Faktoren werden durch ihre jeweiligen, soeben gezeigten Approximationen ersetzt, und die Aussage folgt dann aufgrund der Differenzierbarkeit der Kernfunktion.

Um die jeweilige Ordnung des Bias nachzuweisen, wird zunächst für (1.19) gezeigt

$$
\Big| \sum_{i=1}^{n} m(x_{i,n}) \frac{1}{b_n} \int_{s_{i-1,n}}^{s_{i,n}} K\big(\frac{u-t}{b_n}\big)\, du - \frac{1}{b_n} \int_0^1 f(u) K\big(\frac{u-t}{b_n}\big) du \Big|
$$

$$\begin{aligned}
&\leq \sum_{i=1}^{n} \frac{1}{b_n} \int_{s_{i-1,n}}^{s_{i,n}} |m(x_{i,n}) - m(u)| K\big(\frac{u-t}{b_n}\big) du \\
&\leq \sum_{i=1}^{n} \frac{1}{n^\rho} \frac{1}{b_n} \int_{s_{i-1,n}}^{s_{i,n}} K\big(\frac{u-t}{b_n}\big) du = O(\frac{1}{n^\rho}) = O(b_n^\rho)\,.
\end{aligned}$$

Also gilt für den Bias

$$\begin{aligned}
\big|\mathbb{E}\hat{m}_n(t) - m(t)\big| &\leq \frac{1}{b_n} \int_0^1 |m(u) - m(t)| K\big(\frac{t-u}{b_n}\big) du + O(\frac{1}{n^\rho}) \\
&\leq \int |m(t - zb_n) - m(t)| K(z) dz + O(\frac{1}{n^\rho}) \\
&\leq \int z^\rho b_n^\rho K(z) dz + O(b_n^\rho) = O(b_n^\rho)
\end{aligned}$$

gleichmäßig in t, aufgrund der Hölder-Stetigkeit von m.
Dieselbe Integralapproximation gilt für (1.20):

$$\begin{aligned}
&\Big|\frac{1}{b_n} \sum_{i=1}^{n} (x_{i,n} - x_{i-1,n}) m(x_{i,n}) K\big(\frac{t - x_{i,n}}{b_n}\big) - \frac{1}{b_n} \int_0^1 m(t) K\big(\frac{u-t}{b_n}\big) du\Big| \\
\leq\ & \sum_{i=1}^{n} \frac{1}{b_n} \int_{s_{i-1,n}}^{s_{i,n}} \big|m(x_{i,n})\big| \big|K\big(\frac{t - x_{i,n}}{b_n}\big) - K\big(\frac{u-t}{b_n}\big)\big| + \big|m(x_{i,n}) - m(u)\big| \big|K\big(\frac{t-u}{b_n}\big)\big| du \\
\leq\ & \sum_{i \in J_t} \big((\frac{1}{nb_n})^2 + \frac{1}{n^\rho b_n}\big) = O(b_n^\rho)\,.
\end{aligned}$$

Die Rechnungen sind somit analog wie im Falle von (1.19).
Hinsichtlich des Nadaraya-Watson-Schätzers (1.21) gilt, dass die Gewichte sich zu eins aufsummieren, und daher gilt für den Bias

$$bias(\hat{m}_n^{NW}(t)) = \mathbb{E}\hat{m}_n^{NW}(t) - m(t) = \frac{\sum_{i=1}^{n} (m(x_{i,n}) - m(t)) K(\frac{x_{i,n}-t}{b_n})}{\sum_{i=1}^{n} K(\frac{x_{i,n}-t}{b_n})}\,.$$

Da nur Terme mit $|x_{i,n} - t| \leq b_n$ zur Summe beitragen, gilt unmittelbar

$$bias(\hat{m}_n^{NW}(t)) = O(b_n^\rho)$$

falls m hölderstetig der Ordnung ρ ist. □

Beweis von Lemma 1.2.4

Der Beweis erfolgt nahezu analog zum Beweis von Lemma 1.1.6.
Auch hier wird angenommen, dass $k > 0$ gilt. Es gelten die Bezeichnungen des Beweises von Lemma 1.1.6, mit $Z_n(t) := \hat{m}_n(t) - \mathbb{E}\hat{m}_n(t)$.

$$\begin{aligned}
&\mathbb{P}(\max_{t\in T_{n,k}} |Z_n(t) - Z_n(\theta)| \geq c\cdot(k\cdot h_n)^\rho) \\
\leq\ & \mathbb{P}(\max_{t\in T_{n,k}} |Z_n(t) - Z_n(\theta)| \geq \eta_0 + \eta_1 + ... + \eta_j) \text{ with } \sum_{l=0}^{j} \eta_l \leq c\cdot(kh_n)^\rho
\end{aligned}$$

$$\leq \sum_{t\in D_{n,k,0}} \mathbb{P}(|Z_n(t) - Z_n(\theta)| \geq \eta_0) + \sum_{l=1}^{\infty}\sum_{t\in D_{n,k,l}} \mathbb{P}(|Z_n(t) - Z_n(t_{l-1}(t))| > \eta_l)\,. \tag{1.27}$$

Wähle hier η_l, so dass $\eta_l := \tilde{c}(kh_n)^\rho 2^{-\frac{\gamma l}{\rho r}}$. Dann gilt

$$\sum_{l=0}^{j} \eta_l \overset{j\to\infty}{\to} \tilde{c}(kh_n)^2 \frac{1}{1-2^{-\frac{\gamma}{\rho r}}}\,.$$

Dann folgt für $\tilde{c} \leq c\cdot(1-2^{-\frac{\gamma}{\rho r}})$, dass $\sum \eta_l \leq c\cdot(kh_n)^\rho$.

Für $t \in D_{n,k,0}$ gilt mit Lemma 1.2.3, für $r \geq 2$

$$\mathbb{P}(|Z_n(t) - Z_n(\theta)| \geq \eta_0) \leq \frac{\mathbb{E}\big(|Z_n(t) - Z_n(\theta)|^r\big)}{\eta_0^r} \leq C\frac{(nb_n)^{-r/2}}{\eta_0^r}$$

$$\leq C\tilde{c}^{-r}|k|^{-\rho r}\frac{(nb_n)^{-r/2}}{h_n^{\rho r}}.$$

Für $t \in D_{n,k,l}$, mit $|t - t_{l-1}(t)| \leq 2^{-l}h_n$ gilt:

$$\begin{aligned}
&\mathbb{P}(|Z_n(t) - Z_n(t_{l-1}(t))| \geq \eta_l) \\
\leq\ & \frac{\mathbb{E}\big(|Z_n(t) - Z_n(t_{l-1}(t))|^r\big)}{\eta_l^r} \\
\leq\ & C\frac{\left|\frac{2^{-l}h_n}{b_n}\right|^{\rho r}(nb_n)^{-r/2}}{\eta_l^r} = \qquad \text{und mit } h_n \asymp b_n \\
=\ & 2^{(\frac{\gamma}{\rho}-r)\,l}|k|^{-\rho r}\tilde{C}\frac{(nb_n)^{-r/2}}{h_n^{\rho r}}\,.
\end{aligned}$$

Also gilt insgesamt, für hinreichend kleines γ, welches von der Anzahl der benötigten Momente abhängt

$$\begin{aligned}
& \mathbb{P}(\max_{t\in T_{n,k}} |Z_n(t) - Z_n(\theta)| \geq c \cdot (k \cdot h_n)^\rho) \\
\leq\ & 2\mathbb{P}(|Z_n(t) - Z_n(\theta)| > \eta_0) + \sum_{l=1}^{\infty} (2^l + 1)\mathbb{P}(|Z_n(t) - Z_n(t_{l-1}(t))| > \eta_l) \\
=\ & O\big(|k|^{-\rho r} \frac{(nb_n)^{-r/2}}{h_n^{\rho r}}\big) \frac{1}{1 - 2^{1-r+\frac{\gamma}{\rho}}} =: |k|^{-\rho r} O\big(\frac{(nb_n)^{-r/2}}{h_n^{\rho r}}\big)\,.
\end{aligned}$$

$\square$

2 Blockwise Bootstrap für den geschätzten empirischen Prozess bei schwach abhängigen Beobachtungen

2.1 Empirische Prozesse mit geschätztem Parameter und Goodness-of-Fit

Bei der Analyse von Zeitreihendaten ist es oft von Interesse zu überprüfen, ob den beobachteten Daten eine bestimmte stationäre Verteilung zugrundeliegt. Dies kann mittels Hypothesentests erfolgen, bei denen in einem bestimmten Sinn die empirische Verteilungsfunktion der Beobachtungen $X_1, ..., X_n$ und die theoretisch angenommene verglichen werden.
Oftmals wird hierzu aber nicht eine konkrete Verteilungsfunktion F spezifiziert, vielmehr ist die Hypothese von Interesse, dass F einer bestimmten Familie $\mathcal{F}$ von Verteilungsfunktionen angehört.
Um dennoch die empirische Verteilungsfunktion der Beobachtungen mit einer geeigneten Funktion vergleichen zu können, werden Komponenten, die in der Beschreibung der vermuteten Familie auftreten, anhand der Beobachtungen geschätzt und die so gewonnene Verteilungsfunktion für den Vergleich herangezogen.
Im folgenden wird angenommen, dass es sich bei $\mathcal{F}$ um eine parametrische Familie von Verteilungsfunktionen handelt, wobei der Parameter θ $p-$dimensional sei.

Die Nullhypothese in diesem Fall lautet folgendermaßen:

$$\mathcal{H}_0 : F \in \mathcal{F} := \{F_\theta : \theta \in \Theta \subseteq \mathbb{R}^p\}.$$

Die tatsächliche Verteilungsfunktion unter $\mathcal{H}_0$ wird im weiteren Verlauf mit F_θ bezeichnet und deren Schätzung erfolgt nun über die Schätzung des Parameters θ. Der Schätzer für θ werde mit $\hat{\theta}_n$ bezeichnet.
Der sogenannte geschätzte empirische Prozess wird nun folgendermaßen definiert:

Definition 2.1.1 *Der auf einem Wahrscheinlichkeitsraum $(\Omega, \mathcal{A}, \mathbb{P})$ definierte stochastische Prozess $(W_n(x))_{x\in\bar{\mathbb{R}}}$ mit*

$$W_n(x) = \sqrt{n}\Big(\frac{1}{n}\sum_{t=1}^{n} \mathbb{1}_{\{X_t \leq x\}} - F_{\hat{\theta}_n}(x)\Big)$$

heißt geschätzter empirischer Prozess.

$W_n(x)$ wird als ein Element des Raumes $D(\bar{\mathbb{R}})$, den cadlag-Funktionen auf $\bar{\mathbb{R}} := [-\infty, \infty]$ betrachtet, ausgestattet mit der Supremumsnorm und der σ-Algebra $\mathcal{P}$, welche durch die endlichdimensionalen Projektionen $\pi_{S_n} : D \to \mathbb{R}^n$ erzeugt wird.

Um zu überprüfen, ob den beobachteten Daten tatsächlich eine Verteilung aus der angenommenen Familie zugrundeliegt, werden häufig Teststatistiken angewendet, die auf dem geschätzten empirischen Prozess beruhen. Diese sind messbare Abbildungen von $(D(\bar{\mathbb{R}}), \mathcal{P})$ nach $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$.
Konvergiert der Prozess W_n für $n \to \infty$ in Verteilung gegen einen Prozess mit Pfaden in einem separablen Teilraum von D, und sind die Teststatistiken stetige Funktionale, so kann aus der Verteilungskonvergenz von W_n mit Hilfe des Continuous Mapping Theorems die Verteilungskonvergenz der verwendeten Teststatistiken abgeleitet werden. Somit können bei bekannter Grenzverteilung asymptotisch gültige kritische Werte für die entsprechenden Teststatistiken konstruiert werden.
Häufig verwendete Tests hierbei sind solche, welche auf der Supremumsabweichung zwischen empirischer und geschätzter Verteilungsfunktion beruhen, etwa der Kolmogorov-Smirnov-Test, sowie Tests vom Cramér-von-Mises-Typ, die eine

gewichtete L^2-Abweichung unter $\mathcal{H}_0$ benutzen. Das Grenzverhalten von Teststatistiken, welche auf dem geschätzten empirischen Prozess beruhen, wurde erstmals von Darling [18] und von Kac, Kiefer und Wolfowitz [42] betrachtet. Eine Übersicht über gängige Teststatistiken, welche auf empirischen Prozessen beruhen, findet sich in D'Agostino und Stephens [19].
Das Grenzverhalten des empirischen Prozesses mit geschätztem Parameter selbst wurde erstmals von Durbin [23] untersucht, wobei dort die Verteilungskonvergenz gegen einen zentrierten Gaußprozess nachgewiesen wurde.
Bei Vorliegen stark mischender Daten wurde in Rao und Krishnaiah [66] ebenfalls die Verteilungskonvergenz des geschätzten empirischen Prozesses gegen einen zentrierten Gaußprozess gezeigt.

Die Verteilung von Teststatistiken, welche auf dem geschätzten empirischen Prozess beruhen, hängt in der vorliegenden Situation von der zugrundeliegenden unbekannten Verteilungsfunktion sowie von der unbekannten Abhängigkeitsstruktur der Beobachtungen ab. Daher können auf direktem Wege keine asymptotisch gültigen kritischen Werte zum Testen der Nullhypothese gewonnen werden.
Um dennoch zu kritischen Werten für die interessierende Teststatistik zu gelangen, kann der geschätzte empirische Prozess (2.1.1) auf Grundlage der beobachteten Daten durch ein geeignetes Bootstrap-Verfahren imitiert werden. Diese Art von Resampling-Verfahren wurde von Efron [27] vorgestellt um Verteilungseigenschaften von Statistiken zu imitieren, welche auf unabhängigen und identisch verteilten Beobachtungen basieren.
Imitiert man den Originalprozess, welcher auf den tatsächlichen Beobachtungen $X_1, ..., X_n$ beruht, durch einen Bootstrap-Prozess, der mithilfe der Bootstrap-Stichprobe $X_1^*, ..., X_n^*$ gewonnen wird, so sind die Parameter der Verteilungen der Teststatistiken, die auf dem Bootstrap-Prozess beruhen, bekannt. Für die Bootstrap-Statistik können daher kritische Werte ermittelt werden. Dies kann erfolgen durch analytische Berechnung der Verteilung oder, für den Fall, dass dies nicht praktikabel ist, durch Approximation der Verteilung der Teststatistik durch wiederholtes Simulieren unter Nutzung der beobachteten Daten.

Asymptotische Resultate zum Bootstrap für den gewöhnlichen empirischen Prozess, welcher empirische und eine vollständig spezifizierte Verteilungsfunktion ver-

gleicht, wurden in Bickel und Freedman [10] erzielt, für den Fall unabhängiger und identisch verteilter Beobachtungen, unter Nutzung des von Efron vorgestellten Bootstrap-Verfahrens. In Giné und Zinn [33] wurde, ebenso für i.i.d. Beobachtungen, dieses Bootstrap-Verfahren für empirische Prozesse, welche durch Funktionen indiziert sind, untersucht.
Was den geschätzten empirischen Prozess anbelangt, so ist es naheliegend, die durch die Schätzung der Parameter der vermuteten Verteilungsfunktion gewonnene Information zur Generierung der Bootstrap-Stichprobe zu verwenden, um den Originalprozesses durch einen Bootstrap-Prozess nachzubilden.
Bei Vorliegen unabhängig und identisch verteilter Beobachtungen wurde dies zunächst von Stute et al. [73] betrachtet. Dort wird ebenfalls davon ausgegangen, dass die vermutete Verteilungsfamilie von einem endlichdimensionalen Parameter θ abhängt, so dass durch dessen Schätzung unabhängige Zufallsgrößen mit Verteilungsfunktion $F_{\hat{\theta}_n}$ generiert werden können. Es konnte gezeigt werden, dass Originalprozess und Bootstrap-Prozess fast sicher die gleiche Grenzverteilung besitzen.
In Babu und Rao [6] wurde unter ähnlichen Voraussetzungen zusätzlich ein nichtparametrisches Bootstrap-Verfahren angewendet. Der hierbei entstehende geschätzte empirische Prozess weist allerdings einen zusätzlichen Bias-Term auf. Durch Einführung einer Bias-Korrektur kann jedoch gezeigt werden, dass der um die Bias-Korrektur veränderte Prozess dieselbe Grenzverteilung wie der Originalprozess besitzt.

Ziel der aktuellen Arbeit ist nun zu zeigen, dass unter Vorliegen abhängiger Beobachtungen der geschätzte empirische Prozess mit einem geeigneten Bootstrap-Verfahren gut imitiert wird. Um die Abhängigkeit der Beobachtungen zu beschreiben, wird das von Doukhan und Louhichi [24] eingeführte Konzept der ψ-Weak Dependence genutzt.
Es wird zunächst das Grenzverhalten des Originalprozesses unter den gegebenen Voraussetzungen untersucht. Sodann wird gezeigt, dass auch der Bootstrap-Prozess in Wahrscheinlichkeit dieselbe Grenzverteilung besitzt. Die Anwendung des kritischen Wertes der Bootstrap-Statistik auf die Originalstatistik führt damit asymptotisch zu einem Niveau-α-Test.

Liegen keine unabhängigen Beobachtungen vor, so muss durch das verwendete Bootstrap-Verfahren auch die Abhängigkeitsstruktur des Prozesses geeignet nachgebildet werden, weswegen die Generierung einer Bootstrap-Stichprobe mit zugrundeliegender Verteilungsfunktion $F_{\hat{\theta}_n}$ ohne zusätzliche Modellannahmen nicht auf den abhängigen Fall übertragen werden kann.
Die Verwendung eines modellbasierten Bootstrap-Verfahrens unter der Annahme eines bestimmten Zeitreihenmodells erweist sich jedoch als problematisch, da die Eigenschaften der stationären Verteilung im allgemeinen Fall nicht hinreichend greifbar sind, ebenso wie Effekte, die durch das Schätzen von Innovationen auftreten.
Um diese beim modellbasierten Bootstrap zutage tretenden Probleme zu umgehen, wurde in der vorliegenden Arbeit stattdessen ein Block-Bootstrap-Verfahren angewendet, wie in Künsch [43] für stationäre, abhängige Beobachtungen vorgeschlagen. Hierbei werden aus den Beobachtungen $X_1, ..., X_n$ $k := k_n$ zusammenhängende Blöcke gezogen, die jeweils die Länge $l := l_n$ besitzen, wobei die jeweils ersten Indizes der Blöcke unabhängig aus den Zahlen 1 bis $n - l + 1$ gezogen werden. Dieses sogenannte Moving-Block-Bootstrap-Verfahren wurde ebenfalls in Liu und Singh [47] untersucht.
Weitere Block-Bootstrap-Verfahren wurden in Carlstein [15], Politis und Romano [59] und Politis und Romano [59] vorgestellt, siehe Lahiri [45] für eine Übersicht über die verschiedenen Verfahren.
Im Zusammenhang mit empirischen Prozessen wurden Blockwise-Bootstrap-Verfahren für das Resampling in Naik-Nimbalkar und Rajarshi [53] und Bühlmann [13] für den gewöhnlichen empirischen Prozess, basierend auf α-mischenden Zufallsgrößen, verwendet. Es wurde gezeigt, dass unter bestimmten Voraussetzungen an die Summierbarkeit des Mischungskoeffizienten sowie an Blockanzahl und -länge der auf der Bootstrap-Stichprobe basierende empirische Prozess fast sicher gegen denselben zentrierten Gaußprozess konvergiert wie der Originalprozess. Ein analoges Resultat wurde in Peligrad [58] unter schwächeren Voraussetzungen an den Mischungskoeffizienten erzielt, ebenso wie für assoziierte Folgen von Beobachtungen.
In Bühlmann [14] wurde des weiteren ein allgemeineres Resultat erzielt, welche die obige Methode erfolgreich auf funktionenindizierte Prozesse anwendet, welchen β-mischende Zufallsgrößen zugrundeliegen. Dies wurde ebenfalls in Radulovic [65]

unter schwächeren Bedingungen an den Mischungskoeffizienten betrachtet.
Was die Anwendung dieses Ansatzes auf den hier betrachteten Prozess (2.1.1) anbelangt, so tritt jedoch, ähnlich wie in Babu und Rao [6] bei der Verwendung des nichtparametrischen Bootstrap für unabhängige Beobachtungen, beim entstehenden geschätzten empirischen Prozess, welcher auf der Bootstrap-Stichprobe beruht, ein zusätzlicher Bias-Term auf. Daher wird hier in analoger Weise wie in Babu und Rao [6] eine Bias-Korrektur vorgenommen und gezeigt, dass der so modifizierte Prozess in Wahrscheinlichkeit dieselbe Grenzverteilung wie der Originalprozess besitzt. Die Beweismethoden sind teilweise an das Vorgehen in Bühlmann [14] und Naik-Nimbalkar und Rajarshi [53] angelehnt.
Das genutzte Abhängigkeitskonzept, welches für die zugrundeliegenden Beobachtungen verwendet wird, ist jedoch allgemeiner als das Mixing-Konzept.
Es hat sich gezeigt, dass gewisse Prozessklassen wie etwa innovationsgesteuerte Prozesse, deren Innovationen eine diskrete Verteilung besitzen, nicht mischend sind. Diese Thematik wurde etwa in Chernick [16] und Andrews [3] diskutiert.
In dieser Arbeit wird daher das in Doukhan und Louhichi [24] eingeführte Konzept der schwachen Abhängigkeit (ψ-weak dependence) eingesetzt. Dieses deckt eine breite Klasse wichtiger Prozesse ab, welche auch die mischenden Prozesse einschließt. Modelle, die diesem Abhängigkeitskonzept genügen, wurden unter anderem in Ango Nze, Bühlmann und Doukhan [4] und Ango Nze und Doukhan [5] untersucht.
Im Folgenden wird vorausgesetzt, dass die Beobachtungen $X_1, ..., X_n$ schwach abhängig sind in der Form, dass die Kovarianzen gewisser Funktionen über Teilvektoren aus dem zugrundeliegenden Prozess in einem bestimmten Sinne gegen Null konvergieren, falls die dazwischenliegende Zeitlücke groß wird.
Dieses Konzept der schwachen Abhängigkeit im Sinne von Doukhan und Louhichi [24] lautet in seiner allgemeinen Form folgendermaßen:

Definition 2.1.2 *Die Folge $(X_n)_{n\in\mathbb{N}}$ von Zufallsvariablen heißt $(\zeta, \mathcal{F}, \psi)$-schwach abhängig, falls eine Klasse $\mathcal{F}$ reellwertiger Funktionen, eine Nullfolge $\zeta = (\zeta_r)_{r\in\mathbb{N}}$ und eine Funktion ψ mit Argumenten $(h, k, u, v) \in \mathcal{F}^2 \times \mathbb{N}^2$ existieren, so dass für jedes u-Tupel $(i_1, ..., i_u)$ und jedes v-Tupel $(j_1, ..., j_v)$ mit $i_1 \leq ... \leq i_u < i_u + r \leq j_1 \leq ... \leq j_v$ gilt*

$$|\operatorname{cov}(h(X_{i_1}, ..., X_{i_u}), k(X_{j_1}, ..., X_{j_v}))| \leq \psi(h, k, u, v) \cdot \zeta_r$$

für alle Funktionen $h, k \in \mathcal{F}$, die auf $\mathbb{R}^u$ bzw. $\mathbb{R}^v$ definiert sind.

Um die Verteilungskonvergenz des geschätzten empirischen Prozesses nachzuweisen, wird nun zunächst formuliert, unter welchen Voraussetzungen an die schwache Abhängigkeit der Beobachtungen und an den Parameterschätzer $\hat{\theta}_n$ die Konvergenz des geschätzten empirischen Prozesses gegen einen zentrierten Gaußprozess gezeigt werden kann.

2.2 Verteilungskonvergenz des geschätzten empirischen Prozesses

2.2.1 Verteilungskonvergenz von Prozessen

Der folgende Abschnitt befasst sich mit der Erläuterung der Verteilungskonvergenz von Prozessen auf $D(\bar{\mathbb{R}})$.
Was die Metrisierung anbelangt, so wird der Raum $D(\bar{\mathbb{R}})$ mit der uniformen Metrik ausgestattet, d.h.

$$||x - y||_\infty = \sup_{t\in\bar{\mathbb{R}}} |x(t) - y(t)|.$$

Während die Verwendung der uniformen Metrik insofern anschaulich ist, als die Stetigkeit von bestimmten Funktionalen dann unmittelbar ersichtlich ist, so ist der empirische Prozess bezüglich der zugehörigen Borel-σ-Algebra nicht messbar (S. Pollard [61], S. 65).
Für festes t gilt jedoch, dass $W_n(t)$ eine $\mathcal{A} - \mathcal{B}(\mathbb{R})$-messbare Zufallsvariable ist. Um die Messbarkeit des empirischen Prozesses zu gewährleisten, wird der Raum

$D(\bar{\mathbb{R}})$ mit der σ-Algebra $\mathcal{P}$ ausgestattet, welche durch die endlichdimensionalen Projektionen erzeugt wird. Diese ist äquivalent ist zur σ-Algebra, die durch die abgeschlossenen Kugeln auf$(D(\bar{\mathbb{R}}), ||\cdot||_\infty)$ erzeugt wird, wie in Pollard [61], S. 66, geschildert.
Mit der uniformen Metrik ausgestattet ist $D(\bar{\mathbb{R}})$ zwar nicht separabel, aber es wird sich zeigen, dass der Grenzprozess im betrachteten Fall fast sicher stetig und somit Element eines separablen Teilraums von $D(\bar{\mathbb{R}})$ ist, was ausreichend ist, um über die Anwendung des Continuous Mapping Theorems die Verteilungskonvergenz von entsprechenden Funktionalen abzuleiten.

Es gilt nun zunächst, die Verteilungskonvergenz der auf einem Wahrscheinlichkeitsraum $(\Omega, \mathcal{A}, \mathbb{P})$ definierten Prozesse W_n gegen ein Grenzelement G, definiert auf einem Wahrscheinlichkeitsraum $(\Omega', \mathcal{A}', \mathbb{P}')$ zu charakterisieren (Vereinfachend wird im folgenden das Maß auf $(\Omega', \mathcal{A}')$ ebenfalls mit $\mathbb{P}$ bezeichnet).
Dies geschieht durch die schwache Konvergenz der jeweils auf $(D(\bar{\mathbb{R}}), \mathcal{P})$ induzierten Bildmaße. Hinsichtlich dieser Bildmaße gelten folgende Bezeichnungen:
Für $A \in \mathcal{P}$ gilt:

$$\mathbf{P}_n(A) := \mathbb{P}(W_n \in A) \quad \text{und} \quad \mathbf{P}(A) := \mathbb{P}(G \in A).$$

Für die in der nachfolgenden Definition verwendeten Bezeichnungen wird folgende Vereinbarung getroffen:

$$\mathbf{P}_n f := \mathbb{E}(f(W_n)) := \int f(W_n)\, d\mathbb{P} \quad \text{bzw.} \quad \mathbf{P} f := \mathbb{E}(f(G)) := \int f(G)\, d\mathbb{P}.$$

Die Verteilungskonvergenz auf dem metrischen Raum $(D(\bar{\mathbb{R}}), ||\cdot||_\infty)$ wird nun folgendermaßen charakterisiert (vgl. Pollard [61], S. 65 und Billingsley [11], S. 25)

Definition 2.2.1 *Bezeichne $(W_n)_{n\in\mathbb{N}}$ eine Folge von $\mathcal{A}-\mathcal{P}$-messbaren Abbildungen von Ω nach $(D(\bar{\mathbb{R}}), ||\cdot||_\infty)$ und G eine $\mathcal{A}'-\mathcal{P}$-messbare Abbildung von Ω' nach $(D(\bar{\mathbb{R}}), ||\cdot||_\infty)$. Dann konvergiert die Folge W_n in Verteilung gegen G, wenn für jede stetige, beschränkte, borelmessbare Funktion f auf $(D(\bar{\mathbb{R}}), \mathcal{P}, ||\cdot||_\infty)$ gilt:*

$$\mathbb{E}\big(f(W_n)\big) \overset{n\to\infty}{\to} \mathbb{E}\big(f(G)\big).$$

Ebenso konvergiert eine Folge von Wahrscheinlichkeitsmaßen $(\mathbf{P}_n)_{n\in\mathbb{N}}$ auf $\mathcal{P}$ schwach gegen $\mathbf{P}$, falls $\mathbf{P}_n f \overset{n\to\infty}{\to} \mathbf{P} f$ für jede stetige, beschränkte, borelmessbare Funktion auf $(D(\bar{\mathbb{R}}), \mathcal{P}, ||\cdot||_\infty)$.

W_n konvergiert also in Verteilung gegen G, wenn die auf $(D(\bar{\mathbb{R}}), \mathcal{P})$ durch W_n induzierten Maße $\mathbf{P}_n$ schwach gegen $\mathbf{P}$ konvergieren.

Falls das Grenzelement G bestimmten Bedingungen genügt, kann die Verteilungskonvergenz von W_n gegen G sogar mithilfe von gleichmäßig stetigen Funktionen auf $(D(\bar{\mathbb{R}}), ||\cdot||_\infty)$ charakterisiert werden. Hierzu wird zunächst die vollständige Regularität eines Punktes im metrischen Raum $(D(\bar{\mathbb{R}}), ||\cdot||_\infty)$, ausgestattet mit der σ-Algebra $\mathcal{P}$, definiert (vgl. Pollard [61], S. 67).

Definition 2.2.2 *Ein Punkt $x \in D(\bar{\mathbb{R}})$ wird vollständig regulär bezüglich der Metrik $||\cdot||_\infty$ und der σ-Algebra $\mathcal{P}$ genannt, wenn für jede Umgebung V von x eine gleichmäßig stetige und $\mathcal{P}$-messbare Funktion g existiert mit $g(x) = 1$ und $g \leq I_V$.*

Bezeichne weiter

$$\mathcal{C}_b := \{f\colon (D[\bar{\mathbb{R}}], \mathcal{P}, ||\cdot||_\infty) \to (\mathbb{R}, \mathcal{B}(\mathbb{R}), |\cdot|), \text{ messbar, gleichmäßig stetig und beschränkt}\}. \tag{2.1}$$

Mithilfe dieser Definition und der oben eingeführten Bezeichnungen kann nun die Verteilungskonvergenz von W_n gegen G folgendermaßen charakterisiert werden (vgl. Pollard [61], S. 69):

Korollar 2.2.1 *Falls gilt*

$$\mathbf{P}_n f = \mathbb{E}\big(f(W_n)\big) \overset{n\to\infty}{\to} \mathbb{E}\big(f(G)\big) = \mathbf{P} f$$

für jede Funktion f mit $f \in \mathcal{C}_b$, und falls G mit Wahrscheinlichkeit 1 in einer separablen Menge vollständig regulärer Punkte liegt, so gilt

$$W_n \overset{\mathcal{D}}{\to} G.$$

Um Korollar 2.2.1 anwenden zu können, muss also gewährleistet sein, dass der Grenzprozess mit Wahrscheinlichkeit 1 in einer separablen Menge vollständig regulärer Punkte liegt. Zunächst gilt, dass jeder Punkt in $(D(\bar{\mathbb{R}}), \mathcal{P}, ||\cdot||_\infty)$ vollständig regulär ist (s. Pollard [61], S. 90). Es wird gezeigt werden, dass G mit Wahrscheinlichkeit 1 in $C(\bar{\mathbb{R}}) \subseteq D(\bar{\mathbb{R}})$ liegt. Die Menge $C(\bar{\mathbb{R}})$ ist separabel und $\mathcal{P}$-messbar (vgl. Pollard [61], Problem IV.8). Kann also die Stetigkeit fast sicher des Grenzprozesses nachgewiesen werden, so folgt die Verteilungskonvergenz von W_n gegen G durch Anwendung von Korollar 2.2.1. Aus der Verteilungskonvergenz von W_n gegen ein Grenzelement, welches mit Wahrscheinlichkeit 1 in $C(\bar{\mathbb{R}})$ liegt, kann dann ebenfalls die Verteilungskonvergenz von Funktionalen, welche auf diesem Teilraum stetig sind, gefolgert werden. Näher wird dies in Abschnitt 2.4 beleuchtet.
Im Folgenden werden nun zunächst die Voraussetzungen an den Prozess der Beobachtungen sowie den Schätzer des Parameters θ und anschließend das Resultat bezüglich des Grenzverhaltens von W_n formuliert.

2.2.2 Voraussetzungen und Grenzverhalten des geschätzten empirischen Prozesses

Um die Konvergenz von W_n gegen einen Gaußprozess zu zeigen, werden nun eine Reihe von Voraussetzungen getroffen.
Diese betreffen die Abhängigkeitsstruktur des Prozesses der Beobachtungen X_t sowie die Eigenschaften des Parameterschätzers $\hat{\theta}_n$.
Hierzu wird zunächst definiert, für eine Funktion $h : \mathbb{R}^d \to \mathbb{R}^k$:

$$\mathrm{Lip}(h) := \sup \left\{ \frac{||h(X_1, ..., X_d) - h(Y_1, ..., Y_d)||_2}{||(X_1, ..., X_d)^T - (Y_1, ..., Y_d)^T||_2} \right\}.$$

$||\cdot||_2$ bezeichnet hierbei die 2-Norm auf $\mathbb{R}^d$ bzw. $\mathbb{R}^k$, welche für den Fall $d = 1$ (oder $k = 1$) der 1-Norm entspricht. Eine Funktion h mit $\mathrm{Lip}(h) < \infty$ werde als lipschitzstetig bezeichnet.

Es gelten nun folgende Annahmen:

Annahme $\mathcal{A}$:

(i) Gegeben seien Beobachtungen $X_1, ..., X_n$, welche einem stochastischen Prozess $(X_t)_{t\in\mathbb{Z}}$ entstammen mit:
$(X_t)_{t\in\mathbb{Z}}$ ist ein reellwertiger, strikt stationärer Prozess, definiert auf einem Wahrscheinlichkeitsraum $(\Omega, \mathcal{A}, \mathbb{P})$. $(X_t)_{t\in\mathbb{Z}}$ ist schwach abhängig in folgendem Sinne:
Für $(\zeta_r)_{r\in\mathbb{N}} = \rho^r$, $\rho < 1$, gilt:

$$\forall\, s_1 < s_2 < ... < s_u < s_u + r = t_1 < ... < t_v,\ v \leq 4:$$

$$\big|\operatorname{cov}\big(f(X_{s_1}, ..., X_{s_u}), g(X_{t_1}, ..., X_{t_v})\big)\big| \leq \sqrt{\mathbb{E}f(X_{s_1}, ..., X_{s_u})^2} \cdot \mathrm{Lip}(g) \cdot K \cdot \zeta_r$$

für alle $f \in L_2(\mathbb{R}^u)$ und $g \in L_2(\mathbb{R}^v)$, $\mathrm{Lip}(g) < \infty$ und eine Konstante $K \in [0, \infty)$.

(ii) Die eindimensionalen Randverteilungen von (X_t) besitzen die Verteilungsfunktion $F_\theta(x)$, mit $\theta \in \Theta \subseteq \mathbb{R}^p$, wobei F_θ lipschitzstetig in x ist und:
$\dot{F}_\theta(x) := \frac{\partial}{\partial\theta} F_\theta(x) \in \mathbb{R}^p$ ist gleichmäßig stetig auf $\mathbb{R} \times U$, wobei U offene Umgebung von θ ist, und besitzt endliche Grenzwerte für $x \to \pm\infty$.

(iii) Für den Schätzer $\hat{\theta}_n$ für θ gelte, für eine messbare Funktion:
$\ell : \mathbb{R} \times \Theta \to \mathbb{R}^p$

$$\sqrt{n}(\hat{\theta}_n - \theta) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \ell(X_i, \theta) + \varepsilon_n, \quad \varepsilon_n \xrightarrow{\mathbb{P}} 0.$$

ℓ besitze die folgenden Eigenschaften:
$\big(\mathbb{E}[\ell(X_0, \theta)]\big)^T = (0, .., 0), \quad \mathbb{E}[||\ell(X_0, \theta)||_2^{2+\delta}] < \infty$ für ein $\delta > 0$,
$\mathrm{Lip}(\ell_c) \leq c^\gamma,$ für ein $\gamma > 0$.

Die gestutzte Version $\ell_c(\cdot, \theta)$ von $\ell(\cdot, \theta)$ sei hierbei folgendermaßen definiert:

$$\ell_c(X_0, \theta) \ := \ \ell(X_0, \theta) \cdot \mathbb{1}\{|X_0| \leq c\}$$

$$
\begin{aligned}
&+ \ell(c,\theta) \cdot \mathbb{1}\{X_0 > c\} \\
&+ \ell(-c,\theta) \cdot \mathbb{1}\{X_0 < -c\}. \qquad (2.2)
\end{aligned}
$$

Voraussetzung (ii) impliziert, dass $||\dot{F}_\theta||_\infty := \sup_{x\in\mathbb{R}} ||\dot{F}_\theta(x)||_2$ beschränkt ist.
Des weiteren folgt aufgrund von Voraussetzung (ii) und (iii), dass die Funktion $\ell_c^T(X,\theta) \cdot \dot{F}_\theta(x)$ lipschitzstetig bezüglich X ist, denn:

$$
\begin{aligned}
& |\ell_c^T(X,\theta) \cdot \dot{F}_\theta(x) - \ell_c^T(Y,\theta) \cdot \dot{F}_\theta(x)| \\
= \; & |(\ell_c(X,\theta) - \ell_c(Y,\theta))^T \cdot \dot{F}_\theta(x)| \\
\leq \; & ||\ell_c(X,\theta) - \ell_c(Y,\theta)||_2 \cdot ||\dot{F}_\theta(x)||_2 \\
\leq \; & c^\gamma \cdot ||\dot{F}_\theta||_\infty \cdot |X - Y|.
\end{aligned}
$$

Der Nachweis der asymptotischen Linearität des Parameterschätzers ist eine in diesem Zusammenhang übliche Forderung, siehe Rao und Krishnaiah [66], Stute et al. [73] und Babu und Rao [6]. Sie wird beispielshalber erfüllt von M-Schätzern, siehe hierzu Lahiri [44].
Die Funktion ℓ erfüllt im Allgemeinen nicht die benötigte Lipschitzbedingung bezüglich X_k. Daher wird die Lipschitzstetigkeit für eine bezüglich der X_k gestutzten Version ℓ_c vorausgesetzt, was gemeinsam mit der Forderung an die Momente von ℓ die spätere Beweisführung ermöglicht.

Die Voraussetzung des exponentiellen Abfallens des Abhängigkeitskoeffizienten kann in Abhängigkeit von der konkret vorliegenden Situation vermutlich abgeschwächt werden. In der Tat ist lediglich zu gewährleisten, dass die Summierbarkeit bestimmter Potenzen der Abhängigkeitskoeffizienten erfüllt ist.
Da allerdings ohnehin eine Vielzahl in der Praxis bedeutender Prozesse die oben geforderte Eigenschaft besitzen, stellt dies keine wesentliche Einschränkung dar und erleichtert die Lesbarkeit und somit die Anwendbarkeit der Resultate.
In Neumann und Paparoditis [55] wurde gezeigt, dass die unter $\mathcal{A}$ (i) getroffene Voraussetzung an das Abhängigkeitsverhalten des Prozesses von wichtigen Prozessklassen wie $AR(p)$- und $ARCH(p)$-Prozessen erfüllt wird.

Das folgende Theorem beinhaltet die Konvergenz des geschätzten empirischen Prozesses in Verteilung gegen einen zentrierten Gaußprozess G mit fast sicher

stetigen Pfaden unter den soeben getroffenen Annahmen.

Theorem 2.2.1 *Es gelte $\mathcal{A}$. Dann gilt:*

$$W_n \xrightarrow{\mathcal{D}} G,$$

wobei G ein zentrierter Gaußprozess ist mit fast sicher stetigen Pfaden und Kovarianzfunktion

$$\operatorname{cov}(G(x_i), G(x_j)) = \sum_{k \in \mathbb{Z}} \operatorname{cov}\big(z_\theta(X_0, x_i), z_\theta(X_k, x_j)\big),$$

wobei

$$z_\theta(X_k, x) := \mathbb{1}_{\{X_k \leq x\}} - F_\theta(x) - \ell^T(X_k, \theta) \cdot \dot{F}_\theta(x).$$

2.2.3 Beweis der Verteilungskonvergenz des Originalprozesses

Darstellung des geschätzten empirischen Prozesses und Vorgehensweise

Zunächst wird unter Ausnutzung der in Abschnitt 2.2.2 unter $\mathcal{A}$ genannten Voraussetzungen eine Darstellung des Prozesses W_n gezeigt, wie sie üblicherweise in diesem Rahmen verwendet wird, siehe etwa Rao und Krishnaiah [66], Stute et al. [73] und Babu und Rao [6].
Der Prozess wird hierbei in einen zentrierten Anteil zerlegt und einen Restterm, der das asymptotische Verhalten nicht beeinflusst, wie im folgenden zu zeigen sein wird.
Es gilt zunächst unter Verwendung der Taylorschen Formel aufgrund der Existenz von $\dot{F}_\theta(x)$, für ein $\tilde{\theta}_n$ zwischen $\hat{\theta}_n$ und θ:

$$\begin{aligned}
W_n(x) &= \sqrt{n} \cdot [F_n(x) - F_\theta(x) + F_\theta(x) - F_{\hat{\theta}_n}(x)] \\
&= \sqrt{n} \cdot [F_n(x) - F_\theta(x)] + \sqrt{n} \cdot [(\theta - \hat{\theta}_n)^T \cdot \dot{F}_{\tilde{\theta}_n}(x)] \\
&= \sqrt{n} \cdot [F_n(x) - F_\theta(x)] + \sqrt{n} \cdot [(\theta - \hat{\theta}_n)^T \cdot \dot{F}_\theta(x)] \\
&\quad + \sqrt{n} \cdot (\theta - \hat{\theta}_n)^T \cdot [\dot{F}_{\tilde{\theta}_n}(x) - \dot{F}_\theta(x)].
\end{aligned}$$

Mit Voraussetzung (iii) gilt

$$\sqrt{n} \cdot (\hat{\theta}_n - \theta) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \ell(X_i, \theta) + \varepsilon_n.$$

Gemäß Lemma 2.5.3 ist $\frac{1}{\sqrt{n}} \sum\limits_{i=1}^{n} \ell(X_i, \theta)$ $\mathbb{P}$-beschränkt. Damit gilt auch insbesondere die Konvergenz von $\hat{\theta}_n$ und $\tilde{\theta}_n$ gegen θ in Wahrscheinlichkeit.
Aufgrund von $\hat{\theta}_n \xrightarrow{\mathbb{P}} \theta$ befindet sich $\tilde{\theta}_n$ für hinreichend großes n in U und aufgrund der gleichmäßigen Stetigkeit von $\dot{F}_\theta(x)$ auf $\mathbb{R} \times U$ gilt $\dot{F}_{\tilde{\theta}_n}(x) - \dot{F}_\theta(x) \xrightarrow{\mathbb{P}} 0$, gleichmäßig in x.
Daher konvergiert mit dem Satz von Cramér-Slutsky der Ausdruck

$$\sqrt{n} \cdot (\theta - \hat{\theta}_n)^T \cdot [\dot{F}_{\tilde{\theta}_n}(x) - \dot{F}_\theta(x)]$$

in Wahrscheinlichkeit gegen 0, gleichmäßig in x.

Insgesamt gilt also folgende Darstellung für $W_n(x)$:

$$\begin{aligned}
W_n(x) &= \sqrt{n} \cdot [F_n(x) - F_\theta(x)] - \frac{1}{\sqrt{n}} \sum_{k=1}^{n} \ell^T(X_k, \theta) \cdot \dot{F}_\theta(x) + \tilde{\varepsilon}_n(x) \\
&= \frac{1}{\sqrt{n}} \sum_{k=1}^{n} [\mathbb{1}_{\{X_k \leq x\}} - F_\theta(x) - \ell^T(X_k, \theta) \cdot \dot{F}_\theta(x)] + \tilde{\varepsilon}_n(x) \\
&=: U_n(x) - \frac{1}{\sqrt{n}} \sum_{k=1}^{n} \ell^T(X_k, \theta) \cdot \dot{F}_\theta(x) + \tilde{\varepsilon}_n(x) \\
&=: \frac{1}{\sqrt{n}} \sum_{k=1}^{n} z_\theta(X_k, x) + \tilde{\varepsilon}_n(x) \\
&=: Z_n(x) + \tilde{\varepsilon}_n(x), \qquad (2.3)
\end{aligned}$$

wobei gilt, dass $\sup\limits_{x \in \bar{\mathbb{R}}} \tilde{\varepsilon}_n(x) = o_{\mathbb{P}}(1)$.

Mit Korollar 2.2.1 kann, falls die Pfade des Grenzprozesses mit Wahrscheinlichkeit 1 stetig sind, die Verteilungskonvergenz von W_n gegen G auf $(D[-\infty, \infty], \mathcal{P})$, ausgestattet mit der σ-Algebra $\mathcal{P}$ nun folgendermaßen charakterisiert werden:

$$\begin{aligned}
W_n \xrightarrow{\mathcal{D}} G \quad \Leftrightarrow \quad & \forall f : D(\bar{\mathbb{R}}) \to \mathbb{R} \text{ gleichmäßig stetig und beschränkt gilt:} \\
& \mathbb{E} f(W_n) \overset{n \to \infty}{\to} \mathbb{E} f(G).
\end{aligned}$$

Hierzu werden die folgenden Prozesse $\widetilde{W}_n$ und $\widetilde{G}$ definiert, die zwischen (später festzulegenden) Gitterpunkten $\{x_0, ..., x_M\}$, $M \in \mathbb{N}$, konstant sind:

$$\begin{aligned} \widetilde{W}_n(x) &= W_n(x_i) \text{ für alle } x \in (x_{i-1}, x_i],\ i = 1, ..., M, \\ \widetilde{G}(x) &= G(x_i) \text{ für alle } x \in (x_{i-1}, x_i],\ i = 1, ..., M. \end{aligned}$$

Damit gilt:

$$\begin{aligned} & |\mathbb{E}f(W_n) - \mathbb{E}f(G)| \\ \leq\ & |\mathbb{E}f(W_n) - \mathbb{E}f(\widetilde{W}_n)| + |\mathbb{E}f(\widetilde{W}_n) - \mathbb{E}f(\widetilde{G})| + |\mathbb{E}f(\widetilde{G}) - \mathbb{E}f(G)| \\ =:\ & D_1 + D_2 + D_3. \end{aligned} \tag{2.4}$$

In Betrachtung des Terms D_2 definieren wir weiterhin

$$\vec{W}_n := (W_n(x_1), ..., W_n(x_M))^T \text{ und } \vec{G} := (G(x_1), ..., G(x_M))^T \tag{2.5}$$

sowie

$$\tilde{f}(\vec{W}_n) := f(\widetilde{W}_n) \text{ und } \tilde{f}(\vec{G}) := f(\widetilde{G}), \text{ wobei } \tilde{f} : \mathbb{R}^M \to \mathbb{R}.$$

Da f gleichmäßig stetig und beschränkt ist, besitzt auch $\tilde{f}$ diese Eigenschaften. Daraus folgt, dass die Verteilungskonvergenz von $\widetilde{W}_n$ gegen $\widetilde{G}$ äquivalent ist zur Konvergenz von $\mathbb{E}\tilde{f}(\vec{W}_n)$ gegen $\mathbb{E}\tilde{f}(\vec{G})$ und dies wiederum zur Verteilungskonvergenz von $\vec{W}_n$ gegen $\vec{G}$. (Siehe hierzu Pollard [61], Kapitel V.)

Um die Verteilungskonkvergenz des geschätzten empirischen Prozesses nachzuweisen, wird also zunächst die Konvergenz der endlichdimensionalen Verteilungen und anschließend die stochastische Äquikontinuität des Prozesses sowie die Existenz des Grenzprozesses G gezeigt.

Konvergenz der endlichdimensionalen Verteilungen

Für den Beweis der Konvergenz der endlichdimensionalen Verteilungen des Prozesses W_n ist der folgende zentrale Grenzwertsatz für Dreiecksschemata von schwach abhängigen Zufallsgrößen aus Neumann und Paparoditis [55] (Theorem 6.1) von Bedeutung.

Zum Nachweis der Konvergenz der endlichdimensionalen Verteilungen des Originalprozesses wird dieser Satz für eine Folge von Zufallsvariablen angewendet.

Theorem 2.2.2 (Neumann und Paparoditis [55])
Sei $(X_{n,k})_{k=1,...,n}, n \in \mathbb{N}$, ein Dreiecksschema von zeilenweise unabhängigen Zufallsvariablen mit $\mathbb{E}X_{n,k} = 0$ und $\mathbb{E}X_{n,k}^2 \leq C < \infty$. Außerdem gelte für alle $\varepsilon > 0$

$$\frac{1}{n}\sum_{k=1}^{n} \mathbb{E}X_{n,k}^2 \cdot \mathbb{1}\{|X_{n,k}|/\sqrt{n} > \varepsilon\} \overset{n\to\infty}{\to} 0$$

und

$$\operatorname{var}(X_{n,1} + ... + X_{n,n})/n \overset{n\to\infty}{\to} \sigma^2 \in [0, \infty).$$

Für $n \geq n_0$ existiert eine monotone nichtwachsende und summierbare Folge $(\zeta_r)_{r\in\mathbb{N}}$, so dass, für alle Indizes $1 \leq s_1 < s_2 < ... < s_u < s_u + r = t_1 \leq t_2 \leq n$, die folgenden oberen Schranken für Kovarianzen gelten:
Für alle messbaren und quadratisch integrierbaren Funktionen $f : \mathbb{R}^u \to \mathbb{R}$ gilt:

$$|\operatorname{cov}(f(X_{n,s_1}, ..., X_{n,s_u}), X_{n,t_1})| \leq \sqrt{\mathbb{E}f^2(X_{n,s_1}, ..., X_{n,s_u})} \cdot \zeta_r$$

sowie für alle messbaren und beschränkten Funktionen $f : \mathbb{R}^u \to \mathbb{R}$

$$|\operatorname{cov}(f(X_{n,s_1}, ..., X_{n,s_u}), X_{n,t_1}X_{n,t_2})| \leq ||f||_\infty \cdot \zeta_r.$$

Dann gilt

$$\frac{1}{\sqrt{n}}(X_{n,1} + ... + X_{n,n}) \overset{\mathcal{D}}{\to} \mathcal{N}(0, \sigma^2).$$

Um die Konvergenz von D_2 (siehe (2.4)) gegen Null zu zeigen, ist zunächst die Konvergenz der endlichdimensionalen Verteilungen von Z_n (siehe (2.3)) gegen diejenigen von G zu zeigen. Mit dem Satz von Cramér-Slutsky konvergieren somit auch die endlichdimensionalen Verteilungen von W_n gegen die eines Gaußprozesses.
Setze hierzu, ähnlich wie oben

$$\vec{Z}_n := (Z_n(x_1), ..., Z_n(x_M))^T.$$

Der Satz von Cramér-Wold erlaubt es, die Verteilungskonvergenz auf den univariaten Fall zu reduzieren. Zu beweisen ist also zunächst folgendes Lemma

Lemma 2.2.1
Sei $c = (c_1, ..., c_M)^T \in \mathbb{R}^M$ *beliebig.*
Dann gilt unter Voraussetzung $\mathcal{A}$

$$c^T \vec{Z}_n \xrightarrow{\mathcal{D}} c^T \vec{G} \sim N(0, c^T \Sigma_{x_1,...,x_M} c),$$

wobei

$$\Sigma_{x_1,...,x_M}(i,j) = \sum_{k \in \mathbb{Z}} \operatorname{cov}\big(z_\theta(X_0, x_i), z_\theta(X_k, x_j)\big).$$

Beweis:
Definiere, für $k = 1, ..., n$,

$$T_k := \sum_{i=1}^{M} c_i z_\theta(X_k, x_i).$$

Dann gilt

$$c^T \vec{Z}_n = \frac{1}{\sqrt{n}} \sum_{k=1}^{n} \sum_{i=1}^{M} c_i z_\theta(X_k, x_i) = \frac{1}{\sqrt{n}} \sum_{k=1}^{n} T_k.$$

Es wird nun der oben formulierte zentrale Grenzwertsatz für schwach abhängige Zufallsvariablen auf die Folge T_k angewendet. Hierzu ist zu sicherzustellen, dass die Voraussetzungen des zentralen Grenzwertsatzes erfüllt sind.

- $\mathbb{E}(T_k) = 0$ ist klar, denn es gilt

  $$\mathbb{E}(\mathbb{1}_{\{X_k \leq x_i\}} - F_\theta(x_i)) = 0$$

  und

  $$\mathbb{E}(\ell^T(X_k, \theta) \cdot \dot{F}_\theta(x_i)) = 0$$

  gemäß Voraussetzung $\mathcal{A}$ (iii).

- $\mathbb{E}(T_k^2) \leq C < \infty$ für ein $C \in \mathbb{R}$ gilt, da zum einen $\mathbb{E}(\mathbb{1}_{\{X_k \leq x_i\}} - F_\theta(x_i))$ und zum anderen $\dot{F}_\theta(x)$ und $\mathbb{E}(||\ell(X_0, \theta)||_2^{2+\delta})$ beschränkt sind gemäß Voraussetzung $\mathcal{A}$ (ii) und (iii). Damit gilt auch $\mathbb{E}(||\ell(X_0, \theta)||_2^2) < \infty$ und die

Ungleichung von Cauchy-Schwarz liefert somit die Endlichkeit von $\mathbb{E}(T_k^2)$:

$$\begin{aligned}
\mathbb{E}\big(T_k^2\big) &= \mathbb{E}\Big(\sum_{i=1}^{M}\sum_{j=1}^{M} c_i c_j z_\theta(X_k, x_i) z_\theta(X_k, x_j)\Big) \\
&= \sum_{t=1}^{M}\sum_{r=1}^{M} c_i c_j \mathbb{E}\Big(\big(\mathbb{1}_{\{X_k \le x_i\}} - F_\theta(x_i)\big)\cdot\big(\mathbb{1}_{\{X_k \le x_j\}} - F_\theta(x_j)\big) \\
&\qquad + \big(\mathbb{1}_{\{X_k \le x_i\}} - F_\theta(x_i)\big)\cdot\big(\ell^T(X_k,\theta)\cdot \dot{F}_\theta(x_j)\big) \\
&\qquad + \big(\mathbb{1}_{\{X_k \le x_j\}} - F_\theta(x_j)\big)\cdot\big(\ell^T(X_k,\theta)\cdot \dot{F}_\theta(x_i)\big) \\
&\qquad + \big(\ell^T(X_k,\theta)\cdot \dot{F}_\theta(x_i)\big)\cdot\big(\ell^T(X_k,\theta)\cdot \dot{F}_\theta(x_j)\big)\Big) \\
&\le \sum_{t=1}^{M}\sum_{r=1}^{M} c_i c_j \Big(1 + 2\cdot ||\dot{F}_\theta||_\infty \cdot \mathbb{E}(||\ell(X_k,\theta)||_2) \\
&\qquad + ||\dot{F}_\theta||_\infty^2 \cdot \mathbb{E}(||\ell(X_k,\theta)||_2^2)\Big) < \infty.
\end{aligned}$$

- Die Lindeberg-Bedingung ist erfüllt, da die T_k aufgrund der strikten Stationarität der X_k identisch verteilt sind. Die dann ausreichende quadratische Integrierbarkeit gilt gemäß des zuletzt ausgeführten Punktes.
- Um die Konvergenz von $\sigma_n^2 = \frac{1}{n}\operatorname{var}\big(T_1 + ... + T_n\big)$ gegen $\sigma^2 := c^T \Sigma_{x_1,\dots,x_M} c$ zu zeigen, wird die absolute Konvergenz der Reihe $\sum_{k\in\mathbb{Z}} \operatorname{cov}(T_0, T_k)$ gezeigt, denn dann gilt

$$\begin{aligned}
\sigma_n^2 &= \frac{1}{n}\operatorname{var}(\sum_{k=1}^{n} T_k) = \frac{1}{n}\sum_{k=1}^{n}\sum_{l=1}^{n}\operatorname{cov}(T_k, T_l) \\
&= \frac{1}{n}\sum_{k=-(n-1)}^{n-1}(n - |k|)\operatorname{cov}(T_0, T_k) \\
&= \sum_{k\in\mathbb{Z}} \frac{(n-|k|)_+}{n}\operatorname{cov}(T_0, T_k) \\
&\overset{n\to\infty}{\to} \sum_{k\in\mathbb{Z}} \operatorname{cov}(T_0, T_k) \quad \text{mit dem Satz von Lebesgue}
\end{aligned}$$

Da gilt

$$\sigma_n^2 = c^T \cdot [\operatorname{cov}(\frac{1}{\sqrt{n}}\sum_{k=1}^{n} z_\theta(X_k, x_i), \frac{1}{\sqrt{n}}\sum_{k=1}^{n} z_\theta(X_k, x_j))]_{(i,j)} \cdot c$$

ist durch den Nachweis von $\sigma_n^2 \to \sigma^2$ ebenso die Konvergenz der Kovarianzen von W_n gegen diejenigen von G gezeigt.

Um die absolute Konvergenz von $\sum_{k\in\mathbb{Z}} \operatorname{cov}(T_0, T_k)$ nachzuweisen beobachten wir zunächst

$$\sum_{k\in\mathbb{Z}} \operatorname{cov}(T_0, T_k) = \sum_{k=-\infty}^{\infty} \operatorname{cov}(T_0, T_{|k|}), \tag{2.6}$$

was aufgrund der Stationarität des Prozesses $(X_t)_{t\in\mathbb{Z}}$ gilt.

Weiter gilt

$$\begin{aligned}
& \operatorname{cov}(T_0, T_k) \\
= \; & \operatorname{cov}\Big(\sum_{i=1}^{M} c_i \cdot \big(\mathbb{1}\{X_0 \le x_i\} - F_\theta(x_i) - \ell^T(X_0,\theta)\cdot \dot{F}_\theta(x_i)\big), \\
& \qquad \sum_{i=1}^{M} c_i \cdot \big(\mathbb{1}\{X_k \le x_i\} - F_\theta(x_i) - \ell^T(X_k,\theta)\cdot \dot{F}_\theta(x_i)\big)\Big) \\
= \; & \sum_{i=1}^{M}\sum_{j=1}^{M} c_i c_j \cdot \underbrace{\operatorname{cov}(\mathbb{1}\{X_0 \le x_i\} - F_\theta(x_i), \mathbb{1}\{X_k \le x_j\} - F_\theta(x_j))}_{=:(i)} \\
& - \sum_{i=1}^{M}\sum_{j=1}^{M} c_i c_j \cdot \underbrace{\operatorname{cov}(\mathbb{1}\{X_0 \le x_i\} - F_\theta(x_i), \ell^T(X_k,\theta)\cdot \dot{F}_\theta(x_j))}_{=:(ii\,a)} \\
& - \sum_{i=1}^{M}\sum_{j=1}^{M} c_i c_j \cdot \underbrace{\operatorname{cov}(\ell^T(X_0,\theta)\cdot \dot{F}_\theta(x_i), \mathbb{1}\{X_k \le x_j\} - F_\theta(x_j))}_{=:(ii\,b)} \\
& + \sum_{i=1}^{M}\sum_{j=1}^{M} c_i c_j \cdot \underbrace{\operatorname{cov}(\ell^T(X_0,\theta)\cdot \dot{F}_\theta(x_i), \ell^T(X_k,\theta)\cdot \dot{F}_\theta(x_j))}_{=:(iii)}.
\end{aligned}$$

Es wird nun die absolute Konvergenz der Reihe bei (2.6) unter Ausnutzung der schwachen Abhängigkeit der X_i und der Eigenschaften des Parameterschätzers $\hat{\theta}_n$ gezeigt, was impliziert, dass die Reihe und somit σ_n^2 einen Grenzwert besitzt. Hierzu werden die mit $(i), (iia), (iib)$ und (iii) bezeichneten Terme separat betrachtet.

zu (i):

Unter Einführung der Bezeichnung

$$\bar{Y}_0(x_i) := \mathbb{1}\{X_0 \leq x_i\} - F_\theta(x_i) \tag{2.7}$$

erhält man zunächst

$$|\operatorname{cov}(\mathbb{1}\{X_0 \leq x_i\} - F_\theta(x_i), \mathbb{1}\{X_k \leq x_j\} - F_\theta(x_j))| = |\operatorname{cov}(\bar{Y}_0(x_i), \bar{Y}_k(x_j))|.$$

Um die schwache Abhängigkeit der X_k ausnutzen zu können, muss die Lipschitzstetigkeit in X des zweiten innerhalb der Kovarianz unter (i) auftretenden Terms gegeben sein. Da es sich hierbei allerdings um eine nichtstetige Funktion handelt, wird folgende lipschitzstetige Glättung $\widetilde{\bar{Y}}_{(\cdot)}(x)$ von $\bar{Y}_{(\cdot)}(x)$ definiert (siehe Beweis von Theorem 6 in Doukhan und Louhichi [24]):

$$\widetilde{\bar{Y}}_k(x) := \mathbb{1}\{X_k \leq x\} + \mathbb{1}\{x < X_k \leq x + a\} \cdot \left(1 - \frac{X_k - x}{a}\right) - F_\theta(x).$$

$\widetilde{\bar{Y}}_{(\cdot)}(x)$ ist lipschitzstetig mit Lipschitzkonstante $\frac{1}{a}$. (Siehe auch Lemma (2.5.1)).

Da die Funktionen $\bar{Y}_k(x)$ und $\widetilde{\bar{Y}}_k(x)$ quadratisch integrierbar sind, folgt somit aufgrund der schwachen Abhängigkeit der X_k, für $k \geq 0$ (siehe (2.6)), gemäß Lemma 2.5.1:

$$\begin{aligned} & |\operatorname{cov}(\bar{Y}_0(x_i), \bar{Y}_k(x_j)| \\ \leq\ & |\operatorname{cov}(\bar{Y}_0(x_i), \widetilde{\bar{Y}}_k(x_j))| \\ & + |\operatorname{cov}(\bar{Y}_0(x_i), \bar{Y}_k(x_j) - \widetilde{\bar{Y}}_k(x_j))| \\ \leq\ & \text{const} \cdot \zeta_k^{\frac{1}{2}}, \end{aligned} \tag{2.8}$$

wobei im letzten Schritt a in Abhängigkeit von k so gewählt wird, dass für jede Zeitlücke k beide Summanden von der gleichen Ordnung sind.

zu $(ii\,a)$**:**

Um wiederum die schwache Abhängigkeit der X_k ausnutzen zu können, muss die Funktion $\ell(\cdot, \theta)$ in einen lipschitzstetigen Anteil und einen Restterm zerlegt werden. An dieser Stelle kommt zum Tragen, dass vorausgesetzt wurde, dass $\ell(\cdot, \theta)$ lipschitzstetig ist, sofern die Argumente der Norm

nach beschränkt sind. $\ell_c(\cdot,\theta)$ ersetzt. (siehe (2.2))

$$\begin{aligned}
& |\operatorname{cov}(\mathbb{1}\{X_0 \le x_i\} - F_\theta(x_i), \ell^T(X_k,\theta)\cdot \dot{F}_\theta(x_i))| \\
\le\ & |\operatorname{cov}(\mathbb{1}\{X_0 \le x_i\} - F_\theta(x_i), \ell_c^T(X_k,\theta)\cdot \dot{F}_\theta(x_j))| \\
& + |\operatorname{cov}(\mathbb{1}\{X_0 \le x_i\} - F_\theta(x_i), \ell^T(X_k,\theta)\cdot \dot{F}_\theta(x_j) - \ell_c^T(X_k,\theta)\cdot \dot{F}_\theta(x_j))|.
\end{aligned} \tag{2.9}$$

Zum einen gilt:

$$\begin{aligned}
& |\operatorname{cov}(\mathbb{1}\{X_0 \le x_i\} - F_\theta(x_i), \ell_c^T(X_k,\theta)\cdot \dot{F}_\theta(x_j))| \\
\le\ & \sqrt{\mathbb{E}\big[(\mathbb{1}\{X_0 \le x_i\} - F_\theta(x_i))^2\big]}\cdot c^\gamma \cdot ||\dot{F}_\theta||_\infty \cdot K \cdot \zeta_k \\
\le\ & c^\gamma \cdot ||\dot{F}_\theta||_\infty \cdot K \cdot \zeta_k,
\end{aligned}$$

da aufgrund der Voraussetzungen $||\dot{F}_\theta||_\infty$ beschränkt ist.

Der zweite Summand bei (2.9) kann abgeschätzt werden durch

$$\begin{aligned}
& |\operatorname{cov}(\mathbb{1}\{X_0 \le x_i\} - F_\theta(x_i), (\ell^T(X_k,\theta) - \ell_c^T(X_k,\theta))\cdot \dot{F}_\theta(x_j))| \\
\le\ & \mathbb{E}[|\ell^T(X_k,\theta)\cdot \dot{F}_\theta(x_j)|\cdot \mathbb{1}\{|X_k| > c\}] \\
\le\ & \big(\mathbb{E}[|\ell^T(X_k,\theta)\cdot \dot{F}_\theta(x_j)|^2]\big)^{\frac{1}{2}} \cdot \big(\mathbb{E}[\mathbb{1}\{|X_k| > c\}]\big)^{\frac{1}{2}} \\
\le\ & \big(\mathbb{E}[||\ell^T(X_k,\theta)||_2^2]\big)^{\frac{1}{2}} \cdot ||\dot{F}_\theta||_\infty \cdot \big(\mathbb{E}[\mathbb{1}\{|X_k| > c\}]\big)^{\frac{1}{2}} \\
\le\ & \big(\mathbb{E}[||\ell^T(X_k,\theta)||_2^2]\big)^{\frac{1}{2}} \cdot ||\dot{F}_\theta||_\infty \cdot \frac{\big(\mathbb{E}(|X_k|^2)\big)^{\frac{1}{2}}}{c}.
\end{aligned}$$

Man erhält also für (2.9)

$$\begin{aligned}
& |\operatorname{cov}(\mathbb{1}\{X_0 \le x_i\} - F_\theta(x_i), \ell^T(X_k,\theta)\cdot \dot{F}_\theta(x_j))| \\
\le\ & ||\dot{F}||_\infty \cdot \big(c^\gamma \cdot K \cdot \zeta_k + \big(\mathbb{E}[||\ell^T(X_k,\theta)||_2^2]\big)^{\frac{1}{2}} \cdot \big(\mathbb{E}(|X_k|^2)\big)^{\frac{1}{2}} \cdot c^{-1}\big) \\
\le\ & \text{const}\cdot \zeta_k^{\frac{1}{1+\gamma}},
\end{aligned} \tag{2.10}$$

wobei dies durch Ausbalancieren der beiden Summanden in der vorletzten Zeile erzielt wird.

zu $(ii\,b)$**:**

Mit den oben eingeführten Bezeichnungen gilt für diesen Term, für $k \geq 0$

$$\begin{aligned} & |\operatorname{cov}(\ell^T(X_0,\theta)\cdot \dot{F}_\theta(x_j), \mathbb{1}\{X_k \leq x_i\} - F_\theta(x_j))| \\ \leq\ & |\operatorname{cov}(\ell^T(X_0,\theta)\cdot \dot{F}_\theta(x_j), \widetilde{\bar{Y}}_{x_i}(X_k))| \\ & + |\operatorname{cov}(\ell^T(X_0,\theta)\cdot \dot{F}_\theta(x_j), \bar{Y}_{x_i}(X_k) - \widetilde{\bar{Y}}_{x_i}(X_k))|. \end{aligned} \tag{2.11}$$

Der erste Summand kann durch die Lipschitzstetigkeit von $\widetilde{\bar{Y}}_{x_i}(\cdot)$ mithilfe der schwachen Abhängigkeit der X_k abgeschätzt werden.
Es gilt:

$$\begin{aligned} & |\operatorname{cov}(\ell^T(X_0,\theta)\cdot \dot{F}_\theta(x_j), \tilde{Y}_{x_i}(X_k))| \\ \leq\ & \big(\mathbb{E}\big[(\ell^T(X_0,\theta)\cdot \dot{F}_\theta(x_j))^2\big]\big)^{\frac{1}{2}} \cdot \frac{1}{a}\cdot K\cdot \zeta_k \\ \leq\ & \big(\mathbb{E}(||\ell^T(X_0,\theta)||_2^2)\big)^{\frac{1}{2}} \cdot ||\dot{F}_\theta||_\infty \cdot \frac{1}{a}\cdot K\cdot \zeta_k, \end{aligned}$$

aufgrund der Beschränktheit von $||\dot{F}||_\infty$ und $\mathbb{E}(||\ell(X_k,\theta)||_2^2)$.

Für den zweiten Term unter (2.11) gilt des weiteren

$$\begin{aligned} & |\operatorname{cov}(\ell^T(X_0,\theta)\cdot \dot{F}_\theta(x_j), Y_{x_i}(X_k) - \tilde{Y}_{x_i}(X_k))| \\ \leq\ & |\mathbb{E}[\ell^T(X_0,\theta)\cdot \dot{F}_\theta(x_i)\cdot(-\mathbb{1}\{x_i < X_k \leq x_i + a\}(1 - \frac{X_k - x_i}{a}))]| \\ & + |\underbrace{\mathbb{E}[\ell^T(X_0,\theta)\cdot \dot{F}_\theta(x_j)]}_{=0}\cdot\mathbb{E}[(-\mathbb{1}\{x_i < X_k \leq x_i + a\}(1 - \frac{X_k - x_i}{a}))]| \\ \leq\ & \big(\mathbb{E}(||\ell^T(X_0,\theta)||_2^2)\big)^{\frac{1}{2}} \cdot ||\dot{F}_\theta||_\infty \cdot \big(\mathbb{E}[\mathbb{1}\{x_i < X_k \leq x_i + a\}]\big)^{\frac{1}{2}} \\ =\ & \big(\mathbb{E}(||\ell^T(X_0,\theta)||_2^2)\big)^{\frac{1}{2}} \cdot ||\dot{F}_\theta||_\infty \cdot \big((F_\theta(x_i + a) - F_\theta(x_i))\big)^{\frac{1}{2}} \\ \leq\ & \big(\mathbb{E}(||\ell^T(X_0,\theta)||_2^2)\big)^{\frac{1}{2}} \cdot ||\dot{F}_\theta||_\infty \cdot \big(C\cdot a\big)^{\frac{1}{2}}. \end{aligned}$$

Somit gilt insgesamt für (2.11)

$$\begin{aligned} & |\operatorname{cov}(\mathbb{1}\{X_0 \leq x_i\} - F_\theta(x_j), \ell^T(X_k,\theta)\cdot \dot{F}_\theta(x_j))| \\ \leq\ & \sqrt{\mathbb{E}(||\ell^T(X_k,\theta)||_2^2)} \cdot ||\dot{F}_\theta||_\infty \cdot \big(\frac{1}{a}\cdot K\cdot \zeta_k + \sqrt{C\cdot a}\big) \\ \leq\ & \text{const}\cdot \zeta_k^{\frac{1}{3}}, \end{aligned} \tag{2.12}$$

wobei wiederum der letzte Schritt durch Ausbalancieren der Summanden durch die Wahl von a in Abhängigkeit von k erzielt wird.

zu (iii)

Hier wird wiederum die Funktion ℓ durch deren lipschitzstetige Variante ersetzt. Man erhält für (iii), für $k \geq 0$

$$\begin{aligned}
& |\operatorname{cov}(\ell^T(X_0,\theta)\cdot \dot{F}_\theta(x_i), \ell^T(X_k,\theta)\cdot \dot{F}_\theta(x_j))| \\
\leq\quad & |\operatorname{cov}(\ell^T(X_0,\theta)\cdot \dot{F}_\theta(x_i), \ell_c^T(X_k,\theta)\cdot \dot{F}_\theta(x_j))| \\
& + |\operatorname{cov}(\ell^T(X_0,\theta)\cdot \dot{F}_\theta(x_i), (\ell^T(X_k,\theta)-\ell_c^T(X_k,\theta))\cdot \dot{F}_\theta(x_j))| \\
\leq\quad & \big(\mathbb{E}\big[||\ell^T(X_0,\theta)||_2^2\big]\big)^{\frac{1}{2}}\cdot ||\dot{F}_\theta||_\infty \cdot c^\gamma \cdot ||\dot{F}_\theta||_\infty \cdot K\cdot \zeta_k \\
& + \mathbb{E}\big[|\ell^T(X_0,\theta)\cdot \dot{F}_\theta(x_i)|\cdot |\ell^T(X_k,\theta)\cdot \dot{F}_\theta(x_j)|\cdot \mathbb{1}\{|X_k|>c\}\big] \\
\leq\quad & \big(\mathbb{E}\big[||\ell^T(X_0,\theta)||_2^2\big]\big)^{1/2}\cdot ||\dot{F}_\theta||_\infty^2 \cdot c^\gamma \cdot K\cdot \zeta_k \\
& + \big(\mathbb{E}\big[|\ell^T(X_0,\theta)\cdot \dot{F}_\theta(x_i)|^2\big]\big)^{\frac{1}{2}}\big(\mathbb{E}\big[|\ell^T(X_k,\theta)\cdot \dot{F}_\theta(x_j)|^2\cdot \mathbb{1}\{|X_k|>c\}\big]\big)^{\frac{1}{2}} \\
\leq\quad & \big(\mathbb{E}\big[||\ell^T(X_0,\theta)||_2^2\big]\big)^{\frac{1}{2}}\cdot ||\dot{F}_\theta||_\infty^2 \cdot c^\gamma \cdot K\cdot \zeta_k \\
& + \big(\mathbb{E}\big[|\ell^T(X_0,\theta)\cdot \dot{F}_\theta(x_i)|^2\big]\big)^{\frac{1}{2}} \\
& \quad \cdot\big(\mathbb{E}\big[|\ell^T(X_k,\theta)\cdot \dot{F}_\theta(x_j)|^{2\tilde{\delta}}\big]\big)^{\frac{1}{2\tilde{\delta}}}\cdot \big(\mathbb{E}\big[\mathbb{1}\{|X_k|>c\}\big]\big)^{\frac{\tilde{\delta}-1}{2\tilde{\delta}}}
\end{aligned}$$

für $\tilde{\delta} > 1$.

Für δ mit $2+\delta = 2\tilde{\delta}$ gilt somit weiter, für $\delta > 0$

$$\begin{aligned}
\leq\quad & \big(\mathbb{E}\big[||\ell^T(X_0,\theta)||_2^2\big]\big)^{\frac{1}{2}}\cdot ||\dot{F}_\theta||_\infty^2 \cdot c^\gamma \cdot K\cdot \zeta_k \\
& + \big(\mathbb{E}\big[|\ell^T(X_0,\theta)\cdot \dot{F}_\theta(x_i)|^2\big]\big)^{\frac{1}{2}} \\
& \quad \cdot\big(\mathbb{E}\big[|\ell^T(X_k,\theta)\cdot \dot{F}_\theta(x_j)|^{2+\delta}\big]\big)^{\frac{1}{2+\delta}}\cdot \big(\mathbb{E}\big[\mathbb{1}\{|X_k|>c\}\big]\big)^{\frac{\delta/2}{2+\delta}} \\
\leq\quad & K\cdot \big(\mathbb{E}\big[||\ell(X_0,\theta)||_2^2\big]\big)^{\frac{1}{2}}\cdot ||\dot{F}_\theta||_\infty^2 \cdot c^\gamma \cdot \zeta_k \\
& + \big(\mathbb{E}\big[|\ell^T(X_0,\theta)\cdot \dot{F}_\theta(x_i)|^2\big]\big)^{\frac{1}{2}}\cdot \big(\mathbb{E}\big[||\ell^T(X_k,\theta)||_2^{2+\delta}\cdot ||\dot{F}_\theta||_2^{2+\delta}\big]\big)^{\frac{1}{2+\delta}} \\
& \quad \cdot\big(\mathbb{E}\big[\mathbb{1}\{|X_k|>c\}\big]\big)^{\frac{\delta/2}{2+\delta}} \\
=\quad & K\cdot \big(\mathbb{E}\big[||\ell(X_0,\theta)||_2^2\big]\big)^{1/2}\cdot ||\dot{F}_\theta||_\infty^2 \cdot c^\gamma \cdot \zeta_k \\
& + \big(\mathbb{E}\big[|\ell^T(X_0,\theta)\cdot \dot{F}_\theta(x_i)|^2\big]\big)^{\frac{1}{2}}\cdot ||\dot{F}_\theta||_\infty\cdot \big(\mathbb{E}\big[||\ell^T(X_k,\theta)||_2^{2+\delta}\big]\big)^{\frac{1}{2+\delta}} \\
& \quad \cdot\big(\mathbb{P}(|X_k|>c)\big)^{\frac{\delta/2}{2+\delta}} \\
\leq\quad & K\cdot \big(\mathbb{E}\big[||\ell(X_0,\theta)||_2^2\big]\big)^{1/2}\cdot ||\dot{F}_\theta||_\infty^2 \cdot c^\gamma \cdot \zeta_k
\end{aligned}$$

$$+ \left(\mathbb{E}\left[|\ell^T(X_0,\theta)\cdot \dot{F}_\theta(x_i)|^2\right]\right)^{\frac{1}{2}} \cdot ||\dot{F}_\theta||_\infty \cdot \left(\mathbb{E}(||\ell^T(X_k,\theta)||_2^{2+\delta}\right)^{\frac{1}{2+\delta}}$$
$$\cdot \frac{\left(\mathbb{E}[|X_k|^2]\right)^{\frac{\delta/2}{2+\delta}}}{c^{\delta/(2+\delta)}}.$$

Die obigen Abschätzungen erfolgen durch wiederholtes Anwenden der Hölder-Ungleichung und durch Nutzung der Markov-Ungleichung im letzten Schritt.

Minimiert wird der letzte Ausdruck, wenn beide Summanden von derselben Ordnung sind. Dies kann durch die Wahl von c, jener Konstanten, bei welcher $\ell(\cdot,\theta)$ bezüglich des Argumentes gestutzt wird, erreicht werden.

Man erhält hierbei

$$c = \text{const}\cdot\zeta_k^{-\frac{2+\delta}{\gamma(2+\delta)+\delta}}$$

und daher gilt abschließend

$$\begin{aligned} & |\operatorname{cov}(\ell^T(X_0,\theta)\cdot\dot{F}_\theta(x_i), \ell^T(X_k,\theta)\cdot\dot{F}_\theta(x_j))| \\ \leq \quad & \text{const}\cdot\zeta^{\frac{\delta}{\gamma(2+\delta)+\delta}}. \end{aligned} \tag{2.13}$$

Aufgrund der Glättung der Funktion $\ell(\cdot,\theta)$ und der Abschätzung der Wahrscheinlichkeit, dass $|X_k|$ eine bestimmte Schranke nicht überschreitet, erhöht sich in diesem Schritt die Anforderung an die Momente von $\ell(X_0,\theta)$ zu $\mathbb{E}||\ell(X_0,\theta)||_2^{2+\delta}$, wobei δ positiv ist, aber beliebig klein sein kann. Für X_k selbst ergibt sich keine neue Momentenbedingung. Unter der Voraussetzung des exponentiellen Abfallens von ζ_k bleibt die Summierbarkeit der betrachteten Reihe erhalten, aber es ist ersichtlich, dass an dieser Stelle wiederum die Summierbarkeit von $\zeta_k^{\delta/(\gamma(2+\delta)+\delta)}$ ausreichend ist, um die gewünschte Konvergenz zu erzielen.

Insgesamt gilt also für die hinsichtlich der Konvergenz von σ_n^2 betrachtete Reihe, mit (2.8), (2.10), (2.12) und (2.13):

$$\begin{aligned}
&\sum_{k\in\mathbb{Z}} |\operatorname{cov}(T_0, T_k)| \\
\leq\ & \sum_{k\in\mathbb{Z}} \sum_{i=1}^{M} \sum_{j=1}^{M} |c_i| \cdot |c_j| \cdot \text{const} \cdot \big[\zeta_k^{\frac{1}{2}} + \zeta_k^{\frac{1}{1+\gamma}} + \zeta_k^{\frac{1}{3}} + \zeta_k^{\frac{\delta}{\gamma(2+\delta)+\delta}}\big].
\end{aligned}$$

An dieser Stelle wird exponentielles Abfallen von ζ_k nicht benötigt, sondern es muss lediglich die Summierbarkeit der auftretenden Potenzen $\max\{\zeta^{\frac{1}{1+\gamma}}, , \zeta_k^{\frac{1}{3}}, \zeta_k^{\frac{\delta}{\gamma(2+\delta)+\delta}}\}$ von ζ_k gewährleistet sein.

Somit ist die Konvergenz $\sigma_n^2 \to \sigma^2$ und damit die Konvergenz der Kovarianzen von Z_n gegen diejenigen von G gezeigt.

- Sei jetzt $f : \mathbb{R}^u \to \mathbb{R}$ quadratisch integrierbar bzgl. der Verteilung von $(T_{s_1}, ..., T_{s_u})$, $s_1 < ... < s_u < s_u + r = t_1$.
  Dann gilt zunächst

$$\begin{aligned}
&|\operatorname{cov}(f(T_{s_1}, ..., T_{s_u}), T_{t_1})| \\
\leq\ & \sum_{i=1}^{M} |c_i| \underbrace{|\operatorname{cov}(f(T_{s_1}, ..., T_{s_u}), \bar{Y}_{t_1}(x_i))|}_{=:(i)} \\
&+ \sum_{i=1}^{M} |c_i| \underbrace{|\operatorname{cov}(f(T_{s_1}, ..., T_{s_u}), \ell^T(X_{t_1}, \theta) \cdot F_\theta(x_i)|}_{=:(ii)}
\end{aligned}$$

Die beiden Terme werden analog zum obigen Vorgehen analysiert und man erhält im Falle von (i)

$$\begin{aligned}
&|\operatorname{cov}(f(T_{s_1}, ..., T_{s_u}), \bar{Y}_{t_1}(x_i))| \\
\leq\ & |\operatorname{cov}(f(T_{s_1}, ..., T_{s_u}), \widetilde{\bar{Y}}_{t_1}(x_i)| \\
&+ |\operatorname{cov}(f(T_{s_1}, ..., T_{s_u}), \bar{Y}_{t_1}(x_i) - \widetilde{\bar{Y}}_{t_1}(x_i))| \\
\leq\ & \big(\mathbb{E}[f^2(T_{s_1}, ..., T_{s_u})]\big)^{\frac{1}{2}} \cdot (\frac{1}{a} \cdot K \cdot \zeta_r + (C \cdot a)^{\frac{1}{2}}) \\
\leq\ & \big(\mathbb{E}[f^2(T_{s_1}, ..., T_{s_u})]\big)^{\frac{1}{2}} \cdot (K \cdot C)^{\frac{1}{3}} \cdot \zeta_r^{\frac{1}{3}}
\end{aligned}$$

und für (ii), wobei hier die aus der Glättung von $\ell(\cdot, \theta)$ resultierende Lipschitzstetigkeit von ℓ ausgenutzt wird:

$$
\begin{aligned}
& |\operatorname{cov}(f(T_{s_1}, ..., T_{s_u}), \ell^T(X_{t_1}, \theta) \cdot \dot{F}_\theta(x_i)| \\
\leq\ & |\operatorname{cov}(f(T_{s_1}, ..., T_{s_u}), \ell_c^T(X_{t_1}, \theta) \cdot \dot{F}_\theta(x_i)| \\
& + |\operatorname{cov}(f(T_{s_1}, ..., T_{s_u}), (\ell^T(X_{t_1}, \theta) - \ell_c^T(X_{t_1}, \theta)) \cdot \dot{F}_\theta(x_i)|
\end{aligned}
$$

Der erste Term kann nach oben abgeschätzt werden durch

$$
K \cdot \left(\mathbb{E}[f^2(T_{s_1}, ..., T_{s_u})]\right)^{\frac{1}{2}} \cdot ||\dot{F}_\theta||_\infty \cdot c^\gamma \cdot \zeta_k \,.
$$

Der zweite kann weiter abgeschätzt werden mit

$$
\begin{aligned}
& |\operatorname{cov}(f(T_{s_1}, ..., T_{s_u}), (\ell^T(X_{t_1}, \theta) - \ell_c^T(X_{t_1}, \theta)) \cdot \dot{F}_\theta(x_i)| \\
\leq\ & |\mathbb{E}[f(T_{s_1}, ..., T_{s_u}) \cdot (\ell^T(X_{t_1}, \theta) - \ell_c^T(X_{t_1}, \theta)) \cdot \dot{F}_\theta(x_i)]| \\
& + |\mathbb{E}[f(T_{s_1}, ..., T_{s_u})] \cdot \mathbb{E}[\ell^T(X_{t_1}, \theta) - \ell_c^T(X_{t_1}, \theta) \cdot \dot{F}_\theta(x_i)]| \\
\leq\ & 2 \cdot \left(\mathbb{E}[f^2(T_{s_1}, ..., T_{s_u})]\right)^{\frac{1}{2}} \cdot \left(\mathbb{E}[(\ell^T(X_{t_1}, \theta) \cdot \dot{F}_\theta(x_i) \cdot \mathbb{1}\{|X_{t_1}| > c\})^2]\right)^{\frac{1}{2}} \\
\leq\ & 2 \cdot \left(\mathbb{E}[f^2(T_{s_1}, ..., T_{s_u})]\right)^{\frac{1}{2}} \cdot \left(\mathbb{E}[||\ell^T(X_{t_1}, \theta)||_2^{2+\delta}]\right)^{\frac{1}{1+\delta/2} \cdot \frac{1}{2}} \\
& \cdot ||\dot{F}_\theta||_\infty^{2+\delta} \cdot \left(\mathbb{E}[\mathbb{1}\{|X_{t_1}| > c\}]\right)^{\frac{\delta/2}{1+\delta/2} \cdot \frac{1}{2}} \\
\leq\ & 2 \cdot \left(\mathbb{E}[f^2(T_{s_1}, ..., T_{s_u})]\right)^{\frac{1}{2}} \cdot \left(\mathbb{E}[||\ell^T(X_{t_1}, \theta)||_2^{2+\delta}]\right)^{\frac{1}{2+\delta}} \\
& \cdot ||\dot{F}_\theta||_\infty^{2+\delta} \cdot \left(\mathbb{E}[|X_{t_1}|^2]\right)^{\frac{\delta/2}{2+\delta}} \cdot c^{-\frac{\delta}{2+\delta}}
\end{aligned}
$$

Insgesamt gilt

$$
\begin{aligned}
& |\operatorname{cov}(f(T_{s_1}, ..., T_{s_u}), \ell^T(X_{t_1}, \theta) \cdot \dot{F}_\theta(x_i)| \\
\leq\ & \text{const} \cdot \zeta_k^{\frac{\delta}{\gamma(2+\delta)+\delta}} \,,
\end{aligned}
$$

was wiederum durch Ausbalancieren des ersten und des zweiten Summanden erreicht wird.

Somit erhält man

$$
\begin{aligned}
& \operatorname{cov}(f(T_{s_1}, ..., T_{s_u}), T_{t_1}) \\
\leq\ & \sum_{t=1}^{M} |c_t| \cdot \text{const} \cdot \left(\zeta_r^{\frac{1}{3}} + \zeta_k^{\frac{\delta}{\gamma(2+\delta)+\delta}}\right) . \qquad (2.14)
\end{aligned}
$$

- Sei $f : \mathbb{R}^u \to \mathbb{R}$ messbar und beschränkt. Dann gilt

$$
\begin{aligned}
& |\operatorname{cov}(f(T_{s_1}, ..., T_{s_u}), T_{t_1} \cdot T_{t_2})| \\
\leq \quad & \sum_{i=1}^{M} \sum_{j=1}^{M} |c_i||c_j| \underbrace{|\operatorname{cov}(f(T_{s_1}, ..., T_{s_u}), \bar{Y}_{t_1}(x_i) \cdot \bar{Y}_{t_2}(x_j))|}_{=:(i)} \\
& + \sum_{i=1}^{M} \sum_{j=1}^{M} |c_i||c_j| \cdot \underbrace{|\operatorname{cov}\left(f(T_{s_1}, ..., T_{s_u}), \bar{Y}_{t_1}(x_i) \cdot (\ell^T(X_{t_2}, \theta) \cdot \dot{F}_\theta(x_j))\right)|}_{=:(ii)} \\
& + \sum_{j=1}^{M} \sum_{i=1}^{M} |c_i||c_j| \cdot \underbrace{|\operatorname{cov}\left(f(T_{s_1}, ..., T_{s_u}), \bar{Y}_{t_2}(x_j) \cdot (\ell^T(X_{t_1}, \theta)) \cdot \dot{F}_\theta(x_i)\right)|}_{=:(iii)} \\
& + \sum_{i=1}^{M} \sum_{j=1}^{M} |c_i||c_j| \cdot |\operatorname{cov}\left(f(T_{s_1}, ..., T_{s_u}),\right. \\
& \qquad \left. (\ell^T(X_{t_1}, \theta) \cdot \dot{F}_\theta(x_i)) \cdot (\ell^T(X_{t_2}, \theta) \cdot \dot{F}_\theta(x_j))\right)| \,.
\end{aligned}
$$

Da die Funktion $\bar{Y}_{(\cdot)}(x_i) \cdot \bar{Y}_{(\cdot)}(x_j)$ nicht lipschitzstetig ist, wird wieder eine Glättung vorgenommen, hier durch $\widetilde{\bar{Y}}_{(\cdot)}(x_i) \cdot \widetilde{\bar{Y}}_{(\cdot)}(x_j)$.
Diese Funktion ist lipschitzstetig, denn es gilt

$$
\begin{aligned}
& |\widetilde{\bar{Y}}_1(x_i)\widetilde{\bar{Y}}_2(x_j) - \widetilde{\bar{Y}}_3(x_i)\widetilde{\bar{Y}}_4(x_j)| \\
& \leq |\widetilde{\bar{Y}}_1(x_i) \cdot (\widetilde{\bar{Y}}_2(x_j) - \widetilde{\bar{Y}}_4(x_j))| + |\widetilde{\bar{Y}}_4(x_j) \cdot (\widetilde{\bar{Y}}_3(x_i) - \widetilde{\bar{Y}}_1(x_i))| \\
& \leq ||\widetilde{\bar{Y}}_{(\cdot)}(x_i)||_\infty \cdot \frac{1}{a} \cdot |X_2 - X_4| + ||\tilde{Y}_{x_j}||_\infty \cdot \frac{1}{a} \cdot |X_3 - X_1| \\
& \leq \frac{1}{a} \cdot (|X_2 - X_4| + |X_3 - X_1|) \\
& = \frac{1}{a} \cdot ||(X_1, X_2)^T - (X_3, X_4)^T||_1 \leq \frac{1}{a} \cdot \sqrt{2} \cdot ||(X_1, X_2)^T - (X_3, X_4)^T||_2 \,.
\end{aligned}
$$

Hierbei bezeichnen, wie üblich, $|| \cdot ||_2$ die 2−Norm auf $\mathbb{R}^2$ sowie $|| \cdot ||_1$ die 1-Norm auf $\mathbb{R}^2$.

Für den Fall der Funktionen
$\widetilde{\bar{Y}}_{(\cdot)}(x) \cdot \left(\ell_c^T(\cdot, \theta) \cdot \dot{F}_\theta(x)\right)$ und
$(\ell_c^T(\cdot, \theta) \cdot \dot{F}_\theta(x_i)) \cdot (\ell_c^T(\cdot, \theta) \cdot \dot{F}_\theta(x_j))$
gelten ähnliche Überlegungen. Hierbei macht man sich zunutze, dass aufgrund der Lipschitzstetigkeit der Stutzung der Funktion $\ell_c(\cdot, \theta)$ die Werte

dieser Funktion beschränkt sind, denn für ein beliebiges Argument gilt

$$||\ell_c(X,\theta) - \ell_c(0,\theta)||_2 \leq c^\gamma \cdot |X| \leq c^{1+\gamma}, \text{ falls } |X| \leq c \text{ sowie}$$
$$||\ell_c(X,\theta)||_2 = ||\ell(c,\theta)||_2 \text{ bzw. } ||\ell(-c,\theta)||_2 \leq ||\ell_c(0,\theta)||_2 + c^{1+\gamma}, \text{ falls } |X| > c.$$

Damit gilt zum einen

$$\begin{aligned}
& ||\widetilde{\bar{Y}}_1(x_i) \cdot (\ell_c^T(X_2,\theta) \cdot \dot{F}_\theta(x_j)) - \widetilde{\bar{Y}}_3(x_i) \cdot (\ell_c^T(X_4,\theta) \cdot \dot{F}_\theta(x_j))|| \\
\leq\ & ||\tilde{Y}_{(\cdot)}(x_i)||_\infty \cdot c^\gamma \cdot\ ||\dot{F}_\theta||_\infty \cdot\ |X_2 - X_4| + ||\ell_c(\cdot,\theta)||_\infty \cdot ||\dot{F}_\theta||_\infty \cdot\ \frac{1}{a} \cdot\ |X_3 - X_1| \\
\leq\ & ||\dot{F}_\theta||_\infty \cdot\ \max\{c^\gamma, ||\ell_c(\cdot,\theta)||_\infty \cdot \frac{1}{a}\} \cdot (|X_2 - X_4| + |X_1 - X_3|) \\
\leq\ & ||\dot{F}_\theta||_\infty \cdot\ \max\{c^\gamma, ||\ell_c(\cdot,\theta)||_\infty \cdot \frac{1}{a}\} \cdot \sqrt{2} \cdot\ ||(X_1,X_2)^T - (X_3,X_4)^T||_2\,.
\end{aligned}$$

Analog zeigt man

$$\begin{aligned}
& |(\ell_c^T(X_1,\theta) \cdot\ F_\theta(x_i)) \cdot (\ell_c^T(Z_1,\theta) \cdot\ F_\theta(x_j)) \\
& \quad -\ (\ell_c^T(X_2,\theta) \cdot\ F_\theta(x_i)) \cdot (\ell_c^T(Z_2,\theta) \cdot\ F_\theta(x_j))| \\
\leq\ & ||\dot{F}_\theta||_\infty^2 \cdot\ ||\ell_c(\cdot,\theta)||_\infty \cdot c^\gamma \cdot \sqrt{2} \cdot ||(X_1,X_3)^T - (X_2,X_4)^T||_2\,.
\end{aligned}$$

Es wird nun der Ausdruck (i) betrachtet. Es gilt

$$\begin{aligned}
& |\operatorname{cov}(f(T_{s_1},...,T_{s_u}), \bar{Y}_{t_1}(x_i) \cdot \bar{Y}_{t_2}(x_j))| \\
\leq\ & |\operatorname{cov}(f(T_{s_1},...,T_{s_u}), \widetilde{\bar{Y}}_{t_1}(x_i) \cdot \widetilde{\bar{Y}}_{t_2}(x_j))| \\
& \quad + |\operatorname{cov}(f(T_{s_1},...,T_{s_u}), Y_{t_1}(x_i) \cdot Y_{t_2}(x_j) - \widetilde{\bar{Y}}_{t_1}(x_i) \cdot \widetilde{\bar{Y}}_{t_2}(x_j))|.
\end{aligned}$$

Weiter gilt unter Zuhilfenahme von Lemma 2.5.1 (iv)

$$\begin{aligned}
& |\operatorname{cov}(f(T_{s_1},...,T_{s_u}), \bar{Y}_{t_1}(x_i) \cdot \bar{Y}_{t_2}(x_j) - \widetilde{\bar{Y}}_{t_1}(x_i) \cdot \widetilde{\bar{Y}}_{t_2}(x_j))| \\
\leq\ & \mathbb{E}\big[|f(T_{s_1},...,T_{s_u}) \cdot (\bar{Y}_{t_1}(x_i) \cdot \bar{Y}_{t_2}(x_j) - \widetilde{\bar{Y}}_{t_1}(x_i) \cdot \tilde{\bar{Y}}_{t_2}(x_j))|\big] \\
& \quad +\ \mathbb{E}[|f(T_{s_1},...,T_{s_u})|] \cdot \mathbb{E}[|\bar{Y}_{t_1}(x_i) \cdot \bar{Y}_{t_2}(x_j) - \widetilde{\bar{Y}}_{t_1}(x_i) \cdot \widetilde{\bar{Y}}_{t_2}(x_j)|] \\
\leq\ & 2 \cdot ||f||_\infty \cdot \mathbb{E}[|\bar{Y}_{t_1}(x_i) \cdot \bar{Y}_{x_j}(X_{t_2}) - \widetilde{\bar{Y}}_{x_i}(X_{t_1}) \cdot \widetilde{\bar{Y}}_{t_2}(x_j)|] \\
\leq\ & 2 \cdot ||f||_\infty \cdot 3 \cdot C \cdot a.
\end{aligned}$$

Man erhält somit für (i)

$$\begin{aligned}
& |\operatorname{cov}(f(T_{s_1},...,T_{s_u}), \bar{Y}_{t_1}(x_i)\cdot \bar{Y}_{t_2}(x_j))| \\
\leq\ & ||f||_\infty \cdot \frac{\sqrt{2}}{a}\cdot K\cdot \zeta_r + 6\cdot ||f||_\infty \cdot C\cdot a \\
=\ & ||f||_\infty \cdot (6\cdot\sqrt{2}\cdot C\cdot K)^{1/2}\cdot \zeta_r^{1/2}
\end{aligned}$$

wiederum durch Ausbalancieren der beiden Summanden.

<u>zu (ii):</u>
Aufgrund der Voraussetzungen an die Beobachtungen und die Funktion $\ell(\cdot,\theta)$ erhält man

$$\begin{aligned}
& |\operatorname{cov}(f(T_{s_1},...,T_{s_u}), \bar{Y}_{t_1}(x_i)\cdot(\ell^T(X_{t_2},\theta)\cdot \dot{F}_\theta(x_j)))| \\
\leq\ & |\operatorname{cov}(f(T_{s_1},...,T_{s_u}), \widetilde{\bar{Y}}_{t_1}(x_i)\cdot(\ell_c^T(X_{t_2},\theta)\cdot \dot{F}_\theta(x_j)))| \\
& +\ |\operatorname{cov}(f(T_{s_1},...,T_{s_u}), (\bar{Y}_{t_1}(x_i)-\widetilde{\bar{Y}}_{t_1}(x_i))\cdot(\ell^T(X_{t_2},\theta)\cdot \dot{F}_\theta(x_j)))| \\
& +\ |\operatorname{cov}(f(T_{s_1},...,T_{s_u}), (\bar{Y}_{t_1}(x_i)\cdot(\ell^T(X_{t_2},\theta)\cdot \dot{F}_\theta(x_j)-\ell_c^T(X_{t_2},\theta)\cdot \dot{F}_\theta(x_j)))| \\
\leq\ & ||f||_\infty\cdot\ ||\dot{F}_\theta||_\infty\cdot\ \max\{c^\gamma, ||\ell_c(\cdot,\theta)||_\infty\cdot\frac{1}{a}\}\cdot\sqrt{2}\cdot K\cdot\zeta_r \\
& +\ 2\cdot\mathbb{E}\big[|f(T_{s_1},...,T_{s_u})\cdot(\bar{Y}_{t_1}(x_i)-\widetilde{\bar{Y}}_{t_1}(x_i))\cdot\ell^T(X_{t_2},\theta)\cdot\dot{F}_\theta(x_j)|\big] \\
& +\ 2\cdot\mathbb{E}\big[|f(T_{s_1},...,T_{s_u})\cdot\bar{Y}_{t_1}(x_i)\cdot(\ell^T(X_{t_2},\theta)\cdot\dot{F}_\theta(x_j)-\ell_c^T(X_{t_2},\theta)\cdot\dot{F}_\theta(x_j))|\big] \\
\leq\ & ||f||_\infty\cdot||\dot{F}_\theta||_\infty\cdot\max\{c^\gamma, ||\ell_c(\cdot,\theta)||_\infty\cdot\frac{1}{a}\}\cdot\sqrt{2}\cdot K\cdot\zeta_r \\
& +\ 2\cdot||f||_\infty\cdot(F_\theta(x_i+a)-F_\theta(x_i))^{1/2}\cdot\Big(\mathbb{E}[||\ell(X_{t_2},\theta)||_2^2]\Big)^{1/2}\cdot||\dot{F}_\theta(x_j)||_2 \\
& +\ 2\cdot||f||_\infty\cdot\Big(\mathbb{E}[||\ell(X_{t_2},\theta)||_2^2]\Big)^{1/2}\cdot||\dot{F}_\theta(x_j)||_2\cdot\big(\mathbb{E}[|X_{t_2}|^2]\big)^{1/2}\cdot c^{-1} \\
\leq\ & \text{const}\cdot\big(\max\{c^\gamma, ||\ell_c(\cdot,\theta)||_\infty\cdot\frac{1}{a}\}\cdot\zeta_r + a^{1/2}+c^{-1}\big) \\
\leq\ & \text{const}\cdot\big(2\cdot c^{\gamma+1}\cdot\zeta_r + 2a^{1/2}+c^{-1}\big) \\
=\ & \text{const}\cdot\zeta_r^{\frac{1}{4+\gamma}}
\end{aligned}$$

für Konstanten $\tilde{D}$ und $\tilde{C}$, wobei o.B.d.A. angenommen wird, dass $c \geq \max\{1, \ell(0,\theta)\}$. Der letzte Schritt wird wiederum durch Ausbalancieren der drei Summanden erreicht.

Dies gilt analog für (iii).

zu (iv):

In Analogie zu den oben getätigten Ausführungen erhält man folgende Abschätzung:

$$\begin{aligned}
& |\operatorname{cov}(f(T_{s_1},...,T_{s_u}), \ell^T(X_{t_1},\theta)\cdot \dot{F}_\theta(x_i))\cdot(\ell^T(X_{t_2},\theta)\cdot \dot{F}_\theta(x_j)))| \\
\leq\ & ||f||_\infty \cdot ||\dot{F}_\theta||_\infty^2 \cdot\ ||\ell_c(\cdot,\theta)||_\infty \cdot\ c^\gamma \cdot K \cdot \zeta_r \\
& +\ 4\cdot ||f||_\infty \cdot ||\dot{F}_\theta||_\infty^2 \cdot\ \big(\mathbb{E}[||\ell(X_{t_1},\theta)||_2^2]\big)^{\frac{1}{2}} \\
& \quad \cdot\big(\mathbb{E}[||\ell(X_{t_2},\theta)||_2^{2+\delta}]\big)^{\frac{1}{2+\delta}} \cdot\ \big(\mathbb{E}[|X_{t_2}|^2]\big)^{\frac{\delta/2}{2+\delta}} \cdot\ c^{-\frac{\delta}{2+\delta}} \\
\leq\ & \text{const}\cdot \zeta_r^{\frac{\delta}{\gamma(2+\delta)+\delta}}.
\end{aligned}$$

Also gilt für messbare und beschränkte Funktionen $f:\mathbb{R}^u\to\mathbb{R}$

$$\begin{aligned}
& |\operatorname{cov}(f(T_{s_1},...,T_{s_u}), T_{t_1}\cdot T_{t_2})| \\
\leq\ & \text{const}\Big(\sum_{i=1}^M\sum_{j=1}^M |c_i||c_j|\cdot\zeta_r^{\frac{1}{2}} \\
& +\sum_{i=1}^M\sum_{j=1}^M |c_i||c_j|\cdot\zeta_r^{\frac{1}{1+4\gamma}} \\
& +\sum_{i=1}^M\sum_{j=1}^M |c_i||c_j|\cdot\zeta_r^{\frac{\delta}{\gamma(2+\delta)+\delta}},\Big)
\end{aligned} \tag{2.15}$$

was summierbar ist unter der Voraussetzung, dass $\max\{\zeta_r^{1/2}, \zeta_r^{\frac{1}{1+4\gamma}}, \zeta_r^{\frac{\delta}{\gamma(2+\delta)+\delta}}\}$ summierbar ist.

Insgesamt können die unter (2.14) und (2.15) auftretenden Ausdrücke majorisiert werden durch

$$\sqrt{\mathbb{E}(f^2(T_{s_1},...,T_{s_u}))}\cdot\tilde{\zeta}_r \text{ bzw. } ||f||_\infty\cdot\tilde{\zeta}_r,$$

wobei $\tilde{\zeta}_r$ eine summierbare Folge ist.

Damit sind die Voraussetzungen des zentralen Grenzwertsatzes in Neumann und Paparoditis [55] erfüllt und die Konvergenz der endlichdimensionalen Verteilungen des Prozesses

$$\frac{1}{\sqrt{n}}\sum_{k=1}^n z_\theta(X_k,x)$$

gegen diejenigen von $\vec{G} \sim N(0, \Sigma_{x_1,\dots,x_m})$ ist gezeigt. Mit dem Satz von Cramér-Slutsky folgt damit ebenfalls die Konvergenz der endlichdimensionalen Verteilungen von W_n gegen $N(0, \Sigma_{x_1,\dots,x_m})$. $\square$

Gleichgradige Stetigkeit in Wahrscheinlichkeit auf einem Gitter

Nachdem die Konvergenz von $D_2 = |\mathbb{E}f(\widetilde{W}_n) - \mathbb{E}f(\widetilde{G})|$ gezeigt ist, wird nun die Konvergenz von $D_1 = |\mathbb{E}f(W_n) - \mathbb{E}f(\widetilde{W}_n)|$ nachgewiesen.

Hierzu überlegt man sich, dass aufgrund der gleichmäßigen Stetigkeit von f gilt:

$$\forall \mu > 0 : \exists \nu := \nu(\mu) > 0 : ||W_n - \widetilde{W}_n||_\infty \leq \nu \Rightarrow |f(W_n) - f(\widetilde{W}_n)| \leq \mu.$$

Falls im Gegenzug gilt:

$$||W_n - \widetilde{W}_n||_\infty > \nu,$$

so folgt aufgrund der Beschränktheit von f

$$|f(W_n) - f(\widetilde{W}_n)| \leq 2 \cdot ||f||_\infty.$$

Somit kann D_1 folgendermaßen abgeschätzt werden

$$\begin{aligned} D_1 &= |\mathbb{E}f(W_n) - \mathbb{E}f(\widetilde{W}_n)| \leq \mathbb{E}|f(W_n) - f(\widetilde{W}_n)| \\ &\leq \mu \cdot \mathbb{P}(||W_n - \widetilde{W}_n||_\infty \leq \nu) + 2 \cdot ||f||_\infty \cdot \mathbb{P}(||W_n - \widetilde{W}_n||_\infty > \nu). \end{aligned}$$

Letztere Wahrscheinlichkeit muss also klein sein, etwa kleiner als $\eta := \mu/(2 \cdot ||f||_\infty)$, so dass der gesamte Ausdruck kleiner wird als 2μ, für hinreichend großes n.
Zu zeigen ist nun, dass das Gitter $\mathcal{G}_{\nu,\eta} := \{x_0, ..., x_M\}$ in Abhängigkeit von η und ν so gewählt werden kann, dass dies erfüllt ist.
Weiter gilt

$$\begin{aligned} & \mathbb{P}(||W_n - \widetilde{W}_n||_\infty > \nu) \\ = \; & \mathbb{P}(\max_{i=1,\dots,M} \sup_{t \in (x_{i-1}, x_i]} |W_n(x_i) - W_n(t)| > \nu) \\ \leq \; & \mathbb{P}(\max_{i=1,\dots,M} \sup_{t \in (x_{i-1}, x_i]} |U_n(x_i) - U_n(t) \\ & \qquad + F_\theta(x_i) - F_{\hat{\theta}_n}(x_i) - (F_\theta(t) - F_{\hat{\theta}_n}(t))| > \nu) \end{aligned}$$

$$\begin{aligned} \leq \quad & \mathbb{P}(\max_{i=1,\dots,M} \sup_{t\in(x_{i-1},x_i]} |U_n(x_i) - U_n(t)| > \frac{\nu}{2}) \\ & + \mathbb{P}(\max_{i=1,\dots,M} \sup_{t\in(x_{i-1},x_i]} |F_\theta(x_i) - F_{\hat{\theta}_n}(x_i) - (F_\theta(t) - F_{\hat{\theta}_n}(t))| > \frac{\nu}{2}) \end{aligned}$$

(Siehe (2.3) zur verwendeten Bezeichnung.)
Es wird nun zunächst ein Gitter gefunden, sodass die zweite Wahrscheinlichkeit für genügend großes n unter der geforderten Schranke bleibt. Hinsichtlich der ersten Wahrscheinlichkeit wird ein separates Gitter konstruiert, wobei gezeigt wird, dass beide Wahrscheinlichkeiten für genügend großes n unter der geforderten Schranke bleiben, falls als Gitter $\mathcal{G}_{\nu,\eta}$ die Vereinigung der beiden Gitter gewählt wird.

Es wird nun zunächst der zweite Teil des Ausgangsprozesses betrachtet (siehe hierzu (2.3)). Hierbei gilt folgendes Lemma

Lemma 2.2.2 *Für beliebige $\nu > 0$ gilt:*
Es existiert ein Gitter $\tilde{\mathcal{G}} = \{y_0, ..., y_N\}$ mit $-\infty = y_0 < .... < y_{N-1} < y_N = \infty$, sodass für alle $\eta > 0$

$$\mathbb{P}(\max_{1\leq i\leq N} \sup_{t\in(y_{i-1},y_i]} |\sqrt{n} \cdot \left[F_\theta(y_i) - F_{\hat{\theta}_n}(y_i) - (F_\theta(t) - F_{\hat{\theta}_n}(t))\right]| > \frac{\nu}{2}) \leq \frac{\eta}{2}$$

für alle $n \geq n_0(\eta)$. Dies gilt ebenso für jede endliche Obermenge des Gitters.

Beweis:
Gemäß (2.3) gilt

$$\begin{aligned} & \sqrt{n} \cdot (F_{\hat{\theta}_n}(t) - F_\theta(t)) \\ = \quad & \frac{1}{\sqrt{n}} \sum_{k=1}^{n} \ell^T(X_k, \theta) \cdot \dot{F}_\theta(t) + \tilde{\varepsilon}_n(t), \end{aligned}$$

wobei $\sup_{t\in\mathbb{R}} \tilde{\varepsilon}_n(t) = o_{\mathbb{P}}(1)$.

Aus der Anwendung des zentralen Grenzwertsatzes im Abschnitt zuvor folgt, dass auch $\frac{1}{\sqrt{n}} \sum_{k=1}^{n} \ell(X_k, \theta)$ asymptotisch normalverteilt ist und damit $\mathbb{P}$-beschränkt, da für diesen Teil des Prozesses die Voraussetzungen des zentralen Grenzwertsatzes ebenso erfüllt sind.

Es gilt also aufgrund der $\mathbb{P}$-Beschränktheit, dass zu jedem $\eta > 0$ ein $a := a(\eta) \in \mathbb{R}$ existiert und ein $n_0(\eta)$, sodass für $n \geq n_0(\eta)$

$$\mathbb{P}(||\frac{1}{\sqrt{n}} \sum_{k=1}^{n} \ell(X_k, \theta)||_2 > a) \leq \frac{\eta}{4},$$

wobei $n_0(\eta)$ so gewählt ist, dass gleichzeitig erfüllt ist, dass

$$\mathbb{P}(2 \cdot \sup_{t \in \mathbb{R}} |\tilde{\varepsilon}_n(t)| > \frac{\nu}{4}) \leq \frac{\eta}{4}.$$

Wähle nun, für noch zu bestimmendes N, y_1 und y_{N-1} so, dass für alle $t \leq y_1$ gilt

$$||\dot{F}_\theta(t) - \dot{F}_\theta(y_1)||_2 < \frac{\nu}{8a}$$

bzw. für alle $t \geq y_{N-1}$

$$||\dot{F}_\theta(t) - \dot{F}_\theta(y_{N-1})||_2 < \frac{\nu}{8a}.$$

Dies kann gewährleistet werden aufgrund der Endlichkeit von $\lim_{t \to \pm\infty} ||\dot{F}_\theta(t)||_2$.
Wird das Gitter $\tilde{\mathcal{G}}$ um endlich viele Punkte erweitert, wobei die Gitterpunkte des erweiterten Gitters mit $\tilde{y}_0, ..., \tilde{y}_{\tilde{N}}$ bezeichnet werden, so erhält man hier als Abschätzung für $t \leq y_1$ und $\tilde{y}_i \leq y_1$

$$||\dot{F}_\theta(t) - \dot{F}_\theta(\tilde{y}_i)||_2 < \frac{\nu}{4a} \tag{2.16}$$

und für $t \geq \tilde{y}_{\tilde{N}-1}$ und $\tilde{y}_i \geq y_{N-1}$

$$||\dot{F}_\theta(t) - \dot{F}_\theta(\tilde{y}_{N-i})||_2 < \frac{\nu}{4a}. \tag{2.17}$$

Betrachte nun wieder die Wahl der noch nicht festgelegten Gitterpunkte des Gitters $\tilde{\mathcal{G}}$.
Aufgrund der gleichmäßigen Stetigkeit von $\dot{F}_\theta$ existiert ein $\varepsilon := \varepsilon(\frac{\nu}{4a})$, sodass gilt

$$|x - y| \leq \varepsilon \Rightarrow ||\dot{F}_\theta(x) - \dot{F}_\theta(y)||_2 < \frac{\nu}{4a}.$$

Wähle nun die Gitterpunkte $y_2 < ... < y_{N-2}$ (mit $N = N(\nu)$) so, dass

$$|y_i - y_{i-1}| \leq \varepsilon \text{ für } i = 2, ..., N_1.$$

Damit gilt, für $n \geq n_0(\nu)$:

$$\mathbb{P}(\max_{1 \leq i \leq N} \sup_{t \in (y_{i-1}, y_i]} |\sqrt{n} \cdot \left[F_\theta(y_i) - F_{\hat{\theta}_n}(y_i) - (F_\theta(t) - F_{\hat{\theta}_n}(t))\right]| > \frac{\nu}{2})$$

$$
\begin{aligned}
&= \mathbb{P}(\max_{1\le i\le N} \sup_{t\in(y_{i-1},y_i]} |\frac{1}{\sqrt{n}}\sum_{k=1}^{n} \ell^T(X_k,\theta)\cdot(\dot{F}_\theta(y_i)-\dot{F}_\theta(t))| + |\tilde{\varepsilon}_n(y_i)-\tilde{\varepsilon}_n(t)| > \frac{\nu}{2}) \\
&= \mathbb{P}(||\frac{1}{\sqrt{n}}\sum_{k=1}^{n} \ell(X_k,\theta)||_2 \cdot \max_{1\le i\le N} \sup_{t\in(y_{i-1},y_i]} ||\dot{F}_\theta(y_i)-\dot{F}_\theta(t)||_2 > \frac{\nu}{4}) \\
&\quad + \mathbb{P}(\max_{1\le i\le N} \sup_{t\in(y_{i-1},y_i]} |\tilde{\varepsilon}_n(y_i)-\tilde{\varepsilon}_n(t)| > \frac{\nu}{2}) \\
&\le \mathbb{P}(||\frac{1}{\sqrt{n}}\sum_{k=1}^{n} \ell(X_k,\theta)||_2 \cdot \frac{\nu}{4a} > \frac{\nu}{4}) + \mathbb{P}(\sup_{t\in\mathbb{R}} |\tilde{\varepsilon}_n(y_i)-\tilde{\varepsilon}_n(t)| > \frac{\nu}{4}) \\
&= \mathbb{P}(||\frac{1}{\sqrt{n}}\sum_{k=1}^{n} \ell(X_k,\theta)||_2 > a) + \mathbb{P}(2\cdot\sup_{t\in\mathbb{R}} |\tilde{\varepsilon}_n(t)| > \frac{\nu}{4}) \le \frac{\eta}{2}.
\end{aligned}
$$

Hier ist offensichtlich, dass dieselbe Aussage für jede endliche Obermenge des Gitters $\tilde{\mathcal{G}} = \{y_0, ..., y_N\}$ gilt, denn die Abschätzung in der vorletzten Zeile gilt auch dann gemäß (2.16) und (2.17). □

Um jetzt die Existenz eines Gitters $\mathcal{G}_{\nu,\eta} = \{x_0, ..., x_M\}$ zu zeigen, mit $\tilde{\mathcal{G}} \subseteq \mathcal{G}_{\nu,\eta}$, sodass gilt

$$\mathbb{P}(\max_{i=1,...,M} \sup_{t\in(x_{i-1},x_i]} |U_n(x_i)-U_n(t)| > \frac{\nu}{2}) \le \frac{\eta}{2},$$

wird in mehreren Schritten vorgegangen. Zunächst wird ein feineres Gitter konstruiert, sodass die Wahrscheinlichkeit, dass die Zuwächse des Prozesses zwischen den Punkten des feinen Gitters größer als bestimmte Konstanten sind, kleiner als $\frac{\eta}{4}$ ist.

Dann wird gezeigt, dass unter der Voraussetzung, dass die Zuwächse auf dem feinen Gitter klein sind, auch die Zuwächse zwischen Kontinuum und feinem Gitter klein sind und die Zuwächse zwischen feinem Gitter und einem zu wählenden gröberen Gitter klein sind. Diese Vorgehensweise entspricht derjenigen in Neumann und Paparoditis [55] zum Nachweis der stochastischen Äquikontinuität des dort betrachteten Prozesses.

Definiere hierzu eine Folge von Gittern $\tilde{\mathcal{G}}_j = \{z_{j,0}, ..., z_{j,2^j}\}$, $j \in \mathbb{N}$, auf $\bar{\mathbb{R}}$ mit jeweils $2^j + 1$ Gitterpunkten folgendermaßen:

$$
z_{j,k} = \begin{cases} -\infty, & k = 0 \\ F_\theta^{-1}(\frac{k}{2^j}), & 1 \le k \le 2^j - 1 \\ \infty, & k = 2^j \end{cases}
$$

wobei $F_\theta^{-1}(x) := \inf\{t : F_\theta(t) = x\}$. Zwischen diesen Gitterpunkten sind die Zuwächse der Verteilungsfunktion F_θ aufgrund deren Stetigkeit gleich 2^{-j}.
Setze weiterhin $\mathcal{G}_j := \tilde{\mathcal{G}}_j \cup \tilde{\mathcal{G}}$ (mit $\tilde{\mathcal{G}}$ wie in Lemma 2.2.2), wobei bezeichnet werde

$$\mathcal{G}_j =: \{x_{j,0}, ..., x_{j,M_j}\}, \text{ mit } M_j \leq 2^j + 1 + N.$$

Offensichtlich gilt für $j_1 \leq j_2$, dass $\mathcal{G}_{j_1} \subseteq \mathcal{G}_{j_2}$. J_n wird nun so gewählt, dass gilt

$$2^{J_n - 1} < n \leq 2^{J_n}.$$

Für noch festzulegendes J_0 werden nun die Gitter $\mathcal{G}_{J_0}, ..., \mathcal{G}_{J_n}$ betrachtet.
Hierzu werden zunächst für jedes Gitter $\mathcal{G}_{J_0+1}, ..., \mathcal{G}_{J_n}$ zugehörige Zuwachshöhen λ_j definiert:

$$\lambda_j := D \cdot 2^{-\alpha j}$$

Folgendes Lemma besagt nun, dass höchstens mit einer Wahrscheinlichkeit, welche beliebig klein gewählt werden kann, die Zuwächse von U_n auf $\mathcal{G}_j$, $j = J_0 + 1, ..., J_n$, größer als λ_j sind, für jeweils geeignet gewähltes D und α.

Lemma 2.2.3
Es existieren α und $D := D(\eta)$, sodass für $\lambda_j := D \cdot 2^{-\alpha j}$ gilt:

$$\mathbb{P}(\max_{j \in \{J_0+1,...,J_n\}, k=1,...,M_j} |U_n(x_{j,k-1}) - U_n(x_{j,k})| > \lambda_j) \leq \frac{\eta}{2}. \tag{2.18}$$

Beweis:
Zum Beweis der Aussage wird folgendes Hilfsresultat benötigt, welches besagt, wie die vierten Momente der Zuwächse des Prozesses U_n beschränkt werden können (siehe Lemma 2.5.2). Hierbei gelte $\mathbb{P}^{X_0}((x,y]) := P(X_0 \in (x,y])$:

$$\begin{aligned} \mathbb{E}[|U_n(x) - U_n(y)|^4] &\leq C_\psi \cdot \big(\mathbb{P}^{X_0}((x,y]) + \frac{1}{n}\big)^{1+\psi} \\ &= (F_\theta(y) - F_\theta(x) + \frac{1}{n})^{1+\psi}. \end{aligned} \tag{2.19}$$

Es gilt zunächst unter Anwendung der Bonferroni- und der Markov-Ungleichung:

$$\begin{aligned} &\mathbb{P}(\max_{j \in \{J_0+1,...,J_n\}, k \in \{1,...,M_j\}} |U_n(x_{j,k}) - U_n(x_{j,k-1})| > \lambda_j) \\ \leq &\sum_{j=J_0+1}^{J_n} \sum_{k=1}^{M_j} \mathbb{P}(|U_n(x_{j,k}) - U_n(x_{j,k-1})| > \lambda_j) \end{aligned}$$

$$\leq \sum_{j=J_0+1}^{J_n} \sum_{k=1}^{M_j} C_\psi \cdot \frac{\mathbb{E}[|U_n(x_{j,k}) - U_n(x_{j,k-1})|^4]}{D^4 \cdot 2^{-4\alpha j}}$$

und unter Anwendung des Hilfsresultates (2.19)

$$\begin{aligned}
&\leq \sum_{j=J_0+1}^{J_n} \sum_{k=1}^{M_j} C_\psi \cdot \frac{(F_\theta(x_{j,k}) - F_\theta(x_{j,k-1}) + \frac{1}{n})^{1+\psi}}{D^4 \cdot 2^{-4\alpha j}} \\
&\leq \sum_{j=J_0+1}^{J_n} \frac{C_\psi \cdot 2^{2\psi - \psi j}}{D^4 \cdot 2^{-4\alpha j}} \sum_{k=1}^{M_j} (F_\theta(x_{j,k}) - F_\theta(x_{j,k-1}) + \frac{1}{n}),
\end{aligned}$$

wobei im vorigen Schritt ausgenutzt wurde, dass $1/n < 2^{1-J_n} \leq 2^{1-j}$, für $j = J_0 + 1, ..., J_n$,

$$\begin{aligned}
&\leq \frac{C_\psi}{D^4} \cdot 2^{2\psi} \sum_{j=J_0+1}^{J_n} 2^{(4\alpha-\psi)j} (1 + \underbrace{\frac{2^j}{n}}_{\leq 2^{J_n}/n} + \frac{N}{n}) \\
&\leq (1 + \frac{2^{J_n} + N}{n}) \cdot \frac{C_\psi}{D^4} \cdot 2^{2\psi} \cdot 2^{(4\alpha-\psi)(J_0+1)} \cdot \sum_{j=0}^{J_n - J_0 - 1} 2^{(4\alpha-\psi)j} \qquad (2.20) \\
&\leq (1 + 2 \cdot \frac{2^{J_n - 1} + \frac{N}{2}}{n}) \cdot \frac{C_\psi}{D^4} \cdot 2^{2\psi} \cdot 2^{(4\alpha-\psi)(J_0+1)} \cdot S_{\alpha,\psi} \\
&\leq \frac{\eta}{2},
\end{aligned}$$

was erfüllt ist, falls α so gewählt wird, dass $4\alpha < \psi$ und damit $S_{\alpha,\psi} < \infty$ gilt, und D genügend groß gewählt wird. □

Das nächste Lemma besagt nun, dass die Zuwächse des Prozesses zwischen einem beliebigen Punkt $t \in \mathcal{G}_j$, $j \in \{J_0 + 1, ..., J_n\}$ und dem t am nächsten gelegenen Gitterpunkt aus $\mathcal{G}_{J_0}$ für geeignetes J_0 unter einer beliebig kleinen aber festen Schranke bleiben, falls die Zuwächse innerhalb der Punkte der Gitter $\mathcal{G}_j$, $j = J_0 + 1, ..., J_n$, die Schranken λ_j nicht überschreiten. Dies gilt für hinreichend großes J_0, welches von der Größe der Schranken λ_j und damit von den Konstanten D und α abhängt.

Lemma 2.2.4
Falls gilt:

$$\max_{j\in\{J_0+1,\dots,J_n\},k=1,\dots,M_j} |U_n(x_{j,k-1}) - U_n(x_{j,k})| \leq \lambda_j \tag{2.21}$$

mit $\lambda_j = D \cdot 2^{-\alpha j}$, *(*$\alpha$, D, *sodass die Bedingungen für Lemma 1 erfüllt sind), dann existiert ein* $J_0 := J_0(\nu, \alpha, D)$, *sodass für* $t \in \mathcal{G}_j$, $j \in \{J_0+1, ..., J_n\}$, *mit* $t \in (x_{J_0,i-1}, x_{J_0,i}]$

$$|U_n(x_{J_0,i}) - U_n(t)| \leq \frac{\nu}{4}. \tag{2.22}$$

Beweis:
Betrachte zunächst den Fall, dass das Gitter $\mathcal{G}_j$ nur aus Punkten des dyadischen Gitters $\tilde{\mathcal{G}}_j$ besteht.
Sei $t := x_{j,k}$ ein Punkt beliebig auf $\mathcal{G}_j$, für $j \geq J_0 + 1$.
Dann existiert ein $i \in \{1, ..., 2^{J_0}\}$ mit $t \in (x_{J_0,i-1}, x_{J_0,i}]$.
Für $t = x_{J_0,i}$ gilt $|U_n(x_{J_0,i}) - U_n(t)| = 0$.
Falls $t < x_{J_0,i}$, so existieren Intervalle $[x_{j_l,k_l-1}, x_{j_l,k_l}]$ mit $j_l > J_0$ und $l = 1, ..., L$, wobei die j_l paarweise verschieden sind, sodass:
$[t, x_{J_0,i}] = \bigcup_{l=1}^{L} [x_{j_l,k_l-1}, x_{j_l,k_l}]$, d.h. $[t, x_{J_0,i}]$ kann als Vereinigung von Intervallen aus Gittern $\mathcal{G}_j$, $j \in \{J_0+1, ..., J_n\}$ paarweise verschiedener Feinheit dargestellt werden. Daher kann abgeschätzt werden:

$$\begin{aligned}
& |U_n(x_{J_0,i}) - U_n(t)| \\
\leq\ & \sum_{l=1}^{L} |U_n(x_{j_l,k_l}) - U_n(x_{j_l,k_l-1})| \leq \sum_{l=1}^{L} \lambda_{j_l} \\
=\ & D \cdot \sum_{l=1}^{L} 2^{-\alpha_{j_l}} \leq D \cdot \sum_{j=J_0+1}^{J_n} 2^{-\alpha j} \leq 2^{-\alpha J_0} \cdot D \cdot \sum_{j=1}^{J_n} 2^{-\alpha j} \\
\leq\ & \frac{\nu}{4},
\end{aligned}$$

falls J_0 genügend groß gewählt wird.
Sind nun noch weitere Gitterpunkte zugelassen, so erweitert sich die obige Vereinigung von Intervallen höchstens um ein Intervall, welches die Entfernung von $x_{j,k}$ zum nächstgelegenen Gitterpunkt des dyadischen Gitters in $(x_{J_0,i-1}, x_{J_0,i}]$ beschreibt. In obiger Summe ergänzt sich dann ein Summand der Größe $2^{-\alpha J_n}$ und

J_0 kann wiederum entsprechend gewählt werden. $\square$

Im nächsten Schritt wird nun gezeigt, dass die Zuwächse des Prozesses U_n zwischen einem beliebigen Punkt t aus $\mathbb{R}$ und dem t am nächsten gelegenen Gitterpunkt aus $\mathcal{G}_{J_n}$ unter einer vorgegebenen Schranke bleiben, falls n genügend groß gewählt wird und die Zuwächse zwischen den Gitterpunkten der Gitter $\mathcal{G}_j$ kleiner oder gleich λ_j sind.

Lemma 2.2.5
Falls gilt:

$$\max_{j\in\{J_0+1,\dots,J_n\},k=1,\dots,M_j} |U_n(x_{j,k-1}) - U_n(x_{j,k})| \leq \lambda_j \tag{2.23}$$

mit $\lambda_j = D\cdot 2^{-\alpha j}$, ($\alpha$, D, sodass die Bedingungen für Lemma 1 erfüllt sind), dann existiert ein $n_1 := n_1(D,\alpha,\nu)$ und ein Gitterpunkt $x_{J_n}(t) \in \mathcal{G}_{J_n}$, sodass für $n \geq n_1$ gilt

$$\max_{i=1,\dots,M_{J_0}} \sup_{t\in(x_{J_0,i-1},x_{J_0,i}]} |U_n(t) - U_n(x_{J_n}(t))| \leq \frac{\nu}{4}.$$

Beweis:
Sei jetzt t beliebig aus $\mathbb{R}$. Dann gilt: $\exists i(t) : t \in (x_{J_n,i(t)-1}, x_{J_n,i(t)}]$ und es gilt:

$$\begin{aligned}
& U_n(x_{J_n,i(t)-1}) - \frac{1}{\sqrt{n}}\sum_{k=1}^n \mathbb{P}(x_{J_n,i(t)-1} < X_k \leq x_{J_n,i(t)}) \\
= \; & \frac{1}{\sqrt{n}}\Big(\sum_{k=1}^n \mathbb{1}\{X_k \leq x_{J_n,i(t)-1}\} - F_\theta(x_{J_n,i(t)-1}) - \mathbb{P}(x_{J_n,i(t)-1} < X_k \leq x_{J_n,i(t)})\Big) \\
\leq \; & \frac{1}{\sqrt{n}}\Big(\sum_{k=1}^n \mathbb{1}\{X_k \leq x_{J_n,i(t)-1}\} - F_\theta(x_{J_n,i(t)-1}) - \mathbb{P}(x_{J_n,i(t)-1} < X_k \leq t)\Big) \\
= \; & \frac{1}{\sqrt{n}}\Big(\sum_{k=1}^n \mathbb{1}\{X_k \leq t\} - F_\theta(t)\Big) = U_n(t) \\
\leq \; & \frac{1}{\sqrt{n}}\Big(\sum_{k=1}^n \mathbb{1}\{X_k \leq x_{J_n,i(t)}\} - \big(F_\theta(x_{J_n,i(t)}) - \mathbb{P}(x_{J_n,i(t)-1} < X_k \leq x_{J_n,i(t)})\big)\Big) \\
= \; & U_n(x_{J_n,i(t)}) + \frac{1}{\sqrt{n}}\sum_{k=1}^n \mathbb{P}(x_{J_n,i(t)-1} < X_k \leq x_{J_n,i(t)}).
\end{aligned}$$

Man erhält also

$$U_n(x_{J_n,i(t)-1}) - \frac{1}{\sqrt{n}}\sum_{k=1}^n \mathbb{P}(x_{J_n,i(t)-1} < X_k \leq x_{J_n,i(t)})$$

$$\begin{aligned} &\leq U_n(t) \\ &\leq U_n(x_{J_n,i(t)}) + \frac{1}{\sqrt{n}} \sum_{k=1}^{n} \mathbb{P}(x_{J_n,i(t)-1} < X_k \leq x_{J_n,i(t)}). \end{aligned}$$

Wegen

$$\frac{1}{\sqrt{n}} \sum_{k=1}^{n} \mathbb{P}(x_{J_n,i(t)-1} < X_k \leq x_{J_n,i(t)}) \leq \sqrt{n} \cdot 2^{1-J_n} \leq \frac{2}{\sqrt{n}}$$

folgt damit

$$-\frac{2}{\sqrt{n}} \leq U_n(t) - U_n(x_{J_n,i(t)-1}) \leq U_n(x_{J_n,i(t)}) - U_n(x_{J_n,i(t)-1}) + \frac{2}{\sqrt{n}}$$

und aufgrund von $|U_n(x_{J_n,i(t)}) - U_n(x_{J_n,i(t)-1})| \leq \lambda_{J_n}$ gilt

$$\begin{aligned} & -\frac{2}{\sqrt{n}} \leq U_n(t) - U_n(x_{J_n,i(t)-1}) \leq \lambda_{J_n} + \frac{2}{\sqrt{n}} \\ \Rightarrow \quad & -(\frac{2}{\sqrt{n}} + \lambda_{J_n}) \leq U_n(t) - U_n(x_{J_n,i(t)-1}) \leq \lambda_{J_n} + \frac{2}{\sqrt{n}} \\ \Rightarrow \quad & |U_n(t) - U_n(x_{J_n,i(t)-1})| \leq \lambda_{J_n} + \frac{2}{\sqrt{n}} \leq \frac{\nu}{4} \end{aligned}$$

für genügend großes n, welches aufgrund der Wahl der Gitterpunkte aus $\mathcal{G}_{J_n}$ nicht von t abhängig ist.
Bezeichne mit $x_{J_n}(t)$ den Punkt $x_{J_n,i(t)-1}$.
Da t beliebig gewählt war, gilt somit

$$\sup_{t \in \mathbb{R}} |U_n(t) - U_n(x_{J_n(t)})| \leq \frac{\nu}{4},$$

woraus die Behauptung folgt. □

Nun wird zusammengefasst, dass mit der Wahl von J_0 aus Lemma (2.2.4) der Prozess U_n für genügend großes n gut auf dem groben Gitter $\mathcal{G}_{J_0}$ approximiert werden kann.

Lemma 2.2.6 *Es gelte $\mathcal{A}$. Dann existiert ein Gitter $\mathcal{G}_{\nu,\eta}$ mit $-\infty = x_0 < ... < x_M = \infty$, sodass für $n \geq n_1(\nu, \eta)$ gilt:*

$$\mathbb{P}(\max_{1 \leq i \leq M} \sup_{t \in (x_{i-1}, x_i]} |U_n(x_i) - U_n(t)| > \frac{\nu}{2}) \leq \frac{\eta}{2}.$$

Beweis:
Setze nun $\mathcal{G}_{J_0} =: \{x_0, ..., x_M\}$, λ_j wie oben, wobei D, α und J_0 so gewählt seien, dass die Aussagen der vorangegangenen Lemmata gelten.
Damit gilt insgesamt für $n \geq n_1$:

$$\begin{aligned}
& \mathbb{P}(\max_{1\leq i\leq M} \sup_{t\in(x_{i-1},x_i]} |U_n(x_i) - U_n(t)| > \frac{\nu}{2}) \\
\leq\ & \mathbb{P}(\max_{1\leq i\leq M} \sup_{t\in(x_{i-1},x_i]} |U_n(x_i) - U_n(x_{J_n}(t))| + |U_n(x_{J_n}(t)) - U_n(t)| > \frac{\nu}{2}) \\
\leq\ & \mathbb{P}(\{\max_{1\leq i\leq M} \sup_{t\in(x_{i-1},x_i]\cap\mathcal{G}_{J_n}} |U_n(x_i) - U_n(t))| > \frac{\nu}{4}\} \\
& \qquad \cup \{\max_{1\leq i\leq M} \sup_{t:\, t\in(x_{i-1},x_i]} |U_n(x_{J_n}(t)) - U_n(t)| > \frac{\nu}{4}\}) \\
=\ & 1 - \mathbb{P}(\{\max_{1\leq i\leq M} \sup_{t\in(x_{i-1},x_i]\cap\mathcal{G}_{J_n}} |U_n(x_i) - U_n(t)| \leq \frac{\nu}{4}\} \\
& \qquad \cap \{\max_{1\leq i\leq M} \sup_{t:\, (x_{i-1},x_i]} |U_n(x_{J_n}(t)) - U_n(t)| \leq \frac{\nu}{4}\}) \\
\leq\ & 1 - \mathbb{P}(\max_{\substack{j=J_0+1,..,J_n \\ k=1,...,2^j}} \mathbb{P}\big(|U_n(x_{j,k}) - U_n((x_{j,k-1})| \leq \lambda_j\big) \\
\leq\ & 1 - (1 - \frac{\eta}{2}) = \frac{\eta}{2}.
\end{aligned}$$

□

Setze nun abschließend

$$\mathcal{G}_{\nu,\eta} := \mathcal{G}_{J_0}.$$

Somit gilt für $\mathcal{G}_{\nu,\eta}$, mit $n \geq \max\{n_0, n_1\}$

$$\begin{aligned}
& \mathbb{P}(\max_{1\leq i\leq M} \sup_{t\in(x_{i-1},x_i]} |W_n(x_i) - W_n(t)| > \nu) \\
\leq\ & \mathbb{P}(\max_{1\leq i\leq M} \sup_{t\in(x_{i-1},x_i]} |U_n(x_i) - U_n(t)| > \frac{\nu}{2}) \\
& + \mathbb{P}(\max_{1\leq i\leq M} \sup_{t\in(x_{i-1},x_i]} |\frac{1}{\sqrt{n}} \sum_{k=1}^{n} \ell^T(X_k,\theta) \cdot (\dot{F}_\theta(x_i) - \dot{F}_\theta(t))| + |\tilde{\varepsilon}_n(t)| > \frac{\nu}{2}) \\
\leq\ & \eta
\end{aligned}$$

und somit ist die stochastische Äquikontinuität von W_n nachgewiesen. □

Approximierbarkeit des Grenzprozesses auf einem Gitter und Stetigkeit

Um die Konvergenz des Terms $|\mathbb{E}f(\widetilde{G}) - \mathbb{E}f(G)|$ gegen Null zu zeigen, ist nun die Existenz eines Grenzelementes $G \in D(\bar{\mathbb{R}})$ zu zeigen, welches ebenso wie W_n auf einem Gitter approximiert werden kann.
Es muss dann ebenso wie im letzten Schritt gezeigt werden, dass für alle η und ν ein Gitter $\{x_0, ..., x_M\}$ existiert, sodass

$$\mathbb{P}\big(\max_{1\leq i\leq M} \sup_{t\in(x_{i-1},x_i]} |G(x_i) - G(t)| > \delta\big) \leq \eta.$$

Die Argumentation hier ist analog zu derjenigen in Neumann und Paparoditis [55].
Zunächst weiß man aufgrund der Approximierbarkeit von W_n auf einem Gitter, dass eine Folge von Gittern $\mathcal{G}^{(N)} := \{x_0, ..., x_{M_N}\}$ mit $\mathcal{G}^{(N)} \subseteq \mathcal{G}^{(N+1)}$ existiert, die folgendes erfüllt:

$$\mathbb{P}\big(\max_{1\leq i\leq M_N} \sup_{t\in(x_{i-1},x_i]} |W_n(x_i) - W_n(t)| > \frac{1}{N}\big) \leq \frac{1}{N}\,, \tag{2.24}$$

falls $n \geq n_N$.
Bezeichnet man mit $\mathcal{G}^{(\infty)}$ die Vereinigung aller Gitter $\mathcal{G}^{(N)}$, so existiert gemäß dem Theorem von Kolmogorov (s. Bauer [8]) ein reellwertiger stochastischer Prozess $\widetilde{G}$, definiert auf $\mathcal{G}^{(\infty)}$, dessen endlichdimensionale Verteilungen zentrierte Normalverteilungen mit Kovaranzmatrix $(\Sigma_{x_1,...,x_n}(i,k))_{1\leq i,k\leq n}$ sind.
Aufgrund der Konvergenz der endlichdimensionalen Verteilungen von W_n gegen diejenigen von $\widetilde{G}$ gilt auch für jedes beliebige N':

$$\begin{aligned}&\mathbb{P}\big(\max_{1\leq i\leq M_N} \max_{t\in(x_{i-1},x_i]\cap\mathcal{G}^{(N')}} |W_n(x_i) - W_n(t)| > \frac{1}{N}\big)\\ &\overset{n\to\infty}{\to} \mathbb{P}\big(\max_{1\leq i\leq M_N} \max_{t\in(x_{i-1},x_i]\cap\mathcal{G}^{(N')}} |\widetilde{G}(x_i) - \widetilde{G}(t)| > \frac{1}{N}\big)\,.\end{aligned}$$

Aufgrund von Eigenschaft (2.24) gilt $\forall N' \geq N$ und $n \geq n_N$ auch

$$\mathbb{P}\big(\max_{1\leq i\leq M_N} \max_{t\in(x_{i-1},x_i]\cap\mathcal{G}^{(N')}} |W_n(x_i) - W_n(t)| > \frac{1}{N}\big) \leq \frac{1}{N}$$

und damit folgt auch für den Grenzprozess

$$\mathbb{P}\big(\max_{1\leq i\leq M_N} \max_{t\in(x_{i-1},x_i]\cap\mathcal{G}^{(N')}} |\widetilde{G}(x_i) - \widetilde{G}(t)| > \frac{1}{N}\big) \leq \frac{1}{N} \quad \forall N' \geq N.$$

Aufgrund der Stetigkeit von unten des Maßes $\mathbb{P}$ gilt auch

$$\mathbb{P}\big(\max_{1\le i\le M_N}\max_{t\in(x_{i-1},x_i]\cap\mathcal{G}^{(\infty)}}|\widetilde{G}(x_i)-\widetilde{G}(t)|>\frac{1}{N}\big)\le\frac{1}{N}\quad\forall N'\ge N.$$

Zu zeigen bleibt nun noch, dass diese Relation erhalten bleibt, wenn t beliebig aus $(x_{i-1},x_i]$ zugelassen ist.
Hierzu muss der auf $\mathcal{G}^{(\infty)}$ definierte Prozess $\widetilde{G}$ geeignet zu einem Prozess auf $\bar{\mathbb{R}}$ fortgesetzt werden.
Sei t beliebig aus $\mathbb{R}$. Dann existiert eine Folge

$$x_{i_N-1}\le t\le x_{i_N}\quad(x_{i_N})_{N\in\mathbb{N}},\, x_{i_N}\in\mathcal{G}^{(N)},$$

sodass gilt, dass $x_{i_N-1}\le x_{i_{N+1}-1}$ $\forall N$ und daher besitzt x_{i_N-1} einen Grenzwert $x_{i_\infty-1}$ aus $\mathcal{G}^{(\infty)}$, wobei gilt $x_{i_\infty-1}\in[x_{i_N-1},x_{i_N}]$, $\forall N\in\mathbb{N}$. Es gilt daher

$$\mathbb{P}\big(|\widetilde{G}(x_{i_\infty-1})-\widetilde{G}(x_{i_N})|>\frac{1}{N}\big)\le\frac{1}{N}\,,$$

woraus folgt

$$|\widetilde{G}(x_{i_\infty-1})-\widetilde{G}(x_{i_N})|\overset{N\to\infty}{\to}0\text{ in }\mathbb{P}\text{-Wahrscheinlichkeit.}$$

Setze $G(t):=G(x_{i_\infty-1})$. Daher gilt für den so definierten Prozess G:

$$\mathbb{P}\big(\max_{1\le i\le M_N}\max_{t\in\mathbb{R}:t\in(x_{i-1},x_i]}|\widetilde{G}(x_i)-\widetilde{G}(t)|>\frac{1}{N}\big)\le\frac{1}{N}\quad\forall N.\tag{2.25}$$

Dies bedeutet zum einen die gleichgradige Stetigkeit in Wahrscheinlichkeit auf der Folge von Gittern $\mathcal{G}^{(N)}$ und zum anderen auch die Stetigkeit der Pfade des Prozesses mit Wahrscheinlichkeit eins, denn wären die Pfade von G nicht fast sicher stetig, dann würde gelten

$$\exists\varepsilon>0:\mathbb{P}(\exists x\in\mathbb{R}:\exists N:\,\forall\delta>0:\exists y_\delta:\,|x-y_\delta|\le\delta,\,|G(x)-G(y)|>\frac{1}{N})>\varepsilon$$

und, für hinreichend großes N'

$$\mathbb{P}(\exists x\in\mathbb{R}:\exists N':\,\forall\delta>0:\exists y:\,|x-y_\delta|\le\delta,\,|G(x)-G(y_\delta)|>\frac{2}{N'})>\frac{1}{N'}.\tag{2.26}$$

Falls $x\in\mathcal{G}^{(N')}$, so wird letztere Annahme durch Ausnutzung von (2.25) unmittelbar zum Widerspruch geführt. Gilt $x\notin\mathcal{G}^{(N')}$, so existiert ein Punkt $x_{i_{N'}}$ aus $\mathcal{G}^{(N')}$, sodass für hinreichend kleines δ gilt, dass $x,y_\delta\in(x_{i_{N'-1}},x_{i_{N'}}]$ und damit

$|x - y_\delta| \leq |x - x_{i_{N'}}| + |y_\delta - x_{i_{N'}}|$. Wiederum unter Verwendung von (2.25) ist damit die Nichtgültigkeit von (2.26) gezeigt.
Somit ist die fast sichere Stetigkeit der Pfade des Prozesses G gezeigt.

Es ist also die Verteilungskonvergenz des geschätzten empirischen Prozesses gegen einen zentrierten Gaußprozess mit fast sicher stetigen Pfaden gezeigt. □

2.2.4 Konvergenz des Originalprozesses unter fixierten Alternativen

Um sicherzustellen, dass Tests basierend auf dem geschätzten empirischen Prozess nichttriviale Mächtigkeit besitzen, betrachten wir nun die Grenzverteilung des Prozesses W_n, falls die Nullhypothese nicht erfüllt ist.
Wir analysieren die Situation, dass die gemeinsame Verteilungsfunktion F der Beobachtungen $X_1, ..., X_n$ nicht in der parametrischen Klasse $\mathcal{F}$ liegt:

$$\mathcal{H}_1 : \ F \notin \mathcal{F}.$$

Um das Verhalten des Prozesses W_n für n gegen unendlich analysieren zu können, müssen folgende Voraussetzungen hinsichtlich des Parameterschätzers $\hat{\theta}_n$ und der Beobachtungen $X_1, .., X_n$ getroffen werden:
Zunächst fordern wir, dass $\hat{\theta}_n$ gegen einen fixierten Wert $\bar{\theta}$ konvergiert, auch wenn die Randverteilung der $X_1, ..., X_n$ nicht in der parametrischen Klasse $\mathcal{F}$ liegt:

$$\hat{\theta}_n \xrightarrow{\mathbb{P}} \bar{\theta}.$$

Hinsichtlich der Abhängigkeitsstruktur des Prozesses $X_1, ..., X_n$ nehmen wir des weiteren an, dass $\mathcal{A}$ (i) erfüllt ist.

Aus der Gültigkeit von $\mathcal{H}_1$ ergibt sich, dass eine Funktion g existiert, welche ungleich der Nullfunktion ist, sodass

$$F = F_{\bar{\theta}} + g.$$

Der Prozess W_n besitzt nun in Analogie zu (2.3) die folgende Darstellung

$$W_n(x) \ = \ \sqrt{n} \cdot \left[F_n(x) - F_{\bar{\theta}}(x) - g(x)\right] + \sqrt{n} \cdot \left[F_{\bar{\theta}}(x) - F_{\hat{\theta}_n}(x)\right] + \sqrt{n} \cdot g(x).$$

$$
\begin{aligned}
&= \sqrt{n} \cdot [F_n(x) - F_{\bar{\theta}}(x) - g(x)] + \sqrt{n} \cdot [(\bar{\theta} - \hat{\theta}_n)^T \cdot \dot{F}_{\bar{\theta}}(x)] \\
&\quad + \sqrt{n} \cdot (\bar{\theta} - \hat{\theta}_n)^T \cdot [\dot{F}_{\tilde{\theta}_n}(x) - \dot{F}_{\bar{\theta}}(x)] + \sqrt{n} \cdot g(x) \qquad (2.27)
\end{aligned}
$$

und somit folgt

$$
\begin{aligned}
\frac{1}{\sqrt{n}} \cdot W_n(x) - g(x) &= [F_n(x) - F_{\bar{\theta}}(x) - g(x)] + [(\bar{\theta} - \hat{\theta}_n)^T \cdot \dot{F}_{\bar{\theta}}(x)] \\
&\quad + (\bar{\theta} - \hat{\theta}_n)^T \cdot [\dot{F}_{\tilde{\theta}_n}(x) - \dot{F}_{\bar{\theta}}(x)].
\end{aligned}
$$

Aufgrund der Voraussetzungen konvergiert

$$
\frac{1}{\sqrt{n}} \cdot W_n(x) - g(x)
$$

in Wahrscheinlichkeit gegen Null, gleichmäßig für $x \in \bar{\mathbb{R}}$.

Der Wert einer auf $\frac{1}{\sqrt{n}} \cdot W_n$ basierenden Teststatistik wird also üblicherweise in Wahrscheinlichkeit gegen eine positive Konstante konvergieren, weswegen ebenfalls die Wahrscheinlichkeit, dass der Wert der Teststatistik eine beliebige positive Konstante überschreitet, gegen 1 konvergiert.

2.3 Imitation des Originalprozesses durch einen Bootstrap-Prozess

Das Ziel beim Bootstrapping ist es nun, anhand der beobachteten Daten $X_1, ..., X_n$ neue Daten zu generieren, und anhand dieser neu generierten Daten den auf der Originalstichprobe beruhenden geschätzten empirischen Prozess so zu imitieren, dass der imitierte Prozess bei Gültigkeit der Nullhypothese in Wahrscheinlichkeit Verteilungskonvergenz gegen denselben Grenzprozess wie der Originalprozess aufweist. Falls $\mathcal{H}_0$ nicht gilt, soll das Bootstrap-Verfahren dennoch die Situation unter $\mathcal{H}_0$ imitieren. Dies sichert, dass der verwendete Test nichttriviale Mächtigkeit besitzt und somit in der Lage ist, Alternativen zu erkennen (siehe hierzu Hall und Wilson [35]).

Um die verwendete Begriffsbildung zu präzisieren, werden folgende Annahmen getroffen:

Die Bootstrap-Stichprobe $X_1^*, ..., X_n^* := X_{n1}, ..., X_{nn}$ stelle hierbei ein Dreiecksschema von Zufallsgrößen dar. Definiert sei dieses ebenso wie der Prozess der Beobachtungen $(X_t)_{t\in\mathbb{Z}}$ und die Prozesse W_n auf einem hinreichend großen gemeinsamen Wahrscheinlichkeitsraum. Vereinfachend werde dieser, ebenso wie im ersten Teil, mit $(\Omega, \mathcal{A}, \mathbb{P})$ bezeichnet. Ebenfalls seien auf diesem Raum die noch näher zu beschreibenden Bootstrap-Prozesse W_n^* definiert.
Mit $\mathbb{P}_n^*$ werde weiterhin die auf die Originalstichprobe bedingte Wahrscheinlichkeit bezeichnet:

$$\mathbb{P}_n^*(\,\cdot\,) := \mathbb{P}(\,\cdot\,|(X_1, ..., X_n)).$$

Ebenso bezeichnet $\mathbb{E}_n^*$ die entsprechende bedingte Erwartung.
Um letztlich die Konsistenz des Bootstrap-Verfahrens für bestimmte Teststatistiken nachzuweisen, wird nun zunächst die Konvergenz der Verteilung des Bootstrap-Prozesses W_n^*, bedingt auf die Originalstichprobe, in Wahrscheinlichkeit nachgewiesen.
Um den geschätzten empirischen Prozess hinreichend gut nachbilden zu können, müssen, wie aus dem ersten Teil der Arbeit ersichtlich wurde, Bootstrap-Verfahren angewendet werden, die in geeigneter Weise die Abhängigkeitsstruktur der beobachteten Größen nachbilden.
Ein Vorgehen wie es im Falle unabhängiger Beobachtungen angewendet wurde, um die durch die Schätzung der Parameters θ gewonne Information zu nutzen, ist hierbei allerdings nicht möglich, da zwar Beobachtungen mit zugrundeliegender Verteilungsfunktion $F_{\hat{\theta}_n}$ generiert werden können, wobei diese Verteilungsfunktion die Rolle der Funktion F_θ beim Originalprozess übernimmt. Allerdings ist es auf diesem Wege nicht möglich, ohne zusätzliche Modellannahmen an die Struktur des beobachteten Prozesses das Abhängigkeitsverhalten geeignet zu imitieren. Die Anwendung eines modellbasierten Verfahrens gestaltet sich schwierig, da hierbei die Eigenschaften der stationären Verteilung der Beobachtungen im allgemeinen ebensowenig hinreichend greifbar sind wie Effekte, die durch das Schätzen von Innovationen auftreten, wie bereits eingangs erwähnt.
Um dies zu umgehen, wird die Methode des Blockwise Bootstrap angewendet. Die Bootstrap-Stichprobe $X_1^*, ..., X_n^*$ wird gewonnen, indem aus der Originalstichprobe $k := k_n$ Blöcke der Länge $l := l_n$ unabhängig gezogen werden, wobei sowohl die Anzahl der Blöcke als auch deren Länge in geeigneter Weise gegen unendlich strebt und gilt $n = k \cdot l$, was vereinfachend für alle n angenommen wird. Für den

Fall, dass n nicht Vielfaches von l ist, wird der letzte gezogene Block entsprechend gekürzt.
Dieser Ansatz wurde im Zusammenhang mit empirischen Prozessen bereits von Naik-Nimbalkar und Rajarshi [53], Bühlmann [13], Peligrad [58], Bühlmann [14] und Radulovic [65] verwendet, wie bereits eingangs erwähnt.
Der zugehörige geschätzte empirische Prozess für die Beobachtungen $X_1^*, ..., X_n^*$ lautet:

$$\widehat{W}_n^*(x) = \sqrt{n} \cdot (F_n^*(x) - F_{\hat{\theta}_n^*}(x)),$$

wobei $F_n^*(\cdot)$ die empirische Verteilungsfunktion der Bootstrap-Stichprobe $X_1^*, ..., X_n^*$ bezeichnet.
Geht man nun, um zu einer linearisierten Darstellung des geschätzten empirischen Prozesses zu gelangen, so vor wie im Falle des Originalprozesses, so erhält man:

$$\widehat{W}_n^*(x) = \sqrt{n} \cdot \big(F_n^*(x) - F_{\hat{\theta}_n}(x) + F_{\hat{\theta}_n}(x) - F_{\hat{\theta}_n^*}(x)\big).$$

Während für den zweiten Teil des Prozesses, $F_{\hat{\theta}_n}(x) - F_{\hat{\theta}_n^*}(x)$, unter geeigneten Voraussetzungen an die Klasse $\mathcal{F}$ und den Parameterschätzer wieder eine asymptotisch lineare Darstellung erzielt werden kann, ist der erste Teil, $F_n^*(x) - F_{\hat{\theta}_n}(x)$, aufgrund der Tatsache, dass das Bootstrap-Verfahren nichtparametrisch ist, nicht zentriert. Es muss also eine zusätzliche Korrektur vorgenommen werden, wie dies auch in Babu und Rao [6] für den Fall unabhängiger Daten bei Verwendung eines nichtparametrischen Bootstrap-Verfahrens durchgeführt wurde. Mit $\tilde{F}_n(x)$ wird hierzu der auf die Originalstichprobe bedingte Erwartungswert von $F_n^*(x)$ bezeichnet, siehe hierzu auch Naik-Nimbalkar und Rajarshi [53].

Es gelten die beiden Darstellungen

$$\begin{aligned} F_n^*(x) &= \frac{1}{k}\sum_{i=1}^{k}\frac{1}{l}\sum_{j=1}^{l}\mathbb{1}\{X_{S_i+j} \leq x\} \\ &=: \frac{1}{k}\sum_{i=1}^{k} H_{S_i}(x) \end{aligned}$$

und

$$\begin{aligned}\tilde{F}_n(x) &= E_n^*(F_n^*(x)) = E_n^*\big[\frac{1}{l}\sum_{j=1}^{l}(\mathbb{1}\{X_{S_1+j} \leq x)\}\big] \\ &= \frac{1}{n-l+1}\cdot\frac{1}{l}\sum_{r=0}^{n-l}\sum_{j=1}^{l}\mathbb{1}\{X_{r+j} \leq x\}.\end{aligned}$$

Weiter kann $\widehat{W}_n^*(x)$ nunmehr geschrieben werden als

$$\begin{aligned}\widehat{W}_n^*(x) &= \sqrt{n}\big(F_n^*(x) - \tilde{F}_n(x) + F_{\hat{\theta}_n}(x) - F_{\hat{\theta}_n^*}(x) + \tilde{F}_n(x) - F_{\hat{\theta}_n}(x)\big) \\ &=: \sqrt{n}\big(F_n^*(x) - \tilde{F}_n(x) + F_{\hat{\theta}_n}(x) - F_{\hat{\theta}_n^*}(x)\big) + B_n(t) \\ &=: W_n^*(x) + B_n(x).\end{aligned}$$

$B_n(\cdot)$ bezeichnet den zusätzlichen Bias-Anteil, der durch dei Verwendung eines nichtparametrischen Verfahrens entsteht, während der Prozess $W_n^*(x)$ eine dem Originalprozess entsprechende Darstellung besitzt, wie später gezeigt werden wird.

Insgesamt wird im folgenden gezeigt werden, dass unter $\mathcal{H}_0$

$$W_n^*(x) = \widehat{W}_n^*(x) - B_n(x)$$

in Wahrscheinlichkeit Verteilungskonvergenz gegen denselben Gaußprozess G aufweist wie der Originalprozess. Falls $\mathcal{H}_0$ nicht erfüllt ist, erfolgt unter geeigneten Bedingungen an $\hat{\theta}_n$ ebenfalls die Konvergenz gegen einen zentrierten Gaußprozess, welcher im folgenden mit $\bar{G}$ bezeichnet wird.

2.3.1 Voraussetzungen und Konvergenz des Bootstrap-Prozesses

Um die Verteilungskonvergenz des Bootstrap-Prozesses in Wahrscheinlichkeit gegen einen zentrierten Gaußprozess nachzuweisen wird, wie im letzten Abschnitt geschildert, bewiesen, dass gilt

$$W_n^* \to \bar{G} \text{ in } \mathbb{P}\text{-Wahrscheinlichkeit.}$$

$\bar{G}$ ist unter $\mathcal{H}_0$ derselbe Grenzprozess wie unter Theorem 2.2.1.
Hierzu ist zu zeigen, dass für jede Teilfolge $(m_n)_{n\in\mathbb{N}}$ der natürlichen Zahlen eine Teilfolge $(m'_n)_{n\in\mathbb{N}}$ existiert, sodass mit Wahrscheinlichkeit 1 gilt, dass

$$\mathbb{E}^*_{m'_n}\big(f(W^*_{m'_n})(\omega_{m'_n})\big) \overset{n\to\infty}{\to} \mathbb{E}\big(f(\bar{G})\big) \tag{2.28}$$

für alle gleichmäßig stetigen, $\mathcal{P}$-messbare Funktionen $f\colon D(\bar{\mathbb{R}}) \to \mathbb{R}$. (vgl. [55], Theorem 4.1, zur Verteilungskonvergenz in Wahrscheinlichkeit).
Dies wird auf ähnliche Weise gezeigt wie für den Originalprozess.

Zunächst wird die Konvergenz der endlichdimensionalen Verteilungen von W^*_n gegen diejenigen von $\bar{G}$ gezeigt. Ähnlich wie beim Originalprozess wird die Konvergenz der endlichdimensionalen Verteilungen auf den univariaten Fall reduziert, allerdings mithilfe des modifizierten Satzes von Cramér-Wold (siehe Pauls [57], Satz 8.2). Hieraus wird gefolgert, dass für jede Teilfolge der natürlichen Zahlen wiederum eine Teilfolge existiert, sodass der zwischen Punkten eines geeigneten Gitters konstante Prozess $\widetilde{W}^*_n$ für diese Teilfolge fast sicher in Verteilung gegen den entsprechenden Grenzprozess $\widetilde{\bar{G}}$ konvergiert. (siehe (2.4) zur Bezeichnungsweise). Damit folgt auch die Verteilungskonvergenz von $\widetilde{W}^*_n$ gegen $\widetilde{\bar{G}}$ in Wahrscheinlichkeit.
Des weiteren wird ebenfalls die gleichgradige Stetigkeit in Wahrscheinlichkeit von W^*_n auf einem geeignet gewählten Gitter gezeigt und schließlich wird insgesamt die Verteilungskonvergenz in Wahrscheinlichkeit bewiesen.
Um die Konvergenz des Bootstrap-Prozesses auch zu gewährleisten, falls die Nullhypothese nicht erfüllt ist, wird ebenso wie für die Konvergenz des Originalprozesses unter $\mathcal{H}_1$, vorausgesetzt, dass der Schätzer für den Parameter $\hat{\theta}_n$ gegen einen Wert $\bar{\theta} \in \Theta$ konvergiert.
Durch das Bootstrap-Verfahren wird nun somit die Situation imitiert, dass $F_{\bar{\theta}}$ die wahre Verteilungsfunktion der Originalstichprobe ist. Der entsprechende Grenzprozess ist zentriert, während der Grenzprozess im Falle des Originalprozesses nichtzentriert ist, wie in Abschnitt 2.2.4 erläutert.
Allgemein werde im folgenden Abschnitt die tatsächliche Verteilungsfunktion der Beobachtungen mit $F(\cdot)$ bezeichnet, wobei für $F(\cdot)$ zumindest die Lipschitzstetigkeit gefordert wird.

Insgesamt müssen folgende Voraussetzungen hinsichtlich des Originalprozesses, des Bootstrap-Verfahrens und des Parameterschätzers erfüllt sein:

Annahme $\mathcal{A}^*$:
Es gelte $\mathcal{A}$(i) sowie

$(ia)^*$ Für die Blocklänge l der Bootstrap-Stichprobe gelte

$$l \asymp n^{1/2-\varepsilon} \text{ für ein } \varepsilon > 0 \text{ mit } \tfrac{1}{4} < \varepsilon < \tfrac{1}{2}.$$

$(ib)^*$ Es seien S_i, $i = 1, ..., k$, unabhängig und gleichverteilt auf der Menge der Zahlen $\{0, ..., n-l\}$. Die Bootstrap-Stichprobe wird gewonnen durch

$$X_1^*, ..., X_n^* = X_{S_1+1}, ..., X_{S_1+l}, ..., X_{S_k+l}.$$

$(ii)^*$ Es gilt $\hat{\theta}_n \xrightarrow{\mathbb{P}} \bar{\theta}$.

$(iii)^*$ Für den Schätzer $\hat{\theta}_n^*$ für $\hat{\theta}_n$ gelte, für eine messbare Funktion $\ell\colon \mathbb{R} \times \Theta \to \mathbb{R}^p$ $(k \leq p)$

$$\sqrt{n}(\hat{\theta}_n^* - \hat{\theta}_n) = \frac{1}{\sqrt{k}} \sum_{i=1}^{k} \Big\{ \frac{1}{\sqrt{l}} \sum_{j=1}^{l} \ell(X_{S_i+j}, \bar{\theta}) - \mathbb{E}_n^*\big(\frac{1}{\sqrt{l}} \sum_{j=1}^{l} \ell(X_{S_1+j}, \bar{\theta})\big)\Big\} + \varepsilon_n^*.$$

Für ε_n^* gelte, dass $\varepsilon_n^* \xrightarrow{\mathbb{P}_n^*} 0$ in $\mathbb{P}$-Wahrscheinlichkeit, d.h. $\mathbb{P}_n^*(||\varepsilon_n^*||_2 > \mu) \xrightarrow{\mathbb{P}} 0$ für jedes $\mu > 0$.
ℓ besitze die folgenden Eigenschaften:

$$\mathbb{E}(||\ell(X_0, \bar{\theta})||_2^4) < \infty, \ \mathrm{Lip}(\ell_c) = c^\gamma, \text{ für ein } \gamma > 0.$$

(ℓ_c ist gestutzte Version von ℓ, siehe Annahme $\mathcal{A}$.)

$(iv)^*$ $\dot{F}_{\bar{\theta}}(x)$ existiert und ist gleichmäßig stetig auf $\mathbb{R} \times U$, wobei U offene Umgebung von $\bar{\theta}$ ist und $\dot{F}_{\bar{\theta}}(x)$ besitzt Grenzwerte für $x \to \pm\infty$.

$(v)^*$ $F_{\tilde{\theta}}(x)$, $\tilde{\theta} \in U$, ist lipschitzstetig in x.

$(vi)^*$ Die Verteilungsfunktion $F(x)$ der Beobachtungen $X_1, X_2, ...$, sofern diese nicht $F_{\bar{\theta}}$ ist, ist lipschitzstetig in x.

Voraussetzung $(iii)^*$ resultiert aus der Verwendung eines nichtparametrischen Bootstrap-Verfahrens, siehe hierzu Babu und Rao [6]. Gültig ist $(iii)^*$ unter bestimmten Bedingungen etwa für M-Schätzer, siehe Lahiri [44]. Für den Fall unabhängiger Daten wurde diese Bedingung in Babu und Singh [7] für L-Schätzer untersucht.

Insgesamt gilt dann in Analogie zu Theorem 2.2.1

Theorem 2.3.1 *Es gelte $\mathcal{A}^*$. Dann gilt:*

$$W_n^*(x) \xrightarrow{D} \bar{G} \text{ in } \mathbb{P}\text{-Wahrscheinlichkeit}$$

wobei $\bar{G}$ ein zentrierter Gaußprozess ist mit

$$\operatorname{cov}(\bar{G}(x_i), \bar{G}(x_j)) = \sum_{k\in\mathbb{Z}} \operatorname{cov}\big(z_{\bar{\theta}}(X_0, x_i), z_{\bar{\theta}}(X_k, x_j)\big),$$

wobei

$$z_{\bar{\theta}}(X_k, x) = \mathbb{1}\{X_k \leq x\} - F_{\bar{\theta}}(x) - \ell^T(X_k, \bar{\theta}) \cdot \dot{F}_{\bar{\theta}}(x).$$

Unter $\mathcal{H}_0$ gilt $\hat{\theta}_n \xrightarrow{\mathbb{P}} \theta$ und somit entspricht $\bar{G}$ in diesem Fall dem Grenzprozess wie in Theorem (2.2.1).
Der Beweis von Theorem 2.3.1 erfolgt im Abschnitt 2.3.2, unter Nutzung der Resultate aus Abschnitt 2.3.2, 2.3.4 und 2.3.2.

2.3.2 Beweis der Verteilungskonvergenz des Bootstrap-Prozesses

Darstellung des Bootstrap-Prozesses

Um zu einer geeigneten Darstellung für den Bootstrap-Prozess W_n^* zu gelangen, geht man so vor wie im Falle des Originalprozesses.
Es gilt:

$$W_n^*(x) \quad = \quad \sqrt{n}(F_n^*(x) - \tilde{F}_n(x)) + \sqrt{n}(F_{\hat{\theta}_n}(x) - F_{\hat{\theta}_n^*}(x)).$$

Weiter gilt aufgrund von $(iii)^*$ und $(iv)^*$

$$\begin{aligned}
\sqrt{n}(F_{\hat\theta_n}(x) - F_{\hat\theta_n^*}(x)) &= \sqrt{n}\cdot(\hat\theta_n - \hat\theta_n^*)\cdot \dot F_{\tilde\theta_n^*}(x) \\
&= \sqrt{n}\cdot(\hat\theta_n - \hat\theta_n^*)\cdot \dot F_{\bar\theta}(x) + \sqrt{n}(\hat\theta_n - \hat\theta_n^*)\cdot(\dot F_{\tilde\theta_n^*}(x) - \dot F_{\bar\theta}(x)) \\
&= \frac{1}{\sqrt{k}}\sum_{i=1}^{k}\Big(\frac{1}{\sqrt{l}}\sum_{j=1}^{l}\ell(X_{S_i+j},\bar\theta)\cdot \dot F_{\bar\theta}(x) \\
&\qquad - \mathbb{E}_n^*(\frac{1}{\sqrt{l}}\sum_{j=1}^{l}\ell(X_{S_i+j},\bar\theta))\cdot \dot F_{\bar\theta}(x)\Big) \\
&\quad + \sqrt{n}\cdot(\hat\theta_n - \theta_n^*)\cdot(\dot F_{\tilde\theta_n^*}(x) - \dot F_{\bar\theta}(x)) + \tilde\varepsilon_n^*(x)
\end{aligned}$$

mit $\tilde\theta_n^*$ zwischen $\hat\theta_n$ und $\hat\theta_n^*$ und

$$\sup_{x\in\bar{\mathbb{R}}}|\tilde\varepsilon_n(x)| \overset{\mathbb{P}_n^*}{\to} 0 \text{ in } \mathbb{P}\text{-Wahrscheinlichkeit.}$$

Aus dem Beweis von Lemma 2.3.11 folgt, dass $||\frac{1}{n}\sum_{i=1}^{n}\ell(X_i^*,\bar\theta) - \mathbb{E}_n^*\ell(X_i^*,\bar\theta)||_2 \overset{\mathbb{P}_n^*}{\to} 0$ in $\mathbb{P}$-Wahrscheinlichkeit, und daher gilt mit Voraussetzung $(iii)^*$, dass

$$\hat\theta_n^* \overset{\mathbb{P}_n^*}{\to} \bar\theta \text{ in } \mathbb{P}\text{-Wahrscheinlichkeit.}$$

Damit kann aufgrund der gleichmäßigen Stetigkeit von $\dot F_{(\cdot)}(\cdot)$ geschrieben werden

$$\begin{aligned}
W_n^*(x) &= \frac{1}{\sqrt{k}}\sum_{i=1}^{k}\Big\{\frac{1}{\sqrt{l}}\sum_{j=1}^{l}\mathbb{1}\{X_{S_i+j}\le x\} - \tilde F_n(x) \\
&\qquad - \big[\ell^T(X_{S_i+j},\bar\theta)\cdot\dot F_{\bar\theta}(x) - \mathbb{E}_n^*(\ell^T(X_{S_i+j},\bar\theta)\cdot\dot F_{\bar\theta}(x))\big]\Big\} + \tilde\varepsilon_n^* \\
&=: U_n^*(x) - \frac{1}{\sqrt{k}}\sum_{i=1}^{k}\Big[\frac{1}{\sqrt{l}}\sum_{j=1}^{l}\ell^T(X_{S_i+j},\bar\theta)\cdot\dot F_{\bar\theta}(x) \\
&\qquad - \mathbb{E}_n^*\Big(\frac{1}{\sqrt{l}}\sum_{j=1}^{l}\ell^T(X_{S_i+j},\bar\theta)\cdot\dot F_{\bar\theta}(x)\Big)\Big] + \tilde\varepsilon_n^* \\
&=: \frac{1}{\sqrt{k}}\sum_{i=1}^{k}\Big\{\frac{1}{\sqrt{l}}\sum_{j=1}^{l}z_{\bar\theta}^*(X_{S_i+j},x)\Big\} + \tilde\varepsilon_n^*(x), \qquad (2.29)
\end{aligned}$$

wobei

$$\begin{aligned}
z_{\bar\theta}^*(X_{S_i+j},x) &:= \mathbb{1}\{X_{S_i+j}\le x\} - \tilde F_n(x) \\
&\qquad - \big[\ell^T(X_{S_i+j},\bar\theta)\cdot\dot F_{\bar\theta}(x) - \mathbb{E}_n^*(\ell^T(X_{S_i+j},\bar\theta)\cdot\dot F_{\bar\theta}(x))\big]. \qquad (2.30)
\end{aligned}$$

Nun wird folgendermaßen vorgegangen:

Zunächst wird, wie oben beschrieben, die Konvergenz der endlichdimensionalen Verteilungen von

$$Z_n^*(x) := \frac{1}{\sqrt{k}} \sum_{i=1}^{k} \left\{ \frac{1}{\sqrt{l}} \sum_{j=1}^{l} z_{\bar{\theta}}^*(X_{S_i+j}, x) \right\} \tag{2.31}$$

in Wahrscheinlichkeit gegen die des Grenzprozesses $\bar{G}$ gezeigt.
Mit dem Satz von Cramér-Slutsky gilt dann ebenso die Konvergenz der endlichdimensionalen Verteilungen von W_n^* gegen diejenigen von $\bar{G}$.

Dann wird die stochastische Äquikontinuität des Prozesses W_n^* nachgewiesen.

Konvergenz der endlichdimensionalen Verteilungen des Bootstrap-Prozesses

In diesem Abschnitt wird die Konvergenz der endlichdimensionalen Verteilungen des Bootstrap-Prozesses W_n^* gegen diejenigen von $\bar{G}$ nachgewiesen. Hierbei wird zunächst die Kovarianzstruktur des Prozesses betrachtet.

Konvergenz der Kovarianzen gegen die Kovarianzen des Originalprozesses
Für die Konvergenz der endlichdimensionalen Verteilungen des Bootstrap-Prozesses Z_n^* wird zunächst gezeigt, dass die Kovarianzstruktur des zentrierten Anteils des Prozesses im Grenzwert dieselbe ist wie diejenige des Grenzprozesses $\bar{G}$. Die Konvergenz der Kovarianzen von Z_n^* gegen diejenigen von $\bar{G}$ gilt aufgrund der getroffenen Voraussetzungen wiederum in $\mathbb{P}$-Wahrscheinlichkeit.

Lemma 2.3.1 *Unter Voraussetzung $\mathcal{A}^*$ gilt für beliebige $(s,t) \in \bar{\mathbb{R}}^2$*

$$\mathrm{cov}_n^*(Z_n^*(s), Z_n^*(t)) = \mathrm{cov}(\bar{G}(s), \bar{G}(t)) + o_{\mathbb{P}}(1).$$

Beweis:
Gemäß obiger Definition gilt

$$\begin{aligned} Z_n^*(s) &= \frac{1}{\sqrt{k}} \sum_{i=1}^{k} \Big\{ \frac{1}{\sqrt{l}} \sum_{j=1}^{l} \mathbb{1}\{X_{S_i+j} \leq s\} - \ell^T(X_{S_i+j}, \bar{\theta}) \cdot \dot{F}_{\bar{\theta}}(s) \\ &\qquad - \mathbb{E}_n^* \big[\frac{1}{\sqrt{l}} \sum_{j=1}^{l} \mathbb{1}\{X_{S_i+j} \leq s\} - \ell^T(X_{S_i+j}, \bar{\theta}) \cdot \dot{F}_{\bar{\theta}}(s) \big] \Big\}. \end{aligned}$$

Unter Verwendung von der unter (2.3) definierten Bezeichnung wird weiter definiert

$$\begin{aligned} D_n(S_i, s) &:= \frac{1}{\sqrt{l}} \sum_{j=1}^{l} \mathbb{1}\{X_{S_i+j} \leq s\} - \ell^T(X_{S_i+j}, \bar{\theta}) \cdot \dot{F}_{\bar{\theta}}(s) - F_{\bar{\theta}}(s) \\ &= \frac{1}{\sqrt{l}} \sum_{j=1}^{l} z_{\bar{\theta}}(X_{S_i+j}, s). \end{aligned}$$

Damit gilt

$$\begin{aligned} \mathrm{cov}_n^*(Z_n^*(s), Z_n^*(t)) &= \mathrm{cov}_n^*\Big(\frac{1}{\sqrt{k}} \sum_{i=1}^{k} D_n(S_i, s), \frac{1}{\sqrt{k}} \sum_{i=1}^{k} D_n(S_i, t)\Big) \\ &= \mathrm{cov}_n^*(D_n(S_1, s), D_n(S_1, t)) \\ &= \underbrace{\mathbb{E}_n^*(D_n(S_1, s) \cdot D_n(S_1, t))}_{=:(i)} - \underbrace{\mathbb{E}_n^*(D_n(S_1, s)) \cdot \mathbb{E}_n^*(D_n(S_1, t))}_{=:(ii)}. \end{aligned}$$

Es wird nun gezeigt, dass der erste Teil des letzten Ausdrucks in Wahrscheinlichkeit gegen die entsprechende Kovarianz des Originalprozessees konvergiert und für den zweiten Teil wird Konvergenz gegen Null in Wahrscheinlichkeit nachgewiesen. (Siehe zur Vorgehensweise auch Bühlmann [14].)

zu (i)

Durch Nulladdition erhält man:

$$\begin{aligned} &\mathbb{E}_n^*[D_n(S_1, s) \cdot D_n(S_1, t)] \\ = \; &\frac{1}{n-l+1} \sum_{u=0}^{n-l} \Big\{ \Big(\frac{1}{\sqrt{l}} \sum_{j=u+1}^{u+l} z_{\bar{\theta}}(X_j, s) \Big) \Big(\frac{1}{\sqrt{l}} \sum_{j=u+1}^{u+l} z_{\bar{\theta}}(X_j, t) \Big) \end{aligned}$$

$$
\begin{aligned}
& \qquad - \mathbb{E}\Big[\Big(\frac{1}{\sqrt{l}}\sum_{j=u+1}^{u+l} z_{\bar\theta}(X_j,s)\Big)\Big(\frac{1}{\sqrt{l}}\sum_{j=u+1}^{u+l} z_{\bar\theta}(X_j,t)\Big)\Big] \\
& + \mathbb{E}\Big[\Big(\frac{1}{\sqrt{l}}\sum_{j=u+1}^{u+l} z_{\bar\theta}(X_j,s)\Big)\Big(\frac{1}{\sqrt{l}}\sum_{j=u+1}^{u+l} z_{\bar\theta}(X_j,t)\Big)\Big]\Big\} \\
=: \; & \frac{1}{n-l+1}\sum_{u=0}^{n-l}\Big\{Q_u(s,t) \\
& \qquad + \mathbb{E}\big[\big(\frac{1}{\sqrt{l}}\sum_{j=u+1}^{u+l} z_{\bar\theta}(X_j,s)\big)\big(\frac{1}{\sqrt{l}}\sum_{j=u+1}^{u+l} z_{\bar\theta}(X_j,t)\big)\big]\Big\} \qquad (2.32)
\end{aligned}
$$

mit

$$
\begin{aligned}
Q_u(s,t) \;=\; & \Big(\frac{1}{\sqrt{l}}\sum_{j=u+1}^{u+l} z_{\bar\theta}(X_j,s)\Big)\Big(\frac{1}{\sqrt{l}}\sum_{j=u+1}^{u+l} z_{\bar\theta}(X_j,t)\Big) \\
& - \mathbb{E}\Big[\Big(\frac{1}{\sqrt{l}}\sum_{j=u+1}^{u+l} z_{\bar\theta}(X_j,s)\Big)\Big(\frac{1}{\sqrt{l}}\sum_{j=u+1}^{u+l} z_{\bar\theta}(X_j,t)\Big)\Big]\Big\}.
\end{aligned}
$$

Nun wird gezeigt, dass das arithmetische Mittel über die Q_u in Wahrscheinlichkeit gegen Null konvergiert, während der Ausdruck in der letzten Zeile für n gegen unendlich gegen die entsprechende Kovarianz des Grenzprozesses $\bar{G}$ konvergiert.

Lemma 2.3.2 *Unter den Voraussetzungen von Theorem 2.3.1 gilt die Konvergenz von $(n-l+1)^{-1}\sum_{u=0}^{n-l} Q_u(s,t)$ gegen Null in W'keit, für beliebiges $(s,t) \in \bar{\mathbb{R}}^2$.*

Beweis:
Bezeichne mit

$$
\tilde{z}(X_i,s,X_j,t) \;:=\; z_{\bar\theta}(X_i,s)\cdot z_{\bar\theta}(X_j,t) - \mathbb{E}\big[z_{\bar\theta}(X_i,s)\cdot z_{\bar\theta}(X_j,t)\big].
$$

Ausmultiplizieren von $Q_u(s,t)$ ergibt damit:

$$
\begin{aligned}
Q_u(s,t) \;&=\; \frac{1}{l}\sum_{i=u+1}^{u+l}\sum_{j=u+1}^{u+l}\tilde{z}(X_i,s,X_j,t) \\
&=\; \frac{1}{l}\sum_{i=u+1}^{u+l}\Big\{\tilde{z}(X_i,s,X_i,t) + \sum_{j=1}^{l-1}[\tilde{z}(X_i,s,X_{i+j},t) + \tilde{z}(X_{i+j},s,X_i,t)]\Big\}.
\end{aligned}
$$

Durch Zusammenfassung gleicher Terme erhält man somit, wie in Künsch [43], Abschnitt 3.2, für das arithmetische Mittel über die Q_u

$$\begin{aligned}
&\frac{1}{n-l+1}\sum_{u=0}^{n-l} Q_u \\
=\ &\frac{1}{n-l+1}\sum_{i=1}^{n} w(i)\Big\{\tilde{z}(X_i,s,X_i,t) \\
&\qquad + \sum_{j=1}^{l-1}\min\{w(i),l-j\}\big(\tilde{z}(X_i,s,X_{i+j},t)+\tilde{z}(X_{i+j},s,X_i,t)\big)\Big\} \\
=\ &\frac{1}{n-l+1}\sum_{i=1}^{n} w(i)\Big\{\tilde{z}(X_i,s,X_i,t) \\
&\qquad + \sum_{j=1}^{l-1}\min\{w(i),l-j\}\big(\tilde{z}(X_i,s,X_{i+j},t)+\tilde{z}(X_{i+j},s,X_i,t)\big)\Big\}.
\end{aligned}$$

Für die Gewichtsfunktion $w(\cdot)$ gilt aufgrund der Vielfachheit der auftretenden Terme folgendes:

$$l\cdot w(i) = i\cdot \mathbb{1}\{i\le l\} + l\cdot\mathbb{1}\{l<i\le n-l\} + (n-(i-1))\cdot\mathbb{1}\{n-l<i\le n\}, \quad (2.33)$$

woraus insbesondere folgt, dass $0 < w(\cdot) \le 1$.

Abgeschätzt werden soll jetzt die Wahrscheinlichkeit, dass der letztgenannte Ausdruck betragsmäßig größer als μ ist, wobei $\mu > 0$ beliebig aber fest gewählt sei. Diese Wahrscheinlichkeit kann wiederum beschränkt werden durch Anwendung der Tschebyschow-Ungleichung, und aufgrund der Tatsache, dass der letztgenannte Term zentriert ist, gilt

$$\begin{aligned}
&\mathbb{P}\Big(\Big|\frac{1}{n-l+1}\sum_{u=0}^{n-l} Q_u\Big| > \mu\Big) \\
\le\ &\frac{1}{\mu^2(n-l+1)^2}\cdot\mathbb{E}\Big(\Big\{\sum_{i=1}^{n} w(i)\cdot\tilde{z}(X_i,s,X_i,t) \\
&\qquad + \sum_{j=1}^{l-1}\min\{w(i),l-j\}\cdot[\tilde{z}(X_i,s,X_{i+j},t)+\tilde{z}(X_{i+j},s,X_i,t)]\Big\}^2\Big)
\end{aligned}$$

$$
\begin{aligned}
\leq \quad & \frac{1}{\mu^2(n-l+1)^2} \sum_{i_1=1}^{n} \sum_{i_2=1}^{n} \Big| \operatorname{cov}\Big(w(i_1) \cdot \tilde{z}(X_{i_1}, s, X_{i_1}, t) + \\
& + \sum_{j=1}^{l-1} \min\{w(i_1), l-j\} \cdot [\tilde{z}(X_{i_1}, s, X_{i_1+j}, t) + \tilde{z}(X_{i_1+j}, s, X_{i_1}, t)], \\
& \qquad w(i_2) \cdot \tilde{z}(X_{i_2}, s, X_{i_2}, t) + \\
& + \sum_{j=1}^{l-1} \min\{w(i_2), l-j\} \cdot [\tilde{z}(X_{i_2}, s, X_{i_2+j}, t) + \tilde{z}(X_{i_2+j}, s, X_{i_2}, t)] \Big) \Big|
\end{aligned}
$$

Weiter kann aufgrund der Linearität der Kovarianz ein Ausmultiplizieren erfolgen, des weiteren kann der Betrag der Summe durch die Summe der Beträge über die Kovarianzen abgeschätzt werden. Dann kann jeweils die Gewichtsfunktion als konstanter Faktor ausgeklammert werden und man erhält schließlich als Abschätzung nach oben aufgrund der Stationarität der Beobachtungen

$$
\begin{aligned}
\leq \quad & \frac{1}{\mu^2(n-l+1)^2} \sum_{i=-(n-1)}^{n-1} (n-|i|) \cdot \Big\{ \big| \operatorname{cov}\big(\tilde{z}(X_1, s, X_1, t), \tilde{z}(X_{1+|i|}, s, X_{1+|i|}, t)\big) \big| \\
& + \sum_{j=1}^{l-1} \Big[\big| \operatorname{cov}\big(\tilde{z}(X_1, s, X_{1+j}, t), \tilde{z}(X_{1+|i|}, s, X_{1+|i|}, t)\big) \big| \\
& \qquad + \big| \operatorname{cov}\big(\tilde{z}(X_{1+j}, s, X_1, t), \tilde{z}(X_{1+|i|}, s, X_{1+|i|}, t)\big) \big| \Big] \\
& + \sum_{j=1}^{l-1} \big| \operatorname{cov}\big(\tilde{z}(X_1, s, X_1, t), \tilde{z}(X_{1+|i|}, s, X_{1+|i|+j}, t)\big) \big| \\
& + \sum_{j_1=1}^{l-1} \sum_{j_2=1}^{l-1} \Big[\big| \operatorname{cov}\big(\tilde{z}(X_1, s, X_{1+j_1}, t), \tilde{z}(X_{1+|i|}, s, X_{1+|i|+j_2}, t)\big) \big| \\
& \qquad + \big| \operatorname{cov}\big(\tilde{z}(X_{1+j_1}, s, X_1, t), \tilde{z}(X_{1+|i|}, s, X_{1+|i|+j_2}, t)\big) \big| \Big] \\
& + \sum_{j=1}^{l-1} \big| \operatorname{cov}\big(\tilde{z}(X_1, s, X_1, t), \tilde{z}(X_{1+|i|+j}, s, X_{1+|i|}, t)\big) \big| \\
& + \sum_{j_1=1}^{l-1} \sum_{j_2=1}^{l-1} \Big[\big| \operatorname{cov}\big(\tilde{z}(X_1, s, X_{1+j_1}, t), \tilde{z}(X_{1+|i|+j_2}, s, X_{1+|i|}, t)\big) \big| \\
& \qquad + \big| \operatorname{cov}\big(\tilde{z}(X_{1+j_1}, s, X_1, t), \tilde{z}(X_{1+|i|+j_2}, s, X_{1+|i|}, t)\big) \big| \Big] \Big\}
\end{aligned}
\tag{2.34}
$$

Unter der äußeren Summe tritt die Summe über insgesamt $[1+(2l-2)]^2 = (2l-1)^2$ einzelne Kovarianzen.

Jede einzelne der Kovarianzen ist von der Struktur

$$\begin{aligned} &\operatorname{cov}(\tilde{z}(X_1, s, X_{1+j_1}, t), \tilde{z}(X_{1+|i|}, s, X_{1+|i|+j_2}, t)) \\ = &\sum_{\substack{(v_1,v_2,v_3,v_4)\\ \in\{1,2\}^4}} \operatorname{cov}\big(f_{v_1}(X_1, s) \cdot f_{v_2}(X_{1+j_1}, t), f_{v_3}(X_{1+|i|}, s) \cdot f_{v_4}(X_{1+|i|+j_2}, t)\big), \end{aligned}$$

wobei $f_1(X, s) := \mathbb{1}\{X \leq s\}$, $f_2(X, s) := \ell^T(X, \bar{\theta}) \cdot \dot{F}_{\bar{\theta}}(s)$.

Falls nun für obige Kovarianzen unter (2.34) gilt, dass $1 + j_1 \geq 1 + |i|$, so kann die Eigenschaft der schwachen Abhängigkeit hier nicht ausgenutzt werden, da zwischen den in der Kovarianz auftretenden Termen keine Zeitlücke besteht. Diese Kovarianzen können unter Anwendung der Ungleichung von Cauchy-Schwarz durch eine Konstante abgeschätzt werden. Die entsprechenden Terme sind endlich, falls gilt, dass

$$||\ell(X_0, \bar{\theta})||_2^4$$

endlich ist, wie in Voraussetzung $\mathcal{A}^*$ gefordert. Vereinfachend wird diese Abschätzung für die Kovarianzen benutzt, falls $|i| \leq l-1$. Hiervon betroffen sind für festes i insgesamt $(2l-1)^2$ Terme, und dies für $|i|$ von 0 bis $l-1$. Die Summe über all diese Kovarianzen kann also nach oben abgeschätzt werden durch

$$(2l-1) \cdot (2l-1)^2 \cdot \text{const}\,.$$

Insgesamt kann also für $|i| < l$ der Term (2.34) abgeschätzt werden durch

$$O(l^3 \cdot (n-l+1)^{-1}).$$

Dies geht gegen Null, sofern gewährleistet ist, dass

$$l^3 = o(n) \Leftrightarrow n^{3/2-3\varepsilon-1} = o(1) \Leftrightarrow \frac{1}{2} - 3\varepsilon < 0 \Leftrightarrow \frac{1}{6} < \varepsilon.$$

Für festes i mit $|i| \geq l$ können alle Kovarianzen unter Ausnutzung der schwachen Abhängigkeit der Beobachtungen abgeschätzt werden. Gemäß den Überlegungen, welche zum Nachweis der Konvergenz der endlichdimensionalen Verteilungen des Originalprozesses angestellt wurden, können diese Kovarianzen in Abhängigkeit der Zeitlücke $|i| - l + 1$ abgeschätzt werden durch

$$\operatorname{cov}(\tilde{z}(X_1, s, X_l, t), \tilde{z}(X_{1+|i|}, s, X_{1+|i|+r}, t)) \leq \text{const} \cdot \tilde{\zeta}^{|i|-l+1},$$

wobei $\tilde{\zeta} = \max\{\zeta^{\frac{1}{2+\gamma}}, \zeta^{\frac{1}{2}}\}$.
Auch für diese Abschätzung muss die Voraussetzung

$$||\ell(X_0, \bar{\theta})||_2^4 < \infty$$

erfüllt sein.

Die äußere Summe bei (2.34) läuft auch in diesem Fall über $O(l^2)$ Terme, und man erhält insgesamt für (2.34) im Fall $|i| \geq l$ die Ordnung

$$O(n^{-1} \cdot l^2).$$

Aufgrund der Voraussetzung $l = O(n^{1/2-\varepsilon})$ geht also dieser Teil des Ausdrucks gegen Null. Die Abschätzungen erfolgten gleichmäßig in $(s,t) \in \bar{\mathbb{R}}^2$, sodass insgesamt Lemma 2.3.2 bewiesen ist. □

Um zu zeigen, dass $\mathbb{E}_n^*[D_n(S_1, s) \cdot D_n(S_1, t)]$ insgesamt in Wahrscheinlichkeit gegen die entsprechende Kovarianz des Originalprozesses konvergiert, ist nun noch folgendes Hilfsresultat zu zeigen

Lemma 2.3.3 *Unter Voraussetzung $\mathcal{A}^*$ gilt für $(s,t) \in \bar{\mathbb{R}}$ beliebig, dass,*

$$\begin{aligned} & \frac{1}{n-l+1} \sum_{u=0}^{n-l} \mathbb{E}\Big[\big(\frac{1}{\sqrt{l}} \sum_{j=u+1}^{u+l} z_{\bar{\theta}}(X_j, s)\big)\big(\frac{1}{\sqrt{l}} \sum_{j=u+1}^{u+l} z_{\bar{\theta}}(X_j, t)\big)\Big] \\ \overset{n\to\infty}{\to} \quad & \operatorname{cov}(\bar{G}(s), \bar{G}(t)). \end{aligned}$$

Beweis:

$$\begin{aligned} & \frac{1}{n-l+1} \sum_{u=0}^{n-l} \mathbb{E}\Big[\frac{1}{\sqrt{l}} \sum_{j=u+1}^{u+l} \big(z_{\bar{\theta}}(X_j, s)\big) \frac{1}{\sqrt{l}} \sum_{j=u+1}^{u+l} \big(z_{\bar{\theta}}(X_j, t)\big)\Big] \\ = \quad & \frac{1}{l} \sum_{j=1}^{l} \sum_{r=1}^{l} \mathbb{E}\big(z_{\bar{\theta}}(X_j, s) \cdot z_{\bar{\theta}}(X_r, t)\big) \end{aligned}$$

$$
\begin{aligned}
&= \sum_{i=-(l-1)}^{l-1} \frac{(l-|i|)}{l} \operatorname{cov}\big(z_{\bar\theta}(X_1,s)\cdot z_{\bar\theta}(X_{1+|i|},t)\big) \\
&\overset{n\to\infty}{\to} \sum_{i\in\mathbb{Z}} \operatorname{cov}\big(z_{\bar\theta}(X_1,s), z_{\bar\theta}(X_{1+|i|},t)\big) = \operatorname{cov}(\bar G(s), \bar G(t)). \qquad \square
\end{aligned}
$$

Zum Beweis von Lemma 2.3.1 ist nun gemäß (2.32) noch zu zeigen, dass gilt

$$\mathbb{E}_n^*(D_n(S_1,s))\cdot \mathbb{E}_n^*(D_n(S_1,t)) \overset{\mathbb{P}}{\to} 0.$$

Gemäß Künsch, (3.14) gilt: (Analog zu den Betrachtungen im Beweis von Lemma 2.3.2)

$$\mathbb{E}_n^*[D_n(S_1,s)] = \sqrt{l}(n-l+1)^{-1}\sum_{j=1}^n w(j)\cdot z_{\bar\theta}(X_j,s)$$

wobei $|w(j)| \le 1$ (siehe (2.33)).
Daher gilt:

$$
\begin{aligned}
&\mathbb{P}\Big(\sqrt{l}(n-l+1)^{-1}\Big|\sum_{j=1}^n w(j)\cdot z_{\bar\theta}(X_j,s)\Big| > \mu\Big) \\
\le\ & \frac{1}{\mu^2}\operatorname{var}\big(\sqrt{l}(n-l+1)^{-1}\sum_{j=1}^n w(j)\cdot z_{\bar\theta}(X_j,s)\big) \\
=\ & \mu^{-2}\frac{1}{l(n-l+1)^2}\mathbb{E}\big(\sum_{j=1}^n w(j)\cdot z_{\bar\theta}(X_j,s)\big)^2 \\
\le\ & \mu^{-2}\frac{1}{l(n-l+1)^2}\sum_{i=1}^n\sum_{j=1}^n |\mathbb{E}\big(z_{\bar\theta}(X_i,s)\cdot z_{\bar\theta}(X_j,s)\big)| \\
=\ & \mu^{-2}\frac{1}{l(n-l+1)^2}\sum_{i=-(n-1)}^{i=n-1}(n-|i|)\big|\operatorname{cov}\big(z_{\bar\theta}(X_1,s), z_{\bar\theta}(X_{1+|i|})\big)\big| \\
\le\ & \mu^{-2}\frac{}{l(n-l+1)}\sum_{i=-\infty}^{\infty}\frac{(n-|i|)_+}{n-l+1}\big|\operatorname{cov}\big(z_{\bar\theta}(X_1,s), z_{\bar\theta}(X_{1+|i|})\big)\big| \overset{n\to\infty}{\to} 0
\end{aligned}
$$

für beliebig aber festes $\mu > 0$, wobei die Abschätzungen wiederum gleichmäßig in $(s,t)\in\bar{\mathbb{R}}^2$ gelten.

Damit gilt die Konvergenz der Kovarianzen des Bootstrap-Prozesses gegen diejenigen des Prozesses $\bar{G}$ in Wahrscheinlichkeit, womit Lemma 2.3.1 gezeigt ist. $\square$

Anwendung des zentralen Grenzwertsatzes für Dreiecksschemata

Nachdem nun die Konvergenz in Wahrscheinlichkeit der Kovarianzen des Bootstrap-Prozesses gegen diejenigen des Grenzprozesses $\bar{G}$ nachgewiesen ist, soll nun die Verteilungskonvergenz in Wahrscheinlichkeit der endlichdimensionalen Verteilungen des zentrierten Anteils des Bootstrap-Prozesses nachgewiesen werden. Zunächst wird dies auf den univariaten Fall reduziert. Es wird zunächst gezeigt, dass für jeden festen Vektor $c \in \mathbb{R}^M$ und $x_1, ..., x_M$ die Verteilungkonvergenz von $c^T \vec{W}_n^*$ mit $\vec{W}_n^* = (W_n^*(x_1), ..., W_n^*(x_M))^T$ gegen $c^T \vec{\bar{G}}$ mit $\vec{\bar{G}} = (\bar{G}(x_1), ..., \bar{G}(x_M))^T$ gilt.

Die gewichteten Summen über die einzelnen Blöcke sind hierbei voneinander unabhängige Zufallsgrößen und daher kann der univariate Grenzwertsatz für unabhängige Zufallsgrößen angewendet werden.

Es gilt folgendes Lemma

Lemma 2.3.4 *Seien $M \in \mathbb{N}$, $c = (c_1, ..., c_M)^T$ beliebig $\in \mathbb{R}^M$.*
Dann gilt unter Voraussetzung $\mathcal{A}^*$, mit $\vec{W}_n^* = (W_n^*(x_1), ..., W_n^*(x_M))^T$ und $\vec{\bar{G}} = (\bar{G}(x_1), ..., \bar{G}(x_M))^T$, dass

$$c^T \vec{W}_n^* \xrightarrow{\mathcal{D}} c^T \vec{\bar{G}} \sim N(0, c^T \bar{\Sigma}_{x_1,...,x_M} c) \text{ in } \mathbb{P}\text{-W'keit,}$$

wobei $\bar{\Sigma}_{x_1,...,x_M}(i,j) = \sum_{t \in \mathbb{Z}} \operatorname{cov}\left(z_{\bar{\theta}}(X_0, x_i), z_{\bar{\theta}}(X_t, x_j)\right)$.

Beweis:

Der zentrale Grenzwertsatz wird zunächst auf den zentrierten Anteil Z_n^* des Bootstrap-Prozesses angewendet.

Betrachte $(c_1, ..., c_M)^T \in \mathbb{R}^M$ und $(x_1, ..., x_M)^T \in \mathbb{R}^M$. Es gilt, mit Z_n^* wie in (2.31) und $z_{\bar{\theta}}^*$ wie in (2.30):

$$c^T \vec{Z}_n^* = \frac{1}{\sqrt{n}} \sum_{k=1}^{n} \sum_{r=1}^{M} c_r z_{\bar{\theta}}^*(X_k^*, x_r)$$

$$
\begin{aligned}
&= \frac{1}{\sqrt{k}} \sum_{i=1}^{k} \frac{1}{\sqrt{l}} \sum_{j=1}^{l} \sum_{r=1}^{M} c_r z_{\bar{\theta}}^*(X_{S_i+j}, x_r) \\
&=: \frac{1}{\sqrt{k}} \sum_{i=1}^{k} T_{i,k}^*.
\end{aligned}
$$

(Hierbei gilt, dass $z_{\bar{\theta}}^*$ sowie $T_{i,k}^*$ jeweils von n abhängen.)
Die Zufallsgrößen $T_{i,k}^*$ bilden hierbei ein zeilenweise unabhängig und identisch verteiltes sowie (bedingt auf die Originalstichprobe) zentriertes Dreiecksschema. Es kann somit der zentrale Grenzwertsatz für zeilenweise unabhängige Dreiecksschemata angewendet werden (s. Pruscha [62], S. 348).
Um die Verteilungskonvergenz in Wahrscheinlichkeit nachzuweisen, wird folgendermaßen vorgegangen: Zur Verteilungskonvergenz von $c^T \vec{Z}_n^*$ in Wahrscheinlichkeit ist gemäß der Definition zu Beginn des Abschnittes 2.3.1 zu zeigen, dass für jede Teilfolge $(m_n)_{n\in\mathbb{N}}$ der natürlichen Zahlen $\mathbb{N}$ wiederum eine Teilfolge $(m_n')_{n\in\mathbb{N}}$ existiert, sodass $c^T \vec{Z}_{m_n'}^*$ fast sicher in Verteilung gegen $c^T \vec{\bar{G}}$ konvergiert.
Es ist also hinsichtlich der Lindeberg-Bedingung zu zeigen, dass für jede Teilfolge $(m_n)_{n\in\mathbb{N}} \subseteq \mathbb{N}$ eine Teilfolge $(m_n'')_{n\in\mathbb{N}}$ existiert, sodass

$$
\begin{aligned}
&\mathbb{P}\Big(\omega : \forall \mu > 0 : \frac{1}{k} \cdot \mathbb{E}_{m_n''}^*\Big(\sum_{i=1}^{k} \mathbb{1}\{|T_{i,k}^*| > \sqrt{k} \cdot \mu\} \cdot (T_{1,k}^*)^2 \Big) \overset{n\to\infty}{\to} 0 \Big) = 1 \\
\Leftrightarrow\ &\mathbb{P}\Big(\omega : \forall \mu > 0 : \mathbb{E}_{m_n''}^*\Big(\mathbb{1}\{|T_{1,k}^*| > \sqrt{k} \cdot \mu\} \cdot (T_{1,k}^*)^2 \Big) \overset{n\to\infty}{\to} 0 \Big) = 1, \qquad (2.35)
\end{aligned}
$$

aufgrund der Tatsache, dass die $T_{i,k}^*$ zeilenweise identisch verteilt sind.
Hierbei gilt, dass $k = k(m_n'')$.

Zunächst ist hierzu zu zeigen, dass für jedes beliebige $\mu > 0$ gilt, dass

$$
\mathbb{E}_n^*\Big(\mathbb{1}\{|T_{1,k}^*| > \sqrt{k} \cdot \mu\} \cdot (T_{1,k}^*)^2 \Big) \overset{\mathbb{P}}{\to} 0. \qquad (2.36)
$$

Betrachte dann eine Nullfolge $(\mu_n)_{n\in\mathbb{N}}$. Ist dann $(m_n)_{n\in\mathbb{N}}$ eine beliebige Teilfolge der natürlichen Zahlen, so gilt gemäß Lemma 2.5.4 für eine Teilfolge $m_n^{(1)}$ von m_n

$$
\mathbb{E}_{m_n^{(1)}}^*\Big(\mathbb{1}\{|T_{1,k}^*| > \sqrt{k} \cdot \mu_1\} \cdot (T_{1,k}^*)^2 \Big) \overset{\mathbb{P}-f.s.}{\to} 0.
$$

Weiter gilt

$$
\mathbb{E}_{m_n^{(1)}}^*\Big(\mathbb{1}\{|T_{1,k}^*| > \sqrt{k} \cdot \mu_2\} \cdot (T_{1,k}^*)^2 \Big) \overset{\mathbb{P}}{\to} 0
$$

und demzufolge existiert wiederum eine Teilfolge $(m_n^{(2)})_{n\in\mathbb{N}}$ von $(m_n^{(1)})_{n\in\mathbb{N}}$, sodass

$$\mathbb{E}^*_{m_n^{(2)}}\Big(\mathbb{1}\{|T^*_{1,k}| > \sqrt{k}\cdot\mu_2\}\cdot(T^*_{1,k})^2\Big) \overset{\mathbb{P}-f.s.}{\to} 0.$$

Indem so Teilfolgen $(m_n^{(i)})_{n\in\mathbb{N}}$ mit $(m_n^{(i+1)})_{n\in\mathbb{N}} \subseteq (m_n^{i})_{n\in\mathbb{N}}$ konstruiert werden, kann nun die gesuchte Folge $(m''_n)_{n\in\mathbb{N}}$ gewonnen werden durch:

$$m''_i \;:=\; m_i^{(i)}.$$

Weiter gilt für $\mu' < \mu$, dass $\forall n \in \mathbb{N}$

$$\mathbb{E}^*_n\Big(\mathbb{1}\{|T^*_{1,k}| > \sqrt{k}\cdot\mu\}\cdot(T^*_{1,k})^2\Big) \le \mathbb{E}^*_n\Big(\mathbb{1}\{|T^*_{1,k}| > \sqrt{k}\cdot\mu'\}\cdot(T^*_{i,k})^2\Big)$$

und daher folgt, da μ_n als Nullfolge gewählt ist, Aussage (2.35).
Weiter gilt

$$\begin{aligned}
(s_k^*)^2 &:= \sum_{i=1}^{k}\mathbb{E}^*_n(T^*_{i,n})^2\\
&= \sum_{i=1}^{k}\mathbb{E}^*_n\Big[\big(\frac{1}{\sqrt{l}}\sum_{j=1}^{l}\sum_{t=1}^{M}c_t z^*_{\bar\theta}(X_{S_i+j},x_t)\big)\cdot\big(\frac{1}{\sqrt{l}}\sum_{j=1}^{l}\sum_{t=1}^{M}c_t z^*_{\bar\theta}(X_{S_i+j},x_t)\big)\Big]\\
&= k\cdot\sum_{r=1}^{M}\sum_{t=1}^{M}c_r c_t\cdot\mathrm{cov}^*_n\big(\frac{1}{\sqrt{l}}\sum_{j=1}^{l}z^*_{\bar\theta}(X_{S_1+j},x_r),\frac{1}{\sqrt{l}}\sum_{j=1}^{l}z^*_{\bar\theta}(X_{S_1+j},x_t)\big)\\
&= k\cdot\sum_{r=1}^{M}\sum_{t=1}^{M}c_r c_t\cdot\mathrm{cov}^*_n\big(Z^*_n(x_r),Z^*_n(x_t)\big).
\end{aligned}$$

Gemäß Lemma 2.3.1 gilt somit, dass

$$\begin{aligned}
\frac{1}{k}\cdot(s_k^*)^2 &\overset{n\to\infty}{\to} \sum_{r=1}^{M}\sum_{t=1}^{M}c_r c_t\,\mathrm{cov}(\bar G(x_r),\bar G(x_t))\\
&= \sum_{r=1}^{M}\sum_{t=1}^{M}c_r c_t\sum_{\kappa\in\mathbb{Z}}\mathrm{cov}\big(z_{\bar\theta}(X_0,x_r),z_{\bar\theta}(X_\kappa,x_t)\big)\\
&= c^T\bar\Sigma_{x_1,\dots,x_M}c \text{ in } \mathbb{P}\text{-W'keit.}
\end{aligned}$$

Wiederum existiert also eine Teilfolge $(m'_n)_{n\in\mathbb{N}} \subseteq (m''_n)_{n\in\mathbb{N}}$, sodass

$$\frac{1}{k(m'_n)}\cdot(s_k(m'_n)^*)^2 \overset{n\to\infty}{\to} c^T\bar\Sigma_{x_1,\dots,x_M}c \text{ f.s.}$$

und damit gilt, falls die Lindeberg-Bedingung erfüllt ist, dass

$$c^T Z^*_{m''_n} = \frac{1}{\sqrt{k}} \sum_{i=1}^{k} T^*_{i,k} \xrightarrow{\mathcal{D}} \mathcal{N}(0, c^T \bar{\Sigma}_{x_1,\ldots,x_M} c) \text{ fast sicher.}$$

(Es gilt hier ebenfalls: $k = k(m'_n)$). Da die Teilfolge m_n beliebig gewählt war, folgt damit zunächst die Verteilungskonvergenz von $c^T Z^*_n$ in Wahrscheinlichkeit. Die Verteilungskonvergenz von $\vec{W}^*_n$ folgt nun mit dem Satz von Cramér-Slutsky und damit gilt die Aussage des Lemmas.

Zu zeigen bleibt nun also (2.36).

<u>*Nachweis der Lindeberg-Bedingung:*</u>
Gemäß (2.36) ist zu zeigen, dass

$$\mathbb{E}^*_n(\ \mathbb{1}\{|T^*_{1,k}| > \sqrt{k} \cdot \mu\} \cdot (T^*_{1,k})^2) \xrightarrow{\mathbb{P}} 0.$$

Bezeichne wie in (2.3)

$$z_{\bar{\theta}}(X_i, x_r) = \mathbb{1}\{X_i \leq x_r\} + \ell^T(X_i, \bar{\theta}) \cdot \dot{F}_{\bar{\theta}}(x_r).$$

Dann gilt:

$$\begin{aligned}
& \mathbb{E}^*_n(\ \mathbb{1}\{|T^*_{1,k}| > \sqrt{k} \cdot \mu\} \cdot (T^*_{1,k})^2) \\
= \ & \mathbb{E}^*_n\Big\{\Big(\sum_{j=S_1+1}^{S_1+l} \frac{1}{\sqrt{l}} \sum_{r=1}^{M} c_r z_{\bar{\theta}}(X_j, x_r) - \mathbb{E}^*_n[\sum_{j=S_1+1}^{S_1+l} \frac{1}{\sqrt{l}} \sum_{r=1}^{M} c_r z_{\bar{\theta}}(X_j, x_r)]\Big)^2 \\
& \cdot I\Big(\Big| \sum_{j=S_1+1}^{S_1+l} \frac{1}{\sqrt{l}} \sum_{r=1}^{M} c_r z_{\bar{\theta}}(X_j, x_r) - \mathbb{E}^*_n[\sum_{j=S_1+1}^{S_1+l} \frac{1}{\sqrt{l}} \sum_{r=1}^{M} c_r z_{\bar{\theta}}(X_j, x_r)]\Big| > \sqrt{k} \cdot \mu\Big)\Big\}.
\end{aligned}$$

Gezeigt wird nun, dass die Wahrscheinlichkeit, dass der Wert der oben auftretenden Indikatorfunktion gleich 1 ist, für n gegen unendlich gegen Null strebt. Für diejenigen $\omega \in \Omega$, für welche die Indikatorfunktion gleich Null ist, gilt nämlich, dass der in der Lindeberg-Bedingung auftretende bedingte Erwartungswert gleich Null ist. Somit folgt dann die Konvergenz des bedingten Erwartungswertes gegen Null in $\mathbb{P}$-Wahrscheinlichkeit.

Es gilt

$$\mathbb{E}^*_n\Big(\frac{1}{\sqrt{l}} \sum_{j=S_1+1}^{S_1+l} \sum_{r=1}^{M} c_r z_{\bar{\theta}}(X_j, x_r)\Big) = \frac{1}{n-l+1} \sum_{i=0}^{n-l} \big(\frac{1}{\sqrt{l}} \sum_{j=i+1}^{i+l} \sum_{r=1}^{M} c_r z_{\bar{\theta}}(X_j, x_r)\big)$$

Bezeichne in der folgenden Rechnung $T_j := \sum_{r=1}^{M} c_r z_{\bar\theta}(X_j, x_r)$. Dann gilt weiter

$$
\begin{aligned}
& \mathbb{P}\Big(\big|\frac{1}{\sqrt{l}}\sum_{j=S_1+1}^{S_1+l} T_j - \frac{1}{n-l+1}\sum_{i=0}^{n-l}\big(\frac{1}{\sqrt{l}}\sum_{j=i+1}^{i+l} T_j\big)\big| > \sqrt{k}\cdot\mu\Big) \\
= \; & \mathbb{P}\Big(\bigcup_{u=0}^{n-l}\Big\{\big[\frac{1}{\sqrt{l}}\sum_{j=u+1}^{u+l} T_j - \frac{1}{n-l+1}\sum_{i=0}^{n-l}\big(\frac{1}{\sqrt{l}}\sum_{j=i+1}^{i+l} T_j\big)| > \sqrt{k}\cdot\mu\big] \cap \big[S_1 = u\big]\Big\}\Big) \\
\leq \; & \sum_{u=0}^{n-l}\mathbb{P}\Big(|\frac{1}{\sqrt{l}}\sum_{j=u+1}^{u+l} T_j - \frac{1}{n-l+1}\sum_{i=0}^{n-l}\big(\frac{1}{\sqrt{l}}\sum_{j=i+1}^{i+l} T_j\big)| > \sqrt{k}\cdot\mu\Big)\cdot\frac{1}{n-l+1} \\
\leq \; & \max_{u\in\{0,\dots,n-l\}}\mathbb{P}\Big(|\frac{1}{\sqrt{l}}\sum_{j=u+1}^{u+l} T_j - \frac{1}{n-l+1}\sum_{i=0}^{n-l}\big(\frac{1}{\sqrt{l}}\sum_{j=i+1}^{i+l} T_j\big)| > \sqrt{k}\cdot\mu\Big) \\
\leq \; & \max_{u\in\{0,\dots,n-l\}}\frac{\mathbb{E}\big(\frac{1}{l}|\sum_{j=u+1}^{u+l} T_j - \frac{1}{n-l+1}\sum_{i=0}^{n-l}\sum_{j=i+1}^{i+l} T_j|^2\big)}{k\cdot\mu^2} \\
= \; & \frac{\mathbb{E}\big(\frac{1}{l}|\sum_{j=1}^{l} T_j - \frac{1}{n-l+1}\sum_{i=0}^{n-l}\sum_{j=i+1}^{i+l} T_j|^2\big)}{k\cdot\mu^2}.
\end{aligned}
$$

Der letzte Schritt gilt aufgrund der Stationarität des zugrundeliegenden Prozesses der Beobachtungen.
Nun wird die Ordnung des Zählers betrachtet:
Es gilt

$$
\begin{aligned}
& \mathbb{E}\big(\frac{1}{l}|\sum_{j=u+1}^{u+l}\sum_{r=1}^{M} c_r z_{\bar\theta}(X_j, x_r) - \frac{1}{n-l+1}\sum_{i=0}^{n-l}\sum_{j=i+1}^{i+l}\sum_{r=1}^{M} c_r z_{\bar\theta}(X_j, x_r)|^2\big) \\
= \; & \frac{1}{l}\cdot\mathbb{E}\Big[\big(\sum_{j=u+1}^{u+l}\sum_{r=1}^{M} c_r z_{\bar\theta}(X_j, x_r)\big)\Big]^2 \\
& - \frac{2}{l(n-l+1)}\mathbb{E}\Big[\big(\sum_{j=u+1}^{u+l}\sum_{r=1}^{M} c_r z_{\bar\theta}(X_j, x_r)\big)\big(\sum_{i=0}^{n-l}\sum_{j=i+1}^{i+l}\sum_{r=1}^{M} c_r z_{\bar\theta}(X_j, x_r)\big)\Big] \\
& + \frac{1}{l(n-l+1)^2}\mathbb{E}\Big[\big(\sum_{i=0}^{n-l}\sum_{j=i+1}^{i+l}\sum_{r=1}^{M} c_r z_{\bar\theta}(X_j, x_r)\big)^2\Big].
\end{aligned}
$$

Aufgrund der Voraussetzung an die Momente der Funktion $\ell(\cdot, \bar\theta)$ sind die innerhalb der Summen auftretenden Momente beschränkt. Der erste Term umfasst insgesamt $M^2 \cdot l^2$ Summanden, der zweite $(n-l+1)\cdot l^2 \cdot M^2$ und der letzte

$(n-l+1)^2 \cdot l^2 \dot{M}^2$.
Insgesamt ergibt sich also für den Zähler die Ordnung $O(l)$, während der Nenner die Ordnung k besitzt.
Die Wahrscheinlichkeit kann also insgesamt beschränkt werden durch $O(\frac{l}{k})$ und geht gegen Null unter der Voraussetzung $l = o(k)$, womit Lemma (2.3.4) bewiesen ist. □

Zu zeigen ist nun, dass aus der Aussage dieses Lemmas auch die Konvergenz der endlichdimensionalen Verteilungen des Prozesses W_n^* in Wahrscheinlichkeit folgt. Um dies zu folgern, macht man sich die folgende Version des Satzes von Cramér-Wold zunutze (vgl. Pauls [57], Satz 8.2):

Lemma 2.3.5 *Sei $(\xi_n)_{n\in\mathbb{N}}$ eine straffe Folge von Zufallsvariablen $\xi_n \colon \Omega \to \mathbb{R}^d$ und bezeichne $D := \{\lambda_k \colon\ k \in \mathbb{N}\}$ eine dichte Teilmenge in $\mathbb{R}^d$. Gilt*

$$\xi_n^T \cdot \lambda_k \xrightarrow{\mathcal{D}} \xi^T \cdot \lambda_k \quad \forall \lambda_k \in D,$$

so folgt

$$\xi_n \xrightarrow{\mathcal{D}} \xi \quad \text{in } \mathbb{R}^d.$$

Da in Lemma 2.3.4 die Aussage der Verteilungskonvergenz in Wahrscheinlichkeit für beliebiges $c \in \bar{\mathbb{R}}^M$ gezeigt wurde, gilt dies insbesondere auch für jedes c aus einer abzählbaren dichte Teilmenge des $\bar{\mathbb{R}}^M$. Die Straffheit des betrachteten Zufallsvektors in $\mathbb{R}^M$, welcher aus dem zugrundeliegenden Prozess hervorgeht, indem dieser zwischen Gitterpunkten als konstant angenommen wird, folgt aus den Aussagen des folgenden Abschnittes, der stochastischen Äquikontinuität des Prozesses.
Zu zeigen ist also für die Konvergenz der endlichdimensionalen Verteilungen, dass für jede Teilfolge $(m_n)_{\in\in\mathbb{N}} \subseteq \mathbb{N}$ eine weitere Teilfolge $(m'_n)_{n\in\mathbb{N}} \subseteq \mathbb{N}$ existiert, sodass die Verteilungskonvergenz fast sicher gilt. Dies zeigt man folgendermaßen:
Bezeichne $c^{(1)}, c^{(2)}, \ldots$ eine abzählbare dichte Teilmenge des $\mathbb{R}^M$. Aufgrund der Argumentation im Beweis von Lemma 2.3.4 gilt für eine Teilfolge $m_n^{(1)} \subseteq m_n$ die Verteilungskonvergenz fast sicher von $(c^{(1)})^T \cdot \vec{Z}^*_{m_n^{(1)}}$. Für eine Teilfolge $m^{(2)} \subseteq m_n^{(1)}$ gilt wiederum die Verteilungskonvergenz von $(c^{(2)})^T \cdot \vec{Z}^*_{m^{(2)}}$ fast sicher. Analog gilt für festes $i \in \mathbb{N}$, dass für eine Teilfolge $m_n^{(i)} \subseteq m_n^{(i-1)}$ die fast sichere Verteilungs-

konvergenz von $(c^{(i)})^T \cdot \vec{Z}^*_{m_n^{(i)}}$ gilt. Setze nun als die gesuchte Folge m'_n:

$$m'_i \;:=\; m_i^{(i)}.$$

Für m' gilt demzufolge, dass für alle $c^{(i)}$ mit $i \in \mathbb{N}$ gilt, dass $(c^{(i)})^T \cdot \vec{Z}^*_{m'}$ fast sicher in Verteilung gegen ξ konvergiert. Gemäß dem oben genannten Satz von Cramér-Wold gilt somit auch die Verteilungskonvergenz fast sicher von $\vec{Z}^*_{m'}$ gegen $\vec{\bar{G}}$. Insgesamt gilt also die Verteilungskonvergenz von $\vec{Z}^*_n$ in Wahrscheinlichkeit gegen $\vec{\bar{G}}$, für $n \in \mathbb{N}$.
Somit ist die Konvergenz der endlichdimensionalen Verteilungen des Bootstrap-Prozesses Z^*_n gewährleistet. Mit dem Satz von Cramér-Slutsky folgt somit die Konvergenz der endlichdimensionalen Verteilungen von W^*_n gegen diejenigen von $\bar{G}$, wie in folgendem Lemma zusammengefasst:

Lemma 2.3.6 *Es gelte $\mathcal{A}^*$. Dann folgt*

$$\vec{W}^*_n \xrightarrow{\mathcal{D}} \vec{\bar{G}} \text{ in } \mathbb{P}\text{-Wahrscheinlichkeit.}$$

Stochastische Äquikontinuität des Bootstrap-Prozesses

In Analogie zur Vorgehensweise im Falle des Originalprozesses wird nun die stochastische Äquikontinuität des Bootstrap-Prozesses gezeigt. Um die Äquikontinuität des Bootstrap-Prozesses nachzuweisen, wird in mehreren Schritten vorgegangen.
Ziel ist es, die Existenz eines Gitters mit endlich vielen Punkten auf $[-\infty, \infty]$ zu zeigen, sodass das Maximum der Zuwächse des Prozesses zwischen den Gitterpunkten und dazwischenliegenden Punkten mit hoher Warscheinlichkeit (bedingt auf die Originalstichprobe) klein ist, falls $\omega \in \Omega^{(n)}$. $(\Omega^{(n)})_{n\in\mathbb{N}}$ ist wiederum eine Folge von Teilmengen von Ω, deren Wahrscheinlichkeiten mit $n \to \infty$ gegen 1 geht.

Dies wird gezeigt, indem die zwei Teile des Prozesses separat betrachtet werden, der Anteil, welcher der Differenz zwischen empirischer und tatsächlicher Verteilungsfunktion entspricht sowie der Anteil, der den Effekt der Parameterschätzung imitiert.

In diesem Abschnitt wird folgendes Lemma bewiesen:

Lemma 2.3.7 *Es gelte $\mathcal{A}^*$.*
Dann existiert für alle $\nu > 0$ und $\eta > 0$ ein Gitter $\mathcal{G}^*_{\nu,\eta} := \{x^*_0, ..., x^*_{M^*}\}$ auf $\bar{\mathbb{R}}$, sodass gilt:
Es existiert eine Folge $(\Omega^{(n)})_{n\in\mathbb{N}}$ von Teilmengen von Ω mit $\mathbb{P}(\Omega^{(n)}) \to 1$, sodass gilt, für genügend großes n:

$$\mathbb{P}^*_n\Big(\max_{\{i=1,\dots,M^*\}} \sup_{t\in(x^*_{i-1},x^*_i]} |W^*_n(x^*_i) - W_n(t)| > \nu\Big) \leq \eta$$

für $\omega \in \Omega^{(n)}$.

Betrachte also zunächst

$$U^*_n(t) - U^*_n(s) = \frac{\sqrt{l}}{\sqrt{k}} \sum_{i=1}^{k} (H_{S_i}(t) - H_{S_i}(s)) - (\tilde{F}_n(t) - \tilde{F}_n(s)).$$

Die in der Summe auftretenden Zufallsgrößen sind unabhängig und identisch verteilt und durch 1 beschränkt.
In einem ersten Schritt wird eine Folge von Gittern bis zu einer von n abhängigen Feinheit konstruiert, sodass die entsprechenden Zuwächse zwischen dem feinen und dem groben Gitter mit hoher Wahrscheinlichkeit klein sind. Hierfür wird auf die Chaining-Technik zurückgegriffen. Dazu macht man sich zunutze, dass die Zuwächse zwischen zwei festen Punkten mit hoher (bedingter) Wahrscheinlichkeit eine genügend kleine Schranke nicht überschreiten.
Dann wird gezeigt, dass dies auch beim Übergang zum Kontinuum gilt.

Um zu einer Abschätzung der Wahrscheinlichkeiten zu gelangen, dass zwischen zwei festen Punkten die Zuwächse des Prozesses U^*_n eine bestimmte Schranke nicht überschreiten, wird die Ungleichung von Bennett angewendet. Entscheidend hierfür, ebenso wie für den Übergang vom feinen Gitter auf das Kontinuum, ist die Konvergenz der Varianz der Zuwächse des Bootstrap-Prozesses (bedingt auf die Originalstichprobe) gegen den Grenzwert der Varianz der Zuwächse des Originalprozesses. Für die letztgenannte Problematik ist eine gleichmäßige polynomielle Konvergenzrate in Wahrscheinlichkeit zu zeigen. Dies wird am Ende des Abschnitts in mehreren Schritten durchgeführt.

Die Beweisschritte orientieren sich an der Arbeit von Naik-Nimbalkar und Rajarshi [53], wo der empirische Prozess basierend auf mischenden Zufallsvariablen betrachet wurde.
Unter Nutzung dieses Resultates wird nun zunächst die stochastische Äquikontinuität des ersten Teils des Bootstrap-Prozesses nachgewiesen.

Es gelten folgende Bezeichnungen:
Die Varianz der Zuwächse des Bootstrap-Prozesses U_n^*, bedingt auf die Originalstichprobe, wird nun mit $\hat{\sigma}_n^2(\cdot,\cdot)$ bezeichnet. Es gilt also:

$$\hat{\sigma}_n^2(s,t) := \mathrm{var}_n^*(U_n^*(t) - U_n^*(s)).$$

Weiter gelte

$$\sigma^2(s,t) := \lim_{n\to\infty} \mathrm{var}(U_n(t) - U_n(s)) = \sum_{i\in\mathbb{Z}} \mathrm{cov}(\mathbb{1}\{s < X_1 \leq t\}, \mathbb{1}\{x < X_{1+|i|} \leq t\}).$$

(Der Grenzwert existiert aufgrund der Voraussetzung an die schwache Abhängigkeit der X_i.)
Die betragsmäßige Abweichung von $\hat{\sigma}_n^2(s,t)$ von $\sigma^2(s,t)$ wird mit $\Delta_n(s,t)$ bezeichnet. Damit gilt:

$$\hat{\sigma}_n^2(s,t) \leq \sigma^2(s,t) + \Delta_n(s,t).$$

Setze nun noch

$$\Delta_n := \sup_{(s,t)\in\bar{\mathbb{R}}^2} \Delta_n(s,t). \tag{2.37}$$

Zuwächse des Bootstrap-Prozesses U_n^* zwischen zwei festen Punkten

Was die Zuwächse des Prozesses zwischen zwei festen Punkten anbelangt, gilt unter den hier verwendeten Voraussetzungen ein Analogon zu Lemma 4.1 in Naik-Nimbalkar und Rajarshi [53]. Zum Beweis wird die Bennett-Ungleichung herangezogen, wie sie etwa in Pollard [61], S. 192, zu finden ist, wo auch die in den Voraussetzungen des folgenden Lemmas auftretende Funktion $B(\cdot)$ definiert ist.
Es gilt

Lemma 2.3.8 *Es gelte $\mathcal{A}^*$.*
Dann folgt, für hinreichend großes n und ein $b > 0$:

$$\forall \lambda \in (0,1), \forall \eta > 0 \text{ und } \forall \delta > 0 \text{ mit } \frac{\delta^2}{\eta} > \frac{bn^{-\varepsilon}}{2B^{-1}(\lambda)} \text{ und } \delta^2 \geq \Delta_n \text{ gilt}$$

$$\mathbb{P}_n^*[|U_n^*(t) - U_n^*(s)| > \eta] \leq 2\exp\Big\{ -\frac{\eta^2(\lambda/2)}{2\delta^2}\Big\},$$

falls für die Punkte s und t gilt: $L \cdot |F(t) - F(s)|^{1/2} \leq \delta^2$, für ein $L > 0$.

Beweis:
Um dieses Lemma zu beweisen macht man sich die Unabhängigkeit der einzelnen Blöcke zunutze. Hierzu schreiben wir zunächst:

$$U_n^*(t) - U_n^*(s) = \frac{\sqrt{l}}{\sqrt{k}} \sum_{i=1}^{k} H_{S_i}(t) - H_{S_i}(s) - (\tilde{F}_n(t) - \tilde{F}_n(s)).$$

Weiter gilt aufgrund der Unabhängigkeit der Summanden:

$$\mathrm{var}_n^*(\sum_{i=1}^{k} H_{S_i}(t) - H_{S_i}(s) - (\tilde{F}_n(t) - \tilde{F}_n(s))) = \frac{k}{l} \cdot \hat{\sigma}_n^2(s,t).$$

Die einzelnen Summanden sind unabhängig und beschränkt. Daher kann die Ungleichung von Bennett angewendet werden (s. Pollard [61], S. 192) und es gilt unter der Voraussetzung $l \asymp n^{\frac{1}{2}-\varepsilon}$

$$\begin{aligned}
& \mathbb{P}_n^*(|U_n^*(t) - U_n^*(s)| > \eta) \\
= \; & \mathbb{P}_n^*(|\sum_{i=1}^{k} H_{S_i}(t) - H_{S_i}(s) - (\tilde{F}_n(t) - \tilde{F}_n(s))| > \frac{\sqrt{k}}{\sqrt{l}}\eta) \\
= \; & 2\exp\big\{ -\frac{1}{2}\eta\sqrt{k/l} \cdot \frac{\eta\sqrt{l/k}}{\hat{\sigma}_n^2(s,t)} \cdot B\big(\frac{\eta\sqrt{l/k}}{\hat{\sigma}_n^2(s,t)}\big)\big\}
\end{aligned}$$

Die Funktion $\lambda B(\lambda)$ ist monoton wachsend in λ und daher kann weiter abgeschätzt werden

$$\begin{aligned}
\leq \; & 2\exp\big\{ -\frac{1}{2}\eta\sqrt{k/l} \cdot \frac{\eta\sqrt{l/k}}{\Delta_n + \sigma^2(s,t)} \cdot B\big(\frac{\eta\sqrt{l/k}}{\Delta_n + \sigma^2(s,t)}\big)\big\} \\
& \text{und, falls } \Delta_n < \delta^2 \text{ und } \sigma^2(s,t) < \delta^2 \\
\leq \; & 2\exp\big\{ -\frac{1}{2}\eta\sqrt{k/l} \cdot \frac{\eta\sqrt{l/k}}{2\delta^2} \cdot B\big(\frac{\eta\sqrt{l/k}}{2\delta^2}\big)\big\}
\end{aligned}$$

$$
\begin{aligned}
&= 2\exp\big\{-\frac{1}{2}\cdot\frac{\eta^2}{2\delta^2}\cdot B\big(\frac{\eta\sqrt{l/k}}{2\delta^2}\big)\big\} \\
&= 2\exp\big\{-\frac{1}{2}\cdot\frac{\eta^2}{2\delta^2}\cdot B\big(\frac{\eta b n^{-\varepsilon}}{2\delta^2}\big)\big\}, \qquad (2.38)
\end{aligned}
$$

wegen $l \leq b \cdot n^{-\varepsilon}$, für ein $b > 0$ und n hinreichend groß. Weiter kann dieser Ausdruck nach oben beschränkt werden, indem für festes $\lambda \in (0,1)$ abgeschätzt wird

$$B\big(\frac{\eta b n^{-\varepsilon}}{2\delta^2}\big) \geq \lambda,$$

was erfüllt ist,wenn

$$\frac{\eta b n^{-\varepsilon}}{2\delta^2} < B^{-1}(\lambda) \Leftrightarrow \frac{\delta^2}{\eta} > \frac{b n^{-\varepsilon}}{2\cdot B^{-1}(\lambda)}.$$

Dies ist erfüllbar aufgrund der Eigenschaften der Funktion $B(\cdot)$. ($B(\cdot)$ ist stetig und fallend und es gilt $B(x) \geq (1+\frac{1}{3}x)^{-1}$ für $x > 0$, wie in Pollard [61], S. 193, nachzulesen.)
Mit (2.38) kann somit abschließend abgeschätzt werden

$$\mathbb{P}_n^*(|U_n^*(t) - U_n^*(s)| > \eta) \quad \leq \quad 2\exp\big\{-\frac{1}{2}\cdot\frac{\eta^2}{2\delta^2}\cdot\lambda\big\}.$$

□

Approximation des Bootstrap-Prozesses U_n^* auf einem Gitter

Das zuletzt genannte Lemma (2.3.8) wird nun angewendet auf Teilmengen $\tilde{\Omega}_1^{(n)}$ von Ω, für die gilt, dass Δ_n, also die Abweichung $\sup\limits_{(s,t)\in\bar{\mathbb{R}}^2} |\hat{\sigma}_n^2(s,t) - \sigma^2(s,t)|$, von der Ordnung $O(n^{-r})$, für ein $r > 0$, ist. Auf diesen Teilmengen können die bedingten Wahrscheinlichkeiten für Zuwächse des Prozesses zwischen Kontinuum und den Punkten eines geeigneten Gitters beschränkt werden, was Gegenstand der folgenden beiden Lemmata ist. Die Mengen $\tilde{\Omega}_1^{(n)}$ sind hierbei unabhängig vom gewählten Gitter. Letzteres wird konstruiert aufgrund der Vorgaben an die Größe der Zuwächse und der entsprechenden Wahrscheinlichkeiten.
Gemäß Lemma 2.3.13 konvergiert die Wahrscheinlichkeit der Teilmengen $\tilde{\Omega}_1^{(n)}$ gegen 1 für $n \to \infty$.

Wir konstruieren nun eine Folge von Gittern;
Gemäß Pollard [61], S.143 wird δ_i exponentiell fallend gewählt:

$$\delta_i = 2^{-i}.$$

Um die in Lemma 2 genannte Ungleichung ausnutzen zu können, muss gelten (für beliebig aber fest gewähltes $\lambda \in (0,1)$):

$$\{\delta_i^2 \geq \max\{\Delta_n, bn^{-\varepsilon}/(2B^{-1}(\lambda))\}$$

(Wir können davon ausgehen, dass $\eta < 1$ gilt. Durch die Wahl von δ_i oben werden die Voraussetzungen dann auf alle Fälle nicht verletzt.)
Es ist nun ein $i_{max} =: i_n$ zu bestimmen. Dies wird so gewählt, dass

$$\delta_{i_n}^2 \geq \Delta_n + n^{-\varepsilon}/(2B^{-1}(\lambda)).$$

Für i_n gilt dann:

$$\begin{aligned} 2^{-2i_n} &\geq \Delta_n + n^{-\varepsilon}/(2B^{-1}(\lambda)) \\ \Leftrightarrow i_n &\leq -\ln(\Delta_n + n^{-\varepsilon}/(2B^{-1}(\lambda)))/(2\ln 2) \end{aligned}$$

(i_n ist also von der Größenordnung $konstante \cdot \ln n$, falls Δ_n die Ordnung $O(n^{-r})$ besitzt).
Es wird nun eine Folge von Gittern $\mathcal{G}_i^* = \{x_{i,1}, ..., x_{i,M_i}\}$ mit $x_{i,1} = -\infty$ und $x_{i,M_n} = \infty$ auf $\bar{R}$ konstruiert, sodass für benachbarte Punkte $x_{i,j}$ und $x_{i,j-1}$ des Gitters $\mathcal{G}_i^*$ gilt:

$$\sigma^2(x_{i,j}, x_{i,j-1}) \leq \delta_i^2. \tag{2.39}$$

Dies wird erreicht, indem die Punkte so gewählt werden, dass die Zuwächse der Verteilungsfunktion F zwischen den Gitterpunkten des Gitters $\mathcal{G}_i^*$ von der Höhe

$$\frac{1}{L^2} \cdot 2^{-4i}$$

sind. Dann gilt nämlich:

$$\sigma^2(x_{i,j}, x_{i,j-1}) \leq L(F(x_{i,j}) - F(x_{i,j-1}))^{1/2} \leq 2^{-2i} = \delta_i^2.$$

Die Mächtigkeit des Gitters $\mathcal{G}_i^*$ beträgt demnach:

$$\tilde{c} \cdot 2^{4i} + 1.$$

Wie in Pollard gilt weiter (Der Konstante D entspricht hier $(2/\lambda)^{1/2}$, M_i ist die Mächtigkeit des Gitters $\mathcal{G}_i^*$):

$$\eta_i \quad := \quad \Big(\frac{2}{\lambda}\Big)^{1/2} \cdot \delta_i [\ln(M_{i+1}/\delta_i)]^{1/2}.$$

Damit gilt für $x_{i,j}$ und $x_{i,j-1}$ aus $\mathcal{G}_i^*$, für hinreichend großes i:

$$\mathbb{P}_n^*(|U_n(x_{i,j}) - U_n(x_{i,j-1})| > \eta_i) \leq \frac{\delta_i}{M_{i+1}}.$$

Für die Summe der η_i gilt weiter:

$$\begin{aligned}\sum_{i=1}^{k} \eta_i = \sum_{i=1}^{k} \Big(\frac{2}{\lambda}\Big)^{\frac{1}{2}} \delta_i \underbrace{[2\ln(\tilde{c} \cdot (2^{4i}+1) \cdot 2^i)]^{1/2}}_{O(i)} \\ = \sum_{i=1}^{k} \Big(\frac{2}{\lambda}\Big)^{\frac{1}{2}} 2^{-i} \cdot O(i).\end{aligned}$$

Sei nun t beliebig aus $\mathcal{G}_{i_n}^*$. Wird mit $x_i(t)$ der t am nächsten gelegene Punkt des Gitters $\mathcal{G}_i^*$ bezeichnet, so gilt (gemäß Pollard, S. 143) für einen Index i_0:

$$\mathbb{P}_n^*(\max_{t \in G_{i_n}} |U_n^*(x_{i_0}(t)) - U_n^*(t)| > \sum_{k=i_0}^{i_n} \eta_k) \leq \sum_{k=i_0}^{i_n} \delta_k.$$

Insgesamt gilt also zunächst aufgrund der Summierbarkeit der η_k und δ_k:

Lemma 2.3.9 *Für jedes vorgegebene η-δ-Paar existiert ein Index i_0, sodass für $\omega \in \tilde{\Omega}_1^{(n)}$ mit $\tilde{\Omega}_1^{(n)} \subseteq \Omega$ und $\mathbb{P}(\tilde{\Omega}_1^{(n)}) \overset{n\to\infty}{\to} 1$ gilt:*

$$\mathbb{P}_n^*(\max_{j \in 1,\ldots,M_{i_n}} \max_{t \in [x_{i_0,j-1}, x_{i_0,j}) \cap \mathcal{G}_{i_n}^*} |U_n^*(x_{i_0}(t)) - U_n^*(t)| > \eta) \leq \delta.$$

Wähle hierzu i_0 so, dass für die Summe der η_k und δ_k (wie oben definiert) jeweils η bzw. δ als obere Schranke gilt. Dann kann die Wahrscheinlichkeit entsprechend abgeschätzt werden. (S.o.) □

Nachdem im letzten Lemma gezeigt ist, dass für $\omega \in \tilde{\Omega}_1^{(n)}$ (aufgrund der Kleinheit der Abweichung Δ_n) die Zuwächse zwischen den Punkten des Gitters $\mathcal{G}_{i_0}^*$ und den Punkten des Gitters $\mathcal{G}_{i_n}^*$ mit hoher Wahrscheinlichkeit (bedingt auf die Originalstichprobe) klein sind, gilt es nun noch zu zeigen, dass die Zuwächse zwischen

beliebigen Punkten des Kontinuums und dem nächstgelegenen Gitterpunkt aus $\mathcal{G}^*_{i_n}$ ebenfalls mit hoher Wahrscheinlichkeit klein sind. Dann kann der betrachtete Prozess gut auf dem endlichen Gitter $\mathcal{G}^*_{i_0}$ approximiert werden.
Dies folgt aus einem Analogon zu Lemma 4.3 aus Naik-Nimbalkar und Rajarshi

Lemma 2.3.10 *Für $\mathcal{G}^*_{i_n}$ mit der Mächtigkeit M_{i_n} gilt, für $\omega \in \tilde{\Omega}_1^{(n)}$:*

$$\mathbb{P}^*_n(\sup_{t\in\bar{\mathbb{R}}} |U^*_n(x_{i_n}(t)) - U^*_n(t)| > \eta) \leq \delta.$$

Beweis:
Hierzu wird die gleiche Vorgehensweise gewählt wie in Naik-Nimbalkar und Rajarshi, Lemma 4.3;
Es gilt zunächst: Bei einer Originalstichprobe vom Umfang n nimmt $U^*_n(t) - U^*_n(s)$ höchstens $(n+1)^2$ unterschiedliche Werte an, denn:

Für beliebige Punkte t und t' mit $F_n(t) = F_n(t')$ gilt, dass

$$\mathbb{1}\{X_r \leq t\} = \mathbb{1}\{X_r \leq t'\}.$$

Also gilt ebenfalls:

$$\begin{aligned} & \frac{1}{l}\sum_{r=1}^{l} \mathbb{1}\{X_{i+r} \leq t\} - \frac{1}{l(n-l+1)} \sum_{j=0}^{n-l}\sum_{r=1}^{l} \mathbb{1}\{X_{j+r} \leq t\} \\ = \quad & H_i(t) - \mathbb{E}^*_n(F^*_n(t)) = H_i(t') - \mathbb{E}^*_n(F^*_n(t')) \end{aligned}$$

und damit auch

$$\begin{aligned} & H_{S_i}(t) = H_{S_i}(t'),\ \forall i \in \{1, ..., k\} \\ \text{sowie} \quad & U^*_n(t) = U^*_n(t'). \end{aligned}$$

Die empirische Verteilungsfunktion F_n nimmt nur $n+1$ Werte an, sodass $U^*_n(t) - U^*_n(s)$ nur $(n+1)^2$ Werte annehmen kann.

Also gilt

$$\mathbb{P}^*_n(\sup_{t\in\bar{\mathbb{R}}} |U^*_n(x_{i_n}(t)) - U^*_n(t)| > \eta)$$

$$\leq \quad (n+1)^2 \cdot \sup_{t\in\bar{\mathbb{R}}} \mathbb{P}_n^*(|U_n^*(x_{i_n}(t)) - U_n^*(t)| > \eta).$$

Des weiteren gilt folgende Darstellung:

$$= \sum_{i=1}^{k} \frac{[H_{S_i}(t) - H_{S_i}(s)] - [\tilde{F}_n(t) - \tilde{F}_n(s)]}{\sqrt{l} \cdot \sqrt{k}}.$$

Die Summanden sind beschränkt durch $\frac{\sqrt{l}}{\sqrt{k}}$. Daher kann anhand der Bernstein-Ungleichung (Shorack und Wellner [69], S. 855) abgeschätzt werden:

$$\begin{aligned} & \mathbb{P}_n^*(|U_n^*(x_{i_n}(t)) - U_n^*(t)| > \eta) \\ \leq \quad & 2 \cdot \exp\Big\{ -\frac{1}{2} \frac{\eta^2}{\hat{\sigma}_n^2(x_{i,n}(t), t) + \frac{2}{3}\eta\sqrt{l/k}} \Big\} \\ = \quad & \exp\Big\{ -\frac{1}{2} \frac{\eta^2}{\hat{\sigma}_n^2(x_{i,n}(t), t) + \frac{2}{3}\eta l n^{-1/2}} \Big\}. \end{aligned}$$

Des weiteren kann $\hat{\sigma}_n^2(x_{i,n}(t), t)$ gemäßg (2.37) und (2.39) betragsmäßig abgeschätzt werden durch

$$\hat{\sigma}_n^2(x_{i,n}(t), t) \leq \Delta_n + \sigma^2(x_{i,n}(t), t) \leq \Delta_n + \delta_n^2 \tag{2.40}$$

aufgrund der Wahl der Gitterpunkte.
Daher kann die oben genannte Wahrscheinlichkeit beschränkt werden durch

$$\exp\Big\{ -\frac{1}{2} \frac{\eta^2}{\Delta_n + \delta_n^2 + \frac{2}{3}\eta n^{-1/2}} \Big\}.$$

Δ_n konvergiert stochastisch mit polynomieller Rate gegen Null (siehe Lemma 2.3.13). Somit kann der Ausdruck (2.40) auf den Mengen $\tilde{\Omega}_1^{(n)}$ aufgrund der Eigenschaften von Δ_n und δ_n^2 abgeschätzt werden durch

$$O(n^{-r}), \text{ für ein } r > 0.$$

Insgesamt gilt also:

$$\mathbb{P}_n^*(\sup_{t\in\bar{\mathbb{R}}} |U_n^*(x_{i_n}(t)) - U_n^*(t)| > \eta) \leq 2(n+1)^2 \exp(-Cn^r) \overset{n\to\infty}{\to} 0.$$

□

Stochastische Äquikontinuität des zweiten Teils des Bootstrap-Prozesses

Die stochastische Äquikontinuität des zweiten Teils des Prozesses,

$$\frac{1}{\sqrt{n}}\sum_{j=1}^{n}(\ell(X_j^*,\bar{\theta})-\mathbb{E}_n^*\ell(X_j^*,\bar{\theta}))\cdot\dot{F}_{\bar{\theta}}(\cdot),$$

ist Gegenstand des folgenden Lemmas.

Lemma 2.3.11 *Es existiert eine Folge von Teilmengen $\tilde{\Omega}_2^{(n)} \subseteq \Omega$, wobei $\mathbb{P}(\tilde{\Omega}_2^{(n)}) \overset{n\to\infty}{\to} 1$, sodass für alle $\nu > 0$ und $\eta > 0$ ein Gitter $\tilde{\mathcal{G}}_{\nu,\eta}^* = \{y_0, ..., y_{N^*}\}$ existiert, sodass*

$$\mathbb{P}_n^*\Big(\max_{1\le i\le N^*}\sup_{t\in(y_{i-1},y_i]}\Big|\frac{1}{\sqrt{n}}\sum_{j=1}^{n}(\ell(X_j^*,\bar{\theta})-\mathbb{E}_n^*\ell(X_j^*,\bar{\theta}))^T\cdot(\dot{F}_{\bar{\theta}}(y_i)-\dot{F}_{\bar{\theta}}(t))\Big|>\nu\Big)\le\eta,$$

für hinreichend großes n, für alle $\omega \in \tilde{\Omega}_2^{(n)}$.

Beweis:
Aufgrund der gleichmäßigen Stetigkeit von $\dot{F}_{\bar{\theta}}(\cdot)$ und der Existenz von $\lim_{t\to\pm\infty}||\dot{F}_{\bar{\theta}}(t)||_2$ gilt

$$\max_{1\le i\le N^*}\sup_{t\in(y_{i-1},y_i]}|\dot{F}_{\bar{\theta}}(y_i)-\dot{F}_{\bar{\theta}}(t)|\le a,$$

wobei die Konstante a von der Wahl der Gitterpunkte abhängt. Die Gitterpunkte können so gewählt werden, dass die Konstante a beliebig klein werden kann.
Es gilt also:

$$\begin{aligned}
&\mathbb{P}_n^*\Big(\max_{1\le i\le N^*}\sup_{t\in(y_{i-1},y_i]}\Big|\frac{1}{\sqrt{n}}\sum_{j=1}^{n}(\ell(X_j^*,\bar{\theta})-\mathbb{E}_n^*[\ell(X_j^*,\bar{\theta})])^T\cdot(\dot{F}_{\bar{\theta}}(y_i)-\dot{F}_{\bar{\theta}}(t))\Big|>\nu\Big)\\
=\;&\mathbb{P}_n^*\Big(\max_{1\le i\le N^*}\sup_{t\in(y_{i-1},y_i]}\Big|\frac{1}{\sqrt{n}}\sum_{j=1}^{n}(\ell(X_j^*,\bar{\theta})-\mathbb{E}_n^*[\ell(X_j^*,\bar{\theta})])^T\cdot(\dot{F}_{\bar{\theta}}(y_i)-\dot{F}_{\bar{\theta}}(t))\Big|^2>\nu^2\Big)\\
\le\;&\mathbb{P}_n^*\Big(||\frac{1}{\sqrt{n}}\sum_{j=1}^{n}\ell(X_i^*,\bar{\theta})-\mathbb{E}_n^*[\ell(X_j^*,\bar{\theta})]||_2^2\cdot a^2>\nu^2\Big)\\
\le\;&\frac{a^2\cdot\mathbb{E}_n^*\Big(||\frac{1}{\sqrt{n}}\sum_{j=1}^{n}\ell(X_j^*,\bar{\theta})-\mathbb{E}_n^*[\ell(X_j^*,\bar{\theta})]||_2^2\Big)}{\nu^2}.
\end{aligned}$$

Nun ist die Konvergenz von $\mathbb{E}_n^*\Big(||\frac{1}{\sqrt{n}}\sum\limits_{j=1}^{n}\ell(X_j^*,\bar{\theta})-\mathbb{E}_n^*[\ell(X_j^*,\bar{\theta})]||_2^2\Big)$ in Wahrscheinlichkeit gegen eine Konstante nachzuweisen.

Es gilt:

$$
\begin{aligned}
&\mathbb{E}_n^*\Big(||\frac{1}{\sqrt{n}}\sum_{j=1}^{n}\ell(X_j^*,\bar{\theta})-\mathbb{E}_n^*[\ell(X_j^*,\bar{\theta})]||_2^2\Big)\\
=\;&\mathbb{E}_n^*\Big[\Big(\frac{1}{\sqrt{k}}\sum_{j=1}^{k}\frac{1}{\sqrt{l}}\sum_{r=1}^{l}\ell(X_{S_j+r},\bar{\theta})-\mathbb{E}_n^*[\ell(X_{S_j+r},\bar{\theta})]\Big)^T\\
&\qquad\cdot\Big(\frac{1}{\sqrt{k}}\sum_{j=1}^{k}\frac{1}{\sqrt{l}}\sum_{r=1}^{l}\ell(X_{S_j+r},\bar{\theta})-\mathbb{E}_n^*[\ell(X_{S_j+r},\bar{\theta})]\Big)\Big]\\
=\;&\sum_{\tau=1}^{p}\mathbb{E}_n^*\Big[\Big(\frac{1}{\sqrt{k}}\sum_{j=1}^{k}\frac{1}{\sqrt{l}}\sum_{r=1}^{l}\ell(X_{S_j+r},\bar{\theta})-\mathbb{E}_n^*[\ell(X_{S_j+r},\bar{\theta})]\Big)^2_{(\tau)}\Big]\\
=\;&\sum_{\tau=1}^{p}\frac{1}{k}\sum_{i=1}^{k}\mathbb{E}_n^*\Big[\Big(\frac{1}{\sqrt{l}}\sum_{r=1}^{l}\ell(X_{S_j+r},\bar{\theta})-\mathbb{E}_n^*[\ell(X_{S_j+r},\bar{\theta})]\Big)^2_{(\tau)}\Big]\\
=\;&\sum_{\tau=1}^{p}\mathbb{E}_n^*\Big[\Big(\frac{1}{\sqrt{l}}\sum_{r=1}^{l}\ell(X_{S_1+r},\bar{\theta})-\mathbb{E}_n^*[\ell(X_{S_1+r},\bar{\theta})]\Big)^2_{(\tau)}\Big]\\
=\;&\sum_{\tau=1}^{p}\mathbb{E}_n^*\Big[\Big(\frac{1}{\sqrt{l}}\sum_{r=1}^{l}\ell(X_{S_1+r},\bar{\theta})\cdot\frac{1}{\sqrt{l}}\sum_{t=1}^{l}\ell(X_{S_1+t},\bar{\theta})\Big)_{(\tau)}\Big]\\
&\qquad-\mathbb{E}_n^*\Big[\Big(\frac{1}{\sqrt{l}}\sum_{r=1}^{l}\ell(X_{S_1+r},\bar{\theta})\Big)_{(\tau)}\Big]\cdot\mathbb{E}_n^*\Big[\Big(\frac{1}{\sqrt{l}}\sum_{r=1}^{l}\ell(X_{S_1+r},\bar{\theta})\Big)_{(\tau)}\Big].
\end{aligned}
$$

Der Index (τ) bezeichnet hierbei die τ-te Komponente der betrachteten Vektoren. Die Vorgehensweise zum Nachweis der Konvergenz erfolgt analog zum Beweis der Konvergenz in Wahrscheinlichkeit der Kovarianzen des Bootstrap-Prozesses gegen diejenigen des Originalprozesses. □

Wähle nun $\tilde{\Omega}^{(n)} := \tilde{\Omega}_1^{(n)} \cap \tilde{\Omega}_2^{(n)}$ sowie $\mathcal{G}_{\nu,\eta}^* := \mathcal{G}_{i_0}^* \cup \tilde{\mathcal{G}}_{\nu,\eta}^*$. Dann gilt mit analogen Argumenten wie für die stochastische Äquikontinuität des Originalprozesses Lemma 2.3.7, unter Nutzung von Lemma 2.3.9, 2.3.10 und 2.3.11. □

Konvergenz der Varianz der Zuwächse des Bootstrap-Prozesses U_n^*

In diesem Abschnitt wird gezeigt, dass die Varianz der Zuwächse des Bootstrap-Prozesses U_n^* (siehe (2.29)) in Wahrscheinlichkeit gegen einen Grenzwert konvergiert, welcher demjenigen der Varianz der Zuwächse des zugehörigen Teils des Originalprozesses entspricht.
Bezeichne hierzu mit $\sigma^2(s,t)$ den Grenzwert der Varianz der Differenz des Prozesses U_n an den Punkten s und t (für n gegen unendlich).
Zunächst wird nachgewiesen, dass $\sigma^2(s,t)$ unter den Voraussetzungen an die schwache Abhängigkeit und die Verteilungsfunktion der zugrundeliegenden Beobachtungen stetig in s und t ist.
Es gilt:

$$\begin{aligned}
\sigma^2(s,t) &:= \lim_{n\to\infty} \operatorname{var}[U_n(t) - U_n(s)] \\
&= \lim_{n\to\infty} \operatorname{var}\big[\frac{1}{\sqrt{n}} \sum_{i=1}^{n} \mathbb{1}\{s < X_i \leq t\} - (F(t) - F(s))\big] \\
&= \lim_{n\to\infty} \mathbb{E}\big[(\frac{1}{\sqrt{n}} \sum_{i=1}^{n} Y_i(s,t) - F(s,t))^2\big] \\
&= \lim_{n\to\infty} \Big(\operatorname{var}[Y_1(s,t)] + 2 \cdot \sum_{i=1}^{n-1} \frac{n-i}{n} \cdot \operatorname{cov}[Y_1(s,t), Y_{1+i}(s,t)] \Big) \\
&= \lim_{n\to\infty} \Big(\operatorname{var}[Y_1(s,t)] + 2 \cdot \sum_{i=1}^{\infty} \frac{(n-i)_+}{n} \cdot \operatorname{cov}[Y_1(s,t), Y_{1+i}(s,t)] \Big)
\end{aligned}$$

Zur Abschätzung der Kovarianzen wird Lemma 2.5.1 ausgenutzt. Da die Summe über die Kovarianzen somit absolut konvergent ist, gilt mit dem Satz von Lebesgue

$$\begin{aligned}
&= \operatorname{var}[Y_1(s,t)] + 2 \cdot \sum_{i=1}^{\infty} \cdot \operatorname{cov}[Y_1(s,t), Y_{1+i}(s,t)] \\
&\leq |F(t) - F(s)|^{1/2} \cdot \sum_{i=1}^{\infty} \text{const} \cdot \zeta_i^{\frac{2}{3}} \\
&\leq \tilde{C} \cdot |F(t) - F(s)|^{1/2}.
\end{aligned}$$

Es gilt also:

Lemma 2.3.12 *Sei $(X_i)_{i\in\mathbb{Z}}$ ein Prozess, welcher der Voraussetzung $\mathcal{A}(i)$ genügt. Zudem sei die Verteilungsfunktion der Randverteilungen lipschitzstetig. Dann gilt für $\sigma^2(s,t) := \lim_{n\to\infty} \operatorname{var}(U_n(t) - U_n(s))$, dass*

$$\sigma^2(s,t) \leq \tilde{C} \cdot |F(t) - F(s)|^{\frac{1}{2}}.$$

Wie in Lemma 2.5.1 ersichtlich, ist hier wiederum exponentielles Abnehmen des Abhängigkeitskoeffizienten vorausgesetzt, dies kann allerdings abgeschwächt werden. Lediglich die Summierbarkeit im obigen Sinne muss gewährleistet sein.

Das folgende Lemma beinhaltet nun die gleichmäßige Konvergenz von $\hat{\sigma}_n^2(s,t)$ in Wahrscheinlichkeit gegen $\sigma^2(s,t)$ mit polynomieller Rate (vgl. hierzu Theorem 3.1 und Korollar 3.2 in Naik-Nimbalkar und Rajarshi [53]):

Lemma 2.3.13 *Es gilt unter Voraussetzung $\mathcal{A}^*$:*
Es existiert ein $r > 0$, sodass

$$n^r \cdot \sup_{(s,t)\in\bar{\mathbb{R}}^2} |\hat{\sigma}_n^2(s,t) - \sigma^2(s,t)| \xrightarrow{\mathbb{P}} 0.$$

Beweis:
Dieses wird in mehreren Schritten bewiesen. Zunächst wird die gleichmäßige Konvergenz des deterministischen Anteils gezeigt. Dann wird eine Zerlegung des stochastischen Anteils betrachtet und hierfür die gleichmäßgie Konvergenz in Wahrscheinlichkeit der Ordnung $o_{\mathbb{P}}(n^{-r})$ nachgewiesen, unter Ausnutzung von Monotonie-Argumenten wie in Shorack und Wellner [69], S. 95f.

Es gilt zunächst die in Naik-Nimbalkar und Rajarshi [53], S. 985, verwendete Darstellung:

$$\hat{\sigma}_n^2(s,t) = \frac{l}{n-l+1} \sum_{j=0}^{n-l} \Big\{ \sum_{r=1}^{l} \frac{\mathbb{1}\{s < X_{r+j} \leq t\} - (\tilde{F}_n(t) - \tilde{F}_n(s))}{l} \Big\}^2, \tag{2.41}$$

denn

$$\hat{\sigma}_n^2(s,t) \quad = \quad \operatorname{var}_n^*[U_n^*(t) - U_n^*(s)]$$

$$
\begin{aligned}
&= \operatorname{var}_n^* \Big[\frac{1}{\sqrt{k}} \sum_{i=1}^{k} \frac{1}{\sqrt{l}} \sum_{j=1}^{l} \Big\{ \mathbb{1}\{X_{S_i+j} \le t\} - \mathbb{1}\{X_{S_i+j} \le s\} \\
&\qquad\qquad - \frac{1}{n-l+1} \sum_{u=0}^{n-l} (\mathbb{1}\{X_{u+j} \le t\} - \mathbb{1}\{X_{u+j} \le s\}) \Big\} \Big] \\
&= \operatorname{var}_n^* \big[\frac{1}{\sqrt{l}} \sum_{j=1}^{l} \mathbb{1}\{s < X_{S_1+j} \le t\} - (\tilde{F}_n(t) - \tilde{F}_n(s))\big] \\
&= \frac{1}{l} \cdot \mathbb{E}_n^* \Big[\big(\sum_{j=1}^{l} \mathbb{1}\{s < X_{S_1+j} \le t\} - (\tilde{F}_n(t) - \tilde{F}_n(s)) \big)^2 \Big] \\
&= \frac{l}{n-l+1} \sum_{j=0}^{n-l} \Big\{ \sum_{r=1}^{l} \frac{\mathbb{1}\{s < X_{r+j} \le t\} - (\tilde{F}_n(t) - \tilde{F}_n(s))}{l} \Big\}^2 .
\end{aligned}
$$

Zum Beweis des Lemmas wird die Abweichung von $\hat{\sigma}_n^2(s,t)$ von $\sigma^2(s,t)$ in einen stochastischen und einen deterministischen Anteil zerlegt:

$$
\begin{aligned}
& |\hat{\sigma}_n^2(s,t) - \sigma^2(s,t)| \\
\le\ & |\hat{\sigma}_n^2(s,t) - \mathbb{E}\hat{\sigma}_n^2(s,t)| + |\mathbb{E}\hat{\sigma}_n^2(s,t) - \sigma^2(s,t)|.
\end{aligned}
$$

Nun werden der stochastische und deterministische Anteil separat betrachtet.

Konvergenz des deterministischen Anteils

Zunächst wird die gleichmäßige Konvergenz des deterministischen Anteils $|\sigma^2(s,t) - \hat{\sigma}_n^2(s,t)|$ mit polynomieller Rate nachgewiesen.
Es gilt

Lemma 2.3.14 *Unter Voraussetzung $\mathcal{A}^*$ gilt für ein $r > 0$*

$$\sup_{(s,t)\in\bar{\mathbb{R}}^2} |\mathbb{E}[\hat{\sigma}_n^2(s,t)] - \sigma^2(s,t)| = o(n^{-r}).$$

Beweis:

Es gilt zunächst, mit $F(s,t) := F(t) - F(s)$ und $\tilde{F}_n(s,t) = \tilde{F}_n(t) - \tilde{F}_n(s)$ sowie $Y_i(s,t)$ gemäß (2.65):

$$
\begin{aligned}
\mathbb{E}[\hat{\sigma}_n^2(s,t)] &= \frac{l}{(n-l+1)} \sum_{j=0}^{n-l} \mathbb{E}\Big[\big(\frac{1}{l} \sum_{r=1}^{l} Y_{j+r}(s,t) - F(s,t) \\
&\qquad + F(s,t) - \frac{1}{l} \sum_{r=1}^{l} \tilde{F}_n(s,t) \big)^2 \Big]
\end{aligned}
$$

$$
\begin{aligned}
&= \frac{l}{(n-l+1)} \sum_{j=0}^{n-l} \Big\{ \mathbb{E}\Big[\big(\frac{1}{l}\sum_{r=1}^{l} Y_{j+r} - F(s,t)\big)^2\Big] \\
&\quad + 2 \cdot \mathbb{E}\Big[\big(\frac{1}{l}\sum_{r=1}^{l} Y_{j+r}(s,t) - F(s,t)\big)\big(F(s,t) - \tilde{F}_n(s,t)\big)\Big]\Big\} \\
&\quad + l \cdot \mathbb{E}\big[(F(s,t) - \tilde{F}_n(s,t))^2\big].
\end{aligned}
$$

Die letzten beiden Summanden des Ausdrucks können unter Anwendung der Ungleichung von Cauchy-Schwarz und mithilfe von Lemma 2.5.8 und Lemma 2.5.7 abgeschätzt werden durch

$$
\begin{aligned}
&\leq \frac{2l}{n-l+1} \cdot \sum_{j=0}^{n-l} \Big\{ \mathbb{E}\Big[\big(\frac{1}{l}\sum_{r=1}^{l} Y_{1+r}(s,t) - F(s,t)\big)^2\Big]\Big\}^{1/2} \\
&\qquad\qquad \cdot \Big\{ \mathbb{E}\Big[\big(F(s,t) - \tilde{F}_n(s,t)\big)^2\Big]\Big\}^{1/2} \\
&\quad + l \cdot \mathbb{E}\big[(F(s,t) - \tilde{F}_n(s,t))^2\big] \\
&= l \cdot O(\frac{1}{\sqrt{l}}) \cdot O(\frac{l-1}{n-l+1}) + l \cdot O(\frac{l-1}{n-l+1}) = O(\frac{l^2}{n}).
\end{aligned}
$$

Man erhält also

$$
\begin{aligned}
\mathbb{E}[\hat{\sigma}_n^2(s,t)] &= \frac{l}{(n-l+1)} \sum_{j=0}^{n-l} \mathbb{E}\big[\big(\frac{1}{l}\sum_{r=1}^{l} Y_{j+r}(s,t) - F(s,t)\big)^2\big] + O(\frac{l^2}{n}) \\
&= l \cdot \mathbb{E}\big[\big(\frac{1}{l}\sum_{r=1}^{l} Y_{1+r}(s,t) - F(s,t)\big)^2\big] + O(\frac{l^2}{n}) \\
&= \operatorname{var}[Y_1(s,t)] + 2 \cdot \sum_{i=1}^{l-1} \frac{l-i}{l} \cdot \operatorname{cov}[Y_1(s,t), Y_{1+i}(s,t)] + O(\frac{l^2}{n}).
\end{aligned}
$$

Damit gilt insgesamt für den Bias von $\hat{\sigma}_n^2(s,t)$ unter Ausnutzung der Darstellung für $\sigma^2(s,t)$.

$$
\begin{aligned}
&|\mathbb{E}[\hat{\sigma}_n^2(s,t)] - \sigma^2(s,t)| \\
\leq\ & \Big| \sum_{i=-\infty}^{\infty} \operatorname{cov}(Y_1(s,t), Y_{1+|i|}(s,t)) \Big| \\
&- \sum_{i=-(l-1)}^{l-1} \frac{l-|i|}{l} \cdot \operatorname{cov}(Y_1(s,t), Y_{1+|i|}(s,t)) \Big| + O(\frac{l^2}{n})
\end{aligned}
$$

$$
\begin{aligned}
&\leq \sum_{i=-(l-1)}^{l-1} \frac{|i|}{l} \cdot |\operatorname{cov}(Y_1(s,t), Y_{1+|i|}(s,t))| \\
&\quad + \sum_{|i|\geq l} |\operatorname{cov}(Y_1(s,t), Y_{1+|i|}(s,t))| + O(\frac{l^2}{n}) \\
&\leq \sum_{i=-\infty}^{\infty} \frac{|i|}{l} \cdot \operatorname{cov}(Y_1(s,t), Y_{1+|i|}(s,t))| \\
&\quad + \sum_{|i|\geq l} |\operatorname{cov}(Y_1(s,t), Y_{1+|i|}(s,t))| + O(\frac{l^2}{n}).
\end{aligned}
$$

Mit der Voraussetzung der Summierbarkeit bezüglich i von $\frac{|i|}{l} \cdot |\operatorname{cov}(Y_1(s,t), Y_{1+|i|}(s,t))|$, was unter hinreichenden Bedingungen an die Summierbarkeit des Abhängigkeitskoeffizienten erfüllt ist, und anhand der Voraussetzungen an die Blocklänge l gilt insgesamt die Konvergenz gegen Null des gesamten Ausdrucks.
Der erste Term ist von der Ordnung $O(l^{-1})$, der zweite von der Ordnung $O(\rho^l)$ (unter Voraussetzung des exponentiellen Abfallens des Abhängigkeitskoeffizienten gemäß $\mathcal{A}(i)$). Der letzte ist von der Ordnung $n^{-2\varepsilon}$, unter der geforderten Bedingung $l = O(n^{\frac{1}{2}-\varepsilon})$.
Da die Abschätzungen gleichmäßig in $(s,t) \in \bar{\mathbb{R}}^2$ gelten, folgt somit für ein $r > 0$

$$
\sup_{(s,t)\in\bar{\mathbb{R}}^2} |\mathbb{E}[\sigma^2(s,t)] - \sigma^2(s,t)| = o(n^{-r}).
$$

□

Konvergenz des stochastischen Anteils

Hinsichtlich der Konvergenz des stochastischen Anteils wird zunächst eine Zerlegung von $\hat{\sigma}_n^2(s,t) - \mathbb{E}[\hat{\sigma}_n^2(s,t)]$ vorgenommen.
In Analogie zu Naik-Nimbalkar und Rajarshi [53] (S. 985f) kann der stochastische Anteil folgendermaßen dargestellt werden:

$$
\hat{\sigma}_n^2(s,t) - \mathbb{E}[\hat{\sigma}_n^2(s,t)] = \frac{l}{n-l+1}\Big[\sum_{j=0}^{n-l} V_j(s,t) + V(s,t)\Big],
$$

wobei

$$\begin{aligned} V_j(s,t) &= \Big\{ l^{-1} \sum_{r=1}^{l} [Y_{j+r}(s,t) - (F(t) - F(s))] \Big\}^2 \\ &\quad - \mathbb{E}\Big[\Big\{ l^{-1} \sum_{r=1}^{l} [Y_{j+r}(s,t) - (F(t) - F(s))] \Big\}^2 \Big]. \end{aligned} \tag{2.42}$$

und

$$\begin{aligned} V(s,t) &= -(n-l+1)\big\{[\tilde{F}_n(t) - \tilde{F}_n(s)] - [F(t) - F(s)]\big\}^2 \\ &\quad + (n-l+1)\mathbb{E}\big[\big\{[\tilde{F}_n(t) - \tilde{F}_n(s)] - [F(t) - F(s)]\big\}^2\big]. \end{aligned}$$

(Eine ausführliche Rechnung hierzu findet sich bei Lemma 2.5.5)

Nun wird die gleichmäßige Konvergenz von $l/(n-l+1) \cdot V$ nachgewiesen.

Lemma 2.3.15 *Unter Voraussetzung $\mathcal{A}^*$ gilt*

$$\sup_{(s,t)\in\bar{\mathbb{R}}^2} \frac{l}{n-l+1} \cdot V = o_{\mathbb{P}}(n^{-r}).$$

Beweis:

Es gilt

$$\begin{aligned} & \mathbb{P}\Big(n^r \cdot \frac{l}{n-l+1} \cdot \sup_{(s,t)\in\bar{\mathbb{R}}^2} \big|V\big| > \mu\Big) \\ = \; & \mathbb{P}\Big(n^r \cdot l \cdot \sup_{(s,t)\in\bar{\mathbb{R}}^2} \big|(\tilde{F}_n(t) - \tilde{F}_n(s) - (F(t) - F(s)))^2 \\ & \qquad - \mathbb{E}\big[(\tilde{F}_n(t) - \tilde{F}_n(s) - (F(t) - F(s)))^2\big]\big| > \mu\Big) \\ \leq \; & \mathbb{P}\Big(n^r \cdot l \cdot \sup_{(s,t)\in\bar{\mathbb{R}}^2} \big|\tilde{F}_n(t) - F_n(t) + F_n(s) - \tilde{F}_n(s) \\ & \qquad + F_n(t) - F(t) + F(s) - F_n(s)\big|^2 > \frac{\mu}{2}\Big) \\ & + \mathbb{P}\Big(n^r \cdot l \cdot \sup_{(s,t)\in\bar{\mathbb{R}}^2} \mathbb{E}\big[\big|\tilde{F}_n(t) - F_n(t) + F_n(s) - \tilde{F}_n(s) \\ & \qquad + F_n(t) - F(t) + F(s) - F_n(s)\big|^2\big] > \frac{\mu}{2}\Big) \\ \leq \; & \mathbb{P}\Big(n^r \cdot l \cdot \big\{2 \cdot \sup_{t\in\bar{\mathbb{R}}} |\tilde{F}_n(t) - F_n(t)| + 2 \cdot \sup_{t\in\bar{\mathbb{R}}} |F_n(s) - F(t)|\big\}^2 > \frac{\mu}{2}\Big) \end{aligned}$$

$$
\begin{aligned}
&+ \mathbb{P}\big(n^r \cdot l \cdot \sup_{t\in\bar{\mathbb{R}}} \mathbb{E}\big[\big||\tilde{F}_n(t) - F_n(t)| + |F_n(s) - \tilde{F}_n(s)| \\
&\qquad\qquad + |F_n(t) - F(t)| + |F(s) - F_n(s)|\big|^2\big] > \frac{\mu}{2}\big) \\
=:\ & (i) + (ii) \qquad (2.43)
\end{aligned}
$$

Es wird nun zunächst die Differenz zwischen der empirischen Verteilungsfunktion basierend auf der Bootstrap-Stichprobe und der empirischen Verteilungsfunktion basierend auf der Originalstichprobe betrachtet. Hierbei gilt:

$$
\left|\tilde{F}_n(t) - F_n(t)\right| = O(\frac{l-1}{n-l+1}), \qquad (2.44)
$$

wie in Lemma 2.5.7 gezeigt.
Es gilt des weiteren

$$
\begin{aligned}
\mathbb{E}[(F_n(t) - F(t))^4] &= \mathbb{E}\Big[\Big(\frac{1}{n}\sum_{i=1}^{n} \mathbb{1}\{X_i \le t\} - F(t)\Big)^4\Big], \\
&= O(n^{-2}), \qquad (2.45)
\end{aligned}
$$

wie in Lemma 2.5.8 gezeigt.
Betrachte nun weiter ein Gitter $\{x_0, ..., x_{M_n}\}$ auf $\bar{\mathbb{R}}$, wobei gilt $F(x_i) = \frac{i}{M_n}$ und $x_0 := -\infty$, $x_{M_n} := \infty$. Dann gilt, analog zu Shorack und Wellner [69], S. 95f, für $t \in [x_{i-1}, x_i]$:

$$
F_n(x_{i-1}) - F(x_{i-1}) - \frac{1}{M_n} \le F_n(t) - F(t) \le F_n(x_i) - F(x_i) + \frac{1}{M_n}.
$$

Damit gilt für (i) in (2.43)

$$
\begin{aligned}
&\mathbb{P}\big(n^r \cdot l \cdot \big\{2 \cdot \sup_{t\in\bar{\mathbb{R}}} |\tilde{F}_n(t) - F_n(t)| + 2 \cdot \sup_{t\in\bar{\mathbb{R}}} |F_n(t) - F(t)|\big\}^2 > \frac{\mu}{2}\big) \\
\le\ & \mathbb{P}\big(n^r \cdot l \cdot \text{const} \cdot \frac{(l-1)^2}{(n-l+1)^2} > \frac{\mu}{6}\big) \\
&+ \mathbb{P}\big(n^r \cdot l \cdot \text{const} \cdot \frac{l-1}{n-l+1} \cdot (\max_{i\in\{0,...,M_n\}} |F_n(x_i) - F(x_i)| + \frac{1}{M_n}) > \frac{\mu}{6}\big) \\
&+ \mathbb{P}\big(n^r \cdot l \cdot \text{const} \cdot \big\{\max_{i\in\{0,...,M_n\}} |F_n(x_i) - F(x_i)| + \frac{1}{M_n}\big\}^2 > \frac{\mu}{6}\big) \\
\le\ & \mathbb{P}\big(n^r \cdot l \cdot \text{const} \cdot \frac{(l-1)^2}{(n-l+1)^2} > \frac{\mu}{6}\big) \\
&+ (M_n + 1) \cdot \max_{i\in\{0,...,M_n\}} \mathbb{P}\big(n^r \cdot l \cdot \text{const} \cdot \frac{l-1}{n-l+1} \cdot (|F_n(x_i) - F(x_i)| + \frac{1}{M_n}) > \frac{\mu}{6}\big)
\end{aligned}
$$

$$
\begin{aligned}
&+ (M_n+1)\cdot \max_{i\in\{0,\dots,M_n\}} \mathbb{P}\big(n^r\cdot l\cdot \text{const}\cdot\big\{|F_n(x_i)-F(x_i)|+\frac{1}{M_n}\big\}^2>\frac{\mu}{6}\big)\\
\leq\ & \mathbb{P}(\text{const}\cdot n^{r-\frac{1}{2}-3\varepsilon}>\frac{\mu}{6})\\
&+ O\big(M_n\cdot\frac{n^{2r}\cdot l^4}{n^2}\big)\cdot\big(\mathbb{E}(|F_n(t)-F(t)|^2)+M_n^{-2}\big)\\
&+ O\big(M_n\cdot n^r\cdot l\big)\cdot\big(\mathbb{E}(|F_n(t)-F(t)|^2)+M_n^{-2}\big)\\
&\text{(Anwendung von } c_r\text{- und Markov-Ungleichung)}\\
=\ & \mathbb{P}(\text{const}\cdot n^{r-\frac{1}{2}-3\varepsilon}>\frac{\mu}{6})+O\big(M_n\cdot n^{2r-4\varepsilon-1}+M_n^{-1}\cdot n^{2r-4\varepsilon}\big)\\
&+ O\big(M_n\cdot n^{r-\varepsilon-\frac{1}{2}}+M_n^{-1}\cdot n^{r-\varepsilon+\frac{1}{2}}\big)=o(1),
\end{aligned}
$$

falls $n^{\frac{1}{2}-(r-\varepsilon)} >> M_n >> n^{\frac{1}{2}+(r-\varepsilon)}$.
Dies ist wiederum erfüllbar für $0<r<\varepsilon$. □

Nun wird der erste Summand des stochastischen Anteils betrachtet.

Bezeichne mit

$$V^{(n)}(s,t) := \frac{l}{n-l+1}\sum_{j=0}^{n-l} V_j(s,t). \tag{2.46}$$

Das folgende Lemma beinhaltet die gleichmäßige Konvergenz von (2.46).

Lemma 2.3.16 *Es existiert ein $r>0$, sodass unter $\mathcal{A}^*$ gilt:*

$$\sup_{(s,t)\in\bar{\mathbb{R}}^2} |V^{(n)}(s,t)| = o_{\mathbb{P}}(n^{-r}).$$

Beweis:
Zunächst erfolgt die Betrachtung der zweiten Momente von $\sum_{j=0}^{n-l} V_j(s,t)$.
Unter der Voraussetzung, dass der Prozess $(X_i)_{i\in\mathbb{Z}}$ strikt stationär ist, ist V_j ebenso strikt stationär. Die Eigenschaft der schwachen Abhängigkeit der X_i wird im folgenden für die Abschätzung der Kovarianzen von V_j ausgenutzt, wobei sich die Zeitlücke jeweils um $l-1$ verkleinert.
Insgesamt gilt für die zweiten Momente der Summe über $V_j(s,t)$:

$$
\begin{aligned}
&\mathbb{E}\Big[\big(\sum_{j=0}^{n-l} V_j\big)^2\Big]\\
=\ & \sum_{j=0}^{n-l}\sum_{i=0}^{n-l} |\operatorname{cov}(V_j,V_i)|
\end{aligned}
$$

$$
\begin{aligned}
&\leq \; |\operatorname{var}(V_0)| + 2 \cdot \sum_{i=1}^{n-l} (n-l+1-i) \cdot |\operatorname{cov}(V_0, V_i)|. \\
&\leq \; |\operatorname{var}(V_0)| + 2 \cdot \sum_{i=1}^{l-1} (n-l+1-i) \cdot |\operatorname{cov}(V_0, V_i)| \\
&\quad + 2 \cdot \sum_{l \leq i \leq n-l} (n-l+1-i) \cdot |\operatorname{cov}(V_0, V_i)| \\
&\leq \; 2 \cdot (n-l+1) \cdot \Big(\sum_{i=0}^{l-1} \mathrm{const}_1 \cdot \frac{1}{l^2} + \sum_{l \leq i \leq n-l} \mathrm{const}_2 \cdot (\zeta_{i-l+1})^{1/2} \Big) \\
&= \; O\big((n-l+1) \cdot l^{-1} + (n-l+1)\big) \\
&= \; O\big(n-l+1\big). \qquad (2.47)
\end{aligned}
$$

(Siehe Lemma 2.5.6 zur Abschätzung der Kovarianzen $|\operatorname{cov}(V_0, V_i)|$).

Die getätigten Abschätzungen gelten jeweils gleichmäßig für alle $(s,t) \in \bar{\mathbb{R}}^2$. Betrachte nun wiederum $V_j(s,t)$, wobei analog zu Naik-Nimbalkar und Rajarshi [53] (S. 988) für $s \leq t$ folgende Darstellung gilt:

$$
\begin{aligned}
V_j(s,t) &= \Big\{ \frac{1}{l} \sum_{r=1}^{l} [Y_{j+r}(s,t) - F(s,t)] \Big\}^2 \\
&\quad - \mathbb{E}\Big[\Big\{ \frac{1}{l} \sum_{r=1}^{l} [Y_{j+r}(s,t) - F(s,t)] \Big\}^2 \Big] \\
&= \big(\frac{1}{l} \sum_{r=1}^{l} Y_{j+r}(s,t) \big)^2 - 2 \cdot F(s,t) \cdot \frac{1}{l} \sum_{r=1}^{l} Y_{j+r}(s,t) + F(s,t)^2 \\
&\quad - \mathbb{E}\Big[\big(\frac{1}{l} \sum_{r=1}^{l} Y_{j+r}(s,t) \big)^2 \Big] + 2 \cdot \mathbb{E}\Big[\frac{1}{l} \sum_{r=1}^{l} Y_{j+r}(s,t) \Big] \cdot F(s,t) - F(s,t)^2 \\
&= \big(\frac{1}{l} \sum_{r=1}^{l} Y_{j+r}(s,t) \big)^2 - \mathbb{E}\Big[\big(\frac{1}{l} \sum_{r=1}^{l} Y_{j+r}(s,t) \big)^2 \Big] \\
&\quad - 2 \cdot \big(\frac{1}{l} \sum_{r=1}^{l} Y_{j+r}(s,t) \big) \cdot F(s,t) + 2 \cdot F(s,t)^2 \\
&= \frac{1}{l^2} \big(\sum_{r=1}^{l} Y_{j+r}(s,t) \big)^2 - \frac{1}{l^2} \mathbb{E}\Big[\big(\sum_{r=1}^{l} Y_{j+r}(s,t) \big)^2 \Big] \\
&\qquad - 2 \cdot F(s,t) \cdot \frac{1}{l} \sum_{r=1}^{l} \big(Y_{j+r}(s,t) - F(s,t) \big).
\end{aligned}
$$

Es sei hierbei $Y_j(s,t) := \mathbb{1}\{s < X_j \leq t\}$ und die X_j schwach abhängig wie unter Voraussetzung $\mathcal{A}$ (i) gefordert.

Für $s \geq t$ gilt $V^{(n)} = 0$.
Sei nun $s < t$.
Es sei $\bar{\mathbb{R}}^2$ mit einem zweidimensionalen Gitter $\{x_{(1,n)}, ..., x_{(M_n,n)}\}^2$ versehen, wobei die Gitterpunkte $(x_{(i,n)}, x_{(j,n)})$ gemäß den Zuwächsen der Verteilungsfunktion in Abhängigkeit der noch zu bestimmenden Konstanten M_n folgendermaßen gewählt werden:

$$x_{(i,n)} =: x_i = \begin{cases} -\infty, & i = 0, \\ F^{-1}(\frac{i}{M_n}), & i \in \{1, ..., M_n - 1\}, \\ +\infty, & i = M_n. \end{cases}$$

Es gilt mit diesen Bezeichnungen:

$$\sup_{(s,t)\in\bar{\mathbb{R}}^2} |V^{(n)}(s,t)| = \max_{1\leq i,p\leq M_n} \sup_{(s,t)\in[x_{i-1},x_i]\times[x_{p-1},x_p]} |V^{(n)}(s,t)|.$$

<u>Sei nun $(s,t) \in [x_{i-1}, x_i] \times [x_{p-1}, x_p]$. Es gelte zunächst $i < p$.</u>

Dann gilt aufgrund der Monotonie der Indikatorfunktion folgende Abschätzung nach oben, indem s und t jeweils durch die geeigneten Grenzen der jeweiligen Intervalle ersetzt werden:

$$\begin{aligned} V^{(n)}(s,t) \leq \frac{l}{n-l+1} \sum_{j=0}^{n-l} \Big\{ & \frac{1}{l^2}\big(\sum_{r=1}^{l} Y_{j+r}(x_{i-1}, x_p)\big)^2 - \frac{1}{l^2}\mathbb{E}\big[\big(\sum_{r=1}^{l} Y_{j+r}(x_i, x_{p-1})\big)^2\big] \\ & - 2(\frac{p-1}{M_n} - \frac{i}{M_n})\frac{1}{l}\sum_{r=1}^{l} Y_{j+r}(x_i, x_{p-1}) \\ & + 2(\frac{p}{M_n} - \frac{i-1}{M_n})\frac{1}{l}\sum_{r=1}^{l}(\frac{p}{M_n} - \frac{i-1}{M_n})\Big\}. \end{aligned}$$

(Bei denjenigen Summanden, welche größer oder gleich Null sind, wird die untere Grenze innerhalb der Indikatorfunktion möglichst klein sowie die obere möglichst

groß gewählt, bei den anderen Summanden umgekehrt.)

$$
\begin{aligned}
&= \frac{l}{n-l+1}\sum_{j=0}^{n-l}\Big\{\frac{1}{l^2}\big(\sum_{r=1}^{l} Y_{j+r}(x_{i-1},x_p)\big)^2 - \frac{1}{l^2}\mathbb{E}\Big[\big(\sum_{r=1}^{l} Y_{j+r}(x_{i-1},x_p)\big)^2\Big] \\
&\quad - 2(\frac{p}{M_n} - \frac{i-1}{M_n})\frac{1}{l}\sum_{r=1}^{l}\big(Y_{j+r}(x_{i-1},x_p) - (\frac{p}{M_n} - \frac{i-1}{M_n})\big) \\
&\quad + 2(\frac{p}{M_n} - \frac{i-1}{M_n})\frac{1}{l}\sum_{r=1}^{l} Y_{j+r}(x_{i-1},x_p) \\
&\quad - 2(\frac{p-1}{M_n} - \frac{i}{M_n})\frac{1}{l}\sum_{r=1}^{l} Y_{j+r}(x_i,x_{p-1}) \\
&\quad + \frac{1}{l^2}\mathbb{E}\Big[\big(\sum_{r=1}^{l} Y_{j+r}(x_{i-1},x_p)\big)^2\Big] - \frac{1}{l^2}\mathbb{E}\Big[\big(\sum_{r=1}^{l} Y_{j+r}(x_i,x_{p-1})\big)^2\Big]\Big\}
\end{aligned}
$$

(Hier erfolgten Nulladditionen so, dass die Summanden innerhalb der Summe in der ersten und zweiten Zeile zentriert sind. Die letzten drei Zeilen ergeben sich durch die Korrektur der Veränderungen in Zeile eins und zwei.)

$$
\begin{aligned}
&= \frac{l}{n-l+1}\sum_{j=0}^{n-l}\Big\{\frac{1}{l^2}\big(\sum_{r=1}^{l} Y_{j+r}(x_{i-1},x_p)\big)^2 - \frac{1}{l^2}\mathbb{E}\Big[\big(\sum_{r=1}^{l} Y_{j+r}(x_{i-1},x_p)\big)^2\Big] \\
&\quad - 2(\frac{p}{M_n} - \frac{i-1}{M_n})\frac{1}{l}\sum_{r=1}^{l}\big(Y_{j+r}(x_{i-1},x_p) - (\frac{p}{M_n} - \frac{i-1}{M_n})\big) \\
&\quad + 2(\frac{p}{M_n} - \frac{i-1}{M_n})\frac{1}{l}\sum_{r=1}^{l}\big(Y_{j+r}(x_{i-1},x_p) - Y_{j+r}(x_i,x_{p-1})\big) \\
&\quad + 2\cdot\frac{2}{M_n}\frac{1}{l}\sum_{r=1}^{l} Y_{j+r}(x_i,x_{p-1}) \\
&\quad + \frac{1}{l^2}\mathbb{E}\Big[\big(\sum_{r=1}^{l} Y_{j+r}(x_{i-1},x_p)\big)^2\Big] - \frac{1}{l^2}\mathbb{E}\Big[\big(\sum_{r=1}^{l} Y_{j+r}(x_i,x_{p-1})\big)^2\Big]\Big\} \\
= \;& V^{(n)}(x_{i-1},x_p) \\
&+ \frac{1}{n-l+1}\sum_{j=0}^{n-l}\Big\{2(\frac{p}{M_n} - \frac{i-1}{M_n})\frac{1}{l}\sum_{r=1}^{l}\big(Y_{j+r}(x_{i-1},x_p) - Y_{j+r}(x_i,x_{p-1})\big) \\
&\quad + 2\cdot\frac{2}{M_n}\frac{1}{l}\sum_{r=1}^{l} Y_{j+r}(x_i,x_{p-1}) \\
&\quad + \frac{1}{l^2}\mathbb{E}\Big[\big(\sum_{r=1}^{l} Y_{j+r}(x_{i-1},x_p)\big)^2\Big] - \frac{1}{l^2}\mathbb{E}\Big[\big(\sum_{r=1}^{l} Y_{j+r}(x_i,x_{p-1})\big)^2\Big]\Big\}.
\end{aligned}
$$

Demnach gilt für beliebiges s und t in oben genannten Intervallen

$$\begin{aligned}
&V^{(n)}(s,t) - V^{(n)}(x_{i-1},x_p) \\
&\leq \frac{l}{n-l+1}\sum_{j=0}^{n-l}\Big\{2(\frac{p}{M_n}-\frac{i-1}{M_n})\frac{1}{l}\sum_{r=1}^{l}\big(Y_{j+r}(x_{i-1},x_p)-(Y_{j+r}(x_i,x_{p-1})\big) \\
&\qquad + 2\cdot\frac{2}{M_n}\frac{1}{l}\sum_{r=1}^{l}Y_{j+r}(x_i,x_{p-1}) \\
&\qquad + \frac{1}{l^2}\mathbb{E}\Big[\big(\sum_{r=1}^{l}Y_{j+r}(x_{i-1},x_p)\big)^2\Big] - \frac{1}{l^2}\mathbb{E}\Big[\big(\sum_{r=1}^{l}Y_{j+r}(x_i,x_{p-1})\big)^2\Big]\Big\}.
\end{aligned}$$

Nun gilt:

$$\begin{aligned}
Y_{j+r}(x_{i-1},x_p) - Y_{j+r}(x_i,x_{p-1}) &= \mathbb{1}\{X_{j+r}\in(x_{i-1},x_p]\} - \mathbb{1}\{X_{j+r}\in(x_i,x_{p-1}]\} \\
&= \mathbb{1}\{X_{j+r}\in(x_{i-1},x_i]\} + \mathbb{1}\{X_{j+r}\in(x_{p-1},x_p]\} \\
&= Y_{j+r}(x_{i-1},x_i) + Y_{j+r}(x_{p-1},x_p).
\end{aligned}$$

Daher kann weiter geschrieben werden:

$$\begin{aligned}
&V^{(n)}(s,t) - V^{(n)}(x_{i-1},x_p) \\
&\leq \frac{l}{n-l+1}\sum_{j=0}^{n-l}\Big\{2(\frac{p}{M_n}-\frac{i-1}{M_n})\frac{1}{l}\sum_{r=1}^{l}\big(Y_{j+r}(x_{i-1},x_i)+(Y_{j+r}(x_{p-1},x_p)\big) \\
&\qquad + 2\cdot\frac{2}{M_n}\frac{1}{l}\sum_{r=1}^{l}Y_{j+r}(x_i,x_{p-1}) \\
&\qquad + \frac{1}{l^2}\cdot\mathbb{E}\Big[\big(\sum_{r=1}^{l}Y_{j+r}(x_{i-1},x_p)\big)^2\Big] - \frac{1}{l^2}\mathbb{E}\Big[\big(\sum_{r=1}^{l}Y_{j+r}(x_i,x_{p-1})\big)^2\Big]\Big\} \qquad (2.48) \\
&=: R_{n,1}(x_{i-1},x_p).
\end{aligned}$$

Für jede einzelne Zeile des vorletzten Ausdrucks wird nun Konvergenz gegen Null in Wahrscheinlichkeit der Ordnung n^{-r} nachgewiesen.

Betrachtung der ersten Zeile in (2.48)

Diese Zeile kann zunächst geschrieben werden als

$$\begin{aligned}
&\frac{l}{n-l+1}\sum_{j=0}^{n-l}2(\frac{p-i+1}{M_n})\frac{1}{l}\sum_{r=1}^{l}Y_{j+r}(x_{i-1},x_i)-\mathbb{E}(Y_{j+r}(x_{i-1},x_i))\\
&+\frac{l}{n-l+1}\sum_{j=0}^{n-l}2(\frac{p-i+1}{M_n})\frac{1}{l}\sum_{r=1}^{l}Y_{j+r}(x_{p-1},x_p)-\mathbb{E}(Y_{j+r}(x_{p-1},x_p))\\
&+\frac{l}{n-l+1}\sum_{j=0}^{n-l}2(\frac{p-i+1}{M_n})\frac{1}{l}\sum_{r=1}^{l}2\frac{1}{M_n}\\
&=:2\cdot l\cdot\frac{p-i+1}{M_n}\cdot(A_n+B_n+C_n).
\end{aligned}$$

mit

$$\begin{aligned}
A_n &= \frac{1}{n-l+1}\sum_{j=0}^{n-l}\frac{1}{l}\sum_{r=1}^{l}Y_{j+r}(x_{i-1},x_i)-\mathbb{E}(Y_{j+r}(x_{i-1},x_i)),\\
B_n &= \frac{1}{n-l+1}\sum_{j=0}^{n-l}\frac{1}{l}\sum_{r=1}^{l}Y_{j+r}(x_{p-1},x_p)-\mathbb{E}(Y_{j+r}(x_{p-1},x_p)),\\
C_n &= \frac{1}{n-l+1}\sum_{j=0}^{n-l}\frac{1}{l}\sum_{r=1}^{l}2\frac{1}{M_n}.
\end{aligned}$$

Die Wahrscheinlichkeit, dass $2\cdot l\cdot\frac{p-i+1}{M_n}\cdot(A_n+B_n+C_n)\cdot n^r$ betragsmäßig größer als ein fest vorgegebenes μ wird, kann zunächst mithilfe der Markov-Ungleichung abgeschätzt werden.
Daher werden zunächst die Momente von A_n, B_n und C_n betrachtet. Es gilt unter Verwendung einer Ungleichung für die vierten Momente schwach abhängiger Zufallsvariablen (s. Lemma 2.5.9), mit $\tilde{\zeta}:=\text{const}\cdot\zeta$

$$\begin{aligned}
\mathbb{E}(A_n^4) &= \mathbb{E}\Big[\Big(\frac{1}{n-l+1}\sum_{j=0}^{n-l}\frac{1}{l}\sum_{r=1}^{l}Y_{j+r}(x_{i-1},x_i)-\mathbb{E}\big[Y_{j+r}(x_{i-1},x_i)\big]\Big)^4\Big]\\
&\leq 4!\Big(\frac{1}{(n-l+1)^2}\big(\sum_{v=0}^{l-1}1+\sum_{v=l}^{n-l-1}\zeta_{v-l+1}^{\frac{1}{2}}\big)^2\\
&\qquad+\frac{1}{(n-l+1)^3}\big(\sum_{v=0}^{l-1}(v+1)^2+\sum_{v=l}^{n-l-1}(v+1)^2\zeta_{v-l+1}^{\frac{1}{2}}\big)\Big)
\end{aligned}$$

$$= \quad O(\frac{l^2}{(n-l+1)^2}) = O(n^{-1-2\varepsilon}). \tag{2.49}$$

Dieselbe Überlegung gilt für B_n.
Für

$$C_n = \frac{1}{n-l+1} \sum_{j=0}^{n-l} \frac{1}{l} \sum_{r=1}^{l} 2 \frac{1}{M_n}$$

gilt:

$$C_n = O(\frac{1}{M_n}). \tag{2.50}$$

Insgesamt gilt also mit (2.49) und (2.50), unter Berücksichtigung von $0 \leq (p - i + 1)/M_n \leq 1$:

$$\begin{aligned} & \mathbb{P}\Big(2 \cdot l \cdot \frac{p-i+1}{M_n} \cdot (A_n + B_n + C_n) \cdot n^r > \mu\Big) \\ \leq \quad & \mu^{-4} \cdot \mathbb{E}\Big[\Big(2 \cdot l \cdot \frac{p-i+1}{M_n} \cdot (A_n + B_n + C_n) \cdot n^r\Big)^4\Big] \\ = \quad & O(n^{1-6\varepsilon+4r}) + O(\frac{n^{2-4\varepsilon+4r}}{M_n^4}). \end{aligned} \tag{2.51}$$

Betrachtung der zweiten Zeile in (2.48)

Aufgrund der Beschränktheit durch 1 von $Y_{j+r}(x_i, x_{p-1})$ gilt:

$$\frac{l}{n-l+1} \sum_{j=0}^{n-l} 2 \cdot \frac{2}{M_n} \frac{1}{l} \sum_{r=1}^{l} Y_{j+r}(x_i, x_{p-1}) = O(\frac{l}{M_n}). \tag{2.52}$$

Betrachtung der dritten Zeile in (2.48)

Aufgrund der Tatsache, dass die dort auftretenden quadrierten Erwartungswerte für jedes $j \in \{0, ..., n-l\}$ identisch sind, kann diese Zeile mittels Ausmultiplizieren zunächst geschrieben werden als

$$\frac{1}{l} \sum_{r=1}^{l} \sum_{t=1}^{l} \mathbb{E}\big[Y_r(x_{i-1}, x_p) \cdot Y_t(x_{i-1}, x_p) - Y_r(x_i, x_{p-1}) \cdot Y_t(x_i, x_{p-1})\big]$$

$$
\begin{aligned}
= \quad & \frac{1}{l}\sum_{r=1}^{l}\sum_{t=1}^{l}\mathbb{E}\big[Y_r(x_{i-1},x_i)\cdot Y_t(x_{i-1},x_i)+Y_r(x_{p-1},x_p)\cdot Y_t(x_{p-1},x_p) \\
& + Y_r(x_{i-1},x_i)\cdot Y_t(x_{p-1},x_p)+Y_r(x_{p-1},x_p)\cdot Y_t(x_{i-1},x_i) \\
& + Y_r(x_i,x_{p-1})\cdot Y_t(x_{i-1},x_i)+Y_r(x_i,x_{p-1})\cdot Y_t(x_{p-1},x_p) \\
& + Y_r(x_{i-1},x_i)\cdot Y_t(x_i,x_{p-1})+Y_r(x_{p-1},x_p)\cdot Y_t(x_i,x_{p-1})\big].
\end{aligned}
$$

Exemplarisch erfolgt nun die betragsmäßige Abschätzung des Terms

$$
\Big|\frac{1}{l}\sum_{r=1}^{l}\sum_{t=1}^{l}\mathbb{E}\big[Y_r(x_{i-1},x_i)\cdot Y_t(x_i,x_{p-1})\big]\Big|
$$

durch

$$
\begin{aligned}
& \frac{1}{l}\Big|\sum_{r=1}^{l}\sum_{t=1}^{l}\mathbb{E}\big[Y_r(x_{i-1},x_i)\cdot Y_t(x_i,x_{p-1})\big]\Big| \\
\leq \quad & \frac{1}{l}\sum_{r=1}^{l}\sum_{t=1}^{l}\mathbb{E}\big[Y_r(x_{i-1},x_i)\big] \\
\leq \quad & l\cdot C\cdot\frac{1}{M_n} \\
= \quad & O(\frac{l}{M_n}). \qquad (2.53)
\end{aligned}
$$

Ebenso können auch die anderen Terme abgeschätzt werden. Multipliziert mit n^r ergibt dies ebenfalls die Ordnung $O(n^{1/2-\varepsilon+r})$.
Insgesamt gilt also unter Verwendung von (2.51), (2.52) und (2.53)

$$
\mathbb{E}\big[R_{n,1}(x_i,x_p)^4\big] = O(\frac{l^6}{(n-l+1)^2}+\frac{l^4}{M_n^4})O(n^{1-6\varepsilon})+O(\frac{n^{2-4\varepsilon}}{M_n^4}). \quad (2.54)
$$

Damit gilt unter Anwendung der Markov-Ungleichung

$$
\begin{aligned}
& \mathbb{P}\Big(n^r\cdot|R_n(x_i,x_p)|>\mu\Big) = \frac{n^{4r}\cdot\mathbb{E}\big[R_n(x_i,x_p)^4\big]}{\mu^4} \\
= \quad & O(n^{1-6\varepsilon+4r})+O(\frac{n^{2-4\varepsilon+4r}}{M_n^4}). \qquad (2.55)
\end{aligned}
$$

Hieraus ergibt sich eine Bedingung an ε, denn es soll gelten

$$
n^{1-6\varepsilon+4r}\to 0,
$$

weshalb gelten muss

$$
6\varepsilon-4r>1 \Leftrightarrow r<\frac{1}{4}-\frac{3}{2}\varepsilon,
$$

was unter der Bedingung

$$\frac{1}{4} - \frac{3}{2}\varepsilon > 0 \Leftrightarrow \varepsilon > \frac{1}{6}$$

zu realisieren ist.
Des weiteren ergibt sich aus (2.55) als Bedingung

$$n^{1/2-\varepsilon+r} = o(M_n). \tag{2.56}$$

Analoge Betrachtungen führen zu einer ähnlichen Abschätzung nach unten, wobei gilt

$$\begin{aligned}
& V_n(s,t) - V_n(x_i, x_{p-1}) \\
\geq\ & \frac{l}{n-l+1} \sum_{j=0}^{n-l} \Big\{ -2\Big(\frac{p}{M_n} - \frac{i-1}{M_n}\Big)\frac{1}{l}\sum_{r=1}^{l}\big(Y_{j+r}(x_{i-1}, x_i) + Y_{j+r}(x_{p-1}, x_p)\big) \\
& \qquad - 2 \cdot \frac{2}{M_n}\frac{1}{l}\sum_{r=1}^{l} Y_{j+r}(x_i, x_{p-1}) \\
& \qquad + \frac{1}{l^2}\mathbb{E}\Big[\Big(\sum_{r=1}^{l} Y_{j+r}(x_i, x_{p-1})\Big)^2\Big] - \frac{1}{l^2}\mathbb{E}\Big[\Big(\sum_{r=1}^{l} Y_{j+r}(x_{i-1}, x_p)\Big)^2\Big]\Big\} \\
=:\ & R_{n,2}(x_i, x_{p-1}).
\end{aligned}$$

Also gilt insgesamt für beliebiges $(s,t) \in [x_{i-1}, x_i] \times [x_{p-1}, x_p]$ mit $i < p$:

$$V^{(n)}(x_i, x_{p-1}) + R_{n,2}(x_i, x_{p-1}) \leq V^{(n)}(s,t) \leq V^{(n)}(x_{i-1}, x_p) + R_{n,1}(x_{i-1}, x_p))$$

und daher

$$\begin{aligned}
\sup_{(s,t)\in[x_{i-1},x_i]\times[x_{p-1},x_p]} |V^{(n)}(s,t)| \leq \max\Big\{ & |V^{(n)}(x_i, x_{p-1}) + R_{n,2}(x_i, x_{p-1})|, \\
& |V^{(n)}(x_{i-1}, x_p) + R_{n,1}(x_{i-1}, x_p)|\Big\}.
\end{aligned}$$

Für $i = p$ gilt folgende Abschätzung nach oben:

$$\begin{aligned}
& V^{(n)}(s,t) \\
\leq\ & \frac{l}{n-l+1}\sum_{j=0}^{n-l}\frac{1}{l^2}\Big(\sum_{r=1}^{l} Y_{j+r}(x_{i-1}, x_i)\Big)^2 + 2 \cdot F(x_{i-1}, x_i)
\end{aligned}$$

$$
\begin{aligned}
&= \frac{l}{n-l+1}\sum_{j=0}^{n-l}\frac{1}{l^2}\big(\sum_{r=1}^{l}Y_{j+r}(x_{i-1},x_i)\big)^2 - \mathbb{E}\big[\frac{1}{l^2}\big(\sum_{r=1}^{l}Y_{j+r}(x_{i-1},x_i)\big)^2\big] \qquad (2.57)\\
&\quad + \frac{l}{n-l+1}\sum_{j=0}^{n-l}\Big(2\cdot F(x_{i-1},x_i) + \mathbb{E}\big[\frac{1}{l^2}\big(\sum_{r=1}^{l}Y_{j+r}(x_{i-1},x_i)\big)^2\big]\Big) \qquad (2.58)\\
&=: \tilde{R}_{n,1}(x_{i-1},x_i)
\end{aligned}
$$

Der Ausdruck (2.58) kann abgeschätzt werden durch $\frac{l}{M_n}$.
Hinsichtlich (2.57) werden die zweiten Momente in analoger Weise wie bei der Abschätzung der zweiten Momente von $\sum V_j(s,t)$ abgeschätzt (Siehe hierzu Lemma 2.5.6 und (2.47)).
Es gilt also

$$
\mathbb{E}\big[(\tilde{R}_{n,1}(x_{i-1},x_i))^2\big] = O(\frac{l^3}{n-l+1}) + O(\frac{l^2}{M_n^2}). \qquad (2.59)
$$

Für $i = p$ gilt als Abschätzung nach unten

$$
\begin{aligned}
&V^{(n)}(s,t)\\
&\geq \frac{l}{n-l+1}\sum_{j=0}^{n-l}\Big(-\frac{1}{l^2}\cdot\mathbb{E}\big[\big(\sum_{r=1}^{l}Y_{j+r}(x_{i-1},x_i)\big)^2\big] - 2\cdot\frac{1}{M_n}\cdot\frac{1}{l}\sum_{r=1}^{l}Y_{j+r}(x_{i-1},x_i)\Big)\\
&=: \tilde{R}_{n,2}(x_{i-1},x_i) = O(\frac{l}{M_n}). \qquad (2.60)
\end{aligned}
$$

und somit

$$
\begin{aligned}
&\max_{1\leq i<p\leq M_n}\ \sup_{s,t\in[x_{i-1},x_i]\times[x_{p-1},x_p]}|V^{(n)}(s,t)|\\
\leq\ &\max_{1\leq i<p\leq M_n}\max\Big\{|V^{(n)}(x_i,x_{p-1}) + R_{n,2}(x_i,x_{p-1})|,\\
&\qquad |V^{(n)}(x_{i-1},x_p) + R_{n,1}(x_{i-1},x_p)|,\\
&\qquad |\tilde{R}_{n,1}(x_{i-1},x_i)|, |\tilde{R}_{n,2}(x_{i-1},x_i)|\Big\}.
\end{aligned}
$$

Es gilt weiter anhand obiger Überlegungen (siehe (2.47), (2.46), (2.54), (2.59) und (2.60))

$$
\begin{aligned}
&\mathbb{P}\Big(n^r\cdot\max_{1\leq i,p\leq M_n}\max\{|V^{(n)}(x_i,x_{p-1}) + R_{n,2}(x_i,x_{p-1})|,\\
&\quad |V^{(n)}(x_{i-1},x_p) + R_{n,1}(x_{i-1},x_p)|, |\tilde{R}_{n,1}(x_{i-1},x_i)|, |\tilde{R}_{n,2}(x_{i-1},x_i)|\} > \mu\Big)
\end{aligned}
$$

$$\begin{aligned}
&\leq M_n^2 \cdot \Big\{ \mathbb{P}\Big(n^r \cdot |V^{(n)}(x_i, x_{p-1})| > \mu/2\Big) + \mathbb{P}\Big(n^r \cdot |V^{(n)}(x_{i-1}, x_p)| > \mu/2\Big) \\
&\quad + \mathbb{P}\Big(n^r \cdot |R_{n,1}(x_i, x_{p-1})| > \mu/2\Big) + \mathbb{P}\Big(n^r \cdot |R_{n,2}(x_{i-1}, x_p)| > \mu/2\Big) \\
&\quad + \mathbb{P}\Big(n^r \cdot |\tilde{R}_{n,1}(x_{i-1}, x_i)| > \mu\Big) + \mathbb{P}\Big(n^r \cdot |R_{n,1}(x_{i-1}, x_i)| > \mu\Big) \Big\} \\
&= O(M_n^2 \cdot n^{1/2-3\varepsilon+2r}) + O(M_n^2 \cdot n^{1-6\varepsilon+4r}) + O(n^{2-4\varepsilon} \cdot M_n^{-2}),
\end{aligned}$$

falls (2.56) erfüllt ist. Als Bedingung an M_n ergeben sich hieraus

$$M_n >> n^{\frac{1}{2}-\varepsilon+4r}$$

und hieraus $\varepsilon < \frac{1}{2}$,

$$M_n << n^{\varepsilon-r},$$

woraus folgt $r < \varepsilon$, und

$$M_n << n^{3\varepsilon-\frac{1}{2}-2r},$$

woraus sich $\varepsilon > \frac{1}{6}$ ergibt. Da hiermit gelten muss

$$\min\{\varepsilon - r, 3\varepsilon - \frac{1}{2} - 2r\} > \frac{1}{2} - \varepsilon + 4r,$$

folgt die Bedingung $\varepsilon > 1/4$. □

Verteilungskonvergenz in Wahrscheinlichkeit des Bootstrap-Prozesses

Nun wird gefolgert, dass aus der Konvergenz der endlichdimensionalen Verteilungen und der stochastischen Äquikontinuität, was jeweils in den Abschnitten 2.3.2 und 2.3.2 gezeigt wurde, auch die Verteilungskonvergenz in Wahrscheinlichkeit des Bootstrap-Prozesses folgt, gemäß Lemma 2.3.1.

Zu zeigen ist, dass für alle Teilfolgen $(m_n)_{n\in\mathbb{N}}$ der natürlichen Zahlen eine Teilfolge $(m'_n)_{n\in\mathbb{N}}$ existiert, sodass

$$\mathbb{P}(\forall\, f \in \mathcal{C}_b : \, |\mathbb{E}^*_{m'_n} f(W^*_{m'_n}) - \mathbb{E} f(\bar{G})| \overset{n\to\infty}{\to} 0) = 1,$$

was äquivalent ist dazu, dass

$$\mathbb{P}\big(\forall f \in \mathcal{C}_b,\, \forall \nu > 0 : \exists n_0 = n_0(f, \nu) : \, \forall n \geq n_0 : |\mathbb{E}^*_{m'_n} f(W^*_{m'_n}) - \mathbb{E} f(\bar{G})| \leq \nu\big) = 1.$$

(Siehe (2.1) und (2.28) zu den verwendeten Bezeichnungen.)

Zunächst werden die Testfunktionen $f \in \mathcal{C}_b$ in abzählbar viele Klassen eingeteilt. Für alle $(M, N) \in \mathbb{N}^2$ gelte:

$$\begin{aligned}\mathcal{C}_b^{(N,M)} \quad := \quad \{f \in \mathcal{C}_b : \, ||f||_\infty < N, \, \forall x, y \in D(\bar{\mathbb{R}}) : \\ ||x - y|| < \frac{1}{N} \, \Rightarrow \, |f(x) - f(y)| < \frac{1}{M}\}.\end{aligned}$$

Für jedes f und M existiert somit ein $N_0(f, M)$, sodass $\forall N \geq N_0(f, M)$ gilt, dass $f \in \mathcal{C}_b^{(N,M)}$.
Gemäß Lemma 2.3.7 existiert für alle M und N aus $\mathbb{N}$ ein endliches Gitter $\mathcal{G}^*_{(M,N)} = \{x_0^*, ..., x^*_{\lambda_{(M,N)}}\}$ auf $\bar{\mathbb{R}}$, sodass gilt, dass für jede Teilfolge $(m_n)_{n\in\mathbb{N}}$ von $\mathbb{N}$ eine Teilfolge $(m''_n)_{n\in\mathbb{N}}$ existiert, sodass mit Wahrscheinlichkeit 1

$$\mathbb{P}^*_{m''_n}\Big(\max_{i=1,...,\lambda_{(M,N)}} \sup_{t\in(x^*_{i-1},x^*_i]} |W^*_{m''_n}(x^*_i) - W^*_{m''_n}(t)| > \frac{1}{N}\Big) \leq \frac{1}{MN},$$

für hinreichend großes n.
(Die Abhängigkeit der x_i^* von (M, N) wird zugunsten der besseren Lesbarkeit unterdrückt.)
Definiere gemäß der Definition von $\mathcal{G}^*_{\nu,\eta}$ in Abschnitt 2.3.2, Lemma 2.3.7 weiter eine abzählbare Menge von endlichen Gittern

$$\mathcal{X} := \{\mathcal{G}^*_{(M,N)} : \, (M, N) \in \mathbb{N}^2\}.$$

Durch Konstruktion einer Diagonalfolge wie im Beweis von Lemma 2.3.6 folgt, dass ebenso eine entsprechende Teilfolge $(m'''_n)_{n\in\mathbb{N}}$ existiert, sodass

$$\begin{aligned}\mathbb{P}\Big(\quad & \forall \, \mathcal{G}^*_{(M,N)} \in \mathcal{X} : \, \forall n \geq n_0(M, N) : \\ & \mathbb{P}^*_{m'''_n}\Big(\max_{\{i=1,...,\lambda_{(M,N)}\}} \sup_{t\in(x^*_{i-1},x^*_i]} |W^*_{m'''_n}(x^*_i) - W^*_{m'''_n}(t)| > \frac{1}{N}\Big) \leq \frac{1}{MN}\Big) = 1.\end{aligned}$$

Mit analogen Argumenten existiert wieder eine Teilfolge $(m'_n)_{n\in\mathbb{N}}$, sodass für diese Teilfolge auch die Konvergenz der entsprechenden endlichdimensionalen Verteilungen für alle $\mathcal{G}^*_{(M,N)} \in \mathcal{X}$ mit Wahrscheinlichkeit 1 gilt. Bezeichne die entsprechende Teilmenge von Ω mit Ω^* und betrachte nun ein festes $\omega \in \Omega^*$:
Seien $f \in \mathcal{C}_b$ und $\nu > 0$ beliebig aber fest gewählt. Setze M so, dass $\nu \geq \frac{7}{M}$ und wähle weiter N so, dass gilt

$$\forall x, y \in D(\bar{\mathbb{R}}) : \, ||x - y|| < \frac{1}{N} \, \Rightarrow \, |f(x) - f(y)| < \frac{1}{M}.$$

Dann gilt für das entsprechende Gitter $\mathcal{G}^*_{(M,N)}$, wobei $\widetilde{W}^*_n$ und $\widetilde{\bar{G}}$ konstant sind zwischen den Punkten des Gitters $\mathcal{G}^*_{(M,N)}$

$$\begin{aligned}
& |\mathbb{E}^*_{m'_n}(f(W^*_{m'_n})) - \mathbb{E}^*_{m'_n}(f(\widetilde{W}^*_n))| \\
\leq\ & \frac{1}{M} \cdot \mathbb{P}^*_{m'_n}(||W^*_{m'_n} - \widetilde{W}^*_{m'_n}||_\infty \leq \frac{1}{N}) + 2 \cdot ||f||_\infty \cdot \mathbb{P}^*_{m'_n}(||W^*_{m'_n} - \widetilde{W}^*_{m'_n}||_\infty > \frac{1}{N}) \\
\leq\ & \frac{3}{M}.
\end{aligned}$$

Mit analogen Argumenten wie unter Abschnitt 2.2.3 gilt diese Schranke ebenso für den Grenzprozess und man erhält insgesamt:

$$\begin{aligned}
& |\mathbb{E}^*_{m'_n} f(W^*_{m'_n}) - \mathbb{E} f(\bar{G})| \\
\leq\ & |\mathbb{E}^*_{m'_n}(f(W^*_{m'_n})) - \mathbb{E}^*_{m'_n}(f(\widetilde{W}^*_{m'_n}))| + |\mathbb{E}^*_{m'_n}(f(\widetilde{W}^*_{m'_n})) - \mathbb{E}(f(\widetilde{\bar{G}}))| \\
& + |\mathbb{E}(f(\widetilde{\bar{G}})) - \mathbb{E}(f(\bar{G}))| \\
\leq\ & \frac{7}{M} \leq \nu
\end{aligned}$$

für hinreichend großes n und demnach

$$\mathbb{P}\big(\forall f \in \mathcal{C}_b,\, \forall \nu > 0 : \exists n_0 = n_0(f,\nu) :\, \forall n \geq n_0 : |\mathbb{E}^*_{m'_n} f(W^*_{m'_n}) - \mathbb{E} f(\bar{G})| \leq \nu\big) = 1.$$

Damit ist die Verteilungskonvergenz von W^*_n gegen $\bar{G}$ in Wahrscheinlichkeit gezeigt.

□

2.4 Anwendung auf Teststatistiken und Bootstrap-Konsistenz

Die oben bewiesene Verteilungskonvergenz des Bootstrap-Prozesses gegen denselben Grenzprozess wie im Falle des Originalprozesses kann nun angewendet werden, um kritische Werte für Teststatistiken zu erhalten, die asymptotisch Tests zum gewünschten Signifikanzniveau liefern.
Um die Konsistenz des Bootstrap-Verfahrens nachzuweisen, wird nun gezeigt, dass die Verteilungen der interessierenden Teststatistik H, welche auf der Bootstrap-Stichprobe beruht, gegen die Grenzverteilung der Teststatistik, welche auf

der Originalstichprobe basiert, in folgendem Sinne konvergieren: Es gilt bei stetiger Grenzverteilung der Teststatistik unter $\mathcal{H}_0$

$$\mathbb{P}(\sup_{x\in\mathbb{R}} |\mathbb{P}_n^*(HW_n^* \leq x) - \mathbb{P}(HW_n \leq x)| > \mu) \overset{n\to\infty}{\to} 0, \forall \mu < 0.$$

Dies entspricht der schwachen Konsistenz des Bootstrap-Verfahrens in der Supremums-Metrik, im Sinne von Shao und Tu [68], S. 73. Aufgrund der Konsistenz bezüglich der Supremums-Metrik führt also die Verwendung der entsprechenden Quantile der Bootstrap-Statistik für die Originalstatistik zu einem asymptotischen Niveau-α-Test:

$$\mathbb{P}(HW_n > t_{1-\alpha}^{(n)}) \overset{n\to\infty}{\to} \alpha,$$

wobei $t_{1-\alpha}^{(n)}$ das $(1-\alpha)$-Quantil der Bootstrap-Statistik bezeichnet. Die Quantile der Bootstrap-Teststatistik können in der Praxis durch wiederholtes Simulieren approximiert werden.

Die Übertragung der Verteilungskonvergenz des empirischen Prozesses auf die der Teststatistik liefert das Continuous Mapping Theorem, siehe Pollard [61], Kapitel IV.2.

Theorem 2.4.1 *Gegeben seien metrische Räume $\mathcal{X}$ und $\mathcal{X}'$ sowie darauf σ-Algebren $\mathcal{A}$ bzw. $\mathcal{A}'$. Sei $H : \mathcal{X} \to \mathcal{X}'$ eine $\mathcal{A}/\mathcal{A}'$-messbare Abbildung. Gilt, dass H in jedem Punkt einer separablen, $\mathcal{A}$-messbaren Menge C von vollständig regulären Punkten stetig ist, dann folgt aus $X_n \overset{D}{\to} X$ zusammen mit $\mathbb{P}(X \in C)=1$, dass $HX_n \overset{\mathcal{D}}{\to} HX$.*

Die Pfade von G sind gemäß Theorem 2.2.1 auf einer Menge konzentriert, die die in Theorem 2.4.1 geforderten Eigenschaften erfüllt, denn alle Punkte in $D(\bar{\mathbb{R}})$ und somit insbesondere diejenigen in $C(\bar{\mathbb{R}})$ sind vollständig regulär und es gilt ebenfalls, dass $C(\bar{\mathbb{R}})$ eine vollständige, $\mathcal{P}$-messbare Menge in $(D, \mathbb{P})$ ist.
Bezeichnet $H : D(\bar{\mathbb{R}}) \to \mathbb{R}$ also ein auf $C(\bar{\mathbb{R}})$ stetiges Funktional, so überträgt sich anhand des Continuous Mapping Theorems die Verteilungskonvergenz von W_n gegen G auf die Verteilungskonvergenz von HW_n gegen HG und für den Bootstrap-Prozess gilt die entsprechende Aussage in Wahrscheinlichkeit:

Theorem 2.4.2 *Es gelten $\mathcal{A}$ und $\mathcal{A}^*$. Sei $H : D(\bar{\mathbb{R}}) \to \mathbb{R}$ ein auf $C(\bar{\mathbb{R}})$ stetiges, $\mathbb{P} - \mathcal{B}(\mathbb{R})$-messbares Funktional. Dann gilt*
(i) für jede Teilfolge $(m_n)_{n\in\mathbb{N}}$ von $\mathbb{N}$ existiert eine Teilfolge $(m'_n)_{n\in\mathbb{N}}$, sodass

$$HW_n^* \xrightarrow{\mathcal{D}} H\bar{G} \text{ fast sicher.}$$

Gilt darüber hinaus, dass die Verteilung von HG stetig ist, so folgt unter $\mathcal{H}_0$
(ii)

$$\mathbb{P}\big(\sup_{x\in\mathbb{R}} |\mathbb{P}_n^*(HW_n^* \leq x) - \mathbb{P}(HW_n \leq x)| > \mu\big) \overset{n\to\infty}{\to} 0, \forall \mu > 0.$$

Beweis:
Aussage (i) folgt unmittelbar aus Theorem 2.3.1, dem Continuous Mapping Theorem und den oben ausgeführten Überlegungen. Was (ii) anbelangt, so gilt mit 2.3.1 und dem Continuous Mapping Theorem zunächst, dass für jede Teilfolge $(m_n)_{n\in\mathbb{N}}$ von $\mathbb{N}$ eine Teilfolge $(m'_n)_{n\in\mathbb{N}}$ existiert, sodass

$$\sup_{x\in\mathbb{R}} |\mathbb{P}_{m'_n}^*(HW_{m'_n}^* \leq x) - \mathbb{P}(HG \leq x)| \overset{n\to\infty}{\to} 0 \text{ fast sicher}$$

und damit auch

$$\sup_{x\in\mathbb{R}} |\mathbb{P}_{m'_n}^*(HW_{m'_n}^* \leq x) - \mathbb{P}(HW_{m'_n} \leq x)| \overset{n\to\infty}{\to} 0 \text{ fast sicher,}$$

woraus (ii) folgt. □

Unter der Voraussetzung der stetigen Grenzverteilung der Teststatistik ist also die schwache Konsistenz des Bootstrap-Verfahrens in der Supremums-Metrik gewährleistet.
Damit gilt, wenn mit $t_{1-\alpha}^{(n)}$ das $1-\alpha$-Quantil der Verteilung von HW_n^* bezeichnet wird

$$\begin{aligned}
& \big|\mathbb{P}(HW_n \leq t_{1-\alpha}^{(n)}) - (1-\alpha)\big| \\
\leq\ & \big|\mathbb{P}(HW_n \leq t_{1-\alpha}^{(n)}) - \mathbb{P}_n^*(HW_n^* \leq t_{1-\alpha}^{(n)})\big| \\
& + |\mathbb{P}_n^*(HW_n^* \leq t_{1-\alpha}^{(n)}) - (1-\alpha)| \\
\leq\ & \sup_{x\in\bar{\mathbb{R}}} \big|\mathbb{P}(HW_n \leq x) - \mathbb{P}_n^*(HW_n^* \leq x)\big| \\
& + |\mathbb{P}_n^*(HW_n^* \leq t_{1-\alpha}^{(n)}) - (1-\alpha)| \xrightarrow{\mathbb{P}} 0,
\end{aligned}$$

woraus folgt, dass im deterministischen Sinne

$$\mathbb{P}(HW_n \leq t_{1-\alpha}^{(n)}) \overset{n\to\infty}{\to} 1-\alpha.$$

Falls also die Teststatistik eine stetige Abbildung ist auf der Menge, auf der die Pfade des Grenzprozesses mit Wahrscheinlichkeit 1 konzentriert sind, und falls die Verteilung von HG stetig ist, liefert also die Verwendung des kritischen Wertes der Bootstrap-Statistik einen Test, der asymptotisch das gewünschte Konfidenzniveau erreicht.

2.5 Hilfsresultate

2.5.1 Abschätzung für Kovarianzen durch lipschitzstetige Glättungen

Um betragsmäßige Abschätzungen für auftretenden Kovarianzen zu erhalten, werden lipschitzstetige Glättungen der verwendeten Indikatorfunktionen eingeführt, wie in Doukhan und Louhichi [24] im Beweis von Theorem 6. Dies wird sowohl im Falle des Originalprozesses als auch für den Bootstrap-Prozess benötigt.

Es gelten folgende Bezeichnungen, wie auch zum Teil im Beweis der Konvergenz der endlichdimensionalen Verteilungen in Abschnitt 2.2.3 verwendet:

$$F(s,t) := F(t) - F(s)$$

und

$$\begin{aligned}
Y_i(t) &:= \mathbb{1}\{X_i \leq t\} && (2.61)\\
\bar{Y}_i(t) &:= \mathbb{1}\{X_i \leq t\} - F(t) && (2.62)\\
\widetilde{Y}_i(t) &:= \mathbb{1}\{X_i \leq t\} \\
&\quad + \mathbb{1}\{t < X_i \leq t+a\}\big(1 - \frac{X_i - t}{a}\big) && (2.63)\\
\widetilde{\bar{Y}}_i(s,t) &:= \mathbb{1}\{X_i \leq t\} - F(t) \\
&\quad + \mathbb{1}\{t < X_i \leq t+a\}\big(1 - \frac{X_i - t}{a}\big) && (2.64)\\
Y_i(s,t) &:= \mathbb{1}\{s < X_i \leq t\} && (2.65)\\
\bar{Y}_i(s,t) &:= \mathbb{1}\{s < X_i \leq t\} - F(s,t) && (2.66)
\end{aligned}$$

$$
\begin{aligned}
\widetilde{Y}_i(s,t) &:= \mathbb{1}\{s < X_i \le t\} \\
&\quad + \mathbb{1}\{t < X_i \le t+a\}\big(1 - \frac{X_i - t}{a}\big) \\
&\quad + \mathbb{1}\{s-a < X_i \le s\}\big(1 + \frac{X_i - s}{a}\big) && (2.67)\\
\widetilde{\widetilde{Y}}_i(s,t) &:= \mathbb{1}\{s < X_i \le t\} - F(s,t) \\
&\quad + \mathbb{1}\{t < X_i \le t+a\}\big(1 - \frac{X_i - t}{a}\big) \\
&\quad + \mathbb{1}\{s-a < X_i \le s\}\big(1 + \frac{X_i - s}{a}\big) && (2.68)\\
&=: \mathbb{1}_{1,i} + \mathbb{1}_{2,i} + \mathbb{1}_{3,i}. && (2.69)
\end{aligned}
$$

Die Funktionen $\widetilde{Y}_{(\cdot)}(t)$, $\widetilde{\widetilde{Y}}_{(\cdot)}(t)$, $\widetilde{Y}_{(\cdot)}(s,t)$ und $\widetilde{\widetilde{Y}}_{(\cdot)}(s,t)$ sind lipschitzstetig, und dies gleichmäßig in (s,t), denn

$$
\begin{aligned}
&\sup_{(s,t)\in\mathbb{R}} |\widetilde{\widetilde{Y}}_i(s,t) - \widetilde{Y}_0(s,t)| = \sup_{(s,t)\in\mathbb{R}} |\widetilde{Y}_i(s,t) - \widetilde{\widetilde{Y}}_0(s,t)| \\
= &\sup_{(s,t)\in\mathbb{R}} |\mathbb{1}\{s < X_i \le t\} - \mathbb{1}\{s < X_0 \le t\} \\
&\qquad + \mathbb{1}\{s-a < X_i \le s\}\cdot(1 + \frac{X_i - s}{a}) - \mathbb{1}\{s-a < X_0 \le s\}\cdot(1 + \frac{X_0 - s}{a})| \\
&\qquad + \mathbb{1}\{t < X_i \le t+a\}\cdot(1 - \frac{X_i - t}{a}) - \mathbb{1}\{t < X_0 \le t+a\}\cdot(1 - \frac{X_0 - t}{a}) \\
\le &\ \frac{1}{a}\cdot|X_0 - X_i|.
\end{aligned}
$$

Es gelten somit folgende Kovarianzungleichungen:

Lemma 2.5.1 *Unter Voraussetzung $\mathcal{A}$ sowie $\mathcal{A}^*$ gilt, für $i \ge 0$:*

$$
\begin{aligned}
&(i) && \big|\operatorname{cov}[Y_1(s,t), Y_{1+i}(s,t)]\big| \le (F(t) - F(s))^{\frac{1}{2}} \cdot O(\zeta_i^{\frac{1}{3}}) \\
&(ii) && \sup_{(s,t)\in\mathbb{R}^2} \big|\operatorname{cov}[Y_1(s,t), Y_{1+i}(s,t)]\big| = O(\zeta_i^{\frac{1}{2}})
\end{aligned}
$$

und

$$
\begin{aligned}
&(iii) && \sup_{(s,t)\in\mathbb{R}^2} \big|\operatorname{cov}[\bar{Y}_1(s,t), \bar{Y}_{1+v_1}(s,t)\cdot\bar{Y}_{1+v_2}(s,t)\cdot\bar{Y}_{1+v_3}(s,t)\big| = O(\zeta_{v_1}^{\frac{1}{2}}) \\
&(iv) && \sup_{(s,t)\in\mathbb{R}^2} \big|\operatorname{cov}[\bar{Y}_1(s,t)\cdot\bar{Y}_{1+v_1}(s,t), \bar{Y}_{1+v_2}(s,t)\cdot\bar{Y}_{1+v_3}(s,t)\big| = O(\zeta_{v_2-v_1}^{\frac{1}{2}}) \\
&(v) && \sup_{(s,t)\in\mathbb{R}^2} \big|\operatorname{cov}[\bar{Y}_1(s,t)\cdot\bar{Y}_{1+v_1}(s,t)\cdot\bar{Y}_{1+v_2}(s,t), \cdot\bar{Y}_{1+v_3}(s,t)\big| = O(\zeta_{v_3-v_2}^{\frac{1}{2}}),
\end{aligned}
$$

für $0 \le v_1 \le v_2 \le v_3$.

Beweis:

zu (i):

$\tilde{Y}_{s,t}(X_i)$ ist lipschitzstetig in X_i mit Lipschitzkonstante $1/a$, gleichmäßig für alle $(s,t) \in \mathbb{R}^2$.

Damit gilt aufgrund der Eigenschaft der schwachen Abhängigkeit der X_i:

$$\begin{aligned}
& |\operatorname{cov}\big(Y_1(s,t), Y_{1+i}(s,t)\big)| \\
\leq\ & |\operatorname{cov}\big(Y_1(s,t), \widetilde{Y}_{1+i}(s,t)\big)| \\
& + |\operatorname{cov}\big(\bar{Y}_1(s,t), Y_{1+i}(s,t) - \widetilde{Y}_{1+i}(s,t)\big)| \\
\leq\ & \frac{1}{a} \cdot \big(\mathbb{E}\big[\big(Y_1(s,t)\big)^2\big]\big)^{1/2} \cdot K \cdot \zeta_i \\
& +\ \big(\mathbb{E}\big[(\bar{Y}_{s,t}(X_1))^2\big]\big)^{1/2} \cdot \big(\mathbb{E}\big[(\mathbb{1}\{s-a < X_{1+i} \leq s\}(\frac{X_{1+i}-s}{a}+1) \\
& \qquad + \mathbb{1}\{t < X_{1+i} \leq t+a\}(1 - \frac{X_{1+i}-t}{a}))^2\big]\big)^{1/2} \\
\leq\ & (F(s,t))^{1/2} \cdot (\frac{1}{a} \cdot K \cdot \zeta_i + 2 \cdot \sqrt{C \cdot a}) \\
=\ & (F(t) - F(s))^{\frac{1}{2}} \cdot O(\zeta_i^{\frac{1}{3}}),
\end{aligned}$$

falls $a \asymp \zeta_i^{\frac{2}{3}}$.

Damit ist die Summe über die Kovarianzen unter Voraussetzung $\mathcal{A}$ insbesondere absolut konvergent.

zu (ii):

Analog zu den Überlegungen unter (i) gilt

$$\begin{aligned}
& |\operatorname{cov}\big(Y_1(s,t), Y_{1+i}(s,t)\big)| \\
\leq\ & \frac{1}{a} \cdot \sqrt{\mathbb{E}\big[\big(Y_1(s,t)\big)^2\big]} \cdot K \cdot \zeta_i \\
& +\ \mathbb{E}\big[\mathbb{1}\{s-a < X_{1+i} \leq s\}(\frac{X_{1+i}-s}{a}+1) \\
& \qquad + \mathbb{1}\{t < X_{1+i} \leq t+a\}(1 - \frac{X_{1+i}-t}{a})\big] \\
\leq\ & \frac{1}{a} \cdot K \cdot \zeta_i + 2 \cdot C \cdot a = O(\zeta_i^{\frac{1}{2}}),
\end{aligned}$$

wobei diese Abschätzungen offensichtlich gleichmäßig für $(s,t) \in \bar{\mathbb{R}}^2$ gelten.

zu (iii):
Es gilt

$$\begin{aligned}
& \left|\operatorname{cov}\left(\bar{Y}_1(s,t), \bar{Y}_{1+v_1}(s,t)\cdot\bar{Y}_{1+v_2}(s,t)\cdot\bar{Y}_{1+v_3}(s,t)\right)\right| \\
\leq\ & \left|\operatorname{cov}\left(\bar{Y}_1(s,t), \widetilde{\bar{Y}}_{1+v_1}(s,t)\cdot\widetilde{\bar{Y}}_{1+v_2}(s,t)\cdot\widetilde{\bar{Y}}_{1+v_3}(s,t)\right)\right| \\
& + \left|\operatorname{cov}\Big(\bar{Y}_1(s,t), \bar{Y}_{1+v_1}(s,t)\cdot\bar{Y}_{1+v_2}(s,t)\cdot\bar{Y}_{1+v_3}(s,t) \right. \\
& \qquad\qquad \left. - \widetilde{\bar{Y}}_{1+v_1}(s,t)\cdot\widetilde{\bar{Y}}_{1+v_2}(s,t)\cdot\widetilde{\bar{Y}}_{1+v_3}(s,t)\Big)\right| \\
=\ & \frac{3}{a}\cdot K\cdot\zeta_{v_1} + \left|\mathbb{E}\big(\bar{Y}_{1+v_1}(s,t)\cdot\bar{Y}_{1+v_2}(s,t)\cdot\bar{Y}_{1+v_3}(s,t) \right. \\
& \qquad\qquad \left. - \widetilde{\bar{Y}}_{1+v_1}(s,t)\cdot\widetilde{\bar{Y}}_{1+v_2}(s,t)\cdot\widetilde{\bar{Y}}_{1+v_3}(s,t)\big)\right| \\
=\ & \frac{3}{a}\cdot K\cdot\zeta_{v_1} + \sum_{\substack{1\leq j_1,j_2,j_3\leq 3\\ nichtalle=1}} \mathbb{E}\big(\mathbb{1}_{j_1,1+v_1}\cdot\mathbb{1}_{j_2,1+v_2}\cdot\mathbb{1}_{j_3,1+v_3}\big) \\
=\ & \frac{3}{a}\cdot K\cdot\zeta_{v_1} + \text{const}\cdot C\cdot a = O(\zeta_{v_1}^{\frac{1}{2}}),
\end{aligned}$$

da die Lipschitzkonstante des Produktes aus den drei geglätteten Funktionen aufgrund deren Beschränktheit durch 1 gleich $\frac{3}{a}$ beträgt.
Die Abschätzung von $\mathbb{E}\big(\mathbb{1}_{j_1,1+v_1}\cdot\mathbb{1}_{j_2,1+v_2}\cdot\mathbb{1}_{j_3,1+v_3}\big)$ gilt, da einer der Faktoren von $\mathbb{1}_{j_1,1+v_1}\cdot\mathbb{1}_{j_2,1+v_2}\cdot\mathbb{1}_{j_3,1+v_3}$ aufgrund der Lipschitzstetigkeit von F durch $C\cdot a$ abgeschätzt werden kann und die anderen beiden durch 1 beschränkt sind (denn es gilt, dass j_1, j_2 und j_3 nicht alle gleich 1 sind).

zu (iv):

$$\begin{aligned}
& \left|\operatorname{cov}\left(\bar{Y}_1(s,t)\cdot\bar{Y}_{1+v_1}(s,t), \bar{Y}_{1+v_2}(s,t)\cdot\bar{Y}_{1+v_3}(s,t)\right)\right| \\
\leq\ & \left|\operatorname{cov}\left(\bar{Y}_1(s,t)\cdot\bar{Y}_{1+v_1}(s,t), \widetilde{\bar{Y}}_{1+v_2}(s,t)\cdot\widetilde{\bar{Y}}_{1+v_3}(s,t)\right)\right| \\
& + \left|\operatorname{cov}\Big(\bar{Y}_1(s,t)\cdot\bar{Y}_{1+v_1}(s,t), \bar{Y}_{1+v_2}(s,t)\cdot\bar{Y}_{1+v_3}(s,t) \right. \\
& \qquad\qquad \left. - \widetilde{\bar{Y}}_{1+v_2}(s,t)\cdot\widetilde{\bar{Y}}_{1+v_3}(s,t)\Big)\right| \\
\leq\ & \frac{2}{a}\cdot K\cdot\zeta_{v_2-v_1} + \text{const}\cdot C\cdot a = O(\zeta_{v_2-v_1}^{\frac{1}{2}}).
\end{aligned}$$

zu (v):

$$\left|\operatorname{cov}\left(\bar{Y}_1(s,t)\cdot\bar{Y}_{1+v_1}(s,t)\cdot\bar{Y}_{1+v_2}(s,t), \bar{Y}_{1+v_3}(s,t)\right)\right|$$

$$\begin{aligned}
&\leq \left|\operatorname{cov}\Big(\bar{Y}_1(s,t)\cdot\bar{Y}_{1+v_1}(s,t)\cdot\bar{Y}_{1+v_2}(s,t),\widetilde{\bar{Y}}_{1+v_3}(s,t)\Big)\right| \\
&\quad + \left|\operatorname{cov}\Big(\bar{Y}_1(s,t)\cdot\bar{Y}_{1+v_1}(s,t)\cdot\bar{Y}_{1+v_2}(s,t),\bar{Y}_{1+v_3}(s,t)-\widetilde{\bar{Y}}_{1+v_3}(s,t)\Big)\right| \\
&\leq \frac{2}{a}\cdot K\cdot\zeta_{v_3-v_2}+\text{const}\cdot C\cdot a = O(\zeta_{v_3-v_2}^{\frac{1}{2}}).
\end{aligned}$$

Die Punkte (iv) und (v) werden hierbei in analoger Weise nachgewiesen wie (iii), wobei in allen Fällen die Abschätzungen gleichmäßig in $(s,t)\in\bar{\mathbb{R}}^2$ gelten. □

2.5.2 Hilfsresultate zur Konvergenz des Originalprozesses

Vierte Momente der Zuwächse von U_n

Zum Nachweis der stochastischen Äquikontinuität wird eine Ungleichung für die vierten Momente der Zuwächse des Prozesses U_n (siehe (2.3)) benötigt. Vergleiche hierzu Neumann und Paparoditis [55], Beweis von Theorem 3.1.

Lemma 2.5.2 *Es gelte $\mathcal{A}$. Dann folgt, dass für beliebiges $\psi\in(0,1)$ eine Konstante C_ψ existiert, so dass*

$$\mathbb{E}[|U_n(x)-U_n(y)|^4]\leq C_\psi\cdot\{\mathbb{P}(X_0\in(x,y])+\frac{1}{n}\}^{1+\psi}$$

Beweis:
Bezeichne $B:=(x,y]$ sowie $\mathbb{P}^{X_0}(B):=\mathbb{P}(X_0\in B)$.
Mit Lemma 3 in Doukhan und Louhichi [24] gilt zunächst:

$$\begin{aligned}
&\mathbb{E}[|U_n(x)-U_n(y)|^4] \\
\leq\ & 4!\cdot\Big\{\Big(\sum_{v=0}^{n-1}C_{v,2}(B)\Big)^2+\frac{1}{n}\sum_{v=0}^{n-1}(v+1)^2\cdot C_{v,4}(B)\Big\}.
\end{aligned}$$

Weiter gilt, für ein $\tilde{\rho}<1$

$$C_{v,q}\leq\tilde{C}\cdot\min\{(\mathbb{P}^{X_0}(B),\tilde{\rho}^v)\}=:\tilde{C}\cdot(\mathbb{P}^{X_0}(B)\wedge\tilde{\rho}^v)$$

und

$$\tilde{\rho}^v\leq\mathbb{P}^{X_0}(B)$$

für $v\geq v_0$ mit $v_0=\lceil\frac{\ln(\mathbb{P}^{X_0}(B))}{\ln\tilde{\rho}}\rceil$. Vereinfachend wird die Ganzzahligkeit von $\ln(\mathbb{P}^{X_0}(B))/\ln\tilde{\rho}$ angenommen.

Damit folgt

$$
\begin{aligned}
& \mathbb{E}[|U_n(x) - U_n(y)|^4] \\
\leq\ & 4! \cdot \tilde{C} \cdot \Big\{ \Big(\sum_{v=0}^{n-1} (\mathbb{P}^{X_0}(B) \wedge \tilde{\rho}^v) \Big)^2 + \frac{1}{n} \sum_{v=0}^{n-1} (v+1)^2 \cdot \tilde{C} \cdot (\mathbb{P}^{X_0}(B) \wedge \tilde{\rho}^v) \Big\}
\end{aligned}
$$

und weiter

$$
\begin{aligned}
& \Big(\sum_{v=0}^{n-1} (\mathbb{P}^{X_0}(B) \wedge \tilde{\rho}^v) \Big)^2 + \frac{1}{n} \sum_{v=0}^{n-1} (v+1)^2 \cdot (\mathbb{P}^{X_0}(B) \wedge \tilde{\rho}^v) \\
\leq\ & \Big(\sum_{v=0}^{v_0-1} \mathbb{P}^{X_0}(B) + \sum_{v=v_0}^{n-1} \tilde{\rho}^v \Big)^2 \\
& + \frac{1}{n} \cdot \Big(\sum_{v=0}^{v_0-1} (v+1)^2 \cdot \mathbb{P}^{X_0}(B) + \sum_{v=v_0}^{n-1} (v+1)^2 \cdot \tilde{\rho}^v \Big) \\
\leq\ & \Big(v_0 \cdot \mathbb{P}^{X_0}(B) + \tilde{\rho}^{v_0} \cdot \sum_{v=0}^{n-1-v_0} \tilde{\rho}^v \Big)^2 \\
& + \frac{1}{n} \cdot \Big(\mathbb{P}^{X_0}(B) \cdot \sum_{v=0}^{v_0-1} (v+1)^2 + \tilde{\rho}^{v_0} \cdot \sum_{v=0}^{n-1-v_0} (v+1+v_0)^2 \cdot \tilde{\rho}^v \Big) \\
\leq\ & \Big(\frac{\ln(\mathbb{P}^{X_0}(B))}{\ln \tilde{\rho}} \cdot \mathbb{P}^{X_0}(B) + \mathbb{P}^{X_0}(B) \cdot \sum_{v=0}^{n-1-v_0} \tilde{\rho}^v \Big)^2 \\
& + \frac{1}{n} \cdot \Big(2 \cdot \Big| \frac{\ln(\mathbb{P}^{X_0}(B))}{\ln \tilde{\rho}} \Big|^3 \cdot \mathbb{P}^{X_0}(B) + \mathbb{P}^{X_0}(B) \sum_{v=0}^{n-1-v_0} (v+1+v_0)^2 \cdot \tilde{\rho}^v \Big) \\
\leq\ & C_1 \cdot |\ln(\mathbb{P}^{X_0}(B))|^3 \cdot \big(\mathbb{P}^{X_0}(B)\big)^2 \\
& + \frac{1}{n} \cdot \Big(C_2 \cdot |\ln(\mathbb{P}^{X_0}(B))|^3 \cdot \big(\mathbb{P}^{X_0}(B)\big) \Big) \\
\leq\ & C_\psi \cdot \Big((\mathbb{P}^{X_0}(B))^\psi \cdot \mathbb{P}^{X_0}(B) + \frac{1}{n} \Big) \ (\text{mit } |\ln(\mathbb{P}^{X_0}(B))|^3 \cdot \big(\mathbb{P}^{X_0}(B)\big) \leq C_\psi \cdot (\mathbb{P}^{X_0}(B))^\psi) \\
\leq\ & C_\psi \cdot \Big(\mathbb{P}^{X_0}(B) + \frac{1}{n} \Big)^{1+\psi}.
\end{aligned}
$$

Bemerkung: Zum Nachweis von Lemma 2.5.2 wird das exponentielle Abfallen des Abhängigkeitskoeffizienten benötigt. Zum Verzicht auf diese Annahme müsste eine alternative Abschätzung erfolgen. □

$\mathbb{P}$-Beschränktheit von $\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\ell(X_i,\theta)$

Das folgende Lemma beinhaltet die $\mathbb{P}$-Beschränktheit von $\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\ell(X_i,\theta)$. Diese wird benötigt, um die Gültigkeit der asymptotisch linearisierten Darstellung des Prozesses W_n (siehe (2.3)) zu zeigen. Auch für den Nachweis der stochastischen Äquikontinuität wurde diese als Argument eingesetzt.

Lemma 2.5.3 *Es gelte $\mathcal{A}$. Dann gilt, dass für alle $\mu > 0$ ein $\nu > 0$ existiert, so dass für $n \geq n_0(\mu)$*

$$\mathbb{P}\Big(||\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\ell(X_i,\theta)||_2^2 > \nu\Big) < \mu,$$

d.h. $\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\ell(X_i,\theta)$ ist $\mathbb{P}$-beschränkt.

Beweis:
Es gilt:

$$\begin{aligned}
&\mathbb{P}\Big(||\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\ell(X_i,\theta)||_2^2 > \nu\Big)\\
\leq\ & \frac{1}{\nu}\cdot\mathbb{E}\Big(||\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\ell(X_i,\theta)||_2^2\Big)\\
=\ & \frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{n}\mathbb{E}\big(\ell(X_i,\theta)\cdot\ell^T(X_j,\theta)\big)\\
\leq\ & \frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{t=1}^{p}\operatorname{cov}\big(\ell_{(t)}(X_i,\theta)\cdot\ell_{(t)}^T(X_j,\theta)\big)\\
=\ & \sum_{t=1}^{p}\big(\operatorname{var}(\ell_{(t)}(X_0,\theta)) + 2\sum_{i=1}^{n-1}\frac{n-i}{n}\cdot|\operatorname{cov}(\ell_{(t)}(X_0,\theta),\ell_{(t)}(X_i,\theta))|\big)
\end{aligned}$$

Da die gestutzte Version ℓ_c von ℓ lipschitzstetig ist, folgt auch die Lipschitzstetigkeit der einzelnen Komponenten von ℓ_c. Somit folgt die Konvergenz der Summe aufgrund der schwachen Abhängigkeit der Beobachtungen mit analogen Argumenten wie in Abschnitt 2.2.3. $\square$

2.5.3 Hilfsresultate zur Konvergenz des Bootstrap-Prozesses

Konvergenz in Wahrscheinlichkeit

Um die Verteilungskonvergenz in Wahrscheinlichkeit, wie in Abschnitt 2.3.1 nachzuweisen, wird die folgende bekannte Charakterisierung der stochastischen Konvergenz durch Teilfolgen verwendet, vergleiche hierzu etwa Dudley [22]:

Lemma 2.5.4 *Sei $(X_i)_{i\in\mathbb{N}}$ eine Folge von Zufallsvariablen, definiert auf einem Wahrscheinlichkeitsraum $(\Omega, \mathcal{A}, \mathbb{P})$ mit Werten in einem metrischen Raum (Y, d). Dann gilt, dass $X_i \xrightarrow{\mathbb{P}} X$ für eine Zufallsvariable X, genau dann wenn jede Teilfolge von $(X_i)_{i\in\mathbb{N}}$ eine Teilfolge enthält, welche fast sicher gegen X konvergiert.*

Bedingte Varianz von U_n^*

Gemäß Naik-Nimbalkar und Rajarshi [53], S. 985f gilt folgendes Lemma:

Lemma 2.5.5 *Mit den Bezeichnungen von Lemma 2.3.13 gilt:*

$$\hat{\sigma}_n^2(s,t) - \mathbb{E}[\hat{\sigma}_n^2(s,t)] = \frac{l}{n-l+1}\Big[\sum_{j=0}^{n-l} V_j(s,t) + V(s,t)\Big].$$

Beweis:
Gemäß Naik-Nimbalkar und Rajarshi [53], S. 985f., gelten folgende Überlegungen:

$$\begin{aligned}
&\hat{\sigma}_n^2(s,t) - \mathbb{E}[\hat{\sigma}_n^2(s,t)] \\
= \;& \frac{l}{n-l+1}\sum_{j=0}^{n-l}\Big(\sum_{r=1}^{l}\frac{Y_{j+r}(s,t) - (\tilde{F}_n(t) - \tilde{F}_n(s))}{l}\Big)^2 \\
&- \frac{l}{n-l+1}\sum_{j=0}^{n-l}\mathbb{E}\Big[\Big(\frac{Y_{j+r}(s,t) - (\tilde{F}_n(t) - \tilde{F}_n(s))}{l}\Big)^2\Big] \\
= \;& \frac{l}{n-l+1}\sum_{j=0}^{n-l}\Big\{\big(\frac{1}{l}\sum_{r=1}^{l}Y_{j+r}(s,t)\big)^2 \\
&\qquad - \frac{2}{l}\cdot\big(\sum_{r=1}^{l}Y_{j+r}(s,t)\big)\cdot\tilde{F}_n(s,t) \\
&\qquad + (\tilde{F}_n(s,t))^2\Big\}
\end{aligned}$$

$$
\begin{aligned}
& - \frac{l}{n-l+1} \sum_{j=0}^{n-l} \Big\{ \mathbb{E}\big[\big(\frac{1}{l} \sum_{r=1}^{l} Y_{j+r}(s,t)\big)^2\big] \\
& \qquad\qquad - 2 \cdot \mathbb{E}\big[\frac{1}{l} \cdot \big(\sum_{r=1}^{l} Y_{j+r}(s,t)\big) \cdot \tilde{F}_n(s,t)\big] \\
& \qquad\qquad + \mathbb{E}\big[(\tilde{F}_n(s,t))^2\big] \Big\} \\
= \; & \frac{l}{n-l+1} \sum_{j=0}^{n-l} \big(\frac{1}{l} \sum_{r=1}^{l} Y_{j+r}(s,t)\big)^2 \\
& - \frac{2l}{n-l+1} \sum_{j=0}^{n-l} \frac{1}{l} \cdot \big(\sum_{r=1}^{l} Y_{j+r}(s,t)\big) \cdot \tilde{F}_n(s,t) \\
& + \frac{l}{n-l+1} \sum_{j=0}^{n-l} \big(\tilde{F}_n(s,t)\big)^2 \\
& - \frac{l}{n-l+1} \sum_{j=0}^{n-l} \mathbb{E}\big[\big(\frac{1}{l} \sum_{r=1}^{l} Y_{j+r}(s,t)\big)^2\big] \\
& - \frac{2l}{n-l+1} \sum_{j=0}^{n-l} \mathbb{E}\big[\frac{1}{l} \cdot \big(\sum_{r=1}^{l} Y_{j+r}(s,t)\big) \cdot \tilde{F}_n(s,t)\big] \\
& + \frac{l}{n-l+1} \sum_{j=0}^{n-l} \mathbb{E}\big[\big(\tilde{F}_n(s,t)\big)^2\big] \\
= \; & \frac{l}{n-l+1} \sum_{j=0}^{n-l} \Big\{ \big(\frac{1}{l} \sum_{r=1}^{l} Y_{j+r}(s,t)\big)^2 - 2l \cdot \big(\tilde{F}_n(s,t)\big)^2 + l \cdot \big(\tilde{F}_n(s,t)\big)^2 \Big\} \\
& - \frac{l}{n-l+1} \sum_{j=0}^{n-l} \Big\{ \mathbb{E}\big[\big(\frac{1}{l} \sum_{r=1}^{l} Y_{j+r}(s,t)\big)^2\big] - 2l \cdot \mathbb{E}\big[\big(\tilde{F}_n(s,t)\big)^2\big] \\
& \qquad\qquad + l \cdot \mathbb{E}\big[\big(\tilde{F}_n(s,t)\big)^2\big] \Big\} \\
= \; & \frac{l}{n-l+1} \sum_{j=0}^{n-l} \Big\{ \big(\frac{1}{l} \sum_{r=1}^{l} Y_{j+r}(s,t)\big)^2 - l \cdot \big(\tilde{F}_n(s,t)\big)^2 \\
& \qquad\qquad - 2 \cdot \big(\frac{1}{l} \sum_{r=1}^{l} Y_{j+r}(s,t)\big) \cdot F(s,t) \Big\} \\
& - \frac{l}{n-l+1} \sum_{j=0}^{n-l} \Big\{ \mathbb{E}\big[\big(\frac{1}{l} \sum_{r=1}^{l} Y_{j+r}(s,t)\big)^2\big] - l \cdot \mathbb{E}\big[\big(\tilde{F}_n(s,t)\big)^2\big] \\
& \qquad\qquad + 2l \cdot \tilde{F}_n(s,t) \cdot F(s,t) \Big\}
\end{aligned}
$$

$$
\begin{aligned}
&= \frac{l}{n-l+1}\sum_{j=0}^{n-l}\Big\{\big(\frac{1}{l}\sum_{r=1}^{l}Y_{j+r}(s,t)\big)^2 - 2\cdot\big(\frac{1}{l}\sum_{r=1}^{l}Y_{j+r}(s,t)\big)\cdot F(s,t) \\
&\qquad\qquad + \big(F(s,t)\big)^2\Big\} \\
&\quad - \frac{l}{n-l+1}\sum_{j=0}^{n-l}\Big\{\mathbb{E}\big[\big(\frac{1}{l}\sum_{r=1}^{l}Y_{j+r}(s,t)\big)^2\big] - 2\cdot\mathbb{E}\big[\big(\frac{1}{l}\sum_{r=1}^{l}Y_{j+r}(s,t)\big)\cdot F(s,t)\big] \\
&\qquad\qquad + \big(F(s,t)\big)^2\Big\} \\
&\quad - l\cdot\big(\tilde{F}_n(s,t)\big)^2 + 2l\cdot\tilde{F}_n(s,t)\cdot F(s,t) - l\cdot\big(F(s,t)\big)^2 \\
&\quad - l\cdot\mathbb{E}\big[\big(\tilde{F}_n(s,t)\big)^2\big] - 2l\cdot\mathbb{E}\big[\tilde{F}_n(s,t)\big]\cdot F(s,t) + l\cdot\big(F(s,t)\big)^2 \\
&= \frac{l}{n-l+1}\Big[\sum_{j=0}^{n-l}V_j(s,t) + V(s,t)\Big].
\end{aligned}
$$

□

Lemma 2.5.6 *Es gilt mit den Bezeichnungen im Beweis von Lemma 2.3.13*

$$(i) \qquad |\operatorname{var}(V_0)| = O(l^{-2}).$$

sowie

$$(ii) \qquad |\operatorname{cov}(V_0, V_i)| \le \text{const}\cdot\zeta_{i-l+1}^{\frac{1}{2}}, \ \textit{für } i \ge l.$$

Beweis:

zu (i):

$$\operatorname{var}(V_0) = \mathbb{E}\Big[\big(\frac{1}{l}\sum_{r=1}^{l}\bar{Y}_r(s,t)\big)^4\Big] - \big(\mathbb{E}\big[\big(\frac{1}{l}\sum_{r=1}^{l}\bar{Y}_r(s,t)\big)^2\big]\big)^2.$$

Weiter gilt mit Lemma 3 in Doukhan und Louhichi [24]

$$\mathbb{E}\Big[\big(\frac{1}{l}\sum_{r=1}^{l}\bar{Y}_r(s,t)\big)^4\Big] \le 4!\Big\{\frac{1}{l^2}\Big(\sum_{v=0}^{l-1}C_{v,2}\Big)^2 + \frac{1}{l^3}\sum_{v=0}^{l-1}(v+1)^2\cdot C_{v,4}\Big\}.$$

Wie in Doukhan und Louhichi [24] definiert, gilt hierbei

$$C_{v,q} = \max_{1\le m\le q-1}\ \sup_{(t_1,\dots,t_q)}\Big|\operatorname{cov}\big(\bar{Y}_{t_1}(s,t)\cdot\ldots\cdot\bar{Y}_{t_{m+1}}(s,t), \bar{Y}_{t_m}(s,t)\cdot\ldots\cdot\bar{Y}_{t_q}(s,t)\big)\Big|,$$

wobei $t_{m+1} - t_m = v$.

Hierbei gilt aufgrund der strikten Stationarität von $(X_t)_{t\in\mathbb{Z}}$:

$$\begin{aligned} C_{v,2} &= \operatorname{cov}(\ \bar{Y}_0(s,t), \bar{Y}_v(s,t)) \\ &= \operatorname{cov}(Y_0(s,t), Y_v(s,t)) \\ &\leq \text{const} \cdot \zeta_v \end{aligned} \tag{2.70}$$

sowie

$$\begin{aligned} C_{v,4} = \max\Big\{ & \operatorname{cov}(\bar{Y}_0(s,t), \bar{Y}_{v_1}(s,t) \cdot \bar{Y}_{v_2}(s,t) \cdot \bar{Y}_{v_3}(s,t)), \\ & \operatorname{cov}(\bar{Y}_0(s,t) \cdot \bar{Y}_{v_1}(s,t), \bar{Y}_{v_2}(s,t) \cdot \bar{Y}_{v_3}(s,t)), \\ & \operatorname{cov}(\bar{Y}_0(s,t) \cdot \bar{Y}_{v_1}(s,t) \cdot \bar{Y}_{v_2}(s,t), \bar{Y}_{v_3}(s,t)) \Big\}, \end{aligned}$$

wobei in Zeile 1 gilt $v_1 = v$, in Zeile 2 $v_2 - v_1 = v$ und in Zeile 3 $v_3 - v_2 = v$.

Mit Lemma 2.5.1 (iii), (iv) und (v) gilt

$$C_{v,4} = O(\zeta_v^{\frac{1}{2}}).$$

Unter der in $\mathcal{A}$ gestellten Voraussetzung an die Summierbarkeit der Abhängigkeitskoeffizienten folgt Aussage (i).

<u>zu (ii):</u>
Abkürzend gelte

$$\bar{Y}_j := \bar{Y}_j(s,t) \text{ sowie } \widetilde{\bar{Y}}_j := \bar{Y}_j(s,t).$$

Somit gilt:

$$\begin{aligned} & |\operatorname{cov}(V_0, V_i)| \\ = \ & |\operatorname{cov}\Big(\frac{1}{l^2}\big(\sum_{r=1}^{l} \bar{Y}_{0+r}\big)^2, \frac{1}{l^2}\big(\sum_{r=1}^{l} \bar{Y}_{i+r}\big)^2\Big) \\ \leq \ & \frac{1}{l^4} \sum_{r_1=1}^{l} \sum_{r_2=1}^{l} \sum_{r_3=1}^{l} \sum_{r_4=1}^{l} |\operatorname{cov}(\bar{Y}_{r_1} \cdot \bar{Y}_{r_2}, \bar{Y}_{i+r_3} \cdot \bar{Y}_{i+r_4})| \\ \leq \ & \frac{1}{l^4} \sum_{r_1=1}^{l} \sum_{r_2=1}^{l} \sum_{r_3=1}^{l} \sum_{r_4=1}^{l} \text{const} \cdot \zeta_{i-l+1}^{\frac{1}{2}} \\ = \ & \tilde{\zeta}_{i-l+1}^{\frac{1}{2}} \end{aligned}$$

für $\tilde{\zeta}_i := \text{const} \cdot \zeta_i$, wenn gilt $a \asymp \zeta_{i-l+1}^{\frac{1}{2}}$, unter Ausnutzung von Lemma 2.5.1 (iv).
$\square$

Lemma 2.5.7 *Es gilt*

$$\sup_{t \in \bar{\mathbb{R}}} \left| \tilde{F}_n(t) - F_n(t) \right| = O(\frac{l-1}{n-l+1}).$$

Beweis:
Es gilt

$$\begin{aligned}
& \left| \tilde{F}_n(t) - F_n(t) \right| \\
= & \left| \frac{1}{n-l+1} \sum_{j=0}^{n-l} \frac{1}{l} \sum_{r=1}^{l} \mathbb{1}\{X_{r+j} \leq t\} - \frac{1}{n} \sum_{i=1}^{n} \mathbb{1}\{X_i \leq t\} \right| \\
= & \left| \frac{1}{n-l+1} \sum_{i=1}^{n} w(i) \mathbb{1}\{X_i \leq t\} - \frac{1}{n} \sum_{i=1}^{n} \mathbb{1}\{X_i \leq t\} \right| \\
& (w(i) \text{ definiert wie in (2.33)}) \\
= & \sum_{i=1}^{n} \left| \frac{w(i)}{n-l+1} - \frac{1}{n} \right| \cdot \mathbb{1}\{X_i \leq t\} \\
\leq & \sum_{i=1}^{l} (\frac{1}{n} - \frac{1}{l \cdot (n-l+1)}) \\
& + \sum_{i=l+1}^{n-l} (\frac{1}{n-l+1} - \frac{1}{n}) \\
& + \sum_{i=n-l+1}^{n} (\frac{1}{n} - \frac{1}{l \cdot (n-l+1)}) \\
= & O(\frac{l-1}{n-l+1}), \qquad (2.71)
\end{aligned}$$

denn es gilt wiederum:

$$l \cdot w(i) = i \cdot \mathbb{1}\{i \leq l\} + l \cdot \mathbb{1}\{l < i \leq n-l\} + (n-(i-1)) \cdot \mathbb{1}\{n-l < i \leq n\}$$

und damit

$$\underbrace{\frac{1}{l \cdot (n-l+1)} - \frac{1}{n}}_{<0} \leq \frac{w(i)}{n-l+1} - \frac{1}{n} \leq \underbrace{\frac{1}{n-l+1} - \frac{1}{n}}_{>0}.$$

sowie

$$\Big|\frac{1}{l\cdot(n-l+1)}-\frac{1}{n}\Big| > \Big|\frac{1}{n-l+1}-\frac{1}{n}\Big|.$$

Die Abschätzung (2.71) gilt gleichmäßig für alle $t \in \mathbb{R}$. □

Lemma 2.5.8 *Unter Voraussetzung $\mathcal{A}^*$ gilt*

$$\mathbb{E}\Big[\Big(\frac{1}{n}\sum_{i=1}^{n}\mathbb{1}\{s < X_i \le t\} - F(s,t)\Big)^4\Big] = O(n^{-2}).$$

Beweis:

Das vierte Moment der im Lemma genannten Summe kann wiederum folgendermaßen abgeschätzt werden (siehe Doukhan und Louhichi [24], Lemma 3), unter Verwendung der Bezeichnung (2.62):

$$\begin{aligned}
&\mathbb{E}\Big[\Big(\frac{1}{n}\sum_{i=1}^{n}\bar{Y}_i(s,t)\Big)^4\Big] \\
\le\ & 4!\Big(\frac{1}{n^2}\Big(\sum_{v=0}^{n-1}C_{v,2}\Big)^2 + \frac{1}{n^3}\sum_{v=0}^{n-1}C_{v,4}(v+1)^2\Big).
\end{aligned}$$

Hier gilt also aufgrund der Stationarität von $(X_t)_{t\in\mathbb{Z}}$:

$$\begin{aligned}
C_{v,4} \ =\ \max\Big\{&\Big|\operatorname{cov}\Big(\bar{Y}_1(s,t), \bar{Y}_{1+t_1}(s,t)\cdot\bar{Y}_{1+t_2}(s,t)\cdot\bar{Y}_{1+t_3}(s,t)\Big)\Big|, \\
&\Big|\operatorname{cov}\Big(\bar{Y}_1(s,t)\cdot\bar{Y}_{1+t_1}(s,t), \bar{Y}_{1+t_2}(s,t)\cdot\bar{Y}_{1+t_3}(s,t)\Big)\Big|, \\
&\Big|\operatorname{cov}\Big(\bar{Y}_1(s,t)\cdot\bar{Y}_{1+t_1}(s,t)\cdot\bar{Y}_{1+t_2}(s,t), \bar{Y}_{1+t_3}(s,t)\Big)\Big|\Big\},
\end{aligned} \tag{2.72}$$

wobei in Zeile 1 gilt $t_1 = v$, in Zeile 2 $t_2 - t_1 = v$ und in Zeile 3 $t_3 - t_2 = v$.
Mit Lemma 2.5.1 gilt $C_{v,4} = O(\zeta_v^{\frac{1}{2}})$ (Hier wird $s = -\infty$ gesetzt und somit gilt $Y_i(t) = Y_i(s,t)$).
Für $C_{v,2}$ gilt mit analogen Überlegungen ebenfalls

$$C_{v,2} = O(\zeta_v^{\frac{1}{2}}).$$

Insgesamt gilt also:

$$\begin{aligned}
& \mathbb{E}\Big[\big(\frac{1}{n}\sum_{i=1}^{n}\bar{Y}_i(s,t)\big)^4\Big] \\
\leq\ & 4!\cdot\Big(\frac{1}{n^2}\big(\sum_{v=0}^{n-1}\text{const}\cdot\zeta_v^{\frac{1}{2}}\big)^2+\frac{1}{n^3}\sum_{v=0}^{n-1}\text{const}\cdot\zeta_v^{\frac{1}{2}}\cdot(v+1)^2\Big) \\
=\ & O(n^{-2}).
\end{aligned}$$

□

Zum Beweis der gleichmäßigen Konvergenz von $\hat{\sigma}_n^2$ werden im nächsten Lemma wiederum die vierten Momente einer Summe abgeschätzt.

Lemma 2.5.9 *Unter Voraussetzung $\mathcal{A}^*$ gilt*

$$\begin{aligned}
& \mathbb{E}\Big[\Big(\frac{1}{n-l+1}\sum_{j=0}^{n-l}\frac{1}{l}\sum_{r=1}^{l}\mathbb{1}\{x_{i-1}<X_{j+r}\leq x_i\}-\mathbb{E}(\mathbb{1}\{x_{i-1}<X_{j+r}\leq x_i\})\Big)^4\Big] \\
=\ & O(\frac{l^2}{(n-l+1)^2}).
\end{aligned}$$

Beweis:

Bezeichne im folgenden abkürzend (siehe (2.65)):

$$\bar{Y}_{j+r}:=\bar{Y}_{j+r}(x_{i-1},x_i)=\mathbb{1}\{x_{i-1}<X_{j+r}\leq x_i\}-F(x_{i-1},x_i).$$

Das vierte Moment der im Lemma genannten Summe kann nach Doukhan und Louhichi [24], Lemma 3, wie im vorigen Lemma abgeschätzt werden durch:

$$\begin{aligned}
& \Big(\frac{1}{(n-l+1)^4}\sum_{j=0}^{n-l}\frac{1}{l}\sum_{r=1}^{l}\bar{Y}_{j+r}\Big)^4 \\
\leq\ & 4!\Big(\frac{1}{(n-l+1)^2}\big(\sum_{v=0}^{n-l-1}C_{v,2}\big)^2+\frac{1}{(n-l+1)^3}\sum_{v=0}^{n-l-1}C_{v,4}(v+1)^2\Big).
\end{aligned}$$

Für $C_{v,2}$ und $C_{v,4}$ gilt hier aufgrund der Stationarität des zugrundeliegenden Prozesses

$$
\begin{aligned}
C_{v,4} \;=\; \max\Big\{ & \Big|\operatorname{cov}\Big(\frac{1}{l}\sum_{t_1=1}^{l}\bar{Y}_{t_1}, \big(\frac{1}{l}\sum_{t_2=1}^{l}\bar{Y}_{v_1+t_2}\big)\big(\frac{1}{l}\sum_{t_3=1}^{l}\bar{Y}_{v_2+t_3}\big)\big(\frac{1}{l}\sum_{t_4=1}^{l}\bar{Y}_{v_3+t_4}\big)\Big)\Big|, \\
& \Big|\operatorname{cov}\Big(\big(\frac{1}{l}\sum_{t_1=1}^{l}\bar{Y}_{t_1}\big)\big(\frac{1}{l}\sum_{t_2=1}^{l}\bar{Y}_{v_1+t_2}\big), \big(\frac{1}{l}\sum_{t_3=1}^{l}\bar{Y}_{v_2+t_3}\big)\big(\frac{1}{l}\sum_{t_4=1}^{l}\bar{Y}_{v_3+t_4}\big)\Big)\Big|, \\
& \Big|\operatorname{cov}\Big(\big(\frac{1}{l}\sum_{t_1=1}^{l}\bar{Y}_{t_1}\big)\big(\frac{1}{l}\sum_{t_2=1}^{l}\bar{Y}_{v_1+t_2}\big)\big(\frac{1}{l}\sum_{t_3=1}^{l}\bar{Y}_{v_2+t_3}\big), \frac{1}{l}\sum_{t_4=1}^{l}\bar{Y}_{v_3+t_4}\Big)\Big|\Big\},
\end{aligned}
\tag{2.73}
$$

wobei $0 \leq v_1 \leq v_1 \leq v_2$ und in der ersten Zeile $v_1 = v$ gilt, in der zweiten $v_2 - v_1 = v$ sowie in der dritten $v_3 - v_2 = v$.
Ähnlich wie im Beweis von Lemma 2.5.8 wird auch hier bezeichnet:

$$
\widetilde{\bar{Y}}_j := \widetilde{\bar{Y}}_j(x_{i-1}, x_i)
$$

und weiter

$$
\begin{aligned}
\widetilde{\bar{Y}}_j \;=\;& \mathbb{1}\{x_{i-1} \leq X_j < x_i\} - F(x_{i-1}, x_i) \\
& + \mathbb{1}\{x_{i-1} - a \leq X_j < x_{i-1}\} \cdot \big(1 + \frac{X_j - x_{i-1}}{a}\big) \\
& + \mathbb{1}\{x_i \leq X_j < x_i + a\} \cdot \big(1 - \frac{X_j - x_i}{a}\big) \\
=:\;& \mathbb{1}_{1,j} + \mathbb{1}_{2,j} + \mathbb{1}_{3,j}.
\end{aligned}
$$

(siehe auch (2.69)).
$\widetilde{\bar{Y}}_j$ ist lipschitzstetig mit Lipschitzkontante $\frac{1}{a}$, was ebenso für $l^{-1}\sum_{r=1}^{l}\widetilde{\bar{Y}}_{j+r}$ gilt.

Während die entsprechenden Kovarianzen durch 1 abgeschätzt werden können, falls die Zeitlücke kleiner oder gleich l ist, gilt für $v \geq l$ bei der ersten Kovarianz unter (2.73) folgende Abschätzung.

$$
\begin{aligned}
& \operatorname{cov}\Big(\frac{1}{l}\sum_{t_1=1}^{l}\bar{Y}_{t_1}, \big(\frac{1}{l}\sum_{t_2=1}^{l}\widetilde{\bar{Y}}_{v_1+t_2}\big)\big(\frac{1}{l}\sum_{t_3=1}^{l}\widetilde{\bar{Y}}_{v_2+t_3}\big)\big(\frac{1}{l}\sum_{t_4=1}^{l}\widetilde{\bar{Y}}_{v_3+t_4}\big)\Big) \\
=\;& \frac{1}{l^4}\sum_{t_1}\sum_{t_2}\sum_{t_3}\sum_{t_4}\operatorname{cov}\Big(\bar{Y}_{t_1}, \widetilde{\bar{Y}}_{v_1+t_2}\widetilde{\bar{Y}}_{v_2+t_3}\widetilde{\bar{Y}}_{v_3+t_4}\Big) \\
\leq\;& \frac{3}{a}\cdot K \cdot \zeta_{v_1-l+1}.
\end{aligned}
$$

Nun bleibt noch abzuschätzen

$$\begin{aligned}
&\operatorname{cov}\Big(\frac{1}{l}\sum_{t_1=1}^{l}\bar{Y}_{t_1}, \big(\frac{1}{l}\sum_{t_2=1}^{l}\widetilde{\bar{Y}}_{v_1+t_2}\big)\big(\frac{1}{l}\sum_{t_3=1}^{l}\widetilde{\bar{Y}}_{v_2+t_3}\big)\big(\frac{1}{l}\sum_{t_4=1}^{l}\widetilde{\bar{Y}}_{v_3+t_4}\big) \\
&\qquad - \big(\frac{1}{l}\sum_{t_2=1}^{l}\bar{Y}_{v_1+t_2}\big)\big(\frac{1}{l}\sum_{t_3=1}^{l}\bar{Y}_{v_2+t_3}\big)\big(\frac{1}{l}\sum_{t_4=1}^{l}\bar{Y}_{v_3+t_4}\big)\Big) \\
= \;& \operatorname{cov}\big(\frac{1}{l}\sum_{t_1=1}^{l}\bar{Y}_{t_1}, \frac{1}{l^3}\sum_{t_2}\sum_{t_3}\sum_{t_4}\widetilde{\bar{Y}}_{v_1+t_2}\widetilde{\bar{Y}}_{v_2+t_3}\widetilde{\bar{Y}}_{v_3+t_4} - \bar{Y}_{v_1+t_2}\bar{Y}_{v_2+t_3}\bar{Y}_{v_3+t_4}\big) \\
= \;& \operatorname{cov}\big(\frac{1}{l}\sum_{t_1=1}^{l}\bar{Y}_{t_1}, \frac{1}{l^3}\sum_{t_2}\sum_{t_3}\sum_{t_4}\big(\sum_{\substack{1\le j_1,j_2,j_3\le 3\\ \textit{nicht alle}=1}}\mathbb{1}_{j_1,v_1+t_2}\cdot\mathbb{1}_{j_2,v_2+t_3}\cdot\mathbb{1}_{j_3,v_3+t_4}\big)\big) \\
= \;& \frac{1}{l^4}\sum_{t_1}\sum_{t_2}\sum_{t_3}\sum_{t_4}\sum_{\substack{1\le j_1,j_2,j_3\le 3\\ \textit{nicht alle}=1}}\operatorname{cov}\big(\bar{Y}_{t_1}, \mathbb{1}_{j_1,v_1+t_2}\cdot\mathbb{1}_{j_2,v_2+t_3}\cdot\mathbb{1}_{j_3,v_3+t_4}\big).
\end{aligned}$$

Wiederum gilt:

$$\begin{aligned}
&|\operatorname{cov}\big(\bar{Y}_{t_1}, \mathbb{1}_{j_1,v_1+t_2}\cdot\mathbb{1}_{j_2,v_2+t_3}\cdot\mathbb{1}_{j_3,v_3+t_4}\big)| \\
= \;& |\mathbb{E}(\bar{Y}_{t_1}\cdot\mathbb{1}_{j_1,v_1+t_2}\cdot\mathbb{1}_{j_2,v_2+t_3}\cdot\mathbb{1}_{j_3,v_3+t_4}) - \underbrace{\mathbb{E}(\bar{Y}_{t_1})}_{=0}\cdot\mathbb{E}(\mathbb{1}_{j_1,v_1+t_2}\cdot\mathbb{1}_{j_2,v_2+t_3}\cdot\mathbb{1}_{j_3,v_3+t_4})| \\
\le \;& \mathbb{E}\big(\mathbb{1}_{j_1,v_1+t_2}\cdot\mathbb{1}_{j_2,v_2+t_3}\cdot\mathbb{1}_{j_3,v_3+t_4}\big) \\
\le \;& C\cdot a.
\end{aligned}$$

Also gilt insgesamt

$$\begin{aligned}
&|\operatorname{cov}\Big(\frac{1}{l}\sum_{t_1=1}^{l}Y_{t_1}, \big(\frac{1}{l}\sum_{t_2=1}^{l}\widetilde{\bar{Y}}_{v_1+t_2}\big)\big(\frac{1}{l}\sum_{t_3=1}^{l}\widetilde{\bar{Y}}_{v_2+t_3}\big)\big(\frac{1}{l}\sum_{t_4=1}^{l}\widetilde{\bar{Y}}_{v_3+t_4}\big)\Big)| \\
= \;& O(\zeta_{v-l+1}^{\frac{1}{2}}).
\end{aligned}$$

Analog kann abgeschätzt werden:

$$\begin{aligned}
&|\operatorname{cov}\Big(\big(\frac{1}{l}\sum_{t_1=1}^{l}Y_{t_1}\big)\big(\frac{1}{l}\sum_{t_2=1}^{l}Y_{v_1+t_2}\big), \big(\frac{1}{l}\sum_{t_3=1}^{l}Y_{v_2+t_3}\big)\big(\frac{1}{l}\sum_{t_4=1}^{l}Y_{v_3+t_4}\big)\Big)| \\
= \;& O(\zeta_{v-l+1}^{\frac{1}{2}}),
\end{aligned}$$

falls $a \asymp \zeta_{v-l+1}^{\frac{1}{2}}$.

Die dritte Kovarianz bei $C_{v,4}$ kann abgeschätzt werden, da das Produkt aus den drei Summen durch 1 beschränkt ist und die Summe nach dem Komma wiederum durch eine lipschitzstetige Variante der $Y_{v_3+t_4}$ ersetzt werden kann.

Also ergibt sich in diesem Fall

$$\begin{aligned}
& \Big|\operatorname{cov}\Big(\big(\frac{1}{l}\sum_{t_1=1}^{l}\bar{Y}_{t_1}\big)\big(\frac{1}{l}\sum_{t_2=1}^{l}\bar{Y}_{v_1+t_2}\big)\big(\frac{1}{l}\sum_{t_3=1}^{l}\bar{Y}_{v_2+t_3}\big), \frac{1}{l}\sum_{t_4=1}^{l}\bar{Y}_{v_3+t_4}\Big)\Big| \\
\leq\ & \Big|\operatorname{cov}\Big(\big(\frac{1}{l}\sum_{t_1=1}^{l}\bar{Y}_{t_1}\big)\big(\frac{1}{l}\sum_{t_2=1}^{l}\bar{Y}_{v_1+t_2}\big)\big(\frac{1}{l}\sum_{t_3=1}^{l}\bar{Y}_{v_2+t_3}\big), \frac{1}{l}\sum_{t_4=1}^{l}\tilde{Y}_{v_3+t_4}\Big)\Big| \\
& + \Big|\operatorname{cov}\Big(\big(\frac{1}{l}\sum_{t_1=1}^{l}\bar{Y}_{t_1}\big)\big(\frac{1}{l}\sum_{t_2=1}^{l}\bar{Y}_{v_1+t_2}\big)\big(\frac{1}{l}\sum_{t_3=1}^{l}\bar{Y}_{v_2+t_3}\big), \frac{1}{l}\sum_{t_4=1}^{l}\bar{Y}_{v_3+t_4} - \frac{1}{l}\sum_{t_4=1}^{l}\widetilde{\bar{Y}}_{v_3+t_4}\Big)\Big| \\
\leq\ & \frac{1}{a}\cdot K \\
& + \mathbb{E}\Big(\Big|\big(\frac{1}{l}\sum_{t_1=1}^{l}\bar{Y}_{t_1}\big)\big(\frac{1}{l}\sum_{t_2=1}^{l}\bar{Y}_{v_1+t_2}\big)\big(\frac{1}{l}\sum_{t_3=1}^{l}\bar{Y}_{v_2+t_3}\big)\Big|\cdot\Big|\frac{1}{l}\sum_{t_4=1}^{l}\bar{Y}_{v_3+t_4} - \frac{1}{l}\sum_{t_4=1}^{l}\widetilde{\bar{Y}}_{v_3+t_4}\Big|\Big) \\
& + \mathbb{E}\Big(\Big|\big(\frac{1}{l}\sum_{t_1=1}^{l}\bar{Y}_{t_1}\big)\big(\frac{1}{l}\sum_{t_2=1}^{l}\bar{Y}_{v_1+t_2}\big)\big(\frac{1}{l}\sum_{t_3=1}^{l}\bar{Y}_{v_2+t_3}\big)\Big|\Big)\cdot\mathbb{E}\Big(\Big|\frac{1}{l}\sum_{t_4=1}^{l}\bar{Y}_{v_3+t_4} - \frac{1}{l}\sum_{t_4=1}^{l}\widetilde{\bar{Y}}_{v_3+t_4}\Big|\Big) \\
\leq\ & \big(\frac{1}{a}\cdot K\cdot\zeta_{v-l+1} + 2\cdot C\cdot a\big) = O(\zeta_{v-l+1}^{\frac{1}{2}}),
\end{aligned}$$

falls $a \asymp \zeta_{v-l+1}^{\frac{1}{2}}$.

Von der gleichen Ordnung ist $C_{v,2}$. Insgesamt gilt also, für $\tilde{\zeta}_{v-l+1} = \text{const}\cdot\zeta_{v-l+1}$:

$$\begin{aligned}
& \mathbb{E}\Big(\Big|\frac{1}{n-l+1}\sum_{j=0}^{n-l}\frac{1}{l}\sum_{r=1}^{l}\mathbb{1}\{x_{i-1} < X_{j+r} \leq x_i\} - \mathbb{E}(\mathbb{1}\{x_{i-1} < X_{j+r} \leq x_i\})\Big|^4\Big) \\
\leq\ & 4!\Big(\frac{1}{(n-l+1)^2}\big(\sum_{v=0}^{l-1}1 + \sum_{v=l}^{n-l-1}\tilde{\zeta}_{v-l+1}^{\frac{1}{2}}\big)^2 \\
& \qquad + \frac{1}{(n-l+1)^3}\big(\sum_{v=0}^{l-1}(v+1)^2 + \sum_{v=l}^{n-l-1}(v+1)^2\cdot\tilde{\zeta}_{v-l+1}^{\frac{1}{2}}\big)\Big) \\
\leq\ & 4!\Big(\frac{1}{(n-l+1)^2}\big(l+c_1\big)^2 + \frac{1}{(n-l+1)^3}\big(O(l^3)+c_2\big)\Big) \\
=\ & O(\frac{l^2}{(n-l+1)^2}),
\end{aligned}$$

falls $l/(n-l+1) = o(1)$. □

3 Ausblick

In dieser Arbeit wurden sowohl Fragestellungen der nichtparametrischen Kurvenschätzung behandelt als auch Bootstrap-Verfahren für den geschätzten empirischen Prozess betrachtet. Was den ersten Teil der Arbeit anbelangt, so haben sich die erzielten Konvergenzraten als asymptotisch optimal erwiesen unter den gestellten Voraussetzungen, wobei globale Forderungen an die Glattheit der Kurve gestellt wurden. Die erzielten Konvergenzraten hängen von der Glattheit der Kurve ab und gelten, wenn die verwendete Bandbreite von geeigneter Ordnung ist. So kann also in einem nächsten Schritt untersucht werden, wie in den jeweiligen Modellen die datengesteuerte Wahl der Bandbreite erfolgen kann, sodass die Konvergenzrate asymptotisch, siehe hierzu Hesse [40] zur datengesteuerten Wahl der Bandbreite in Dekonvolutionsmodellen bei der Schätzung der Kurve selbst.
Bei der Modalwertschätzung in Dekonvolutionsmodellen wurde hinsichtlich der Fehlerverteilung zum einen vorausgesetzt, dass diese bekannt ist, zum anderen wurde die gewöhnliche Glattheit (ordinary smoothness) angenommen. Während die erzielten Resultate sich vermutlich unmittelbar auf den Fall übertragen lassen, dass die Abnahme der charakteristischen Funktion der Fehlerdichte mit logarithmischer Rate erfolgt, so existieren bereits Arbeiten zur Kurvenschätzung in Dekonvolutionsmodellen, welche weiter von diesen Voraussetzungen abweichen. So wurde in Johannes [41] das Problem der Dichteschätzung im Dekonvolutionsmodell mit unbekannter Fehlerverteilung beleuchtet. Ebenso kann die Voraussetzung aufgeweicht werden, dass die charakteristische Funktion der Fehlerdichte von Null weg beschränkt ist. So wurde in Delaigle und Meister [20] ein Kurvenschätzer vorgeschlagen, dessen Konvergenzrate nicht durch Nullstellen in der charakteristischen Funktion der Fehlerverteilung beeinflusst wird, wie es etwa bei Verteilungen auftritt, die kompakten Träger besitzen. Auch das Problem der Modalwertschätzung kann unter diesen allgemeineren Bedingungen untersucht werden.

Was den geschätzten empirischen Prozess anbelangt, so wurde zunächst die Konvergenz des Originalprozesses unter $\mathcal{H}_0$ nachgewiesen. Es wurde ebenso das Verhalten des Prozesses unter einer fixierten Alternative betrachtet. In diesem Zusammenhang kann weiter untersucht werden, wie sich der Prozess unter geeigneten lokalen Alternativen verhält, die mit einer bestimmten Rate gegen die Nullhypothese konvergieren. In Durbin [23] wurden hierzu bereits Überlegungen für den geschätzten empirischen Prozess bei unabhängigen Daten angestellt, in Quessy und Mailhot [64] wurden verschiedene lokale Alternativen bei Tests auf normalverteilte Beobachtungen vorgeschlagen. Ebenfalls bietet sich an, das Verhalten des geschätzten empirischen Prozesses zu untersuchen, wenn die Beobachtungen schwach abhängig wie im betrachteten Modell sind, jedoch keine stetige Verteilungsfunktion besitzen. Unter Unabhängigkeit wurde in Henze [38] der auf Zähldaten beruhende geschätzte empirische Prozess betrachtet, in Szücs [74] wurden sowohl stetige als auch diskrete Verteilungen für die zugrundeliegenden Beobachtungen zugelassen, wobei hier jeweils parametrische Bootstrap-Verfahren angewendet wurden. Bei Vorliegen schwach abhängiger Beobachtungen bietet sich nun wiederum die Imitation durch ein Block-Bootstrap-Verfahren an, dessen Konsistenz es zu untersuchen gilt.

Symbolverzeichnis

Allgemeines

sup	Supremum
max	Maximum
inf	Infimum
min	Minimum
$\limsup$	Limes Superior
argmax	Stelle, an welcher Funktion Maximum annimmt, $y = \operatorname{argmax}_{x \in D} f(x) \Leftrightarrow f(y) \geq f(x) \, \forall x \in D$
supp (f)	Träger der Funktion f
$f * g$	Faltung von f und g
ϕ_X	charakteristische Funktionen der Verteilung von X
$H(f, g)$	Hellinger-Abstand der Wahrscheinlichkeitsdichten f und g
$\rho(f, g)$	Hellinger-Affinität der Wahrscheinlichkeitsdichten f und g
$\lambda^{[n]}$	n-dimensionales Lebesgue-Maß
$\lfloor a \rfloor$	Abrundungsfunktion, größte ganze Zahl kleiner oder gleich a
$\lceil a \rceil$	Aufrundungsfunktion, kleinste ganze Zahl größer oder gleich a
$\mathbb{1}\{A\}$, $\mathbb{1}_{\{A\}}$	Indikatorfunktion der Menge A
const	konstant
$a \wedge b$	$\min\{a, b\}$
$a \vee b$	$\max\{a, b\}$
$\mathbb{R}$	Menge der reellen Zahlen
$\mathbb{N}$	Menge der positiven natürlichen Zahlen

$\mathcal{B}(\mathbb{R}^d)$	Sigma-Algebra der Borelmengen auf $\mathbb{R}^d$
$\bar{\mathbb{R}}$	kompaktifizierte reelle Achse, $\bar{\mathbb{R}} = [-\infty, \infty]$
$\mathbb{E}(\cdot)$	Erwartungswert
$\mathbb{E}_n^*(\cdot)$	bedingte Erwartung, $\mathbb{E}_n^*(\cdot) = \mathbb{E}(\cdot\|(X_1, ..., X_n))$
$\mathbb{P}_n^*(\cdot)$	bedingte Wahrscheinlichkeit, $\mathbb{P}_n^*(\cdot) = \mathbb{P}(\cdot\|(X_1, ..., X_n))$
$\mathrm{var}(X)$	Varianz einer Zufallsvariablen X
$\mathrm{cov}(X, Y)$	Kovarianz zweier Zufallsvariablen X und Y
$\mathrm{var}_n^*(X)$	Varianz, bedingt auf die Originalstichprobe
$\mathrm{cov}_n^*(X, Y)$	Kovarianz, bedingt auf die Originalstichprobe
$X_n \to X$ $\mathbb{P}$-f.s.	$\mathbb{P}$-fast sichere Konvergenz
$X_n \xrightarrow{\mathbb{P}} X$	$\mathbb{P}$-stochastische Konvergenz
$X_n \xrightarrow{\mathcal{D}} X$	Konvergenz in Verteilung
$\|\| \cdot \|\|_\infty$	Supremums-Norm
$\|\| \cdot \|\|_2$	Eudklidische Norm
$\| \cdot \|$	Betrag
$O(a_n)$, $o(a_n)$	Landau-Notation, $a_n = O(b_n) \Leftrightarrow 0 \leq \limsup_{n\to\infty}(\|a_n/b_n\|) < \infty$ $a_n = o(b_n) \Leftrightarrow \lim_{n\to\infty}(\|a_n/b_n\|) = 0$
$O_{\mathbb{P}}(a_n)$, $o_{\mathbb{P}}(a_n)$	Landau-Notation in Wahrscheinlichkeit $X_n = O_{\mathbb{P}}(a_n) \Leftrightarrow \forall\varepsilon > 0\, \exists\, \delta\colon \mathbb{P}(\|X_n\|/\|a_n\| > \delta) < \varepsilon$ $X_n = o_{\mathbb{P}}(a_n) \Leftrightarrow \forall\varepsilon > 0\colon \mathbb{P}(\|X_n\|/\|a_n\| > \varepsilon) \overset{n\to\infty}{\to} 0.$
$\mathrm{Lip}(h)$	Lipschitzmodulus der Funktion $h\colon \mathbb{R}^d \to \mathbb{R}^k$, $\mathrm{Lip}(h) := \sup\left\{\frac{\|\|h(X_1, ..., X_d) - h(Y_1, ..., Y_d)\|\|_2}{\|\|(X_1, ..., X_d)^T - (Y_1, ..., Y_d)^T\|\|_2}\right\}.$

Nichtparametrische Kurvenschätzung: Daten und Funktionen

Y_i, (Y_i, Z_i)	Beobachtungen im jeweiligen Dekonvolutionsmodell
X_i	Realisierungen der nichtbeobachtbaren Zufallsvariablen X
ε_i	Fehlervariablen im Dekonvolutionsmodell, verteilt wie ε
$Y_{i,n}$	Beobachtungen im Fixed-Design-Modell

$x_{i,n}$	Designpunkte im Fixed-Design-Modell
$\varepsilon_{i,n}$	Fehlervariablen im Fixed-Design-Modell
f_X	Dichte von X
f_ε	Dichte von ε
$f_{X,Z}$	Dichte von (X, Z)
$f_{Y,Z}$	Dichte von (Y, Z)
$f_{Z\|X=x}$	bedingte Dichte von Z unter Bedingung $X = x$
m	Regressionsfunktion im jeweiligen Modell
K	Kernfunktion
θ	Modalwert der interessierenden Kurve
$[\tau_1, \tau_2]$, $[0, 1]$	kompakte Intervalle, innerhalb derer sich die interessierenden Modalwerte befinden
β	Ordnung der Schlechtgestelltheit des Dekonvolutionsproblems
p	Anzahl stetiger und beschränkter Ableitungen der interessierenden Kurve
$\hat{\theta}_n$	Modalwertschätzer im Dekonvolutionsmodell
$\hat{\theta}_n^c$, $\hat{\theta}_n^g$	Modalwertschätzer im Fixed-Design-Modell
b_n	Bandbreite
$K_{b_n}^*$	Dekonvolutionskern
$\hat{f}_{X,n}$	Schätzer für f_X
$\hat{m}_n$	Schätzer für m
$\widetilde{\mathcal{F}}_p$	Funktionenklasse für Dichteschätzung im Dekonvolutionsmodell
$\mathcal{F}_p$	Funktionenklasse für Errors-in-Variables-Regression
$\mathcal{F}_\rho$	Funktionenklasse im Fixed-Design-Regressionsmodell
$f^{(j)}$	j-te Ableitung von f

Bootstrap für empirische Prozesse: Daten und Prozesse

$(\Omega, \mathcal{A}, \mathbb{P})$	zugrundeliegender Wahrscheinlichkeitsraum
$(X_t)_{t\in\mathbb{Z}}$	den Beobachtungen zugrundeliegernder Prozess

$X_1, ..., X_n$	Originalstichprobe
$X_1^*, ..., X_n^*$	Bootstrap-Stichprobe
$\mathcal{F}$	parametrische Klasse von Verteilungsfunktionen
F_θ	Funktion aus $\mathcal{F}$
$\hat{\theta}_n$	Schätzer für θ
$D[\bar{\mathbb{R}}]$	Raum der Cadlag-Funktionen auf $\bar{\mathbb{R}}$
$\mathcal{P}$	Sigma-Algebra auf D
$C[\bar{\mathbb{R}}]$	Stetige Funktionen auf $\bar{\mathbb{R}}$ mit $C[\bar{\mathbb{R}}] \subseteq D[\bar{\mathbb{R}}]$
$\mathbf{P}_n$, $\mathbf{P}$	Wahrscheinlichkeitsmaße auf $(D[\bar{\mathbb{R}}], \mathcal{P})$
$\mathcal{C}^b$	Menge von Funktionen f mit $f\colon (D[\bar{\mathbb{R}}], \mathcal{P}, \|\| \cdot \|\|_\infty) \to (\mathbb{R}, \mathcal{B}(\mathbb{R}), \| \cdot \|)$, messbar, gleichmäßig stetig und beschränkt
l	Blocklänge
k	Blockanzahl
F_n	empirische Verteilungsfunktion, basierend auf Originalstichprobe
F_n^*	empirische Verteilungsfunktion, basierend auf Bootstrap-Stichprobe
$\tilde{F}_n$	$\mathbb{E}_n^*(F_n^*)$
$F(s,t)$	$F(t) - F(s)$
$\tilde{F}_n(s,t)$	$\tilde{F}_n(t) - \tilde{F}_n(s)$
W_n	geschätzter empirischer Prozess: $W_n = \sqrt{n}(F_n - F_{\hat{\theta}_n})$
U_n	gewöhnlicher empirischer Prozess basierend auf Originalstichprobe
Z_n	zentrierter Anteil von W_n
W_n^*	Bootstrap-Imitation von W_n
Z_n^*	zentrierter Anteil von W_n^*
U_n^*	gewöhnlicher empirischer Prozess basierend auf Bootstrap-Stichprobe
$G, \bar{G}$	Grenzprozesse auf Original- bzw. Bootstrap-Seite
$\tilde{F}_n(s,t)$	$\tilde{F}_n(t) - \tilde{F}_n(s)$
$\sigma^2(s,t)$	Limes der Varianz der Zuwächse von U_n, für $n \to \infty$
$\hat{\sigma}_n^2(s,t)$	Varianz der Zuwächse von U_n^*, bedingt auf $(X_1, ..., X_n)$

Literaturverzeichnis

[1] C. Abraham, G. Biau und B. Cadre. Simple estimation of the mode of a multivariate density. *Canad. J. Statist.*, 31, 23–34, 2003.

[2] C. Abraham, G. Biau und B. Cadre. On the asymptotic properties of a simple estimate of the mode. *ESAIM Probab. Statist.*, 8, 1–11, 2004.

[3] D. W. K. Andrews. Non-strong mixing autoregressive processes. *J. Appl. Prob.*, 21, 930–934, 1984.

[4] P. Ango Nze, P. Bühlmann und P. Doukhan. Weak dependence beyond mixing and asymptotics for nonparametric regression. *Ann. Statist.*, 30, 397–430, 2002.

[5] P. Ango Nze und P. Doukhan. Weak dependence: models and applications to econometrics. *Econometric Theory*, 20, 995–1045, 2004.

[6] G. J. Babu und C. R. Rao. Goodness-of-fit tests when parameters are estimated. *Sankyha*, 66, 63–74, 2004.

[7] G. J. Babu und K. Singh. Asymptotic represantiations related to jackknifing and bootstrapping L-statistics. *Sankyha Ser. A*, 46, 195–206, 1984.

[8] H. Bauer. Wahrscheinlichkeitstheorie. De Gruyter, 2002.

[9] P. Bickel und J. Fan. Some problems on estimating a unimodal regression function. *Statist. Sinica*, 6 , 23–45, 1996.

[10] P. J. Bickel und D. A. Freedman. Some asymptotic theory for the bootstrap. *Ann. Statist.*, 9, 1196–1217, 1981.

[11] P. Billingsley. Convergence of probability measures. Wiley, 1968.

[12] M. Sh. Braverman. Independent Random Variables and Rearrangement Invariant Spaces. *London Mathematical Society Lecture Note Series No. 194*, Cambridge University Press, Cambridge, 1994.

[13] P. Bühlmann. Blockwise bootstrapped empirical process for stationary sequences. *Ann. Statist.*, 22, 995–1012, 1994.

[14] P. Bühlmann. The blockwise bootstrap for general empirical processes of stationary sequences. *Stoch. Proc. Appl.*, 58, 247–265, 1995.

[15] E. Carlstein. The use of subseries methods for estimating the variance of a general statistic from a stationary time series. *Ann. Statist.*, 14, 1171-–1179, 1986.

[16] M. R. Chernick. A limit theorem for the maximum of autoregressive processes with uniform marginal distributions. *Ann. Prob.*, 9, 145–149, 1981.

[17] C.-K. Chu und J. S. Marron. Comparison of two bandwidth selectors with dependent errors. *Ann. Statist.*, 19, 1906–1918, 1991.

[18] D. A. Darling. The Cramér-Smirnov test in the parametric case. *Ann. Math. Statist.*, 26, 1–20, 1955.

[19] R. B. D'Agostino und M. A. Stephens. Goodness-of-fit techniques. Marcel Dekker, 1986.

[20] A. Delaigle und A. Meister. Nonparametric function estimation under Fourier-oscillating noise. *Statistica Sinica*, 21, 1065–1092, 2011.

[21] D. Donoho und L. Liu. Geometrizing rates of convergence. *Ann. Statist.*, 19, 668–701, 1991.

[22] R. M. Dudley. Real Analysis and Probability. Chapman und Hall, 1989.

[23] J. Durbin. Weak convergence of the sample distribution function when parameters are estimated. *Ann. Statist.*, 1, 279–290, 1973.

[24] P. Doukhan und S. Louhichi. A new weak dependence condition and application to moment inequalities. *Stoch. Proc. Appl.*, 84, 313–342, 1999.

[25] W. Eddy. Optimum kernel estimators of the mode. *Ann. Statist.*, 8, 870–882, 1980.

[26] W. Eddy. The asymptotic distributions of kernel estimators of the mode. *Z. Wahrsch. Verw. Gebiete*, 59, 279-290, 1982.

[27] B. Efron. Bootstrap methods: another look at the jackknife. *Ann. Statist.*, 7, 1–26, 1979.

[28] J. Fan. On the optimal rates of convergence for nonparametric deconvolution problems, *Ann. Statist.*, 19, 1257–1272, 1991a.

[29] J. Fan. Global behavior of deconvolution kernel estimates. *Statist. Sinica*, 1, 541–551, 1991b.

[30] J. Fan und Y. K. Truong. Nonparametric regression with errors in variables. *Ann. Statist.*, 21, 1900–1925, 1993.

[31] Th. Gasser, H.G. Müller. Estimating regression functions and their derivatives by the kernel method. *Scand. J. Statist.*, 11, 171–185, 1984.

[32] C. Genest und B. Rémillard. Validity of the parametric bootstrap for goodness-of-fit testing in semiparametric models. *Ann. Inst. H. Poincaré Probab. Statist.*, 44, 1096–1127, 2008.

[33] E. Giné und J. Zinn. Bootstrapping general empirical measures. *Ann. Probab.*, 18, 851–869, 1990.

[34] P. Hall, S.N. Lahiri, J. Polzehl. On bandwidth choice in nonparametric regression with both short- and long-range dependent errors. *Ann. Statist.*, 23, 1921–1936, 1995.

[35] P. Hall und S.R. Wilson. Two guidelines for bootstrap hypothesis testing. *Biometrics*, 47, 757–762, 1991.

[36] R. Z. Hasminskii. Lower bounds for the risk of nonparametric estimates of the mode. In *Contributions to Statistics: Jaroslav Hájek Memorial Volume*, J. Jurecková, ed., Academia, Prague, 1979, pp. 91–97, 1979.

[37] S. B. Hedges und P. Shah. Comparison of mode estimation methods and application in molecular clock analysis. *BMC Bioinformatics*, 4:31, 2003.

[38] N. Henze. Empirical-distribution-function goodness-of-fit tests for discrete models. *Canad. J. Statist.*, 24, 81–93, 1996.

[39] E. Herrmann, K. Ziegler. Rates of consistency for nonparametric estimation of the mode in absence of smoothness assumptions. *Statist. Probab. Lett.*, 68, 359–368, 2004.

[40] C. H. Hesse. Data-driven deconvolution. *J. Nonparametr. Stat.*, 10, 343-–373, 1999.

[41] J. Johannes. Deconvolution with unknown error distribution. *Ann. Statist.*, 37, 2301–2323, 2009.

[42] M. Kac, J. Kiefer und J. Wolfowitz. On tests of normality and other tests of goodness of fit based on distance methods. *Ann. Math. Statist.*, 26, 189–211, 1955.

[43] H. R. Künsch. The jackknife and the bootstrap for general stationary observations. *Ann. Statist.*, 17, 1217–1241, 1989.

[44] S. N. Lahiri. Resampling Methods for dependent data. Springer, 2003.

[45] S. N. Lahiri. Theoretical comparisons of block bootstrap methods. *Ann. Statist.*, 27, 386–404, 1992.

[46] E. Liebscher. Estimation of the density and the regression function under mixing conditions. *Statist. Decisions*, 19, 9–26, 2001.

[47] R. Y. Liu und K. Singh. Moving blocks jackknife and bootstrap capture weak dependence. In *Exploring the Limits of Bootstrap* (R. Lepage und L. Billard, eds.), 225—248. Wiley, New York.

[48] M.C. Liu und R. L. Taylor. A consistent nonparametric density estimator for the deconvolution problem. *Canad. J. Statist.*, 17, 427–438, 1989.

[49] A. Meister. On general consistency in deconvolution mode estimation. *J. Statist. Plann. Inf.*, 141, 771–781, 2010.

[50] H.-G. Müller. Kernel Estimators of zeros and of Location and Size of Extrema of Regression Functions. *Scand. J. Stat.*, 12, 221–232, 1985.

[51] H.-G. Müller (1989). Adaptive nonparametric peak estimation. *Ann. Statist.*, 17, 1053–1069, 1989.

[52] H.-G. Müller und U. Stadtmüller. Estimation of heteroscedasticity in regression analysis. *Ann. Statist.*, 15, 610–625, 1987.

[53] U. V. Naik-Nimbalkar und M. B. Rajarshi. Validity of blockwise bootstrap for empirical processes with stationary observations. *Ann. Statist.*, 22, 980–994, 1994.

[54] M. H. Neumann. Optimal change-point estimation in inverse problems. *Scand. J. Stat.*, 24, 503–521, 1997.

[55] M. H. Neumann und E. Paparoditis. Goodness-of-fit tests for markovian time series models: Central limit theory and boostrap approximations. *Bernoulli*, 14, 14–46, 2007.

[56] E. Parzen. On estimation of a probability density and mode. *Ann. Math. Statist.*, 12, 1215–1230, 1962.

[57] T. Pauls. Resampling-Verfahren und ihre Anwendungen in der nichtparametrischen Testtheorie. Dissertation, Universität Düsseldorf, 2003.

[58] M. Peligrad. On the blockwise bootstrap for empirical processes for stationary sequences. *Ann. Probab.*, 26, 877–901, 1998.

[59] D. Politis und J. P. Romano. A circular block resampling procedure for stationary data. In *Exploring the Limits of Bootstrap* (R. Lepage und L. Billard, eds.), 263—270. Wiley, New York, 1992.

[60] D. Politis und J. P. Romano The stationary bootstrap. *J. Amer. Statist. Assoc.*, 89, 1303—1313, 1994.

[61] D. Pollard. Convergence of stochastic processes. Springer, 2008.

[62] H. Pruscha. Vorlesungen über mathematische Statistik. Teubner, 2000.

[63] M. Rachdi und R. Sabre. Consistent estimates of the mode of the probability density function in nonparametric deconvolution problems. *Statist. Probab. Lett.*, 47 , 297–307, 2000.

[64] J.-F. Quessy und M. Mailhot. Asymptotic power of tests of normality under local alternatives. *J. Statist. Plann. Inf.*, 141 , 2787–2802, 2009.

[65] D. Radulovic. The bootstrap for empirical processes based on stationary observations. *Stoch. Proc. Appl.*, 65, 259–279, 1996.

[66] M. S. Rao, und Y. S. R. Krishnaiah. Weak convergence of empirical processes of strong mixing sequences under estimation of parameters. *Sankhya Ser. A*, 50, 26–43, 1988.

[67] J. Romano (1988). On weak convergence and optimality of kernel density estimates of the mode. *Ann. Statist.*, 16, 629–647.

[68] J. Shao und D. Tu. The Jackknife and Bootstrap. Springer, 1995.

[69] G. R. Shorack, und J.A. Wellner. Empirical processes with applications to statistics. Wiley, 1986.

[70] J. Shoung und C. Zhang. Least squares estimators of the mode of a unimodal regression function. *Ann. Statist.*, 29, 648–665, 2001.

[71] L. Stefanski. Rates of convergence of some estimators in a class of deconvolution problems. *Statist. Probab. Lett.*, 9, 229–235, 1990.

[72] L. Stefanski und R. Carroll. Deconvoluting Kernel Density Estimators. *Statistics*, 21, 169–184, 1990.

[73] W. Stute, W. Gonzáles-Manteiga und M. Presedo-Quindimil. Bootstrap based goodness-of-fit-tests. *Metrika*, 40, 243–256, 1993.

[74] G. Szücs. Parametric bootstrap tests for continuous and discrete distributions. *Metrika*, 67, 63–81, 2008.

[75] N. G. Ushakov. Selected topics in characteristic functions. VSP, 1999.

[76] P. Vieu. A note on density mode estimation. *Stat. Probab. Lett.*, 26, 297–307, 1996.

[77] D. Werner. Funktionalanalysis. Springer, 1995.

[78] B. Wieczorek. On optimal estimation of the mode in nonparametric deconvolution problems. *J. Nonparam. Statist.*, 22, 65–80, 2010.

[79] B. Wieczorek und K. Ziegler. On optimal estimation of a non-smooth mode in a nonparametric regression model with α-mixing errors. *J. Statist. Plann. Inf.*, 140, 406-418, 2010.

[80] S. Yang. Maximal moment inequality for partial sums of strong mixing sequences and application. *Acta Math. Sinica*, 23, 1013–1024, 2007.

Ehrenwörtliche Erklärung

Hiermit erkläre ich,

- dass mir die Promotionsordnung der Fakultät bekannt ist,
- dass ich die Dissertation selbst angefertigt habe, keine Textabschnitte oder Ergebnisse eines Dritten oder eigener Prüfungsarbeiten ohne Kennzeichnung übernommen und alle von mir benutzten Hilfsmittel, persönliche Mitteilungen und Quellen in meiner Arbeit angegeben habe,
- dass ich die Hilfe eines Promotionsberaters nicht in Anspruch genommen habe und dass Dritte weder unmittelbar noch mittelbar geldwerte Leistungen von mir für Arbeiten erhalten haben, die im Zusammenhang mit dem Inhalt der vorgelegten Dissertation stehen,
- dass ich die Dissertation noch nicht als Prüfungsarbeit für eine staatliche oder eine andere wissenschaftliche Prüfung eingereicht habe,
- dass ich bei der Auswahl und Auswertung des Materials sowie bei der Herstellung des Manuskriptes nicht von Dritten unterstützt wurde,
- dass ich weder die gleiche noch eine in wesentlichen Teilen ähnliche bzw. eine andere Abhandlung bereits an einer anderen Hochschule als Dissertation eingereicht habe.

Weimar, 11. März 2013 Barbara Wieczorek