

**Bewertung Amerikanischer Optionen  
mit Hilfe von  
regressionsbasierten Monte-Carlo-Verfahren**

**Dissertation**

zur Erlangung des akademischen Grades des  
Doktors der Naturwissenschaften (Dr. rer. nat.)  
der Naturwissenschaftlich-Technischen Fakultäten  
der Universität des Saarlandes

eingereicht im Juli 2007

von

**Dipl.-Math. Nebojša Todorović**

geboren in Schwäbisch Hall

Dekan: Univ.-Prof. Dr.-Ing. T. Herfet

Tag der Einreichung: 13.07.2007

Tag des Kolloquiums: 11.10.2007

Vorsitzender: Univ.-Prof. Dr. E. Albrecht

Gutachter:

1. Univ.-Prof. Dr. M. Kohler,  
Universität des Saarlandes
2. Univ.-Prof. Dr. J. Franke,  
Technische Universität Kaiserslautern
3. Univ.-Prof. Dr. H. Walk,  
Universität Stuttgart

Akadem. Beisitzer: Dr. R. Kirsch

## Danksagung

Mein Dank gilt an erster Stelle Herrn Prof. Dr. Michael Kohler, der mich für die Stochastik begeisterte, und der mich mit dem Thema der Dissertation vertraut machte. Weiter möchte ich ihm recht herzlich für die intensive Betreuung und ständige Gesprächsbereitschaft über den gesamten Zeitraum während des Entstehens dieser Arbeit danken, ohne welche die Arbeit sicher nicht entstanden wäre. Ich danke ihm nicht nur für das mathematische Fachwissen, das er mir vermittelte, sondern auch für die wissenschaftliche Arbeitsweise, die er mich lehrte. Des Weiteren möchte ich recht herzlich Anne Schwartz danken, die gewillt war, eine erste Version dieser Arbeit zu lesen und eine Vielzahl von Verbesserungsvorschlägen vorbrachte. Auch danke ich ihr für ihre Unterstützung bei den Vorträgen in englischer Sprache. Überdies möchte ich mich bei Frau Margot Andreis-Günther und Dipl.-Math. Conny Clausen sowie Dipl.-Inf. Jens Mehnert für die nette, warmherzige und kollegiale Arbeitsatmosphäre bedanken, sowie dafür, dass sie mir hilfreich zur Seite standen. Ebenso geht mein Dank an Herrn Dr. Daniel Egloff für die zahlreichen nützlichen und informativen Gespräche über die Bewertungstheorie für Amerikanische Optionen sowie für seine Einladungen nach Zürich. Besonders Danken möchte ich Herrn Prof. Dr. Adam Krzyżak, der mir einen Aufenthalt in Kanada ermöglichte. Auch danke ich ihm für seine Gastfreundschaft während meines Aufenthaltes an der Concordia University in Montreal. Herrn Prof. Dr. Jürgen Franke sowie Herrn Prof. Dr. Harro Walk danke ich für die Arbeit als Mitberichter.

An dieser Stelle möchte ich mich auch bei all den Personen bedanken, die meinen Aufenthalt in Saarbrücken verschönerten, allen voran bei Anne Schwartz. Auch danke ich ihnen für die netten entstandenen Freundschaften, die auch nach meinem Aufenthalt in Saarbrücken noch bestehen werden.

Nicht zuletzt möchte ich mich vor allem auch bei meiner Familie recht herzlich bedanken, die mich während der Promotion immer unterstützte und mir stets zur Seite stand.

## Abstract

Pricing of American options in discrete time by means of regression-based Monte Carlo methods is considered in this thesis. Thereby, the options are allowed to be based on several underlyings (so-called basket options). It is assumed that the underlyings price processes satisfy the Markov property and consequently are Markov processes. In this dissertation, Monte Carlo methods are used to generate artificial sample paths of these price processes, and subsequently adaptive nonparametric least squares regression estimates are used to estimate the so-called continuation values from these data. The continuation values describe the mean values of the American options for given values of the underlyings at time  $t$ , subject to the constraint that the options are not exercised at time  $t$ , but optimally exercised in the future. Adaptive least squares neural networks as well as adaptive least squares spline estimates are used as nonparametric regression estimates. In conjunction with the pricing of American options, these estimates are theoretically analysed, and results with respect to consistency and rate of convergence are derived. Finally, the pricing of American options on simulated data by means of the estimates is illustrated.

## Zusammenfassung

In der vorliegenden Arbeit wird die Bewertung Amerikanischer Optionen in diskreter Zeit mit Hilfe von regressionsbasierten Monte-Carlo-Verfahren betrachtet. Die Optionen dürfen dabei auf mehreren Underlyings basieren (so genannte Basket-Optionen). Für die Preisprozesse der Underlyings wird angenommen, dass diese die Markov-Eigenschaft besitzen und somit Markov-Prozesse sind. Mit Monte-Carlo-Verfahren werden dann künstliche Stichprobenpfade dieser Preisprozesse erzeugt, um anschließend mit adaptiven nichtparametrischen Kleinste-Quadrate-Regressionsschätzern ausgehend von diesen Stichprobenpfaden die so genannten Kontinuierungswerte zu schätzen. Dabei beschreiben die Kontinuierungswerte die Mittelwerte der Amerikanischen Optionen für gegebene Kurse der Underlyings zur Zeit  $t$  gemäß der Bedingung, dass die Optionen nicht zur Zeit  $t$ , sondern optimal in der Zukunft ausgeübt werden. Als nichtparametrische Regressionsschätzer werden adaptive Kleinste-Quadrate-Splineschätzer sowie adaptive Neuronale-Netze-Schätzer verwendet. Diese Schätzer werden im Zusammenhang mit der Bewertung Amerikanischer Optionen theoretisch analysiert, und Resultate zur Konsistenz und zur Konvergenzgeschwindigkeit werden hergeleitet. Abschließend wird mit den Schätzern das Bewerten Amerikanischer Optionen auf simulierten Daten illustriert.

# Inhaltsverzeichnis

|          |  |           |
|----------|--|-----------|
| <b>1</b> | <b>Einführung</b>  | <b>9</b>  |
| 1.1      | Bekannte Modelle bzw. Verfahren . . . . .  | 9         |
| 1.2      | Regressionsbasierte Monte-Carlo-Verfahren . . . . .  | 13        |
| 1.3      | Neu erzielte Resultate . . . . .   | 20        |
| <b>2</b> | <b>Allgemeine Resultate zur Konsistenz und Konvergenzgeschwindigkeit</b>                             | <b>23</b> |
| 2.1      | Theoretisches Hauptresultat . . . . .  | 23        |
| 2.2      | Resultate für polynomiale Splinefunktionen . . . . .   | 33        |
| 2.3      | Resultate für neuronale Netze . . . . .  | 42        |
| <b>3</b> | <b>Ein alternativer Zugang zur Analyse des Schätzers bei spezieller Re-<br/>gressionsdarstellung</b> | <b>52</b> |
| 3.1      | Ein alternativer Zugang . . . . .  | 52        |
| 3.2      | Beweis von Satz 3.1 . . . . .  | 54        |
| <b>4</b> | <b>Adaptive Wahl von Parametern</b>  | <b>60</b> |
| 4.1      | Adaption . . . . .   | 60        |
| 4.2      | Adaptive Spline- und Neuronale-Netze-Schätzer . . . . .  | 69        |
| <b>5</b> | <b>Anwendung auf simulierte Daten</b>  | <b>81</b> |
| <b>A</b> | <b>Hilfsresultate</b>  | <b>95</b> |

## Abkürzungsverzeichnis

|                                  |   |
|----------------------------------|---|
| $\mathbb{N}$                     | Menge der natürlichen Zahlen  |
| $\mathbb{N}_0$                   | $\mathbb{N} \cup \{0\}$   |
| $\mathbb{Z}$                     | Menge der ganzen Zahlen   |
| $\mathbb{R}$                     | Menge der reellen Zahlen  |
| $\mathbb{R}_+$                   | Menge der nichtnegativen reellen Zahlen                             |
| $e$                              | Eulersche Zahl  |
| $\lceil x \rceil$                | kleinste ganze Zahl größer oder gleich $x$                          |
| $\lfloor x \rfloor$              | größte ganze Zahl kleiner oder gleich $x$                           |
| $\log$                           | natürlicher Logarithmus (Basis $e$ )                                |
| $ A $                            | Kardinalität bzw. Anzahl der Elemente der endlichen Menge $A$       |
| $I_A$                            | Indikatorfunktion der Menge $A$                                     |
| $[a, b], (a, b)$                 | abgeschlossenes bzw. offenes Intervall von $a$ bis $b$              |
| $(a, b], [a, b)$                 | links- bzw. rechtsoffenes Intervall von $a$ bis $b$                 |
| $x_0 = \arg \min_{x \in D} f(x)$ | Kurzschreibweise für $x_0 \in D$ und $f(x_0) = \min_{x \in D} f(x)$ |
| $\ \cdot\ $                      | Euklidische Norm im $\mathbb{R}^d$                                  |
| $\ \cdot\ _\infty$               | Supremumsnorm einer Funktion bzw. eines Vektors                     |
| $\mathbf{P}_{X_t}$               | Verteilung der Zufallsvariablen $X_t$                               |
| $\mathbf{E}$                     | Erwartungswert  |
| $\mathbf{E}Z^2$                  | Kurzschreibweise für $\mathbf{E}(Z^2)$                              |
| $\mathbf{Var}(Z)$                | Varianz der Zufallsvariablen $Z$                                    |

Alle in einem gemeinsamen Zusammenhang auftretenden Zufallsvariablen seien auf dem Wahrscheinlichkeitsraum  $(\Omega, \mathcal{A}, \mathbf{P})$  definiert.

Mit messbaren Funktionen  $f : \mathbb{R}^d \rightarrow \mathbb{R}$  sind die bezüglich der Borelschen  $\sigma$ -Algebren auf  $\mathbb{R}^d$  bzw.  $\mathbb{R}$  messbaren Funktionen gemeint.





# Kapitel 1

## Einführung

In diesem Kapitel wird eine Einführung in die vorliegende Arbeit gegeben. Zuerst werden in Abschnitt 1.1 bekannte Modelle bzw. Verfahren zur Bewertung Amerikanischer Optionen aufgezeigt, wobei die regressionsbasierten Monte-Carlo-Verfahren nicht explizit erläutert werden. Auf diese Verfahren wird separat in Abschnitt 1.2 eingegangen. Abschließend befasst sich Abschnitt 1.3 mit den neu erzielten Resultaten, die aus dieser Arbeit hervorgehen. Zudem wird dort die Vorgehensweise, die dieser Arbeit zugrunde liegt, erläutert.

### 1.1 Bekannte Modelle bzw. Verfahren

Finanzinstrumente, die von einfacheren Finanzinstrumenten wie beispielsweise Aktien oder Währungen abgeleitet werden, werden als Derivate bezeichnet. Ein Beispiel für Derivate sind Optionen. Optionen sind (Finanz-)Kontrakte mit vereinbarten Rechten, die der Optionskäufer vom Optionsverkäufer, auch Stillhalter genannt, erwirbt. Es gibt zwei grundsätzliche Arten von Optionen: Die Kaufoption, auch Call-Option genannt, und die Verkaufsoption, auch Put-Option genannt.

Der Käufer einer Call-Option bzw. einer Put-Option erwirbt sich das Recht, ein bestimmtes, dem Vertrag zugrunde liegendes Gut, auch Underlying genannt, zu einem späteren Zeitpunkt zu einem bereits zum aktuellen Zeitpunkt festgelegten Preis zu kaufen bzw. zu verkaufen. Dieser festgelegte Preis wird als Ausübungspreis bezeichnet. Der Verkäufer einer Call- bzw. Put-Option verpflichtet sich dementsprechend, dem Käufer dieses Underlying zu überlassen bzw. abzunehmen, sobald dieser sein erworbenes Recht ausüben möchte.

Kann die Option nur zu einem fest gewählten Zeitpunkt, dem so genannten Verfallszeitpunkt, ausgeübt werden, so spricht man von einer Europäischen Option. Kann sie dagegen jederzeit bis zum Verfallszeitpunkt ausgeübt werden, spricht man von einer Amerikanischen Option. Die meisten an den Börsen gehandelten Optionen sind Amerikanischen Typs. Eine ausführliche Beschreibung sowie umfangreiche Beispiele zu Optionen können z. B. in Hull (2006) gefunden werden.

Die Investition in Optionen kann aufgrund ganz unterschiedlicher Motivationen erfolgen. Ein Motiv kann zum Beispiel der Wunsch nach Absicherung gegen Kursschwankungen des Underlyings sein, wie beispielsweise bei Optionen auf Wechselkurse. Aber auch das genau entgegengesetzte Motiv, nämlich die Spekulation auf Kursschwankungen des Underlyings, ist ein Grund, in Optionen zu investieren.

Die Finanzmathematik beschäftigt sich seit langem mit der Frage nach dem “fairen Preis” einer Option. Die Idee dabei ist, zunächst ein geeignetes (stochastisches) Modell für den zukünftigen Kurs des Underlyings aufzustellen, um dann im Rahmen dieses Modells den “fairen Preis” der Option zu berechnen. Heutzutage wird in der Praxis meistens die so genannte Black-Scholes-Theorie zur Bewertung von Optionen eingesetzt, die auf Black und Scholes (1973) sowie auf Merton (1973) zurückgeht. Diese entwickelten eine als Black-Scholes-Modell (oder auch Black-Scholes-Merton-Modell) bekannt gewordene Bewertungsmethodik, welche in einer Vielzahl von Büchern zur Finanzmathematik dargestellt wird (siehe z. B. Elliott und Kopp (1999), Korn und Korn (2001) oder Sandmann (2001)). Dabei wird der Kurs  $S_t$  des Underlyings in kontinuierlicher Zeit  $t$  durch eine so genannte geometrische Brownsche Bewegung, die beschrieben wird durch die stochastische Differentialgleichung (SDGL)

$$\frac{dS_t}{S_t} = \mu dt + \sigma dW_t, \quad (1.1)$$

mit Rendite  $\mu \in \mathbb{R}$ , Volatilität  $\sigma \geq 0$  und Brownscher Bewegung bzw. Wiener Prozess  $\{W_t\}_t$ , modelliert. Die Lösung dieser SDGL ist die geometrische Brownsche Bewegung

$$S_t = S_0 \cdot \exp\left(\left(\mu - \frac{1}{2}\sigma^2\right) \cdot t + \sigma \cdot W_t\right), \quad (1.2)$$

wobei  $S_0$  der Kurs des Underlyings zur Zeit  $t = 0$  ist. In diesem Modell wird angenommen, dass der Finanzmarkt vollständig ist, d. h. die Option kann durch Kaufen bzw. Leerverkaufen des Underlyings und eines Bonds, welches einem festverzinslichen Wertpapier entspricht, komplett nachgebildet werden. Unter der Annahme der Arbitragefreiheit (d. h. des Ausschlusses von risikolosen Gewinnen) kann gezeigt werden, dass der “faire Preis” der Option durch das Anfangskapital einer Handelsstrategie, die die Option nachbildet, gegeben ist. Im Rahmen der Black-Scholes-Theorie kann dieser faire Preis durch Lösen einer partiellen Differentialgleichung (bzw. bei Amerikanischen Optionen durch Lösen eines freien Randwertproblems, vgl. z. B. Seydel (2006)) berechnet werden. Im einfachen Modell (1.2) existieren explizite Formeln für den fairen Preis, die so genannten Black-Scholes-Formeln.

Es ist seit Langem bekannt, dass die geometrische Brownsche Bewegung in (1.2) zur Modellierung von in der Praxis beobachteten Kursentwicklungen nur bedingt geeignet ist, da beispielsweise in (1.2) im Vergleich zu den in der Praxis beobachteten Daten starke Schwankungen der Kurse zu selten auftreten. Daher wurden andere Modelle entwickelt. Eine Alternative zum Black-Scholes-Modell ist das Modell der konstanten Elastizität der Varianz, welches unter dem Namen CEV-Modell (von “constant elasticity of variance”)

bekannt ist und durch die SDGL

$$\frac{dS_t}{S_t} = \mu dt + \sigma S_t^{\alpha-1} dW_t \quad (1.3)$$

beschrieben wird (vgl. Cox und Ross (1976)). Hierbei hat der Kurs des Underlyings die Volatilität  $\sigma S_t^{\alpha-1}$ , und es ist einfach zu erkennen, dass das CEV-Modell für  $\alpha = 1$  dem Modell (1.1) und somit der geometrischen Brownschen Bewegung in (1.2) entspricht. Wie in (1.2) wird auch im CEV-Modell der Kurs des Underlyings kontinuierlich modelliert. Solche Modelle werden auch Diffusionsmodelle genannt. In den letzten 25 Jahren wurde eine Vielzahl weiterer Modelle entwickelt, von denen die meisten Diffusionsmodelle sind. Hier werden nur einige der wichtigsten genannt. Eine unmittelbare Verallgemeinerung von (1.3), und somit auch von (1.1), sind die von Dupire (1994) und Derman und Kani (1994) vorgeschlagenen nichtlinearen Markov-Diffusionen, deren Volatilität  $\sigma$  vom Preis  $S_t$  und - im Unterschied zu (1.3) - auch von der Zeit  $t$  abhängt. Dieses Diffusionsmodell wird beschrieben durch die SDGL

$$\frac{dS_t}{S_t} = \mu dt + \sigma(t, S_t) dW_t. \quad (1.4)$$

Alternativ dazu gibt es auch stochastische Volatilitätsmodelle (vgl. Heston (1993) oder Hull und White (1987)), bei denen sowohl der Preis  $S_t$  als auch die Volatilität  $\sigma_t$  durch SDGLen beschrieben werden:

$$\frac{dS_t}{S_t} = \mu dt + \sigma_t dW_t^1, \quad \sigma_t = f(Y_t), \quad dY_t = \alpha_t dt + \gamma_t dW_t^2. \quad (1.5)$$

In (1.5) wird weiter angenommen, dass die Wiener Prozesse  $\{W_t^1\}_t$  und  $\{W_t^2\}_t$  unabhängig voneinander sind. Im Gegensatz zu obigen Modellen ist ein anderer Ansatz die Modellierung des Kurses des Underlyings durch Sprungdiffusionsmodelle. Bei diesen Modellen wird der kontinuierliche Kurs des Underlyings durch Sprünge überlagert. Ein Beispiel hierfür ist das vorgeschlagene Modell von Merton (1976), welches man durch Hinzufügen einer Sprungkomponente  $V_t$  zu (1.4) erhält gemäß

$$\frac{dS_t}{S_t} = \mu dt + \sigma(t, S_t) dW_t + dV_t. \quad (1.6)$$

Modelle, die den Kurs des Underlyings von vornherein nur durch Sprünge beschreiben, heißen Sprungmodelle. Bei diesen wird der Preisprozess  $\{S_t\}_t$  als reiner Sprungprozess (mit stückweisen konstanten Pfaden) angesetzt, wie beispielsweise bei exponentiellen Levy-Prozessen (vgl. z. B. Kapitel 11 in Cont und Tankov (2004)). Allgemeine Sprungprozesse sind z. B. in den Kapiteln 11, 14 und 15 in Cont und Tankov (2004) beschrieben.

Bei der Mehrzahl der vorgestellten Modelle ist die Annahme eines vollständigen Marktes, wie es bei dem Black-Scholes-Modell der Fall ist, nicht gewährleistet. Aufgrund der Unvollständigkeit der Märkte in der Mehrzahl der obigen Modellen kann die Option durch eine Handelsstrategie nicht vollständig nachgebildet werden. Aus diesem Grund lässt sich der Begriff des "fairen Preises" einer Option nicht wie beim Black-Scholes-Modell

unmittelbar aus einer Handelsstrategie ableiten. Es ist aber möglich, dass die Auszahlung einer Option zu einem großen Teil nachgebildet werden kann, wobei ein Restrisiko für den Optionsverkäufer bleibt. Dieser wiederum muss sich überlegen, wie er dieses Restrisiko mit dem Preis der Option verrechnet. Verschiedene Ansätze dazu sind beispielsweise in den Kapiteln 6 und 7 in Föllmer und Schied (2002) oder in Kapitel 10 in Cont und Tankov (2004) aufgeführt.

Der resultierende Preis einer Option ist in diesen Modellen meist durch Lösen von partiellen Differentialgleichungen bzw. durch Lösen von Randwertproblemen bei Amerikanischen Optionen beschreibbar. Alternativ hierzu lässt sich der Preis einer auf wenigen Underlyings basierenden Option durch Baummethode (z. B. Bi- bzw. Trinomialbäume) bestimmen. Der Preis einer allgemeinen Option kann hingegen durch Monte-Carlo-Verfahren bestimmt werden.

Eine Herausforderung für die obigen Modelle stellt das Bewerten so genannter Basket-Optionen, d. h. auf mehreren Aktien basierenden Optionen, dar. Die Herausforderung ist dabei die Modellierung der Abhängigkeitsstruktur der Aktien. Beim Black-Scholes-Modell würde dabei die Korrelationsmatrix einer mehrdimensionalen Brownschen Bewegung gewählt werden, wobei bekannt ist, dass mit dieser Wahl der Korrelationsmatrix starke Kursschwankungen wiederum zu selten auftreten. Alternativ werden bei den obigen Modellen heutzutage meistens so genannte Copula-Modelle verwendet, bei denen die multivariate Verteilungsfunktion als multivariate Funktion, des so genannten Copula, der Verteilungsfunktion der Randverteilungen gewählt wird. Die Abhängigkeitsstruktur wird dann durch die Wahl einer Klasse von Funktionen für das Copula festgelegt. Eine theoretische Einführung zu Copulas kann in Nelson (1999) und eine praktisch orientierte Umsetzung kann in Deutsch und Eller (1999) gefunden werden. An dieser Stelle sei noch erwähnt, dass zum einen die Wahl des richtigen Copula für die betrachteten Underlyings und zum anderen der erheblich größere Aufwand bei der Bestimmung der Preise der Option (beispielsweise durch partielle Differentialgleichungen, deren Bauart sich durch das Copula verkompliziert) Herausforderungen darstellen.

Im Unterschied zu den obigen Verfahren wird bei der Modellierung des Kurses des Underlyings durch Zeitreihenschätzung der Kurs in diskreter Zeit modelliert. Ein Ansatz hier sind beispielsweise die auf Engle (1982) zurückgehenden ARCH-Modelle (von “autoregressive conditional heteroskedasticity”). Dabei hat das Modell der Ordnung  $q$  die Form

$$e_t = \sigma_t Z_t, \quad \sigma_t^2 = \omega + \alpha_1 e_{t-1}^2 + \dots + \alpha_q e_{t-q}^2, \quad (1.7)$$

wobei  $\omega > 0$ ,  $\alpha_1 \geq 0, \dots, \alpha_q \geq 0$  und die  $Z_t$  unabhängig identisch verteilte Zufallsvariablen mit Erwartungswert 0 und Varianz 1 sind. Der Prozess  $\{e_t\}_{t \in \mathbb{Z}}$  in (1.7) heißt dabei ARCH( $q$ ). Möchte man mit diesen ARCH-Prozessen Finanzdaten (wie den Kurs eines Underlyings) passend modellieren, muss eine hohe Modellordnung  $q$  verwendet werden, was wiederum zum Schätzen einer großen Zahl von Parametern führt. Dies kann durch die GARCH-Modelle (von “generalized ARCH”), die eine Erweiterung der ARCH-Modelle

darstellen, umgangen werden. Diese GARCH-Modelle wurden von Bollerslev (1986) vorgeschlagen, und ihre einfachste Version ist das GARCH(1,1)-Modell der Form

$$e_t = \sigma_t Z_t, \quad \sigma_t^2 = \omega + \alpha e_{t-1}^2 + \beta \sigma_{t-1}^2. \quad (1.8)$$

In (1.8) heißt dabei der Prozess  $\{e_t\}_{t \in \mathbb{Z}}$  GARCH(1,1). Sowohl die ARCH- als auch die GARCH-Modelle sind stochastische Volatilitätsmodelle in diskreter Zeit, auch wenn sich ihre Ansätze zur Modellierung von Finanzdaten auf den ersten Blick von denen obigen Diffusionsmodellen unterscheiden. In Nelson (1990) wurde gezeigt, dass die ARCH- sowie GARCH-Prozesse im Grenzübergang gegen einen Diffusionsprozess in kontinuierlicher Zeit konvergieren. Mehr zu ARCH- oder GARCH-Modellen kann z. B. in Franke, Härdle und Hafner (2004) oder Jondeau, Poon und Rockinger (2007) gefunden werden. Dort werden auch einige Verallgemeinerungen des GARCH-Ansatzes, wie beispielsweise die EGARCH-Modelle (von “exponential GARCH”) oder TGARCH-Modelle (von “threshold GARCH”), beschrieben.

In einer Reihe von Arbeiten wurden bereits Methoden der Zeitreihenschätzung zur Bewertung von Optionen angewandt. Dabei wurden in erster Linie Monte-Carlo-Verfahren eingesetzt. Mehr Einzelheiten zur Optionspreisbewertung mittels Zeitreihenschätzung können z. B. in Duan (1995), Duan et al. (2003), Duan, Ritchken und Sun (2005) oder in Stentoft (2005) gefunden werden.

Einen anderen Ansatz für die Bewertung von Optionen liefert das Monte-Carlo-Verfahren. Bei diesem Verfahren werden künstlich Pfade der zugrunde liegenden Preisprozesse erzeugt, um den Preis einer Option zu schätzen. Der erste Versuch zur Nutzung von Monte-Carlo-Verfahren zur Bewertung Europäischer Optionen geht zurück auf Boyle (1977). Seither haben sich sehr viele Arbeiten mit den Monte-Carlo-Verfahren zur Optionsbewertung befasst. Die meisten dieser Arbeiten haben dabei die Preise von Europäischen bzw. von Amerikanischen Optionen, die auf einem oder auch zwei Underlyings basieren können, bewertet. Der nächste Abschnitt beschäftigt sich mit Monte-Carlo-Verfahren, die zur Bewertung Amerikanischer Optionen zu Hilfe genommen werden. Diese Methoden lassen sich mit Ausnahme der dualen Methoden als regressionsbasierte Monte-Carlo-Verfahren beschreiben.

## 1.2 Regressionsbasierte Monte-Carlo-Verfahren

Wie im vorherigen Abschnitt erläutert, werden bei den Monte-Carlo-Verfahren künstlich Pfade der zugrunde liegenden Preisprozesse erzeugt, um den Preis einer Option zu schätzen. Liegt beispielsweise der einfache Fall vor, dass eine Europäische Option bewertet werden soll, dann wird der Preis dieser Option als Erwartungswert bezüglich des risikolosen Maßes (siehe z. B. Karatzas und Shreve (1998) für entsprechende Resultate in vollständigen und arbitragefreien Märkten) dargestellt, und anschließend wird dieser Erwartungswert durch das Stichprobenmittel geschätzt. Es sei hier erwähnt, dass durch eine Reihe von Techniken die Varianz des Stichprobenmittels reduziert werden kann, vgl. z. B. Kapitel 4 in Glasserman (2004).

Liegt nicht oben genannter Fall vor und gilt es, eine Amerikanische Option zu bewerten, dann gestaltet sich die Preisbestimmung schwieriger aufgrund der Tatsache, dass bei einer Amerikanischen Option der Optionskäufer die Option bis zum Verfallszeitpunkt jederzeit ausüben kann. Regressionsbasierte Monte-Carlo-Verfahren bestimmen den Wert einer Amerikanischen Option, indem der optimale Ausübungszeitpunkt durch rekursives Vergleichen des Kurswertes zu einem festen Zeitpunkt mit dem so genannten Kontinuierungswert, dem mittleren zukünftigen Wert der Option, bestimmt wird. Dabei ist der Kontinuierungswert der Wert, den man im Mittel erhält, wenn die Option optimal in der Zukunft ausgeübt wird. Dieser mittlere zukünftige Kurswert wird rekursiv durch Anwenden eines Regressionsschätzers auf eine so genannte Monte-Carlo-Stichprobe des Kurses geschätzt.

Wegen der freien Wahl des Ausübungszeitpunktes bei Amerikanischen Optionen müssen optimale Stoppprobleme gelöst werden. In vollständigen und arbitragefreien Märkten ist der Preis einer Amerikanischen Option mit Verfallszeitpunkt  $T$  gegeben durch den Wert des optimalen Stoppproblems (siehe z. B. Karatzas und Shreve (1998))

$$V_0 = \sup_{\tau \in \mathcal{T}[0, T]} \mathbf{E}\{f_\tau(X_\tau)\}. \quad (1.9)$$

Hierbei ist  $f_t$  die (diskontierte) Auszahlungsfunktion (auch Payoff-Funktion genannt),  $X_t$  ein stochastischer Prozess, der die relevanten Risikofaktoren modelliert, wie beispielsweise die Kurse des zugrunde liegenden Underlyings, und  $\mathcal{T}[0, T]$  ist die Klasse aller Stoppzeiten mit Werten in  $[0, T]$ . Wie in Abschnitt 1.1 erwähnt, wird in der Praxis  $X_t$  oft als eine geometrische Brownsche Bewegung modelliert, wie zum Beispiel im Black-Scholes-Modell, bei dem eine risikoneutrale Bewertung stattfindet (siehe z. B. Kapitel 13 in Steele (2001)). Mehr zur risikoneutralen Bewertung von Optionen in vollständigen sowie in unvollständigen Märkten kann beispielsweise in Bingham und Kiesel (2004) und in der dort aufgeführten Literatur nachgelesen werden.

Eine besondere Herausforderung im Hinblick auf die praktische Optionsbewertung ist, dass es eine rein analytische Lösung nur für einfache Bewertungsfälle oder (aber) Spezialfälle gibt. Bereits bei den Amerikanischen Optionen ist man aufgrund des optimalen Stoppproblems gezwungen, auf numerisch-approximative Lösungsverfahren zurückzugreifen.

Ein erster Schritt in Richtung der numerischen Lösung von (1.9) ist der Übergang von stetiger zu diskreter Zeit. In der Terminologie des Finanzwesens bedeutet dies, dass die Amerikanische Option durch eine Bermuda-Option approximiert wird. Die Konvergenz der zeitdiskreten Approximationen gegen das optimale Stoppproblem im zeitstetigen Falle wird in Lamberton und Pagès (1990) für den Markovschen Fall, aber auch für allgemeinere stochastische Prozesse betrachtet.

In Analogie zu (1.9) kann der Preis  $V_0$  einer Amerikanischen Option im zeitdiskreten Fall definiert werden als die Lösung eines optimalen Stoppproblems

$$V_0 = \sup_{\tau \in \mathcal{T}(0, \dots, T)} \mathbf{E}\{f_\tau(X_\tau)\}. \quad (1.10)$$

Hierbei ist  $f_t$  die (diskontierte) Auszahlungsfunktion,  $X_0, X_1, \dots, X_T$  ist der zugrunde liegende (zeitdiskrete) stochastische Prozess, und  $\mathcal{T}(0, \dots, T)$  ist die Klasse aller  $\{0, \dots, T\}$ -wertigen Stoppzeiten, d. h.  $\tau \in \mathcal{T}(0, \dots, T)$  ist eine messbare Funktion von  $X_0, \dots, X_T$ , für die gilt:

$$\{\tau = \alpha\} \in \mathcal{F}(X_0, \dots, X_\alpha) \text{ für alle } \alpha \in \{0, \dots, T\}.$$

An dieser Stelle sei für zusätzlichen theoretischen Hintergrund zur Bewertung von zeitdiskreten Amerikanischen Optionen (Bermuda-Optionen) z. B. auf Duffie (2001) oder Schweizer (2002) verwiesen.

Explizite Lösungen von (1.10) existieren mit Ausnahme von sehr wenigen Fällen in aller Regel nicht. Es besteht aber eine Vielzahl von numerischen Verfahren, die das optimale Stoppproblem lösen (zweifelloos spätestens seit der Binomialbaum-Methode von Cox, Ross und Rubinstein (1979)), jedes mit seinen Stärken und Schwächen. Die ersten Versuche zur Nutzung von Simulationen in diesem Kontext wurden unternommen von Boessarts (1989), Tilley (1993), Barraquand und Martineau (1995), Broadie und Glassermann (1997), Broadie, Glassermann und Jain (1997) sowie Broadie und Glassermann (1998). Während in Boessarts (1989), Tilley (1993) sowie in Barraquand und Martineau (1995) die Standard-Gitter-Methode verwendet wurde, benutzten Broadie und Glassermann (1997), Broadie, Glassermann und Jain (1997) sowie Broadie und Glassermann (1998) die Stochastische-Gitter-Methode, deren Ansatz der Binomial-Methode ähnelt. An dieser Stelle sei erwähnt, dass beispielsweise in Kapitel 8 in Glasserman (2004) eine Reihe von Verfahren beschrieben wird, die sich mit Ausnahme der dualen Methoden alle als regressionsbasierte Monte-Carlo-Verfahren beschreiben lassen. Auch wird in Glassermans Kapitel 8.6 auf die noch folgenden Ansätze eingegangen.

Eine alternative Idee zu obigen Ansätzen wurde in Carriere (1996) erwähnt. Dort wurde mittels eines rückwärts gerichteten Induktionstheorems gezeigt, dass die Bewertung Amerikanischer Optionen äquivalent ist zur Berechnung einer Anzahl von bedingten Erwartungswerten. Da im Allgemeinen die bedingten Erwartungswerte in der Praxis sehr schwer zu berechnen sind, wurde in Carriere (1996) auch erwähnt, wie man Simulationstechniken mit fortgeschrittenen Regressionsmethoden kombinieren kann, um die bedingten Erwartungswerte zu approximieren. In diesem Kontext wurde auch die nichtparametrische Regression erwähnt. Eine theoretische Analyse des dort vorgestellten Schätzers erfolgte aber nicht.

Tsitsiklis und Van Roy (1999) sowie Longstaff und Schwartz (2001) stellten unabhängig voneinander eine neue und einfachere simulationsbasierte Methode zur Bewertung Amerikanischer Optionen vor. Dieses Verfahren basiert auf der parametrischen Regressions-schätzung der Kontinuierungswerte. Dabei besteht der Unterschied zwischen den beiden vorgestellten Verfahren im Rekursionsschritt. Bei Longstaff und Schwartz (2001) wird dabei bis zum Ende der Laufzeit der Option vorausgeblickt, während Tsitsiklis und Van Roy (1999) nur den nächstfolgenden Schritt betrachten. Die in den beiden Arbeiten vorgeschlagene Regressionsdarstellung für die Kontinuierungswerte  $q_t$  wird weiter unten explizit in geschriebener Form präsentiert. Hierbei sei erwähnt, dass in Tsitsiklis und Van Roy (1999) auch die Konsistenz bei spezieller Bauart von  $q_t$  untersucht wurde,

während in Longstaff und Schwartz (2001) keine Aussagen zur Konsistenz oder zur Konvergenzgeschwindigkeit erfolgten. Eine Aussage zur Konvergenzgeschwindigkeit von dem in Tsitsiklis und Van Roy (1999) vorgestellten Verfahren erfolgte später in Tsitsiklis und Van Roy (2001). Untersuchungen zur Konsistenz und Konvergenzgeschwindigkeit des in Longstaff und Schwartz (2001) vorgestellten Verfahrens erfolgten in Clément, Lambertson und Protter (2002).

Die eingesetzten Verfahren der parametrischen Regressionsschätzung in Tsitsiklis und Van Roy (1999) sowie in Longstaff und Schwartz (2001) setzen die Bauart der Kontinuierungswerte als linear voraus, und somit werden nur die Koeffizienten dieser linearen Funktion geschätzt. Diese Verfahren können allerdings nur dann gute Ergebnisse liefern, wenn sich die zugrunde liegenden Kontinuierungswerte durch Funktionen der vorausgesetzten parametrischen Bauart - also durch lineare Funktionen - gut approximieren lassen. Um diesen Nachteil zu umgehen, wurden in Egloff (2005) Verfahren der nichtparametrischen Regressionsschätzung zur Schätzung der Kontinuierungswerte vorgeschlagen. Dabei wurden die Verfahren aus Tsitsiklis und Van Roy (1999) und Longstaff und Schwartz (2001) bzw. die dort verwendete Regressionsdarstellung für die Kontinuierungswerte kombiniert, indem eine Schrittweite, die die Vorausschau beschreibt, als neuer Parameter eingeführt wurde. Diese für die Kontinuierungswerte  $q_t$  vorgeschlagene Regressionsdarstellung wird ebenfalls unten explizit in geschriebener Form präsentiert. Neben der Vorstellung des neuen Verfahrens beinhaltet Egloff (2005) auch Resultate zur Konsistenz und zur Konvergenzgeschwindigkeit bei den auftretenden Regressionsproblemen. Im Hinblick auf eine Vereinfachung der theoretischen Analyse des Schätzers ist dieser in Egloff (2005) so definiert, dass der Schätzer nur schlecht berechnet werden kann (da die Wahl der Parameter des Schätzers nicht datenabhängig erfolgte). Deshalb sind in Egloff (2005) auch keine Illustrationen des Schätzverfahrens durch Anwendung auf simulierte Daten, wie in Longstaff und Schwartz (2001) und Tsitsiklis und Van Roy (1999), enthalten.

An dieser Stelle sei erwähnt, dass die Verfahren in Tsitsiklis und Van Roy (1999), Longstaff und Schwartz (2001) sowie in Egloff (2005) die frühzeitige Ausübungsregel approximieren und deshalb eine untere Schranke für den wahren optimalen Stoppwert anbieten. Obere Schranken, die auf dem so genannten Martingal-Dualitätsproblem basieren, wurden in Rogers (2002), Haugh und Kogan (2004) sowie in Andersen und Broadie (2004) hergeleitet. Mehr Einzelheiten hierzu sowie weitere Literaturangaben können in Glasserman (2004) gefunden werden. Ein Vergleich von einigen Monte-Carlo-Ansätzen erfolgt in Laprise et al. (2001).

Da sich die in dieser Arbeit vorgestellten Verfahren der vorgeschlagenen Regressionsdarstellung für die Kontinuierungswerte in Egloff (2005) bedienen werden, wird als erstes die formale Schreibweise der Kontinuierungswerte  $q_t$  eingeführt. Anschließend werden die vorgeschlagenen Regressionsdarstellungen der Kontinuierungswerte aus Tsitsiklis und Van Roy (1999), Longstaff und Schwartz (2001) sowie Egloff (2005) präsentiert.



**Kontinuierungswerte  $q_t$** 

Im Folgenden wird angenommen, dass  $X_0, X_1, \dots, X_T$  ein  $\mathbb{R}^d$ -wertiger Markov-Prozess ist. In diesem Markov-Prozess sind alle notwendigen Informationen über die Finanzinvestitionen vorhanden, die sowohl Preise der zugrunde liegenden Underlyings als auch zusätzliche Risikofaktoren beinhalten, die die stochastische Volatilität oder die stochastischen Zinsraten leiten. Weder die Markov-Eigenschaft noch die Form der Auszahlung als Funktion des Zustandes von  $X_t$  stellen eine Einschränkung dar, und können immer durch Einbinden zusätzlicher Variablen erreicht werden. Beispielsweise im Falle von Asiatischen Optionen kann das ‐laufende Minimum‐ als Zusatzvariable in  $X_t$  hinzugefügt werden.

Die Berechnung von (1.10) kann durch die Bestimmung einer optimalen Stopppregel  $\tau^* \in \mathcal{T}(0, \dots, T)$ , die

$$V_0 = \mathbf{E}\{f_{\tau^*}(X_{\tau^*})\} \quad (1.11)$$

genügt, erreicht werden. Sei

$$q_t(x) = \sup_{\tau \in \mathcal{T}(t+1, \dots, T)} \mathbf{E}\{f_\tau(X_\tau) | X_t = x\} \quad (1.12)$$

der Kontinuierungswert, der den Wert der Option zur Zeit  $t$  bei gegebenem  $X_t = x$  beschreibt gemäß der Bedingung, die Option zur Zeit  $t$  zu halten anstatt sie auszuüben. Hierbei ist  $\mathcal{T}(t+1, \dots, T)$  die Klasse aller  $\{t+1, \dots, T\}$ -wertigen Stoppzeiten. Die allgemeine Theorie für optimale Stoppzeiten für Markov-Prozesse, siehe z. B. Chow, Robbins und Siegmund (1971), Neveu (1975), Shiriyayev (1978) oder El Karoui (1981), impliziert, dass

$$\tau^* = \inf\{s \geq 0 : q_s(X_s) \leq f_s(X_s)\} \quad (1.13)$$

eine optimale Stoppzeit ist, d. h.  $\tau^*$  erfüllt (1.11). Deshalb genügt es, die Kontinuierungswerte (1.12) zu berechnen, um das optimale Stoppproblem (1.10) zu lösen. Dabei genügen die Kontinuierungswerte den folgenden dynamischen Programmgleichungen:

$$\begin{aligned} q_T(x) &= 0, \\ q_t(x) &= \mathbf{E}\{\max\{f_{t+1}(X_{t+1}), q_{t+1}(X_{t+1})\} | X_t = x\} \quad (t = 0, 1, \dots, T-1). \end{aligned} \quad (1.14)$$

Dabei ist die erste Beziehung trivial, die zweite wird weiter unten gezeigt. Diese dynamischen Programmgleichungen zeigen, dass das optimale Stoppproblem in diskreter Zeit äquivalent ist zu einer Reihe von Regressionsproblemen.

**Regressionsdarstellungen für  $q_t$** 

Für obige Kontinuierungswerte  $q_t$  wird in Tsitsiklis und Van Roy (1999) die Regressionsdarstellung

$$q_t(x) = \mathbf{E}[\max\{f_{t+1}(X_{t+1}), q_{t+1}(X_{t+1})\} | X_t = x] \quad (1.15)$$

verwendet, während die Verfahren in Longstaff und Schwartz (2001) auf der Regressionsdarstellung

$$q_t(x) = \mathbf{E}[f_{\tau_{t+1}^*}(X_{\tau_{t+1}^*}) | X_t = x] \quad (1.16)$$

mit

$$\tau_{t+1}^* = \inf\{s \geq t + 1 : q_s(X_s) \leq f_s(X_s)\},$$

wobei  $\tau_{t+1}^*$  die optimale Stoppzeit nach dem Zeitpunkt  $t$  ist, basiert. Weiter ist  $f_t$  die (diskontierte) Auszahlungsfunktion für eine Option auf den Preisprozess  $\{X_t\}_{t \in \{0,1,\dots,T\}}$ . In der Tat gilt in Analogie zu (1.13)

$$q_t(x) = \mathbf{E}\{f_{\tau_{t+1}^*}(X_{\tau_{t+1}^*}) | X_t = x\}, \quad \text{wobei} \quad \tau_{t+1}^* = \inf\{s \geq t + 1 : q_s(X_s) \leq f_s(X_s)\}$$

und folglich durch Verwendung der Markov-Eigenschaft von  $\{X_s\}_{s=0,\dots,T}$

$$\begin{aligned} q_t(X_t) &= \mathbf{E} \left\{ f_{t+1}(X_{t+1}) \cdot I_{\{q_{t+1}(X_{t+1}) \leq f_{t+1}(X_{t+1})\}} + f_{\tau_{t+2}^*}(X_{\tau_{t+2}^*}) \cdot I_{\{q_{t+1}(X_{t+1}) > f_{t+1}(X_{t+1})\}} \middle| X_t \right\} \\ &= \mathbf{E}\{\mathbf{E}\{\dots | X_0, \dots, X_{t+1}\} | X_0, \dots, X_t\} \\ &= \mathbf{E}\{\max\{f_{t+1}(X_{t+1}), q_{t+1}(X_{t+1})\} | X_t\}. \end{aligned}$$

Eine Verallgemeinerung der Regressionsdarstellungen (1.15) und (1.16) ist gegeben durch die in Egloff (2005) verwendete Regressionsdarstellung:

$$q_t(x) = \mathbf{E}[\theta_{t+1:w}(f, q)(X_{t+1}, \dots, X_{t+w+1}) | X_t = x], \quad (1.17)$$

wobei

$$\begin{aligned} \theta_{t+1:w}(f, q)(X_{t+1}, \dots, X_{t+w+1}) &= \\ &= \sum_{s=t+1}^{t+w+1} f_s(X_s) \cdot I_{\{f_{t+1}(X_{t+1}) < q_{t+1}(X_{t+1}), \dots, f_{s-1}(X_{s-1}) < q_{s-1}(X_{s-1}), f_s(X_s) \geq q_s(X_s)\}} \\ &+ q_{t+w+1}(X_{t+w+1}) \cdot I_{\{f_{t+1}(X_{t+1}) < q_{t+1}(X_{t+1}), \dots, f_{t+w+1}(X_{t+w+1}) < q_{t+w+1}(X_{t+w+1})\}} \end{aligned}$$

und  $w \in \{0, 1, \dots, T - t - 1\}$ . Wie oben bereits erwähnt, entspricht  $w$  der Schrittweite, die die Vorausschau beschreibt. Mit  $w = 0$  erhält man die Darstellung (1.15), und mit  $w = T - t - 1$  die Darstellung (1.16). Deshalb kann (1.17) als eine Art “Interpolation” von (1.15) und (1.16) interpretiert werden.

Um (1.17) zu zeigen, ist zu beachten, dass  $X_t$  ein Markov-Prozess. Mit der Markov-Eigenschaft folgt:

$$\begin{aligned} q_t(X_t) &= \mathbf{E}\{\max\{f_{t+1}(X_{t+1}), q_{t+1}(X_{t+1})\} | X_t\} \\ &= \mathbf{E}\{f_{t+1}(X_{t+1}) \cdot I_{\{f_{t+1}(X_{t+1}) \geq q_{t+1}(X_{t+1})\}} \\ &+ q_{t+1}(X_{t+1}) \cdot I_{\{f_{t+1}(X_{t+1}) < q_{t+1}(X_{t+1})\}} | X_0, \dots, X_t\} \end{aligned}$$

$$\begin{aligned}
&= \mathbf{E} \left\{ f_{t+1}(X_{t+1}) \cdot I_{\{f_{t+1}(X_{t+1}) \geq q_{t+1}(X_{t+1})\}} \right. \\
&\quad + \mathbf{E} \{ f_{t+2}(X_{t+2}) \cdot I_{\{f_{t+2}(X_{t+2}) \geq q_{t+2}(X_{t+2})\}} \\
&\quad + q_{t+2}(X_{t+2}) \cdot I_{\{f_{t+2}(X_{t+2}) < q_{t+2}(X_{t+2})\}} | X_0, \dots, X_{t+1} \} \cdot I_{\{f_{t+1}(X_{t+1}) < q_{t+1}(X_{t+1})\}} \\
&\quad \left. | X_0, \dots, X_t \right\} \\
&= \mathbf{E} \left\{ \mathbf{E} \{ f_{t+1}(X_{t+1}) \cdot I_{\{f_{t+1}(X_{t+1}) \geq q_{t+1}(X_{t+1})\}} \right. \\
&\quad + f_{t+2}(X_{t+2}) \cdot I_{\{f_{t+2}(X_{t+2}) \geq q_{t+2}(X_{t+2})\}} \cdot I_{\{f_{t+1}(X_{t+1}) < q_{t+1}(X_{t+1})\}} \\
&\quad + q_{t+2}(X_{t+2}) \cdot I_{\{f_{t+2}(X_{t+2}) < q_{t+2}(X_{t+2})\}} \cdot I_{\{f_{t+1}(X_{t+1}) < q_{t+1}(X_{t+1})\}} | X_0, \dots, X_{t+1} \} \\
&\quad \left. | X_0, \dots, X_t \right\} \\
&= \mathbf{E} \left\{ \mathbf{E} \{ f_{t+1}(X_{t+1}) \cdot I_{\{f_{t+1}(X_{t+1}) \geq q_{t+1}(X_{t+1})\}} \right. \\
&\quad + f_{t+2}(X_{t+2}) \cdot I_{\{f_{t+1}(X_{t+1}) < q_{t+1}(X_{t+1}), f_{t+2}(X_{t+2}) \geq q_{t+2}(X_{t+2})\}} \\
&\quad + q_{t+2}(X_{t+2}) \cdot I_{\{f_{t+1}(X_{t+1}) < q_{t+1}(X_{t+1}), f_{t+2}(X_{t+2}) < q_{t+2}(X_{t+2})\}} | X_0, \dots, X_{t+1} \} \left. | X_0, \dots, X_t \right\} \\
&= \mathbf{E} \left\{ f_{t+1}(X_{t+1}) \cdot I_{\{f_{t+1}(X_{t+1}) \geq q_{t+1}(X_{t+1})\}} \right. \\
&\quad + f_{t+2}(X_{t+2}) \cdot I_{\{f_{t+1}(X_{t+1}) < q_{t+1}(X_{t+1}), f_{t+2}(X_{t+2}) \geq q_{t+2}(X_{t+2})\}} \\
&\quad + q_{t+2}(X_{t+2}) \cdot I_{\{f_{t+1}(X_{t+1}) < q_{t+1}(X_{t+1}), f_{t+2}(X_{t+2}) < q_{t+2}(X_{t+2})\}} \left. | X_0, \dots, X_t \right\} \\
&= \dots \\
&= \mathbf{E} \left\{ f_{t+1}(X_{t+1}) \cdot I_{\{f_{t+1}(X_{t+1}) \geq q_{t+1}(X_{t+1})\}} \right. \\
&\quad + f_{t+2}(X_{t+2}) \cdot I_{\{f_{t+1}(X_{t+1}) < q_{t+1}(X_{t+1}), f_{t+2}(X_{t+2}) \geq q_{t+2}(X_{t+2})\}} \\
&\quad + \dots \\
&\quad + f_{t+w+1}(X_{t+w+1}) \\
&\quad \quad \cdot I_{\{f_{t+1}(X_{t+1}) < q_{t+1}(X_{t+1}), \dots, f_{t+w}(X_{t+w}) < q_{t+w}(X_{t+w}), f_{t+w+1}(X_{t+w+1}) \geq q_{t+w+1}(X_{t+w+1})\}} \\
&\quad + q_{t+w+1}(X_{t+w+1}) \\
&\quad \quad \cdot I_{\{f_{t+1}(X_{t+1}) < q_{t+1}(X_{t+1}), \dots, f_{t+w}(X_{t+w}) < q_{t+w}(X_{t+w}), f_{t+w+1}(X_{t+w+1}) < q_{t+w+1}(X_{t+w+1})\}} \\
&\quad \left. | X_0, \dots, X_t \right\} \\
&= \mathbf{E} \{ \theta_{t+1:w}(f, q)(X_{t+1}, \dots, X_{t+w+1}) | X_t \} .
\end{aligned}$$

In diesem Abschnitt wurde auf die regressionsbasierten Monte-Carlo-Verfahren eingegangen. Es hat sich gezeigt, dass das Bewerten Amerikanischer Optionen (im zeitdiskreten Fall) geschachtelten Regressionsproblemen entspricht (wie beispielsweise beim Verfahren bzw. Algorithmus von Tsitsiklis und Van Roy (1999)). Während in Tsitsiklis und Van Roy (1999) und Longstaff und Schwartz (2001) Verfahren der parametrischen Regressions-schätzung eingesetzt wurden, verwendete Egloff (2005) Verfahren der nichtparametrischen Regressions-schätzung. Damit wurde der Nachteil, der durch die Verfahren der parametrischen Regressions-schätzung hervorgerufen wird (siehe oben), umgangen. Da in Egloff (2005) jedoch die Wahl der Parameter nicht datenabhängig erfolgte, war die Bestimmung der Parameter problematisch, weshalb auch keine Simulationen möglich waren. Im nächsten Abschnitt wird ein Verfahren vorgestellt, bei dem dieser Nachteil umgangen

wird.

### 1.3 Neu erzielte Resultate

In diesem Abschnitt wird auf die Resultate eingegangen, die aus den in dieser Arbeit vorgestellten Verfahren hervorgehen. Hierbei sei erwähnt, dass dieser Arbeit die Arbeiten von Egloff, Kohler und Todorovic (2006) sowie Kohler, Krzyżak und Todorovic (2006) zugrunde liegen. Als erstes wird die Verwendung von regressionsbasierten Monte-Carlo-Verfahren mit Einsatz von nichtparametrischen Methoden zur Bewertung Amerikanischer Optionen motiviert. Anschließend folgt eine Auflistung der erzielten Resultate, und zum Abschluss wird auf die der Arbeit zugrunde liegende Vorgehensweise eingegangen.

Wie im vorherigen Abschnitt bereits erläutert, existieren bei den Verfahren in Tsitsiklis und Van Roy (1999) sowie in Longstaff und Schwartz (2001) aufgrund der verwendeten Verfahren (welche auf Methoden der parametrischen Regressionsschätzung basieren) Nachteile, sobald die Bauart der Kontinuierungswerte nicht linear ist. Bei dem Verfahren in Egloff (2005) werden wiederum nichtparametrische Methoden der Regressionsschätzung zur Schätzung der Kontinuierungswerte verwendet. Da der Schätzer jedoch aufgrund der nicht adaptiv gewählten Parameter (d. h. rein unter Verwendung der gegebenen Daten) nicht implementierbar ist, lässt er sich in der Praxis nicht anwenden.

In dieser Arbeit werden regressionsbasierte Monte-Carlo-Verfahren mit Einsatz von nichtparametrischen Methoden der Regressionsschätzung verwendet, um Amerikanische Optionen in diskreter Zeit zu bewerten. Ein Grund für diese Wahl ist, dass sich durch deren Verwendung nicht nur der Mittelwert der Auszahlung bei (optimalem) Ausüben der Option bestimmen lässt, sondern es könnte beispielsweise die ganze Verteilung des Ausübungspreises geschätzt werden. Dies ist im Vergleich zu bisherigen Verfahren zur Optionspreisberechnung (basierend auf partiellen Differentialgleichungen) ein deutlicher Vorteil. Sowohl partielle Differentialgleichungen als auch baumbasierte Methoden zur Bewertung von Amerikanischen Optionen lassen sich im Niedrigdimensionalen problemlos lösen. Im Hochdimensionalen sowie bei allgemeineren stochastischen Prozessen oder Auszahlungsfunktionen wird die numerische Lösung jedoch sehr kompliziert. Des Weiteren ist bei den Baum-Methoden eine richtige Modellierung der korrelierten Struktur der Underlyings grundsätzlich unmöglich, falls vier oder mehr Underlyings zugrunde liegen. Die regressionsbasierten Methoden mit Einsatz von nichtparametrischen Verfahren der Regressionsschätzung sind auch im Hochdimensionalen anwendbar, was ein weiterer Grund für deren Wahl ist. Auch sind sie viel flexibler als die parametrischen Methoden und außerdem weniger zeitaufwendig als die Baum-Methoden.

Weiter wird in dieser Arbeit die Regressionsdarstellung von Egloff (2005) für die Kontinuierungswerte verwendet. Jedoch werden mit den in dieser Arbeit vorgestellten nichtparametrischen Schätzverfahren die nichtparametrischen Schätzverfahren aus Egloff (2005) erheblich verbessert. Durch die Neugenerierung der zur Regressionsschätzung verwendeten Monte-Carlo-Stichproben in jedem Zeitschritt (was vom Aufwand her

im Vergleich zur Regressionsschätzung vernachlässigbar ist) wird die Berechnung der Schätzer soweit vereinfacht, dass sie ohne Probleme implementierbar sind. Des Weiteren werden in dieser Arbeit moderne Verfahren der multivariaten Regressionsschätzung wie Splines oder neuronale Netze mit einer verdeckten Schicht für die Schätzer verwendet, während in Egloff (2005) stückweise Polynome verwendet werden.

Neben der Vorstellung des neuen Verfahrens enthält diese Arbeit auch eine theoretische Analyse der aus den neuen Verfahren entstehenden Schätzer. Dabei werden Resultate zur Konsistenz und zur Konvergenzgeschwindigkeit (unter Glattheitsannahmen an die Kontinuierungswerte) hergeleitet.

Ein weiterer Vorteil des in dieser Arbeit vorgestellten Verfahrens zur Bewertung Amerikanischer Optionen ist, dass für die zugrunde liegenden Preisprozesse nur die Markov-Eigenschaft gefordert wird, und somit dieses Verfahren für alle Preisprozesse, die Markov-Prozesse sind, anwendbar ist. Selbstverständlich können auch Preisprozesse zugrunde gelegt werden, die ursprünglich zwar keine Markov-Prozesse sind, durch geeignete Darstellung aber die Markov-Eigenschaft erlangen können.

Es wird auch gezeigt, dass die aus dieser Arbeit resultierenden Verfahren bei Amerikanischen Optionen (basierend auf Black-Scholes-Modellen) in Simulationen bessere Resultate liefern als die Verfahren in Tsitsiklis und Van Roy (1999) und Longstaff und Schwartz (2001). Dies ist sowohl bei Basket-Optionen der Fall als auch bei Optionen, bei denen die Auszahlung etwas komplizierter ist als bei der Put-Option.

Zusammenfassend lässt sich festhalten, dass in dieser Arbeit ein Monte-Carlo-Algorithmus entwickelt wurde, der in der Lage ist, eine Approximation des Preises (1.10) zu berechnen. Dies ist auch dann möglich, wenn die Option auf einer großen Anzahl korrelierter Underlyings basiert, eine komplizierte Auszahlungsfunktion beinhaltet oder kompliziertere Preisprozesse zugrunde liegen. Die eingeführten Schätzer werden theoretisch hinsichtlich Konsistenz und Konvergenzgeschwindigkeit hin analysiert. Weiter erzielt dieser Algorithmus in Simulationen bessere Resultate als die Algorithmen in Tsitsiklis und Van Roy (1999) und Longstaff und Schwartz (2001).

### **Vorgehensweise in dieser Arbeit**

Im Folgenden wird auf die der Arbeit zugrunde liegende Vorgehensweise eingegangen. In Kapitel 2 wird der rekursive Monte-Carlo-Algorithmus zur Schätzung der Kontinuierungswerte erläutert. Weiter wird ein theoretisches Hauptresultat formuliert und bewiesen, mit dessen Hilfe Resultate zur Konsistenz und zur Konvergenzgeschwindigkeit für die ebenfalls in Kapitel 2 vorgestellten Kleinste-Quadrate-Schätzer (mit fest gewählten Parametern) hergeleitet werden.

Kapitel 3 beschäftigt sich mit einem Spezialfall des theoretischen Hauptresultats aus dem vorherigen Kapitel. Dieser Spezialfall wird mit einem alternativen Zugang zur Analyse des Kleinste-Quadrate-Schätzers, durch den der Beweis vereinfacht wird, verifiziert. Der

alternative Zugang entspricht dabei der Anwendung von Resultaten aus der Theorie der Regressionsschätzung mit Fehlern in der abhängigen Variable.

Da die bisher vorgestellten Kleinste-Quadrate-Schätzer von der Wahl der Parameter abhängen, wird in Kapitel 4 ein Resultat bezüglich der Adaption (d.h. einer rein datenabhängigen Wahl der Parameter) formuliert und bewiesen. Dabei wird die Methode der Unterteilung der Stichprobe verwendet. Anschließend werden adaptive Kleinste-Quadrate-Schätzer definiert, für welche mit Hilfe des vorhergehenden Resultats Resultate bezüglich Konsistenz und Konvergenzgeschwindigkeit hergeleitet werden.

Im fünften und letzten Kapitel wird der in dieser Arbeit vorgestellte Algorithmus mit den Algorithmen in Tsitsiklis und Van Roy (1999) sowie Longstaff und Schwartz (2001) anhand mehrerer (simulierter) praktischer Anwendungsbeispiele verglichen. Zu diesem Zweck wurden insbesondere für die Datenerzeugung sowie für die (adaptive Kleinste-Quadrate-)Schätzergenerierung Programme erstellt. Weiter wurden Programme mit dem in dieser Arbeit vorgestellten Algorithmus sowie den Algorithmen in Tsitsiklis und Van Roy (1999) und Longstaff und Schwartz (2001) erstellt, um Amerikanische Optionen zu bewerten. Die simulierten Anwendungsbeispiele zeigen, dass die in dieser Arbeit neu eingeführten Schätzer nicht nur theoretisch (hinsichtlich Aussagen zur Konsistenz und zur Konvergenzgeschwindigkeit), sondern auch praktisch (hinsichtlich für Anwendungen bei der Optionspreisbewertung relevanter Simulationen) deutlich den Schätzern von Tsitsiklis und Van Roy (1999) und Longstaff und Schwartz (2001) überlegen sind. Zieht man in Betracht, dass die neu eingeführten Schätzer auch den Schätzern aus Egloff (2005) überlegen sind (da diese keinerlei praktische Relevanz haben), so sieht man, dass damit in dieser Arbeit ein wesentlicher Fortschritt bei der Monte-Carlo-Bewertung Amerikanischer Optionen gelungen ist.

# Kapitel 2

## Allgemeine Resultate zur Konsistenz und Konvergenzgeschwindigkeit

In diesem Kapitel werden - im Zusammenhang mit der Schätzung des Wertes einer Amerikanischen Option - Aussagen zur Konsistenz sowie zur Konvergenzgeschwindigkeit eines Kleinste-Quadrate-Schätzers mit festem Parameter hergeleitet. Zuerst wird dazu in Abschnitt 2.1 der Schätzer definiert und ein Resultat hinsichtlich des integrierten quadratischen Fehlers des Schätzers formuliert. Anschließend wird in den Abschnitten 2.2 bzw. 2.3 der Kleinste-Quadrate-Schätzer als der so genannte Kleinste-Quadrate-Splineschätzer bzw. Neuronale-Netze-Schätzer spezifiziert. Mit Hilfe des Resultats aus Abschnitt 2.1 werden für den jeweiligen Schätzer in den Abschnitten 2.2 bzw. 2.3 Resultate zur Konsistenz und Konvergenzgeschwindigkeit erzielt.

### 2.1 Theoretisches Hauptresultat

Wie in Kapitel 1 bereits erwähnt wurde, wird in dieser Arbeit die Regressionsdarstellung von Egloff (2005) für die Kontinuierungswerte  $q_t$  verwendet. Der Vollständigkeit halber erfolgt hier noch einmal die explizite Darstellung:

$$q_t(x) = \mathbf{E}[\theta_{t+1:w}(f, q)(X_{t+1}, \dots, X_{t+w+1}) | X_t = x] \quad (2.1)$$

wobei

$$\begin{aligned} \theta_{t+1:w}(f, q)(X_{t+1}, \dots, X_{t+w+1}) = & \\ & \sum_{s=t+1}^{t+w+1} f_s(X_s) \cdot I_{\{f_{t+1}(X_{t+1}) < q_{t+1}(X_{t+1}), \dots, f_{s-1}(X_{s-1}) < q_{s-1}(X_{s-1}), f_s(X_s) \geq q_s(X_s)\}} \\ & + q_{t+w+1}(X_{t+w+1}) \cdot I_{\{f_{t+1}(X_{t+1}) < q_{t+1}(X_{t+1}), \dots, f_{t+w+1}(X_{t+w+1}) < q_{t+w+1}(X_{t+w+1})\}} \end{aligned} \quad (2.2)$$

und  $w \in \{0, 1, \dots, T - t - 1\}$ .

Leider ist im Allgemeinen der bedingte Erwartungswert in (2.1) in den Anwendungen nicht berechenbar. Die Grundidee von regressionsbasierten Monte-Carlo-Methoden zur

Bewertung von Amerikanischen Optionen ist, die Regressionsschätzer rekursiv auf eine künstlich erzeugte Stichprobe

$$(X_t, \theta_{t+1:w}(f, \hat{q})(X_{t+1}, \dots, X_{t+w+1})) ,$$

die so genannte Monte-Carlo-Stichprobe, anzuwenden, um Schätzer  $\hat{q}_t$  von  $q_t$  zu konstruieren.

### Ein rekursiver Algorithmus zur Schätzung der Kontinuierungswerte

Im Folgenden wird ein Algorithmus beschrieben, der die Kontinuierungswerte  $q_t$  rekursiv durch einen Kleinste-Quadrate-Schätzer schätzt.

Um die Kontinuierungswerte rekursiv schätzen zu können, werden künstlich unabhängige Markov-Prozesse

$$\{X_{i,t}^{(l)}\}_{t=0,\dots,T} \quad (l = 0, 1, \dots, T-1, i = 1, \dots, n) ,$$

die identisch verteilt sind wie  $\{X_t\}_{t=0,\dots,T}$ , erzeugt. Diese Monte-Carlo-Stichprobe wird genutzt, um rekursiv Daten zu generieren und um durch Nutzung der Regressionsdarstellung in (2.1)  $q_t$  zu schätzen.

Gestartet wird mit

$$\hat{q}_{n,T}(x) = 0 \quad (x \in \mathbb{R}^d) .$$

Halte  $t \in \{0, 1, \dots, T-1\}$  fest. Sind die Schätzer  $\hat{q}_{n,t+1}, \dots, \hat{q}_{n,t+w+1}$  von  $q_{t+1}, \dots, q_{t+w+1}$  gegeben, so schätzt man

$$q_t(x) = \mathbf{E}[\theta_{t+1:w}(f, q)(X_{t+1}, \dots, X_{t+w+1}) | X_t = x]$$

durch Anwendung eines Kleinste-Quadrate-Schätzers auf eine “approximative” Stichprobe von

$$(X_t, \theta_{t+1:w}(f, q)(X_{t+1}, \dots, X_{t+w+1})) .$$

Mit der Notation

$$\hat{Y}_{i,t}^{(t)} = \theta_{t+1:w}(f, \hat{q}_n)(X_{i,t+1}^{(t)}, \dots, X_{i,t+w+1}^{(t)}) ,$$

wobei hier die Abhängigkeit bei  $\hat{Y}_{i,t}^{(t)}$  von  $\hat{q}_{n,s}$  mit  $s \in \{t+1, \dots, t+w+1\}$  nicht sofort ersichtlich ist, ist diese “approximative” Stichprobe gegeben durch

$$\{(X_{i,t}^{(t)}, \hat{Y}_{i,t}^{(t)}) : i = 1, \dots, n\} . \quad (2.3)$$

Es ist zu beachten, dass diese Stichprobe von der  $t$ -ten Stichprobe von  $\{X_s\}_{s=0,\dots,T}$  und von  $\hat{q}_{n,t+1}, \dots, \hat{q}_{n,t+w+1}$  abhängt, d. h. für jeden Zeitschritt  $t$  wird eine neue Stichprobe von dem stochastischen Prozess  $\{X_s\}_{s=0,\dots,T}$  genutzt, um die Daten in (2.3) zu definieren.

Bevor der Kleinste-Quadrate-Schätzer  $\hat{q}_{n,t}$  von  $q_t$  mit Hilfe von (2.3) definiert wird (siehe hierzu Kapitel 4.2), wird in einer ersten Betrachtungsweise ein Kleinste-Quadrate-Schätzer für einen fest gewählten Parameter definiert. Konkret bedeutet dies, dass



in diesem Kapitel keine automatische Wahl der Parameter stattfindet und daher alle Aussagen nur für Schätzer mit fest gewählten Parametern gelten, d. h. die Parameter werden unabhängig von den Daten gewählt. Nichtsdestotrotz wird im Folgenden aus notationellen Gründen der Kleinste-Quadrate-Schätzer in Abhängigkeit vom fest gewählten Parameter sowie vom Parameter  $w$  eingeführt. Wegen der zusätzlichen Schreibweise des Kleinste-Quadrate-Schätzers in Abhängigkeit vom fest gewählten Parameter sowie von  $w$  entspricht diese Notation des Schätzers den Notationen in Kapitel 4.

Um einen Schätzer  $q_{n,t}^{w,p}$  mit fest gewählten Parametern  $w$  und  $p$  von  $q_t$  definieren zu können, wird zuerst (analog zu oben) mit

$$q_{n,T}^{w,p}(x) = 0 \quad (x \in \mathbb{R}^d)$$

gestartet. Außerdem seien für festgehaltenes  $t \in \{0, 1, \dots, T-1\}$  die Schätzer  $q_{n,t+1}^{w,p}, \dots, q_{n,t+w+1}^{w,p}$  von  $q_{t+1}, \dots, q_{t+w+1}$  gegeben. Des Weiteren wird in der Definition von  $\hat{Y}_{i,t}^{(t)}$  durch  $\hat{q}_{n,s}^{w,p}$  (mit  $s \in \{t+1, \dots, t+w+1\}$ ) ersetzt, und es folgt die neue Notation

$$Y_{i,t}^{(t)} = \theta_{t+1:w}(f, q_n^{w,p})(X_{i,t+1}^{(t)}, \dots, X_{i,t+w+1}^{(t)}) . \quad (2.4)$$

Als nächstes wird in (2.3)  $\hat{Y}_{i,t}^{(t)}$  durch  $Y_{i,t}^{(t)}$  ersetzt, wobei die dadurch neue “approximative” Stichprobe durch

$$\{(X_{i,t}^{(t)}, Y_{i,t}^{(t)}) : i = 1, \dots, n\} \quad (2.5)$$

gegeben ist. Ausgehend von dieser Stichprobe (2.5) wird der Kleinste-Quadrate-Schätzer  $q_{n,t}^{w,p}$  von  $q_t$  definiert. Zuvor wird noch der in der Definition benötigte so genannte Stützungsoperator eingeführt. Sei  $\beta_n > 0, z \in \mathbb{R}$  und sei

$$T_{\beta_n} z = \max\{-\beta_n, \min\{\beta_n, z\}\}.$$

Hierbei bezeichnet  $T_{\beta_n} z$  den Stützungsoperator am (Schwellen-)Wert  $\beta_n$ .

Bezeichne  $\mathcal{P}_n$  eine nichtleere Parametermenge, und sei  $p \in \mathcal{P}_n$  Parameter dieser Parametermenge. Des Weiteren existiere für jedes  $p \in \mathcal{P}_n$  ein Schätzer  $q_{n,t}^{w,p}$  von  $q_t$ , der gegeben ist durch

$$\tilde{q}_{n,t}^{w,p}(\cdot) = \arg \min_{f \in \mathcal{F}_{n,p}} \left( \frac{1}{n} \sum_{i=1}^n |f(X_{i,t}^{(t)}) - Y_{i,t}^{(t)}|^2 \right) \quad (2.6)$$

und

$$q_{n,t}^{w,p}(x) = T_{\beta_n} \tilde{q}_{n,t}^{w,p}(x) \quad (x \in \mathbb{R}^d), \quad (2.7)$$

wobei (für jedes  $p \in \mathcal{P}_n$ )  $\mathcal{F}_{n,p}$  ein Raum von Funktionen  $f : \mathbb{R}^d \rightarrow \mathbb{R}$  ist.

Aus Gründen der Einfachheit wird in (2.6) die Existenz des Minimums angenommen, wobei keine Eindeutigkeit gefordert wird. Falls das Minimum nicht eindeutig definiert ist,

kann als Kleinste-Quadrate-Schätzer jede Funktion gewählt werden, die das Minimum erreicht, wobei für diese Funktion ebenfalls die theoretischen Resultate gelten. Wird die Existenz des Minimums nicht angenommen, dann kann der Kleinste-Quadrate-Schätzer definiert werden, indem man in Abhängigkeit von der Stichprobe (2.5) diejenige Funktion aus dem zugrunde liegenden Funktionenraum  $\mathcal{F}_{n,p}$  auswählt, für die gilt (siehe dazu Kohler (2000))

$$\frac{1}{n} \sum_{i=1}^n |\tilde{q}_{n,t}^{w,p}(X_{i,t}) - Y_{i,t}|^2 = \min_{f \in \mathcal{F}_{n,p}} \frac{1}{n} \sum_{i=1}^n |f(X_{i,t}) - Y_{i,t}|^2 + o(n^{-1}).$$

Um das Hauptresultat in diesem Abschnitt formulieren zu können, werden so genannte Überdeckungszahlen benötigt, welche im Folgenden eingeführt werden.

Seien  $x_1, \dots, x_n \in \mathbb{R}^d$ , und setze  $x_1^n = (x_1, \dots, x_n)$ . Definiere den Abstand  $d_2(f, g)$  zwischen  $f, g : \mathbb{R}^d \rightarrow \mathbb{R}$  durch

$$d_2(f, g) = \left( \frac{1}{n} \sum_{i=1}^n |f(x_i) - g(x_i)|^2 \right)^{1/2}.$$

Eine  $\epsilon$ -Überdeckung von  $\mathcal{F}$  (hinsichtlich des Abstandes  $d_2$ ) ist eine Menge von Funktionen  $f_1, \dots, f_\kappa : \mathbb{R}^d \rightarrow \mathbb{R}$  mit der Eigenschaft

$$\min_{1 \leq j \leq \kappa} d_2(f, f_j) < \epsilon \quad \text{für alle } f \in \mathcal{F}.$$

Sei  $\mathcal{N}_2(\epsilon, \mathcal{F}, x_1^n)$  die Größe  $\kappa$  der kleinsten  $\epsilon$ -Überdeckung von  $\mathcal{F}$  (hinsichtlich des Abstandes  $d_2$ ), und setze  $\mathcal{N}_2(\epsilon, \mathcal{F}, x_1^n) = \infty$ , falls keine  $\epsilon$ -Überdeckung von  $\mathcal{F}$  mit endlicher Größe existiert.  $\mathcal{N}_2(\epsilon, \mathcal{F}, x_1^n)$  ist die so genannte  $L_2$ - $\epsilon$ -Überdeckungszahl von  $\mathcal{F}$  auf  $x_1^n$ . Der Einfachheit halber wird sie im Folgenden lediglich als Überdeckungszahl bezeichnet.

Des Weiteren sind die unten aufgeführten theoretischen Resultate in der so genannten  $O_{\mathbf{P}}$ -Notation verfasst. Unter dieser Notation versteht man, dass  $a_n = O_{\mathbf{P}}(b_n)$  ist, falls  $\limsup_{n \rightarrow \infty} \mathbf{P}(a_n > c \cdot b_n) = 0$  für eine endliche Konstante  $c$  gilt. Da nun alle benötigten Notationen bekannt sind, folgt das erste Resultat, welches eine Abschätzung für den integrierten quadratischen Fehler des (Regressions-)Schätzers  $q_{n,t}^{w,p}$  gibt, der durch Minimierung des empirischen  $L_2$ -Risikos über einen datenunabhängigen Funktionenraum und anschließender Stützung definiert ist.

**Satz 2.1** *Sei  $\beta_n \geq L > 0$ . Nehme an, dass  $X_0, X_1, \dots, X_T$  ein  $\mathbb{R}^d$ -wertiger Markov-Prozess ist und dass die diskontierte Auszahlungsfunktion  $f_t$  absolut beschränkt ist durch  $L$ . Sei  $q_t$  der Kontinuierungswert definiert durch (1.12). Des Weiteren sei  $w \in \{0, 1, \dots, T-t-1\}$  fest. Der Schätzer  $q_{n,t}^{w,p}$  (mit  $p \in \mathcal{P}_n$  fest) sei definiert durch (2.6) und (2.7) mit Daten definiert in (2.5). Sei  $c_1 > 0$ . Angenommen, es existiert ein  $\delta_{n,t} > 0$  mit den Eigenschaften*

$$\delta_{n,t} \rightarrow 0 \quad (n \rightarrow \infty), \quad \frac{n \cdot \delta_{n,t}}{\beta_n^2} \rightarrow \infty \quad (n \rightarrow \infty)$$

und

$$c_1 \sqrt{n} \delta \geq \int_{\delta/(2^9 \sigma_0)}^{\sqrt{\delta}} \left( \log \mathcal{N}_2 \left( u, \left\{ f - g : f \in \mathcal{F}_{n,p}, \right. \right. \right. \\ \left. \left. \left. \frac{1}{n} \sum_{i=1}^n |f(x_i) - g(x_i)|^2 \leq \delta \right\}, x_1^n \right) \right)^{1/2} du \quad (2.8)$$

für alle  $\delta \geq \delta_{n,t}$  und alle  $g \in \mathcal{F}_{n,p}$ , sowie dass für alle  $x_1, \dots, x_n \in \mathbb{R}^d$  und alle  $\delta \geq 2\beta_n^2 \delta_{n,t}$

$$\frac{0.5 \sqrt{n} \delta}{768 \sqrt{2} \beta_n^2} \geq \int_{0.5\delta/(128\beta_n^2)}^{\sqrt{\delta}} \left( \log \mathcal{N}_2 \left( \frac{u}{4\beta_n}, \left\{ f - q_t : f \in T_{\beta_n} \mathcal{F}_{n,p}, \right. \right. \right. \\ \left. \left. \left. \frac{1}{n} \sum_{i=1}^n |f(x_i) - q_t(x_i)|^2 \leq \frac{\delta}{\beta_n^2} \right\}, x_1^n \right) \right)^{1/2} du. \quad (2.9)$$

Dann gilt

$$\int |q_{n,t}^{w,p}(x) - q_t(x)|^2 \mathbf{P}_{X_t}(dx) \\ = O_{\mathbf{P}} \left( \sum_{s=t}^{T-1} \delta_{n,s} + \max_{s \in \{t, t+1, \dots, T-1\}} \inf_{\substack{f \in \mathcal{F}_{n,p} \\ \|f\|_{\infty} \leq \beta_n + 1}} \int |f(x) - q_s(x)|^2 \mathbf{P}_{X_s}(dx) \right)$$

für alle  $t \in \{0, \dots, T\}$ .

**Bemerkung 2.2** Der rechte Term in letzter Gleichung

$$\inf_{\substack{f \in \mathcal{F}_{n,p} \\ \|f\|_{\infty} \leq \beta_n + 1}} \int |f(x) - q_s(x)|^2 \mathbf{P}_{X_s}(dx)$$

ist der so genannte Approximationsfehler, der angibt, wie gut die Regressionsfunktion  $q_s$  in  $L_2(\mathbf{P}_{X_s})$  durch Funktionen aus  $\mathcal{F}_{n,p}$  mit  $\|f\|_{\infty} \leq \beta_n + 1$  approximiert werden kann. Des Weiteren wird Satz 2.1 in den Abschnitten 2.2 und 2.3 zur Herleitung von Konsistenz- sowie Konvergenzgeschwindigkeitsaussagen verwendet.

**Beweis von Satz 2.1:** Bevor auf den Beweis eingegangen wird, ist zu beachten, dass die Beschränktheit der diskontierten Auszahlungsfunktion  $f_t$  durch  $L |q_t(x)| \leq L$  für alle  $x \in \mathbb{R}^d$  impliziert. Des Weiteren impliziert die Definition von  $q_{n,t}^{w,p}$   $|q_{n,t}^{w,p}(x)| \leq \beta_n$  für alle  $x \in \mathbb{R}^d$ .

Im Folgenden wird gezeigt, dass

$$\int |q_{n,s}^{w,p}(x) - q_s(x)|^2 \mathbf{P}_{X_s}(dx) \\ = O_{\mathbf{P}} \left( \sum_{t=s}^{T-1} \delta_{n,t} + \max_{t \in \{s, s+1, \dots, T-1\}} \inf_{\substack{f \in \mathcal{F}_{n,p} \\ \|f\|_{\infty} \leq \beta_n + 1}} \int |f(x) - q_t(x)|^2 \mathbf{P}_{X_t}(dx) \right) \quad (2.10)$$

für alle  $s \in \{0, \dots, T\}$  gilt.

Für  $s = T$  gilt  $q_{n,T}^{w,p}(x) = 0 = q_T(x)$ , und somit ist die Aussage trivial. Sei  $t < T$ , und nehme an, dass die Aussage für jedes  $s \in \{t+1, \dots, T-1\}$  gilt. Durch Induktion genügt es, (2.10) für  $s = t$  zu zeigen.

Sei

$$\mathcal{D}_{n,t} = \{X_{j,s}^{(r)} : r \geq t+1, s \in \{0, \dots, T-1\}, j \in \{1, \dots, n\}\},$$

und setze

$$q_t^w(X_{1,t}^{(t)}) = \mathbf{E}^*\{Y_{1,t}^{(t)} | X_{1,t} = X_{1,t}^{(t)}\},$$

wobei in  $\mathbf{E}^*\{\cdot | X_{1,t} = X_{1,t}^{(t)}\}$  der bedingte Erwartungswert hinsichtlich  $\mathcal{D}_{n,t}$  genommen wird. D. h. alle Daten, die zur Konstruktion von  $q_{n,t+1}^{w,p}, \dots, q_{n,t+w+1}^{w,p}$  genutzt wurden, sind fest.

Es gilt nun, (2.10) für  $s = t$  zu zeigen. Dazu wird die Fehler-Zerlegung des  $L_2$ -Fehlers betrachtet:

$$\begin{aligned} & \int |q_{n,t}^{w,p}(x) - q_t(x)|^2 \mathbf{P}_{X_t}(dx) \\ &= \int |q_{n,t}^{w,p}(x) - q_t(x)|^2 \mathbf{P}_{X_t}(dx) - 3 \frac{1}{n} \sum_{i=1}^n |q_{n,t}^{w,p}(X_{i,t}^{(t)}) - q_t(X_{i,t}^{(t)})|^2 \\ & \quad + \frac{3}{n} \sum_{i=1}^n |q_{n,t}^{w,p}(X_{i,t}^{(t)}) - q_t(X_{i,t}^{(t)})|^2 - \frac{3}{n} \sum_{i=1}^n |\tilde{q}_{n,t}^{w,p}(X_{i,t}^{(t)}) - q_t(X_{i,t}^{(t)})|^2 \\ & \quad + \frac{3}{n} \sum_{i=1}^n |\tilde{q}_{n,t}^{w,p}(X_{i,t}^{(t)}) - q_t(X_{i,t}^{(t)})|^2 - \frac{6}{n} \sum_{i=1}^n |\tilde{q}_{n,t}^{w,p}(X_{i,t}^{(t)}) - q_t^w(X_{i,t}^{(t)})|^2 \\ & \quad + \frac{6}{n} \sum_{i=1}^n |\tilde{q}_{n,t}^{w,p}(X_{i,t}^{(t)}) - q_t^w(X_{i,t}^{(t)})|^2 \\ &= \sum_{j=1}^4 T_{j,n}. \end{aligned}$$

Es sei hier erwähnt, dass der Beweis vollständig ist, sobald einmal gezeigt wurde, dass für  $j \in \{1, 2, 3, 4\}$  gilt:

$$T_{j,n} = O_{\mathbf{P}} \left( \sum_{s=t}^{T-1} \delta_{n,s} + \max_{s \in \{t, t+1, \dots, T-1\}} \inf_{\substack{f \in \mathcal{F}_{n,p} \\ \|f\|_{\infty} \leq L+1}} \int |f(x) - q_s(x)|^2 \mathbf{P}_{X_s}(dx) \right). \quad (2.11)$$

Im ersten Schritt des Beweises wird  $T_{2,n}$  betrachtet. Da  $q_t$  absolut beschränkt ist durch  $L$ , gilt

$$T_{2,n} \leq 0,$$

was wiederum  $T_{2,n} = O_{\mathbf{P}}(0)$  und (2.11) mit  $j = 2$  impliziert.

Im zweiten Schritt des Beweises wird  $T_{1,n}$  betrachtet. Es ist zu beachten, dass

$$\int |q_{n,t}^{w,p}(x) - q_t(x)|^2 \mathbf{P}_{X_t}(dx) - 3 \frac{1}{n} \sum_{i=1}^n |q_{n,t}^{w,p}(X_{i,t}^{(t)}) - q_t(X_{i,t}^{(t)})|^2 > \delta_{n,t}$$

äquivalent ist zu

$$\begin{aligned} & 2 \int |q_{n,t}^{w,p}(x) - q_t(x)|^2 \mathbf{P}_{X_t}(dx) - 2 \frac{1}{n} \sum_{i=1}^n |q_{n,t}^{w,p}(X_{i,t}^{(t)}) - q_t(X_{i,t}^{(t)})|^2 \\ & > \int |q_{n,t}^{w,p}(x) - q_t(x)|^2 \mathbf{P}_{X_t}(dx) + \frac{1}{n} \sum_{i=1}^n |q_{n,t}^{w,p}(X_{i,t}^{(t)}) - q_t(X_{i,t}^{(t)})|^2 + \delta_{n,t}, \end{aligned}$$

was wiederum äquivalent ist zu

$$\frac{\int |q_{n,t}^{w,p}(x) - q_t(x)|^2 \mathbf{P}_{X_t}(dx) - \frac{1}{n} \sum_{i=1}^n |q_{n,t}^{w,p}(X_{i,t}^{(t)}) - q_t(X_{i,t}^{(t)})|^2}{\delta_{n,t} + \int |q_{n,t}^{w,p}(x) - q_t(x)|^2 \mathbf{P}_{X_t}(dx) + \frac{1}{n} \sum_{i=1}^n |q_{n,t}^{w,p}(X_{i,t}^{(t)}) - q_t(X_{i,t}^{(t)})|^2} > \frac{1}{2},$$

und dass die letzte Ungleichung die Aussage

$$\frac{|\int |q_{n,t}^{w,p}(x) - q_t(x)|^2 \mathbf{P}_{X_t}(dx) - \frac{1}{n} \sum_{i=1}^n |q_{n,t}^{w,p}(X_{i,t}^{(t)}) - q_t(X_{i,t}^{(t)})|^2|}{\delta_{n,t} + \int |q_{n,t}^{w,p}(x) - q_t(x)|^2 \mathbf{P}_{X_t}(dx) + \frac{1}{n} \sum_{i=1}^n |q_{n,t}^{w,p}(X_{i,t}^{(t)}) - q_t(X_{i,t}^{(t)})|^2} > \frac{1}{2}$$

impliziert. Unter Berücksichtigung der letzten vier Beziehungen folgt weiter:

$$\begin{aligned} & \mathbf{P} \left\{ \int |q_{n,t}^{w,p}(x) - q_t(x)|^2 \mathbf{P}_{X_t}(dx) - 3 \frac{1}{n} \sum_{i=1}^n |q_{n,t}^{w,p}(X_{i,t}^{(t)}) - q_t(X_{i,t}^{(t)})|^2 > \delta_{n,t} \right\} \\ & \leq \mathbf{P} \left\{ \exists f \in T_{\beta_n} \mathcal{F}_{n,p} : \int |f(x) - q_t(x)|^2 \mathbf{P}_{X_t}(dx) - 3 \frac{1}{n} \sum_{i=1}^n |f(X_{i,t}^{(t)}) - q_t(X_{i,t}^{(t)})|^2 > \delta_{n,t} \right\} \\ & \leq \mathbf{P} \left\{ \sup_{f \in T_{\beta_n} \mathcal{F}_{n,p}} \frac{|\int |f(x) - q_t(x)|^2 \mathbf{P}_{X_t}(dx) - \frac{1}{n} \sum_{i=1}^n |f(X_{i,t}^{(t)}) - q_t(X_{i,t}^{(t)})|^2|}{\delta_{n,t} + \int |f(x) - q_t(x)|^2 \mathbf{P}_{X_t}(dx) + \frac{1}{n} \sum_{i=1}^n |f(X_{i,t}^{(t)}) - q_t(X_{i,t}^{(t)})|^2} > \frac{1}{2} \right\}. \end{aligned}$$

Die letzte Wahrscheinlichkeit kann mit Hilfe von Lemma A.8 weiter abgeschätzt werden. Dazu setzt man

$$\epsilon = 0.5, \quad \alpha = \delta_{n,t}, \quad B = \beta_n$$

und beachtet, dass mit diesen Notationen einerseits die Voraussetzung (A.4) durch die Annahme (2.9) erfüllt ist sowie andererseits die Voraussetzung (A.3) für  $n$  hinreichend

groß durch die Annahme  $n \cdot \delta_{n,t}/\beta_n^2 \rightarrow \infty$  impliziert wird. Weiter ist noch zu beachten, dass  $|q_t(x)| \leq L$  und  $|q_{n,t}^{w,p}(x)| \leq \beta_n$  für alle  $x \in \mathbb{R}^d$  gilt sowie dass der zu betrachtende Funktionenraum

$$T_{\beta_n} \mathcal{F}_{n,p} = \{T_{\beta_n} \mathcal{F}_{n,p} : f \in \mathcal{F}_{n,p}\}$$

beschränkt ist durch  $\beta_n$ . Die Anwendung von Lemma A.8 liefert nun für  $n$  genügend groß:

$$\begin{aligned} & \mathbf{P} \left\{ \sup_{f \in T_{\beta_n} \mathcal{F}_{n,p}} \frac{|\int |f(x) - q_t(x)|^2 \mathbf{P}_{X_t}(dx) - \frac{1}{n} \sum_{i=1}^n |f(X_{i,t}^{(t)}) - q_t(X_{i,t}^{(t)})|^2|}{\delta_{n,t} + \int |f(x) - q_t(x)|^2 \mathbf{P}_{X_t}(dx) + \frac{1}{n} \sum_{i=1}^n |f(X_{i,t}^{(t)}) - q_t(X_{i,t}^{(t)})|^2} > \frac{1}{2} \right\} \\ & \leq 15 \exp \left( -\frac{n \cdot \delta_{n,t} \cdot 1/4}{512 \cdot 2304 \beta_n^2} \right) \\ & \rightarrow 0 \quad (n \rightarrow \infty), \end{aligned}$$

da nach Voraussetzung  $n \cdot \delta_{n,t}/\beta_n^2 \rightarrow \infty$  für  $n$  gegen unendlich gilt. Insgesamt erhält man für  $T_{1,n}$

$$\mathbf{P}\{T_{1,n} > \delta_{n,t}\} \rightarrow 0 \quad (n \rightarrow \infty),$$

woraus wiederum  $T_{1,n} = O_{\mathbf{P}}(\delta_{n,t})$  folgt. Dies impliziert (2.11) mit  $j = 1$ .

Im dritten Schritt des Beweises wird  $T_{3,n}$  betrachtet. Mit der Beziehung  $a^2 = (a-b+b)^2 \leq 2(a-b)^2 + 2b^2$  folgt für  $T_{3,n}$ :

$$\begin{aligned} T_{3,n} & \leq \frac{6}{n} \sum_{i=1}^n |q_t^w(X_{i,t}^{(t)}) - q_t(X_{i,t}^{(t)})|^2 \\ & = \frac{6}{n} \sum_{i=1}^n |q_t^w(X_{i,t}^{(t)}) - q_t(X_{i,t}^{(t)})|^2 - 18 \int |q_t^w(x) - q_t(x)|^2 \mathbf{P}_{X_t}(dx) \\ & \quad + 18 \int |q_t^w(x) - q_t(x)|^2 \mathbf{P}_{X_t}(dx) \\ & = \sum_{j=5}^6 T_{j,n}. \end{aligned}$$

Im vierten Schritt des Beweises wird  $T_{5,n}$  betrachtet. Konditioniert man  $T_{5,n}$  auf  $\mathcal{D}_{n,t}$ , dann kann analog wie im zweiten Schritt für genügend großes  $n$  mit Lemma A.8 gezeigt werden (wobei hier der Funktionenraum nur eine einzige Funktion enthält, und damit (A.4) trivialerweise erfüllt ist):

$$\begin{aligned} & \mathbf{P}\{T_{5,n} > \delta_{n,t} \mid \mathcal{D}_{n,t}\} \\ & \leq 15 \exp \left( -\frac{n \cdot \delta_{n,t} \cdot 1/4}{512 \cdot 2304 \beta_n^2} \right) \\ & \rightarrow 0 \quad (n \rightarrow \infty), \end{aligned}$$

was wiederum  $T_{5,n} = O_{\mathbf{P}}(\delta_{n,t})$  impliziert.

Im fünften Schritt des Beweises wird  $T_{6,n}$  betrachtet. Proposition 6.4 in Egloff (2005), welche in diesem Kontext besagt

$$\left(|q_t^w - q_t|\right)^{1/2} \leq \sum_{s=t+1}^{t+w+1} \left(|q_{n,s}^{w,p} - q_s|\right)^{1/2}$$

impliziert

$$\left\{ \int |q_t^w(x) - q_t(x)|^2 \mathbf{P}_{X_t}(dx) \right\}^{1/2} \leq \sum_{s=t+1}^{t+w+1} \left\{ \int |q_{n,s}^{w,p}(x) - q_s(x)|^2 \mathbf{P}_{X_s}(dx) \right\}^{1/2}. \quad (2.12)$$

Die Anwendung von (2.12) auf  $T_{6,n}$  führt zu

$$\begin{aligned} T_{6,n} &= 18 \int |q_t^w(x) - q_t(x)|^2 \mathbf{P}_{X_t}(dx) \\ &\leq 18 \left( \sum_{s=t+1}^{t+w+1} \left( \int |q_{n,s}^{w,p}(x) - q_s(x)|^2 \mathbf{P}_{X_s}(dx) \right)^{1/2} \right)^2 \\ &\leq 18(w+1)^2 \cdot \sum_{s=t+1}^{t+w+1} \int |q_{n,s}^{w,p}(x) - q_s(x)|^2 \mathbf{P}_{X_s}(dx), \end{aligned}$$

und mit (2.10) für  $s \geq t+1$  erhält man

$$T_{6,n} = O_{\mathbf{P}} \left( \sum_{s=t+1}^{T-1} \delta_{n,s} + \max_{s \in \{t+1, \dots, T-1\}} \inf_{\substack{f \in \mathcal{F}_{n,p} \\ \|f\|_{\infty} \leq \beta_{n+1}}} \int |f(x) - q_s(x)|^2 \mathbf{P}_{X_s}(dx) \right).$$

Insgesamt implizieren die letzten drei Schritte (2.11) mit  $j = 3$ .

Im sechsten Schritt des Beweises wird  $T_{4,n}$  betrachtet. Hierbei findet Lemma A.1 Anwendung (mit  $\bar{Y}_i = Y_i = Y_{i,t}^{(t)}$ , die konditioniert auf  $\mathcal{D}_{n,t}$  betrachtet werden). Da nach Definition  $Y_{i,t}^{(t)}$  beschränkt ist, ist die Voraussetzung (A.1) in Lemma A.1 erfüllt. Setzt man

$$\delta_n = \delta_{n,t}$$

in Lemma A.1, dann ist die Voraussetzung (A.2) durch die Annahme (2.8) erfüllt. Weiter gilt  $\delta_n \rightarrow 0$  und  $n \cdot \delta_n \rightarrow \infty$  für  $n$  gegen unendlich. Wird  $T_{4,n}$  auf  $\mathcal{D}_{n,t}$  konditioniert, dann gilt mit Aussage von Lemma A.1:

$$\begin{aligned} &\mathbf{P} \left\{ \frac{1}{n} \sum_{i=1}^n |\tilde{q}_{n,t}^{w,p}(X_{i,t}^{(t)}) - q_t^w(X_{i,t}^{(t)})|^2 \right. \\ &\quad \left. > \text{const} \cdot \left( \delta_{n,t} + \inf_{f \in \mathcal{F}_{n,p}} \frac{1}{n} \sum_{i=1}^n |f(X_{i,t}^{(t)}) - q_t^w(X_{i,t}^{(t)})|^2 \right) \middle| \mathcal{D}_{n,t} \right\} \\ &\longrightarrow 0 \quad (n \rightarrow \infty). \end{aligned}$$

Um Lemma A.1 anwenden zu können, wird auf  $X_i = X_{i,t}^{(t)}$  konditioniert, und es ist  $Y_i = Y_{i,t}^{(t)} = \bar{Y}_i$  aufgrund der Definition von  $Y_{i,t}^{(t)}$  und  $\bar{Y}_i$ . Aus diesem Grund verschwindet der Messfehler  $\frac{1}{n} \sum_{i=1}^n |Y_i - Y_{i,t}^{(t)}|^2$  in Lemma A.1. Die obige Aussage impliziert

$$T_{4,n} = O_{\mathbf{P}} \left( \delta_{n,t} + \inf_{f \in \mathcal{F}_{n,p}} \frac{1}{n} \sum_{i=1}^n |f(X_{i,t}^{(t)}) - q_t^w(X_{i,t}^{(t)})|^2 \right).$$

Im siebten Schritt des Beweises wird der empirische  $L_2$ -Fehler

$$\inf_{f \in \mathcal{F}_{n,p}} \frac{1}{n} \sum_{i=1}^n |f(X_{i,t}^{(t)}) - q_t^w(X_{i,t}^{(t)})|^2$$

in obiger Gleichung betrachtet. Es gilt:

$$\begin{aligned} & \inf_{f \in \mathcal{F}_{n,p}} \frac{1}{n} \sum_{i=1}^n |f(X_{i,t}^{(t)}) - q_t^w(X_{i,t}^{(t)})|^2 \\ & \leq \inf_{\substack{f \in \mathcal{F}_{n,p} \\ \|f\|_{\infty} \leq \beta_n + 1}} \left( \frac{1}{n} \sum_{i=1}^n |f(X_{i,t}^{(t)}) - q_t^w(X_{i,t}^{(t)})|^2 - 3 \int |f(x) - q_t^w(x)|^2 \mathbf{P}_{X_t}(dx) \right. \\ & \quad \left. + 3 \int |f(x) - q_t^w(x)|^2 \mathbf{P}_{X_t}(dx) \right) \\ & \leq \sup_{\substack{f \in \mathcal{F}_{n,p} \\ \|f\|_{\infty} \leq \beta_n + 1}} \left( \frac{1}{n} \sum_{i=1}^n |f(X_{i,t}^{(t)}) - q_t^w(X_{i,t}^{(t)})|^2 - 3 \int |f(x) - q_t^w(x)|^2 \mathbf{P}_{X_t}(dx) \right) \\ & \quad + 3 \inf_{\substack{f \in \mathcal{F}_{n,p} \\ \|f\|_{\infty} \leq \beta_n + 1}} \int |f(x) - q_t^w(x)|^2 \mathbf{P}_{X_t}(dx) \\ & = \sum_{j=7}^8 T_{j,n}. \end{aligned}$$

Im achten Schritt des Beweises wird  $T_{7,n}$  betrachtet. Man beachte hierbei, dass  $|f(x)| \leq \beta_n + 1$  und  $|q_t^w(x)| \leq \beta_n$  für alle  $x \in \mathbb{R}^d$  gilt. Nach Voraussetzung gilt  $\beta_n \geq L$ . Analog wie im zweiten Schritt kann für genügend großes  $n$  mit Lemma A.8 gezeigt werden, dass für  $T_{7,n}$  – konditioniert auf  $\mathcal{D}_{n,t}$  – wegen

$$\{f \in \mathcal{F}_{n,p} : \|f\|_{\infty} \leq \beta_n + 1\} \subseteq T_{\beta_n + 1} \mathcal{F}_{n,p}$$

gilt:

$$\begin{aligned} & \mathbf{P} \{T_{7,n} > \delta_{n,t} \mid \mathcal{D}_{n,t}\} \\ & \leq \mathbf{P} \left\{ \sup_{f \in T_{\beta_n + 1} \mathcal{F}_{n,p}} \left( \frac{1}{n} \sum_{i=1}^n |f(X_{i,t}^{(t)}) - q_t^w(X_{i,t}^{(t)})|^2 - 3 \int |f(x) - q_t^w(x)|^2 \mathbf{P}_{X_t}(dx) \right) \right. \\ & \quad \left. > \delta_{n,t} \mid \mathcal{D}_{n,t} \right\} \end{aligned}$$



$$\begin{aligned}
 &\leq \mathbf{P} \left\{ \exists f \in T_{\beta_n+1} \mathcal{F}_{n,p} : \right. \\
 &\quad \left. \frac{1}{n} \sum_{i=1}^n |f(X_{i,t}^{(t)}) - q_t^w(X_{i,t}^{(t)})|^2 - 3 \int |f(x) - q_t^w(x)|^2 \mathbf{P}_{X_t}(dx) > \delta_{n,t} \mid \mathcal{D}_{n,t} \right\} \\
 &\rightarrow 0 \quad (n \rightarrow \infty),
 \end{aligned}$$

was wiederum  $T_{7,n} = O_{\mathbf{P}}(\delta_{n,t})$  impliziert.

Im neunten (und letzten) Schritt des Beweises wird  $T_{8,n}$  (ohne den Faktor 3) betrachtet. Es folgt mit der Beziehung  $(a+b)^2 \leq 2a^2 + 2b^2$ :

$$\begin{aligned}
 &\inf_{\substack{f \in \mathcal{F}_{n,p} \\ \|f\|_{\infty} \leq \beta_n+1}} \int |f(x) - q_t^w(x)|^2 \mathbf{P}_{X_t}(dx) \\
 &= \inf_{\substack{f \in \mathcal{F}_{n,p} \\ \|f\|_{\infty} \leq \beta_n+1}} \int |f(x) - q_t(x) + q_t(x) - q_t^w(x)|^2 \mathbf{P}_{X_t}(dx) \\
 &\leq \inf_{\substack{f \in \mathcal{F}_{n,p} \\ \|f\|_{\infty} \leq \beta_n+1}} \left( 2 \int |f(x) - q_t(x)|^2 \mathbf{P}_{X_t}(dx) + 2 \int |q_t(x) - q_t^w(x)|^2 \mathbf{P}_{X_t}(dx) \right) \\
 &= \inf_{\substack{f \in \mathcal{F}_{n,p} \\ \|f\|_{\infty} \leq \beta_n+1}} \left( 2 \int |f(x) - q_t(x)|^2 \mathbf{P}_{X_t}(dx) \right) + 2 \int |q_t(x) - q_t^w(x)|^2 \mathbf{P}_{X_t}(dx).
 \end{aligned}$$

Analog wie im fünften Schritt erhält man mit der Anwendung von (2.12) und (2.10) für  $s \geq t$  auf den letzten Term in der letzten Gleichung (ohne den Faktor 2)

$$\begin{aligned}
 &\int |q_t(x) - q_t^w(x)|^2 \mathbf{P}_{X_t}(dx) \\
 &= O_{\mathbf{P}} \left( \sum_{s=t}^{T-1} \delta_{n,s} + \max_{s \in \{t, t+1, \dots, T-1\}} \inf_{\substack{f \in \mathcal{F}_{n,p} \\ \|f\|_{\infty} \leq \beta_n+1}} \int |f(x) - q_s(x)|^2 \mathbf{P}_{X_s}(dx) \right).
 \end{aligned}$$

Insgesamt implizieren die letzten vier Schritte (2.11) für  $j = 4$ , und der Beweis ist vollständig.  $\square$

In den folgenden zwei Abschnitten werden Folgerungen aus Satz 2.1 bezüglich der Konsistenz und Konvergenzgeschwindigkeit für zwei unterschiedliche Funktionenräume  $\mathcal{F}_{n,p}$  betrachtet.

## 2.2 Resultate für polynomiale Splinefunktionen

In diesem Kapitel werden polynomiale Splinefunktionen zur Bewertung von Amerikanischen Optionen untersucht. Dazu werden die Funktionenräume  $\mathcal{F}_{n,p}$  mit Hilfe von polynomialen Splinefunktionen (stückweise Polynome mit globalen Glattheitseigenschaften) gewählt. Diese Funktionenräume  $\mathcal{F}_{n,p}$  wurden aus folgenden Gründen gewählt. Die

Splinefunktionen sind numerisch gut zu handhaben, was auf die so genannten Basisplines (kurz: B-Splines) zurückzuführen ist. Sie besitzen gute Approximationseigenschaften im Deterministischen und letztendlich Glattheitseigenschaften (Stetigkeit, Differenzierbarkeit, usw.), die implizieren, dass der resultierende Schätzer ebenfalls glatt ist (was beispielsweise bei der Schätzung von glatten Regressionsfunktionen wünschenswert ist).

Des Weiteren wird in diesem Abschnitt angenommen, dass der stochastische Prozess  $X_t$  ( $t \in \{0, 1, \dots, T\}$ ) beschränkt ist. Üblicherweise findet diese Annahme bei der Modellierung von Finanzprozessen, wie z. B. bei den Diffusions-, Sprungdiffusions- oder Zeitreihen-Modellen, keine Anwendung. In solch einem Fall muss  $X_t$  geeignet lokalisiert werden, d. h. es wird eine Konstante  $A > 0$  mit sehr großem Wert gewählt und  $X_t$  durch seine beschränkte Approximation

$$X_t^A = X_{\min\{t, \tau_A\}} \quad \text{wobei } \tau_A = \inf\{s \geq 0 : X_s \notin [-A, A]^d\}$$

ersetzt. Es sei hier erwähnt, dass der – durch die Lokalisierung – auftretende so genannte Lokalisierungsfehler in vielen in der Praxis genutzten Modellen explizit geschätzt werden kann. Eine Illustration hierzu ist beispielsweise in Kapitel 4 in Egloff, Kohler und Todorovic (2006) zu sehen. Aus Gründen der Einfachheit wird im weiteren Kontext angenommen, dass der stochastische Prozess stetige Pfade besitzt, um so eine zusätzliche Stützung von  $X_t^A$  vernachlässigen zu können.

Wie bereits am Anfang des Kapitels beschrieben, werden im Folgenden polynomiale Splinefunktionen gewählt, um Funktionenräume  $S_{M, \alpha}$  mit  $M \in \mathbb{N}_0$  und  $\alpha \in (0, \infty)$  zu definieren. Wie sich unten herausstellen wird, sind diese Funktionenräume parametrisiert durch  $p = (M, \alpha)$ .

Für  $p = (M, \alpha)$  und  $k \in \mathbb{Z}$  wird  $u_k = k \cdot \alpha$  gesetzt. Sei  $B_{k, M} : \mathbb{R} \rightarrow \mathbb{R}$  der univariate B-Spline vom Grad  $M$  mit Knotenfolge  $(u_l)_{l \in \mathbb{Z}}$  und Support  $\text{supp}(B_{k, M}) = [u_k, u_{k+M+1}]$ . Im Falle von  $M = 0$  ist der B-Spline  $B_{k, 0}$  die Indikatorfunktion vom Intervall  $[u_k, u_{k+1})$ . Für  $M = 1$  erhält man für den zugehörigen B-Spline die so genannte Hut-Funktion, die gegeben ist durch

$$B_{k, 1}(x) = \begin{cases} \frac{x - u_k}{u_{k+1} - u_k} & \text{für } u_k \leq x \leq u_{k+1}, \\ \frac{u_{k+2} - x}{u_{k+2} - u_{k+1}} & \text{für } u_{k+1} < x \leq u_{k+2}, \\ 0 & \text{sonst.} \end{cases}$$

Für eine ausführliche Definition von  $B_{k, M}$  siehe z. B. de Boor (1978) oder Kapitel 14.1 in Györfi et al. (2002). An dieser Stelle sei erwähnt, dass man den in der Literatur angegebenen Parameter  $K$  des B-Splines hier erhält durch

$$2 \cdot \left\lceil \frac{A}{\alpha} \right\rceil := K.$$

Die B-Splines  $B_{k, M}$  sind stückweise univariate Polynome vom Grad  $M$ . Des Weiteren sind für  $M > 0$  die B-Splines global  $(M - 1)$ -mal und zwischen den Knoten  $u_l$   $M$ -mal

stetig differenzierbar. Mit 0-maliger stetiger Differenzierbarkeit ist dabei die Stetigkeit gemeint.

Im Folgenden werden multivariate Splinefunktionen mittels Tensorprodukten von univariaten B-Splines eingeführt. Hierbei definiert man für jeden Multi-Index  $\mathbf{k} = (k_1, \dots, k_d) \in \mathbb{Z}^d$  den multivariaten B-Spline  $B_{\mathbf{k},M} : \mathbb{R}^d \rightarrow \mathbb{R}$  durch

$$B_{\mathbf{k},M}(x^{(1)}, \dots, x^{(d)}) = B_{k_1,M}(x^{(1)}) \cdots B_{k_d,M}(x^{(d)}) \quad (x^{(1)}, \dots, x^{(d)} \in \mathbb{R}).$$

Sei

$$S_{M,\alpha} = \left\{ \sum_{\mathbf{k} \in \mathbb{Z}^d: \text{supp}(B_{\mathbf{k},M}) \cap [-A,A]^d \neq \emptyset} a_{\mathbf{k}} \cdot B_{\mathbf{k},M} : a_{\mathbf{k}} \in \mathbb{R} \right\}$$

die lineare Hülle der multivariaten B-Splines  $B_{\mathbf{k},M}$ , so dass  $\text{supp}(B_{\mathbf{k},M})$  einen nichtleeren Schnitt mit  $[-A, A]^d$  besitzt. Mit den Parametern  $M \in \mathbb{N}_0$  und  $\alpha \in (0, \infty)$  sind nun durch  $S_{M,\alpha}$  die Funktionenräume mit polynomialen Splinefunktionen definiert.

Aus der Definition des so genannten Splineraumes  $S_{M,\alpha}$  folgt, dass die Funktionen in  $S_{M,\alpha}$   $(M - 1)$ -mal stetig differenzierbare stückweise Polynome vom Grad kleiner oder gleich  $M$  (in jeder Koordinate) sind. Des Weiteren sind die Funktionen in  $S_{M,\alpha}$  auf den Definitionsbereichen

$$[u_{k_1}, u_{k_1+1}] \times \dots \times [u_{k_d}, u_{k_d+1}] \quad (\mathbf{k} = (k_1, \dots, k_d) \in \mathbb{Z}^d) \quad (2.13)$$

definiert und verschwinden auf all den Würfeln (2.13), für die ein  $j \in \{1, \dots, d\}$  existiert, so dass entweder

$$k_j > 0 \quad \text{und} \quad u_{k_j-M} > A$$

oder

$$k_j < 0 \quad \text{und} \quad u_{k_j+M+1} < -A.$$

Infolgedessen ist  $S_{M,\alpha}$  ein linearer Funktionenraum bestehend aus stückweisen Polynomen hinsichtlich äquidistanter Partitionen auf  $\mathbb{R}^d$  in Würfel der Kantenlänge  $\alpha$ , die außerhalb einer kompakten Menge verschwinden.

Ausgehend von diesem Funktionenraum  $S_{M,\alpha}$  wird der so genannte Kleinste-Quadrate-Splineschätzer (mit Parameter  $p = (M, \alpha)$ ) gewählt. Als eine erste Folgerung aus Satz 2.1 erhält man eine Aussage zur Konsistenz des Kleinste-Quadrate-Splineschätzers.

**Korollar 2.3 (Konsistenz)** Sei  $\beta_n \geq L > 0$ . Sei  $X_0, X_1, \dots, X_T$  ein  $\mathbb{R}^d$ -wertiger Markov-Prozess, und sei  $X_t \in [-A, A]^d$  fast sicher für ein  $A > 0$  und für alle  $t \in \{0, 1, \dots, T\}$ . Des Weiteren sei die diskontierte Auszahlungsfunktion  $f_t$  absolut beschränkt durch  $L$ , d.h.

$$|f_t(x)| \leq L \quad \text{für} \quad x \in [-A, A]^d \quad \text{und} \quad t \in \{0, 1, \dots, T\}.$$

Sei  $q_t$  der Kontinuierungswert definiert durch (1.12). Seien die Parameter  $M \in \mathbb{N}_0$  und  $\alpha_n \in (0, \infty)$  fest gewählt und sei ebenfalls  $w \in \{0, 1, \dots, T-t-1\}$  fest gewählt. Definiere den Schätzer  $q_{n,t}^{w,p}$  (mit  $p = p_n = (M, \alpha_n)$ ) für alle  $t \in \{0, \dots, T\}$  durch (2.6) und (2.7) mit  $\mathcal{F}_{n,p} = S_{M,\alpha_n}$  sowie mit Daten definiert in (2.5). Es gelte

$$\alpha_n \rightarrow 0 \quad (n \rightarrow \infty) \quad \text{und} \quad \frac{\log(n)}{\alpha_n^d \cdot n} \rightarrow 0 \quad (n \rightarrow \infty).$$

Dann gilt

$$\int |q_{n,t}^{w,p}(x) - q_t(x)|^2 \mathbf{P}_{X_t}(dx) \rightarrow 0 \quad \text{nach Wahrscheinlichkeit}$$

für alle  $t \in \{0, \dots, T\}$ .

**Beweis:** Bevor auf den Beweis eingegangen wird, ist zu beachten, dass o. B. d. A.  $\beta_n = L$  angenommen werden kann, und dass  $|f_t(x)| \leq L$  auch  $|q_t(x)| \leq L$  ( $x \in [-A, A]^d$ ) impliziert.

Des Weiteren sei (in Abhängigkeit von  $A$ )  $D_n \in \mathbb{N}$  (mit  $D_n \rightarrow \infty$  für  $n$  gegen unendlich, siehe unten) die Dimension des linearen Vektorraumes  $\mathcal{F}_{n,p} = S_{M,\alpha_n}$ . Setze  $\delta_{n,t} := c_2 \cdot D_n \cdot \log(n)/n$ . Mit dieser Wahl sind die Bedingungen an  $\delta_{n,t}$  in Satz 2.1 erfüllt. Um dies zu zeigen, wird als erstes (2.8) für  $\delta \geq \delta_{n,t}$  und  $g \in \mathcal{F}_{n,p}$  betrachtet.

Beachtet man, dass  $\mathcal{F}_{n,p}$  ein linearer Vektorraum ist (mit Dimension  $D_n$ ) und für ein  $R \geq \delta$  die Beziehung

$$\left\{ f - g : f \in \mathcal{F}_{n,p}, \frac{1}{n} \sum_{i=1}^n |f(x_i) - g(x_i)|^2 \leq \delta \right\} \subseteq \left\{ f \in \mathcal{F}_{n,p} : \frac{1}{n} \sum_{i=1}^n |f(x_i)|^2 \leq R \right\}$$

gilt (wegen  $g \in \mathcal{F}_{n,p}$ ), dann impliziert Lemma A.4

$$\begin{aligned} & \mathcal{N}_2 \left( u, \left\{ f - g : f \in \mathcal{F}_{n,p}, \frac{1}{n} \sum_{i=1}^n |f(x_i) - g(x_i)|^2 \leq \delta \right\}, x_1^n \right) \\ & \leq \mathcal{N}_2 \left( u, \left\{ f \in \mathcal{F}_{n,p} : \frac{1}{n} \sum_{i=1}^n |f(x_i)|^2 \leq R \right\}, x_1^n \right) \\ & \leq \left( \frac{4\sqrt{R} + u}{u} \right)^{D_n}, \end{aligned}$$

und daraus folgt

$$\begin{aligned} & \int_{\delta/(2^9\sigma_0)}^{\sqrt{\delta}} \left( \log \mathcal{N}_2 \left( u, \left\{ f - g : f \in \mathcal{F}_{n,p}, \frac{1}{n} \sum_{i=1}^n |f(x_i) - g(x_i)|^2 \leq \delta \right\}, x_1^n \right) \right)^{1/2} du \\ & \leq \int_{\delta/(2^9\sigma_0)}^{\sqrt{\delta}} \left( \log \left( \frac{4\sqrt{R} + u}{u} \right)^{D_n} \right)^{1/2} du. \end{aligned}$$

Sei  $\delta > 1/n$ . Da im unteren Integral  $u$  von unten begrenzt wird durch  $u \geq \delta/(2^9\sigma_0)$ , erhält man

$$\begin{aligned}
 & \int_{\delta/(2^9\sigma_0)}^{\sqrt{\delta}} \left( \log \left( \frac{4\sqrt{R} + u}{u} \right)^{D_n} \right)^{1/2} du \\
 & \leq \int_{\delta/(2^9\sigma_0)}^{\sqrt{\delta}} \left( \log \left( \frac{4\sqrt{R} + \delta/(2^9\sigma_0)}{\delta/(2^9\sigma_0)} \right)^{D_n} \right)^{1/2} du \\
 & \leq \int_{\delta/(2^9\sigma_0)}^{\sqrt{\delta}} \left( \log \left( \frac{4\sqrt{R} + 1/(n2^9\sigma_0)}{1/(n2^9\sigma_0)} \right)^{D_n} \right)^{1/2} du \\
 & \leq c_3 \cdot \sqrt{\delta} \sqrt{D_n \cdot \log(n)}.
 \end{aligned}$$

Insgesamt folgt, dass (2.8) impliziert wird durch:

$$\begin{aligned}
 c_1 \sqrt{n} \delta & \geq c_3 \cdot \sqrt{\delta} \sqrt{D_n \cdot \log(n)} \\
 \Leftrightarrow \delta & \geq c_4 \cdot D_n \cdot \frac{\log(n)}{n}.
 \end{aligned}$$

Als nächstes wird für  $x_1, \dots, x_n \in \mathbb{R}^d$  und  $\delta \geq 2L^2\delta_{n,t}$  (2.9) betrachtet. Lemma A.3, Lemma A.6 und Lemma A.12 implizieren für die Überdeckungszahl

$$\begin{aligned}
 & \mathcal{N}_2 \left( \frac{u}{4L}, \left\{ f - q_t : f \in T_L \mathcal{F}_{n,p}, \frac{1}{n} \sum_{i=1}^n |f(x_i) - q_t(x_i)|^2 \leq \frac{\delta}{L^2} \right\}, x_1^n \right) \\
 & \leq \mathcal{N}_2(u/4L, T_L \mathcal{F}_{n,p}, x_1^n) \\
 & \leq 3 \left( \frac{2e(2L)^2}{(u/4L)^2} \cdot \log \frac{3e(2L)^2}{(u/4L)^2} \right)^{D_n+1} \\
 & \leq 3 \left( \frac{192eL^4}{u^2} \right)^{2D_n+2}
 \end{aligned}$$

mit  $L$  und  $D_n$  wie oben. Mit diesem Resultat kann analog zu oben gezeigt werden, dass mit  $\delta > 1/n$  und da im unteren Integral  $u$  von unten begrenzt wird durch  $u \geq 0.5\delta/(128L^2)$  die Beziehung

$$\begin{aligned}
 & \int_{0.5\delta/(128L^2)}^{\sqrt{\delta}} \left( \log \mathcal{N}_2 \left( \frac{u}{4L}, \left\{ f - q_t : f \in T_L \mathcal{F}_{n,p}, \frac{1}{n} \sum_{i=1}^n |f(x_i) - q_t(x_i)|^2 \leq \frac{\delta}{L^2} \right\}, x_1^n \right) \right)^{1/2} du \\
 & \leq c_5 \cdot \sqrt{\delta} \sqrt{D_n \cdot \log(n)}
 \end{aligned}$$

gilt, mit der weiter gefolgert werden kann, dass (2.9) impliziert wird durch:

$$\begin{aligned}
 \frac{0.5\sqrt{n}\delta}{768\sqrt{2}L^2} & \geq c_5 \cdot \sqrt{\delta} \sqrt{D_n \cdot \log(n)} \\
 \Leftrightarrow \delta & \geq c_6 \cdot D_n \cdot \frac{\log(n)}{n}.
 \end{aligned}$$

Mit der Wahl von  $\delta_{n,t} := c_2 \cdot D_n \cdot \log(n)/n$  und  $c_2 = \max\{c_4, c_6\}$  ist (2.8) für alle  $\delta \geq \delta_{n,t}$  und alle  $g \in \mathcal{F}_{n,p}$  erfüllt. Weiter ist für alle  $x_1, \dots, x_n \in \mathbb{R}^d$  und alle  $\delta \geq 2L^2\delta_{n,t}$  (2.9) erfüllt. Die Anwendung von Satz 2.1 liefert

$$\begin{aligned} & \int |q_{n,t}^{w,p}(x) - q_t(x)|^2 \mathbf{P}_{X_t}(dx) \\ &= O_{\mathbf{P}} \left( \sum_{s=t}^{T-1} \frac{c_2 \cdot D_n \cdot \log(n)}{n} + \max_{s \in \{t, t+1, \dots, T-1\}} \inf_{\substack{f \in \mathcal{F}_{n,p} \\ \|f\|_{\infty} \leq L+1}} \int |f(x) - q_s(x)|^2 \mathbf{P}_{X_s}(dx) \right). \end{aligned}$$

Um den linken Term auf der rechten Seite in obiger Gleichung abschätzen zu können, wird als erstes für die Dimension  $D_n$  des Splineraumes eine obere Schranke angegeben. Dazu ist zu beachten, dass der Splineraum  $S_{M, \alpha_n}$  in einer Menge von stückweise Polynomen der Partition (2.13) mit Kantenlänge  $\alpha_n$  enthalten ist. Die Anzahl der Partitionen, auf die die stückweise Polynome aus  $S_{M, \alpha_n}$  nicht verschwinden, beträgt dabei  $[(A - (-A))/\alpha_n]^d$ . Weiter ist zu berücksichtigen, dass weitere  $2(M+1)$  Partitionen in jeder Koordinate (und außerhalb des Definitionsbereichs) benötigt werden, um eine Basis zu bilden. Insgesamt werden somit

$$[(A - (-A))/\alpha_n] + 2(M+1)]^d = [2 \cdot (\lceil A/\alpha_n \rceil + (M+1))]^d$$

Partitionen benötigt. Weiter sind die stückweise Polynome vom Grad kleiner oder gleich  $M$ . Hiermit kann die Dimension des linearen Vektorraumes  $S_{M, \alpha_n}$  nach oben abgeschätzt werden durch

$$D_n \leq \left[ (M+1) \cdot 2 \left( \left\lceil \frac{A}{\alpha_n} \right\rceil + (M+1) \right) \right]^d.$$

Damit folgt für den linken Term in vorletzter Gleichung

$$\frac{c_2 \cdot D_n \cdot \log(n)}{n} \leq \frac{c_2 \cdot [(M+1) \cdot 2(\lceil A/\alpha_n \rceil + (M+1))]^d \cdot \log(n)}{n} \rightarrow 0 \quad (n \rightarrow \infty),$$

da nach Voraussetzung  $\log(n)/(\alpha_n^d \cdot n) \rightarrow 0$  für  $n \rightarrow \infty$  gilt. Für den rechten Term muss noch

$$\max_{s \in \{t, t+1, \dots, T-1\}} \inf_{\substack{f \in \mathcal{F}_{n,p} \\ \|f\|_{\infty} \leq L+1}} \int |f(x) - q_s(x)|^2 \mathbf{P}_{X_s}(dx) \rightarrow 0$$

gezeigt werden. Die Definition von  $q_t$  und die Voraussetzung  $|f_t(x)| \leq L$  für alle  $x \in [-A, A]^d$  implizieren

$$\int |q_t(x)|^2 \mathbf{P}_{X_t}(dx) \leq L^2 < \infty,$$

und deshalb ist  $q_t$  in  $L_2(\mathbf{P}_{X_t})$ . Nach Lemma A.14 kann  $q_t$  (da  $q_t \in L_2(\mathbf{P}_{X_t})$ ) durch stetige Funktionen beliebig approximiert werden. Da mit Splines - sofern der Knotenabstand gegen Null konvergiert - wiederum die stetigen Funktionen beliebig approximierbar auf beschränkten Mengen sind, was der Vollständigkeit halber noch gezeigt wird, folgt die Behauptung des Korollars.

Es bleibt noch zu zeigen, dass Splines stetige Funktionen auf beschränkten Mengen beliebig approximieren. Da für  $x \in [-A, A]^d$   $|q_t(x)| \leq L$  gilt, genügt es, die Behauptung für stetige Funktionen  $r$  mit  $\|r\|_\infty \leq L + 1$  zu zeigen, die auf dem Kompaktum  $[-A, A]^d$  natürlich Lipschitz-stetig sind.

Sei  $M \leq M_{\max(n)} \in \mathbb{N}_0$  und  $\alpha \leq \alpha_{\max(n)} \in (0, \infty)$  mit der Eigenschaft

$$M_{\max(n)} \cdot \alpha_{\max(n)} \rightarrow 0 \quad (n \rightarrow \infty).$$

Sei  $I = \{\mathbf{k} \in \mathbb{Z}^d : \text{supp}(B_{\mathbf{k},M}) \cap [-A, A]^d \neq \emptyset\}$ . Wähle für  $\mathbf{k} \in I$  ein  $u_{\mathbf{k}} \in \text{supp}(B_{\mathbf{k},M})$ , und definiere den so genannten Quasi-Interpolanten  $Qr \in \mathcal{F}_{n,p}$  durch

$$Qr = \sum_{\mathbf{k} \in I} r(u_{\mathbf{k}}) \cdot B_{\mathbf{k},M}.$$

Dann gilt  $\|Qr\|_\infty \leq \|r\|_\infty$ , was impliziert:

$$\begin{aligned} & \inf_{\substack{f \in \mathcal{F}_{n,p} \\ \|f\|_\infty \leq L+1}} \int |f(x) - r(x)|^2 \mathbf{P}_{X_s}(dx) \\ & \leq \int |(Qr)(x) - r(x)|^2 \mathbf{P}_{X_s}(dx) \\ & \leq 4\|r\|_\infty^2 \mathbf{P}_{X_s}(\mathbb{R}^d \setminus [-A, A]^d) + \int_{[-A, A]^d} |(Qr)(x) - r(x)|^2 \mathbf{P}_{X_s}(dx) \\ & \leq 4\|r\|_\infty^2 \mathbf{P}_{X_s}(\mathbb{R}^d \setminus [-A, A]^d) + \sup_{x \in [-A, A]^d} |(Qr)(x) - r(x)|^2. \end{aligned}$$

Da  $X_s \in [-A, A]^d$  fast sicher gilt, verschwindet der erste Term in der letzten Ungleichung. Des Weiteren erhält man für  $c_7 > 0$  konstant und  $x \in [-A, A]^d$

$$\begin{aligned} & |(Qr)(x) - r(x)| \\ & \leq \sum_{\mathbf{k} \in I} |r(u_{\mathbf{k}}) - r(x)| \cdot B_{\mathbf{k},M}(x) \\ & = \sum_{\mathbf{k} \in I: x \in \text{supp}(B_{\mathbf{k},M})} |r(u_{\mathbf{k}}) - r(x)| \cdot B_{\mathbf{k},M}(x) \\ & \leq \sup_{\substack{u, v \in \mathbb{R}^d \\ \|u-v\|_\infty \leq (M+1)\alpha}} |r(u) - r(v)| \cdot \sum_{\mathbf{k} \in I: x \in \text{supp}(B_{\mathbf{k},M})} B_{\mathbf{k},M}(x) \\ & \leq \sup_{\substack{u, v \in \mathbb{R}^d \\ \|u-v\|_\infty \leq (M+1)\alpha}} |r(u) - r(v)| \\ & \leq \sup_{\substack{u, v \in \mathbb{R}^d \\ \|u-v\|_\infty \leq (M_{\max(n)}+1)\alpha_{\max(n)}}} c_7 \|u - v\|_\infty \\ & \leq c_7 \cdot (M_{\max(n)} + 1) \cdot \alpha_{\max(n)}. \end{aligned}$$

Mit Hilfe dieses Resultats und mit  $M_{\max(n)} \cdot \alpha_{\max(n)} \rightarrow 0$  ( $n \rightarrow \infty$ ) kann gefolgert werden:

$$\begin{aligned} & \sup_{x \in [-A, A]^d} |(Qr)(x) - r(x)| \\ & \leq c_7 \cdot (M_{\max(n)} + 1) \cdot \alpha_{\max(n)} \rightarrow 0 \quad (n \rightarrow \infty). \end{aligned}$$

□

Im obigen Korollar wurde gezeigt, dass der  $L_2$ -Fehler des Schätzers  $q_{n,t}^{w,p}$  gegen Null konvergiert nach Wahrscheinlichkeit für einen gegen unendlich konvergierenden Stichprobenumfang der Monte-Carlo-Stichprobe. Im Hinblick auf eine Anwendung mit notwendigem endlichen Stichprobenumfang wäre es schön zu wissen, wie schnell dieser  $L_2$ -Fehler für einen immer größer werdenden Stichprobenumfang gegen Null konvergiert. In der nichtparametrischen Regressionsschätzung ist es wohlbekannt, dass ohne Glattheitsvoraussetzungen an die Regressionsfunktion die Konvergenzrate beliebig klein werden kann (siehe z. B. Cover (1968), Devroye (1982) oder Kapitel 3 in Györfi et al. (2002)). Aus diesem Grund wird hier für die Kontinuierungswerte  $q_t$  eine bestimmte Glattheitseigenschaft angenommen, die in der folgenden Definition eingeführt wird.

**Definition 2.4** Sei  $p = k + \gamma$  mit  $k \in \mathbb{N}_0$  und  $0 < \gamma \leq 1$ , und sei  $C > 0$ . Eine Funktion  $f : \mathbb{R}^d \rightarrow \mathbb{R}$  heißt  $(p, C)$ -glatt, falls für alle  $\alpha_1, \dots, \alpha_d \in \mathbb{N}_0$  mit  $\alpha_1 + \dots + \alpha_d = k$  die partielle Ableitung

$$\frac{\partial^k f}{\partial^{\alpha_1 x^{(1)}} \dots \partial^{\alpha_d x^{(d)}}}$$

von  $f$  existiert und die folgende Eigenschaft hat:

$$\left| \frac{\partial^k f}{\partial^{\alpha_1 x^{(1)}} \dots \partial^{\alpha_d x^{(d)}}}(x) - \frac{\partial^k f}{\partial^{\alpha_1 x^{(1)}} \dots \partial^{\alpha_d x^{(d)}}}(z) \right| \leq C \cdot \|x - z\|^\gamma$$

für alle  $x, z \in \mathbb{R}^d$ .

Als eine weitere Folgerung aus Satz 2.1 erhält man eine Aussage zur Konvergenzgeschwindigkeit des Kleinste-Quadrate-Splineschätzers.

**Korollar 2.5 (Konvergenzgeschwindigkeit)** Sei  $\beta_n \geq L > 0$ . Sei  $X_0, X_1, \dots, X_T$  ein  $\mathbb{R}^d$ -wertiger Markov-Prozess, und sei  $X_t \in [-A, A]^d$  fast sicher für ein  $A > 0$  und für alle  $t \in \{0, 1, \dots, T\}$ . Des Weiteren sei die diskontierte Auszahlungsfunktion  $f_t$  absolut beschränkt durch  $L$ , d.h.

$$|f_t(x)| \leq L \quad \text{für } x \in [-A, A]^d \quad \text{und } t \in \{0, 1, \dots, T\}.$$

Sei  $q_t$  der Kontinuierungswert definiert durch (1.12). Des Weiteren sei  $p, C > 0$  fest und sei  $q_t$   $(p, C)$ -glatt. Weiter sei  $w \in \{0, 1, \dots, T - t - 1\}$  fest gewählt. Ebenfalls seien die Parameter  $M \in \mathbb{N}_0$  und  $\alpha_n \in (0, \infty)$  fest gewählt. Definiere den Schätzer  $q_{n,t}^{w,p}$  (mit  $p = p_n = (M, \alpha_n)$ ) für alle  $t \in \{0, \dots, T\}$  durch (2.6) und (2.7) mit Daten definiert in (2.5) sowie mit  $\mathcal{F}_{n,p} = S_{M, \alpha_n}$ , wobei  $\alpha_n$  gegeben ist durch

$$\alpha_n = c_8 \cdot C^{-2/(2p+d)} \cdot (\log(n)/n)^{1/(2p+d)} \quad (2.14)$$



mit  $c_8 > 0$  konstant. Dann gilt

$$\int |q_{n,t}^{w,p}(x) - q_t(x)|^2 \mathbf{P}_{X_t}(dx) = O_{\mathbf{P}} \left( C^{\frac{2d}{2p+d}} \cdot \left( \frac{\log(n)}{n} \right)^{\frac{2p}{2p+d}} \right)$$

für alle  $t \in \{0, \dots, T\}$ .

**Beweis:** Bevor auf den Beweis eingegangen wird, ist zu beachten, dass o. B. d. A.  $\beta_n = L$  angenommen werden kann. Mit der Beziehung

$$\min_{\substack{f \in \mathcal{F}_{n,p} \\ \|f\|_{\infty} \leq L+1}} \int |f(x) - q_t(x)|^2 \mathbf{P}_{X_t}(dx) \leq \min_{\substack{f \in \mathcal{F}_{n,p} \\ \|f\|_{\infty} \leq L+1}} \sup_{x \in [-A, A]^d} |f(x) - q_t(x)|^2$$

und der Annahme, dass  $q_t$   $(p, C)$ -glatt ist, existiert ein  $f \in \mathcal{F}_{n,p}$  mit

$$\sup_{x \in [-A, A]^d} |f(x) - q_t(x)| \leq c_9 \cdot C \cdot \alpha_n^p,$$

wobei  $\alpha_n$  der Kantenlänge der kubischen Partition entspricht, die in der Definition des Splineraires genutzt wurde (vgl. Theorem 12.8 in Schumaker (1981)). Mit diesen oberen Schranken kann weiterhin gefolgert werden:

$$\begin{aligned} \min_{\substack{f \in \mathcal{F}_{n,p} \\ \|f\|_{\infty} \leq L+1}} \int |f(x) - q_t(x)|^2 \mathbf{P}_{X_t}(dx) &\leq c_9^2 \cdot C^2 \cdot \alpha_n^{2p} \\ &= c_9^2 \cdot C^2 \cdot C^{-\frac{4p}{2p+d}} \cdot (n/\log(n))^{\frac{-2p}{2p+d}} \\ &\leq \text{const} \cdot C^{\frac{2d}{2p+d}} \cdot \left( \frac{\log(n)}{n} \right)^{\frac{2p}{2p+d}}. \end{aligned}$$

Setze  $\delta_{n,t} := c_2 \cdot D_n \cdot \log(n)/n$ , wobei (in Abhängigkeit von  $A$ )  $D_n \in \mathbb{N}$  (mit  $D_n \rightarrow \infty$  für  $n$  gegen unendlich) die Dimension des linearen Vektorraumes  $\mathcal{F}_{n,p} = S_{M, \alpha_n}$  ist. Mit dieser Wahl von  $\delta_{n,t}$  sind die Voraussetzungen in Satz 2.1 erfüllt, was in Korollar 2.3 bereits gezeigt wurde.

Mit  $\delta_{n,t}$  und unter Ausnutzung letzter Ungleichung impliziert Satz 2.1

$$\begin{aligned} &\int |q_{n,t}^{w,p}(x) - q_t(x)|^2 \mathbf{P}_{X_t}(dx) \\ &= O_{\mathbf{P}} \left( \sum_{s=t}^{T-1} c_2 \cdot \frac{D_n \cdot \log(n)}{n} + \max_{s \in \{t, t+1, \dots, T-1\}} \inf_{\substack{f \in \mathcal{F}_{n,p} \\ \|f\|_{\infty} \leq L+1}} \int |f(x) - q_s(x)|^2 \mathbf{P}_{X_s}(dx) \right) \\ &= O_{\mathbf{P}} \left( \sum_{s=t}^{T-1} c_2 \cdot \frac{D_n \cdot \log(n)}{n} + \text{const} \cdot C^{\frac{2d}{2p+d}} \cdot \left( \frac{\log(n)}{n} \right)^{\frac{2p}{2p+d}} \right) \\ &= O_{\mathbf{P}} \left( C^{\frac{2d}{2p+d}} \cdot \left( \frac{\log(n)}{n} \right)^{\frac{2p}{2p+d}} \right) \end{aligned}$$

für alle  $t \in \{0, 1, \dots, T\}$ . Die letzte Gleichung folgt mit (2.14) und aus der Beziehung

$$\frac{D_n \cdot \log(n)}{n} \leq \frac{\left[ (M+1) \cdot 2 \left( \left\lceil \frac{A}{\alpha_n} \right\rceil + (M+1) \right) \right]^d \cdot \log(n)}{n} \leq c_{10} \cdot \alpha_n^{-d} \cdot \frac{\log(n)}{n},$$

wobei  $c_{10} > 0$  konstant.  $\square$

### Bemerkungen.

**Bemerkung 2.6** *Es sei hier betont, dass in Korollar 2.3 und 2.5 keine Voraussetzung an die Verteilung von  $X_t$  gestellt wird, außer dass  $X_t \in [-A, A]^d$  fast sicher gilt. Insbesondere wird nicht die Existenz einer Dichte von  $X_t$  bezüglich des Lebesgue–Borel–Maßes gefordert.*

**Bemerkung 2.7** *Es ist wohlbekannt, dass die optimale Konvergenzrate für die Schätzung von  $(p, C)$ -glatten Funktionen  $n^{-2p/(2p+d)}$  ist (siehe z. B. Stone (1982) oder Kapitel 3 in Györfi et al. (2002)). Deshalb ist die Konvergenzrate in Korollar 2.5 vermutlich auch in diesem Kontext bis auf den logarithmischen Faktor  $(\log(n))^{2p/(2p+d)}$  optimal.*

**Bemerkung 2.8** *Angenommen,  $X_0 = x_0$  fast sicher für ein  $x_0 \in [-A, A]^d$ . Der Preis*

$$V_0 = \max\{f_0(x_0), q_0(x_0)\}$$

*einer Amerikanischen Option (vgl. (1.10) und (1.12)) kann durch*

$$\hat{V}_0 = \max\{f_0(x_0), q_{n,0}^{w,p}(x_0)\}$$

*geschätzt werden. Da die Verteilung von  $X_0$  auf  $x_0$  konzentriert ist, gilt unter den Voraussetzungen in Korollar 2.5 die folgende Fehlerschranke:*

$$\begin{aligned} |\hat{V}_0 - V_0|^2 &= |\max\{f_0(x_0), q_{n,0}^{w,p}(x_0)\} - \max\{f_0(x_0), q_0(x_0)\}|^2 \\ &\leq |q_{n,0}^{w,p}(x_0) - q_0(x_0)|^2 \\ &= O_{\mathbf{P}} \left( C^{\frac{2d}{2p+d}} \cdot \left( \frac{\log(n)}{n} \right)^{\frac{2p}{2p+d}} \right). \end{aligned}$$

In diesem Abschnitt erfolgten Aussagen zur Konsistenz und Konvergenzgeschwindigkeit des Kleinste-Quadrate-Splineschätzers. Der Schätzer wird dabei zum Bewerten Amerikanischer Optionen verwendet. Der nächste Abschnitt präsentiert ebenfalls Aussagen zu Konsistenz und Konvergenzgeschwindigkeit. Dort wird im Unterschied zu diesem Abschnitt ein so genannter Neuronale-Netze-Schätzer betrachtet.

## 2.3 Resultate für neuronale Netze

In diesem Abschnitt werden, im Zusammenhang mit der Schätzung des Wertes einer Amerikanischen Option, neuronale Netze als Funktionenräume zur nichtparametrischen Regressionsschätzung eingesetzt. Das neuronale Netz ist eines von wenigen Standardverfahren, um im Hochdimensionalen die (unbekannte) Regressionsfunktion zu schätzen. Im

Folgendes wird die Klasse der vorwärtsgerichteten neuronalen Netze (mit einer verdeckten Schicht) eingeführt werden. Für eine ausführliche Beschreibung der neuronalen Netze siehe beispielsweise Kapitel 16 in Györfi et al. (2002).

Sei  $\sigma : \mathbb{R} \rightarrow [0, 1]$  eine Sigmoid-Funktion, d. h. es wird angenommen, dass  $\sigma$  monoton wachsend ist und die Eigenschaften

$$\sigma(x) \rightarrow 0 \quad (x \rightarrow -\infty) \quad \text{and} \quad \sigma(x) \rightarrow 1 \quad (x \rightarrow \infty)$$

erfüllt. Ein Beispiel für solch eine Sigmoid-Funktion ist die so genannte logistische Funktion, welche durch

$$\sigma(x) = \frac{1}{1 + e^{-x}} \quad (x \in \mathbb{R})$$

definiert ist. Weitere Beispiele zu Sigmoid-Funktionen können z. B. in Kapitel 16.1 in Györfi et al. (2002) gefunden werden.

Wie oben bereits angedeutet, werden die Kontinuierungswerte, welche den zu schätzenden Regressionsfunktionen entsprechen, durch neuronale Netze mit  $K \in \mathbb{N}$  versteckten Neuronen und einer Sigmoid-Funktion  $\sigma$  geschätzt.

Sei  $\beta_n > 0$  (welches später so gewählt wird, dass  $\beta_n \rightarrow \infty$  ( $n \rightarrow \infty$ )), und sei  $\mathcal{F}_K(\beta_n)$  die Klasse der vorwärtsgerichteten neuronalen Netze, die definiert ist durch

$$\mathcal{F}_K(\beta_n) = \left\{ \sum_{i=1}^K d_i \cdot \sigma(a_i^T x + b_i) + d_0 \quad : \quad a_i \in \mathbb{R}^d, b_i \in \mathbb{R}, \sum_{i=0}^K |d_i| \leq \beta_n \right\}, \quad (2.15)$$

wobei  $\sigma$  die Sigmoid-Funktion von oben ist. Aus der Definition von  $\mathcal{F}_K(\beta_n)$  ist ersichtlich, dass die darin enthaltenen Funktionen durch  $\beta_n$  beschränkt sind. Des Weiteren ist  $\mathcal{F}_K(\beta_n)$  durch  $p = K \in \mathbb{N}$  parametrisiert.

Ausgehend vom Funktionenraum  $\mathcal{F}_K(\beta_n)$  wird der so genannte Neuronale-Netze-Schätzer (mit Parameter  $p = K$ ) gewählt. Das folgende Korollar weist die Konsistenz dieses Schätzers nach.

**Korollar 2.9 (Konsistenz)** Sei  $L > 0$ . Sei  $X_0, X_1, \dots, X_T$  ein  $\mathbb{R}^d$ -wertiger Markov-Prozess, und sei die diskontierte Auszahlungsfunktion  $f_t$  absolut beschränkt durch  $L$ , d.h.

$$|f_t(x)| \leq L \quad \text{für} \quad x \in \mathbb{R}^d \quad \text{und} \quad t \in \{0, 1, \dots, T\}.$$

Sei  $q_t$  der Kontinuierungswert definiert durch (1.12). Sei  $w \in \{0, 1, \dots, T - t - 1\}$  fest gewählt. Des Weiteren sei  $\beta_n > 0$  und  $K_n \in \mathbb{N}$  fest. Definiere den Schätzer  $\hat{q}_{n,t}^{w,p}$  (mit  $p = p_n = K_n$ ) für alle  $t \in \{0, \dots, T\}$  durch (2.6) und (2.7) mit  $\mathcal{F}_{n,p} = \mathcal{F}_{K_n}(\beta_n)$  sowie mit Daten definiert in (2.5). Es gelte

$$\beta_n \rightarrow \infty \quad (n \rightarrow \infty), \quad K_n \rightarrow \infty \quad (n \rightarrow \infty) \quad \text{und} \quad \frac{\beta_n^4 \cdot K_n \cdot \log(n)}{n} \rightarrow 0 \quad (n \rightarrow \infty).$$

Dann gilt

$$\int |q_{n,t}^{w,p}(x) - q_t(x)|^2 \mathbf{P}_{X_t}(dx) \rightarrow 0 \quad \text{nach Wahrscheinlichkeit}$$

für alle  $t \in \{0, \dots, T\}$ .

Als Basis für den Beweis dient Satz 2.1. Um diesen Satz anwenden zu können, müssen die darin enthaltenen Voraussetzungen überprüft werden. Zur Überprüfung von (2.8) und (2.9) wird eine Abschätzung der Überdeckungszahl von  $\mathcal{F}_K(\beta_n)$  benötigt, welche Lemma 2.12 (siehe unten) liefert. Zuvor werden zwei Lemmata erstellt und verifiziert, die als Hilfsmittel für den Beweis von Lemma 2.12 dienen.

**Lemma 2.10** *Seien  $\mathcal{F}$  und  $\mathcal{G}$  zwei Familien von reellen Funktionen auf  $\mathbb{R}^m$ . Falls  $\mathcal{F} \oplus \mathcal{G}$  die Menge von Funktionen  $\{f+g : f \in \mathcal{F}, g \in \mathcal{G}\}$  bezeichnet, dann gilt für jedes  $z_1^n \in \mathbb{R}^{n \cdot m}$  und  $\epsilon, \delta > 0$*

$$\mathcal{N}_2(\epsilon + \delta, \mathcal{F} \oplus \mathcal{G}, z_1^n) \leq \mathcal{N}_2(\epsilon, \mathcal{F}, z_1^n) \mathcal{N}_2(\delta, \mathcal{G}, z_1^n).$$

**Beweis:** Seien  $\{f_1, \dots, f_{\mathcal{K}}\}$  und  $\{g_1, \dots, g_{\Lambda}\}$  eine  $\epsilon$ -Überdeckung und eine  $\delta$ -Überdeckung von  $\mathcal{F}$  und  $\mathcal{G}$  bzw. auf  $z_1^n$  mit minimaler Größe. Dann existiert für jedes  $f \in \mathcal{F}$  und  $g \in \mathcal{G}$  ein  $\kappa \in \{1, \dots, \mathcal{K}\}$  und  $\lambda \in \{1, \dots, \Lambda\}$ , so dass gilt

$$\left( \frac{1}{n} \sum_{i=1}^n |f(z_i) - f_{\kappa}(z_i)|^2 \right)^{1/2} < \epsilon$$

und

$$\left( \frac{1}{n} \sum_{i=1}^n |g(z_i) - g_{\lambda}(z_i)|^2 \right)^{1/2} < \delta.$$

Mit der Dreiecksungleichung für Normen folgt

$$\begin{aligned} & \left( \frac{1}{n} \sum_{i=1}^n |f(z_i) + g(z_i) - (f_{\kappa}(z_i) + g_{\lambda}(z_i))|^2 \right)^{1/2} \\ & \leq \left( \frac{1}{n} \sum_{i=1}^n |f(z_i) - f_{\kappa}(z_i)|^2 \right)^{1/2} + \left( \frac{1}{n} \sum_{i=1}^n |g(z_i) - g_{\lambda}(z_i)|^2 \right)^{1/2} \\ & \leq \epsilon + \delta, \end{aligned}$$

was beweist, dass  $\{f_{\kappa} + g_{\lambda} : 1 \leq \kappa \leq \mathcal{K}, 1 \leq \lambda \leq \Lambda\}$  eine  $(\epsilon + \delta)$ -Überdeckung von  $\mathcal{F} \oplus \mathcal{G}$  auf  $z_1^n$  ist.  $\square$

**Lemma 2.11** *Seien  $\mathcal{F}$  und  $\mathcal{G}$  zwei Familien von reellen Funktionen auf  $\mathbb{R}^m$ , so dass  $|f(x)| \leq M_1$  und  $|g(x)| \leq M_2$  für alle  $x \in \mathbb{R}^m$ ,  $f \in \mathcal{F}$ ,  $g \in \mathcal{G}$ . Falls  $\mathcal{F} \odot \mathcal{G}$  die Menge von Funktionen  $\{f \cdot g : f \in \mathcal{F}, g \in \mathcal{G}\}$  bezeichnet, dann gilt für jedes  $z_1^n \in \mathbb{R}^{n \cdot m}$  und  $\epsilon, \delta > 0$*

$$\mathcal{N}_2(\epsilon + \delta, \mathcal{F} \odot \mathcal{G}, z_1^n) \leq \mathcal{N}_2(\epsilon/M_2, \mathcal{F}, z_1^n) \mathcal{N}_2(\delta/M_1, \mathcal{G}, z_1^n).$$

**Beweis:** Sei  $\{f_1, \dots, f_{\mathcal{K}}\}$  und  $\{g_1, \dots, g_{\Lambda}\}$  eine  $\epsilon/M_2$ -Überdeckung und eine  $\delta/M_1$ -Überdeckung von  $\mathcal{F}$  und  $\mathcal{G}$  bzw. auf  $z_1^n$  mit minimaler Größe. Wegen der Beschränktheit von  $f$  und  $g$  kann o. B. d. A.  $|f_{\kappa}(z)| \leq M_1$  und  $|g_{\lambda}(z)| \leq M_2$  angenommen werden. Weiter existiert für jedes  $f \in \mathcal{F}$  und  $g \in \mathcal{G}$  ein  $\kappa \in \{1, \dots, \mathcal{K}\}$  und  $\lambda \in \{1, \dots, \Lambda\}$ , so dass gilt

$$\left( \frac{1}{n} \sum_{i=1}^n |f(z_i) - f_{\kappa}(z_i)|^2 \right)^{1/2} < \frac{\epsilon}{M_2}$$

und

$$\left( \frac{1}{n} \sum_{i=1}^n |g(z_i) - g_{\lambda}(z_i)|^2 \right)^{1/2} < \frac{\delta}{M_1}.$$

Des Weiteren folgt mit der Dreiecksungleichung für Normen

$$\begin{aligned} & \left( \frac{1}{n} \sum_{i=1}^n |f(z_i) \cdot g(z_i) - f_{\kappa}(z_i) \cdot g_{\lambda}(z_i)|^2 \right)^{1/2} \\ &= \left( \frac{1}{n} \sum_{i=1}^n |f(z_i) \cdot (g_{\lambda}(z_i) + g(z_i) - g_{\lambda}(z_i)) - f_{\kappa}(z_i) \cdot g_{\lambda}(z_i)|^2 \right)^{1/2} \\ &\leq \left( \frac{1}{n} \sum_{i=1}^n |g_{\lambda}(z_i) \cdot (f(z_i) - f_{\kappa}(z_i))|^2 \right)^{1/2} + \left( \frac{1}{n} \sum_{i=1}^n |f(z_i) \cdot (g(z_i) - g_{\lambda}(z_i))|^2 \right)^{1/2} \\ &\leq M_2 \cdot \left( \frac{1}{n} \sum_{i=1}^n |f(z_i) - f_{\kappa}(z_i)|^2 \right)^{1/2} + M_1 \cdot \left( \frac{1}{n} \sum_{i=1}^n |g(z_i) - g_{\lambda}(z_i)|^2 \right)^{1/2} \\ &\leq \epsilon + \delta, \end{aligned}$$

was impliziert, dass  $\{f_{\kappa} \cdot g_{\lambda} : 1 \leq \kappa \leq \mathcal{K}, 1 \leq \lambda \leq \Lambda\}$  eine  $(\epsilon + \delta)$ -Überdeckung von  $\mathcal{F} \odot \mathcal{G}$  auf  $z_1^n$  ist.  $\square$

**Lemma 2.12** Sei  $\mathcal{F}_K(\beta_n)$  definiert durch (2.15), sei  $\epsilon > 0$ , und sei  $x_1^n \in (\mathbb{R}^d)^n$ . Dann gilt

$$\mathcal{N}_2(\epsilon, \mathcal{F}_K(\beta_n), x_1^n) \leq \left( \frac{12e\beta_n(K+1)}{\epsilon} \right)^{(4d+9)K+1}.$$

**Beweis von Lemma 2.12:** Definiere die folgenden Klassen von Funktionen:

$$\begin{aligned} \mathcal{G}_1 &= \{a^T x + b : a \in \mathbb{R}^d, b \in \mathbb{R}\}, \\ \mathcal{G}_2 &= \{\sigma(a^T x + b) : a \in \mathbb{R}^d, b \in \mathbb{R}\}, \\ \mathcal{G}_3 &= \{c \cdot \sigma(a^T x + b) : a \in \mathbb{R}^d, b \in \mathbb{R}, c \in [-\beta_n, \beta_n]\}, \end{aligned}$$

wobei  $\beta_n > 0$  und  $\sigma : \mathbb{R} \rightarrow [0, 1]$  eine Sigmoid-Funktion ist (d. h.  $\sigma$  ist eine nichtfallende Funktion mit der Eigenschaft  $\lim_{x \rightarrow -\infty} \sigma(x) = 0$  und  $\lim_{x \rightarrow \infty} \sigma(x) = 1$ ).

$\mathcal{G}_1$  ist ein linearer Vektorraum der Dimension  $d + 1$ . Lemma A.12 impliziert

$$V_{\mathcal{G}_1^+} \leq d + 2,$$

wobei  $\mathcal{G}^+$  die Menge

$$\begin{aligned}\mathcal{G}^+ &= \{ \{(z, t) \in \mathbb{R}^d \times \mathbb{R} : t \leq g(z)\} : g \in \mathcal{G} \} \\ &\subseteq \{ \{(z, t) \in \mathbb{R}^d \times \mathbb{R} : \alpha \cdot t + g(z) \geq 0\} : g \in \mathcal{G}, \alpha \in \mathbb{R} \}\end{aligned}$$

für alle Subgraphen von Funktionen aus  $\mathcal{G}$  bezeichnet und  $V_{\mathcal{G}^+}$  die so genannte VC-Dimension von  $\mathcal{G}^+$  ist (siehe Definition A.5).

Da  $\sigma$  eine nichtfallende Funktion ist, impliziert Lemma A.13

$$V_{\mathcal{G}^+} \leq d + 2.$$

Des Weiteren folgt mit Lemma A.3 und Lemma A.6 für  $0 < \epsilon < 1/4$

$$\begin{aligned}\mathcal{N}_2(\epsilon, \mathcal{G}_2, x_1^n) &\leq \mathcal{M}_2(\epsilon, \mathcal{G}_2, x_1^n) \\ &\leq 3 \left( \frac{2e}{\epsilon^2} \log \frac{3e}{\epsilon^2} \right)^{d+2} \\ &\leq 3 \left( \frac{3e}{\epsilon^2} \right)^{2d+4}.\end{aligned}$$

Mit Lemma 2.11 gilt für  $0 < \epsilon/2\beta_n < 1/4$  oder äquivalent dazu für  $0 < \epsilon < \beta_n/2$

$$\begin{aligned}\mathcal{N}_2(\epsilon, \mathcal{G}_3, x_1^n) &\leq \mathcal{N}_2\left(\frac{\epsilon}{2}, \{c : |c| \leq \beta_n\}, x_1^n\right) \mathcal{N}_2\left(\frac{\epsilon}{2\beta_n}, \mathcal{G}_2, x_1^n\right) \\ &\leq \frac{2\beta_n}{(\epsilon/2)} \cdot 3 \left( \frac{3e}{(\epsilon/(2\beta_n))^2} \right)^{2d+4} \\ &\leq \left( \frac{12e\beta_n}{\epsilon} \right)^{4d+9}.\end{aligned}$$

Mit Hilfe dieses Resultats und durch mehrmaliges Anwenden von Lemma 2.10 erhält man für  $0 < \epsilon < (K+1) \cdot \beta_n/2$  eine obere Schranke für die Überdeckungszahl von  $\mathcal{F}_K(\beta_n)$ :

$$\begin{aligned}\mathcal{N}_2(\epsilon, \mathcal{F}_K(\beta_n), x_1^n) &\leq \mathcal{N}_2\left(\frac{\epsilon}{K+1}, \{c_0 : |c_0| \leq \beta_n\}, x_1^n\right) \left( \mathcal{N}_2\left(\frac{\epsilon}{K+1}, \mathcal{G}_3, x_1^n\right) \right)^K \\ &\leq \frac{2\beta_n(K+1)}{\epsilon} \left( \frac{12e\beta_n(K+1)}{\epsilon} \right)^{(4d+9)K} \\ &\leq \left( \frac{12e\beta_n(K+1)}{\epsilon} \right)^{(4d+9)K+1}.\end{aligned}$$

Der Beweis ist wegen der Beschränktheit von  $\mathcal{F}_K(\beta_n)$  für  $\epsilon \geq (K+1) \cdot \beta_n/2 \geq \beta_n$  trivial und somit vollständig.  $\square$

**Beweis von Korollar 2.9:** Im Beweis findet Satz 2.1 Anwendung. Setze  $\delta_{n,t} = c_{11} \cdot \beta_n^4 \cdot K_n \cdot \log(n)/n$ . Mit dieser Wahl von  $\delta_{n,t}$  sind die geforderten Eigenschaften in Satz 2.1 erfüllt, was im Folgenden gezeigt wird.

Als erstes wird (2.8) für  $\delta \geq \delta_{n,t}$  und  $g \in \mathcal{F}_{K_n}(\beta_n)$  betrachtet. Lemma 2.12 impliziert für die Überdeckungsanzahl

$$\begin{aligned} & \mathcal{N}_2\left(u, \left\{f - g : f \in \mathcal{F}_{K_n}(\beta_n), \frac{1}{n} \sum_{i=1}^n |f(x_i) - g(x_i)|^2 \leq \delta\right\}, x_1^n\right) \\ & \leq \mathcal{N}_2\left(u, \mathcal{F}_{K_n}(\beta_n), x_1^n\right) \\ & \leq \left(\frac{12e\beta_n(K_n + 1)}{u}\right)^{(4d+9)K_n+1}, \end{aligned}$$

und daraus folgt

$$\begin{aligned} & \int_{\delta/(2^9\sigma_0)}^{\sqrt{\delta}} \left(\log \mathcal{N}_2\left(u, \left\{f - g : f \in \mathcal{F}_{K_n}(\beta_n), \frac{1}{n} \sum_{i=1}^n |f(x_i) - g(x_i)|^2 \leq \delta\right\}, x_1^n\right)\right)^{1/2} du \\ & \leq \int_{\delta/(2^9\sigma_0)}^{\sqrt{\delta}} \left(\log \left(\frac{12e\beta_n(K_n + 1)}{u}\right)^{(4d+9)K_n+1}\right)^{1/2} du. \end{aligned}$$

Sei  $\delta > 1/n$ . Da im unteren Integral  $u$  von unten begrenzt wird durch  $u \geq \delta/(2^9\sigma_0)$ , erhält man

$$\begin{aligned} & \int_{\delta/(2^9\sigma_0)}^{\sqrt{\delta}} \left(\log \left(\frac{12e\beta_n(K_n + 1)}{u}\right)^{(4d+9)K_n+1}\right)^{1/2} du \\ & \leq \int_{\delta/(2^9\sigma_0)}^{\sqrt{\delta}} \left(\log \left(\frac{12e\beta_n(K_n + 1)(2^9\sigma_0)}{\delta}\right)^{(4d+9)K_n+1}\right)^{1/2} du \\ & \leq \int_{\delta/(2^9\sigma_0)}^{\sqrt{\delta}} \left(\log (12e\beta_n(K_n + 1)(2^9\sigma_0)n)^{(4d+9)K_n+1}\right)^{1/2} du \\ & \leq c_{12} \cdot \sqrt{\delta} \sqrt{K_n \cdot \log(n)}. \end{aligned}$$

Insgesamt folgt, dass (2.8) impliziert wird durch:

$$\begin{aligned} c_1 \sqrt{n} \delta & \geq c_{12} \cdot \sqrt{\delta} \sqrt{K_n \cdot \log(n)} \\ \Leftrightarrow \delta & \geq c_{13} \cdot K_n \cdot \frac{\log(n)}{n}. \end{aligned}$$

Als nächstes wird für  $x_1, \dots, x_n \in \mathbb{R}^d$  und  $\delta \geq 2\beta_n^2 \delta_{n,t}$  (2.9) betrachtet. Da alle  $f \in \mathcal{F}_{K_n}(\beta_n)$  absolut beschränkt sind durch  $\beta_n$ , gilt  $\mathcal{F}_{K_n}(\beta_n) = T_{\beta_n} \mathcal{F}_{n,p}$ . Wieder impliziert Lemma 2.12 für die Überdeckungsanzahl

$$\mathcal{N}_2\left(\frac{u}{4\beta_n}, \left\{f - q_t : f \in \mathcal{F}_{K_n}(\beta_n), \frac{1}{n} \sum_{i=1}^n |f(x_i) - q_t(x_i)|^2 \leq \frac{\delta}{\beta_n^2}\right\}, x_1^n\right)$$

$$\begin{aligned} &\leq \mathcal{N}_2\left(u/(4\beta_n), \mathcal{F}_{K_n}(\beta_n), x_1^n\right) \\ &\leq \left(\frac{12e\beta_n(K_n+1)}{u/(4\beta_n)}\right)^{(4d+9)K_n+1}. \end{aligned}$$

Damit und analog zu oben erhält man für das Integral in (2.9) eine obere Schranke gemäß

$$\begin{aligned} &\int_{0.5\delta/(128\beta_n^2)}^{\sqrt{\delta}} \left( \log \mathcal{N}_2\left(\frac{u}{4\beta_n}, \left\{f - q_t : f \in T_{\beta_n} \mathcal{F}_{n,p}, \frac{1}{n} \sum_{i=1}^n |f(x_i) - q_t(x_i)|^2 \leq \frac{\delta}{\beta_n^2}\right\}, x_1^n\right) \right)^{1/2} du \\ &\leq c_{14} \cdot \sqrt{\delta} \sqrt{K_n \cdot \log(n)}. \end{aligned}$$

Mit dieser Abschätzung folgt, dass (2.9) impliziert wird durch:

$$\begin{aligned} \frac{0.5\sqrt{n}\delta}{768\sqrt{2}\beta_n^2} &\geq c_{14} \cdot \sqrt{\delta} \sqrt{K_n \cdot \log(n)} \\ \Leftrightarrow \delta &\geq c_{15} \cdot \frac{\beta_n^4 \cdot K_n \cdot \log(n)}{n}. \end{aligned}$$

Wird  $\delta_{n,t} = c_{11} \cdot \beta_n^4 \cdot K_n \cdot \log(n)/n$  gesetzt - mit  $c_{11} = \max\{c_{13}, c_{15}\}$  - dann ist (2.8) für alle  $\delta \geq \delta_{n,t}$  und alle  $g \in \mathcal{F}_{K_n}(\beta_n)$  erfüllt. Des Weiteren ist für alle  $x_1, \dots, x_n \in \mathbb{R}^d$  und alle  $\delta \geq 2\beta_n^2\delta_{n,t}$  (2.9) erfüllt. Somit sind die Voraussetzungen von Satz 2.1 erfüllt, und es folgt

$$\begin{aligned} &\int |q_{n,t}^{w,p}(x) - q_t(x)|^2 \mathbf{P}_{X_t}(dx) \\ &= O_{\mathbf{P}} \left( \sum_{s=t}^{T-1} \frac{c_{11} \cdot \beta_n^4 \cdot K_n \cdot \log(n)}{n} + \max_{s \in \{t, t+1, \dots, T-1\}} \inf_{\substack{f \in \mathcal{F}_{K_n}(\beta_n) \\ \|f\|_{\infty} \leq \beta_{n+1}}} \int |f(x) - q_s(x)|^2 \mathbf{P}_{X_s}(dx) \right) \end{aligned}$$

für alle  $t \in \{0, 1, \dots, T\}$ . Weiter impliziert die Voraussetzung  $|f_t(x)| \leq L$  für alle  $x \in \mathbb{R}^d$  und  $t \in \{0, 1, \dots, T\}$ , dass  $q_t$  absolut beschränkt ist durch  $L$ . Da  $q_t$  absolut beschränkt ist, liefert Lemma A.15 (wegen  $K_n \rightarrow \infty$  für  $n \rightarrow \infty$ ) das gewünschte Endresultat

$$\max_{s \in \{t, t+1, \dots, T-1\}} \inf_{\substack{f \in \mathcal{F}_{K_n}(\beta_n) \\ \|f\|_{\infty} \leq \beta_{n+1}}} \int |f(x) - q_s(x)|^2 \mathbf{P}_{X_s}(dx) \rightarrow 0.$$

□

Als nächstes wird die Konvergenzgeschwindigkeit des Neuronale-Netze-Schätzers analysiert. Dazu wird als erstes angenommen, dass der stochastische Prozess  $X_t$  ( $t \in \{0, 1, \dots, T\}$ ) beschränkt ist gemäß des zweiten und dritten Abschnittes im Abschnitt 2.2.

Als zweites wird die Glattheit der Kontinuierungswerte eingeschränkt, indem Forderungen an ihre Fourier-Transformierte gestellt werden (siehe unten). Dies erfolgt, da in der nichtparametrischen Regression Glattheitsvoraussetzungen an die Regressionsfunktion



notwendig sind, um nicht-triviale Resultate hinsichtlich der Konvergenzrate herleiten zu können (siehe z. B. Cover (1968), Devroye (1982) oder Kapitel 3 in Györfi et al. (2002)).

Im Folgenden wird die Klasse von Funktionen, die eine Fourier-Transformierte mit endlichem ersten absoluten Moment besitzen, eingeführt. Die Fourier-Transformierte  $\tilde{F}$  von einer Funktion  $f \in L_1(\mathbb{R}^d)$  ist definiert durch

$$\tilde{F}(\omega) = \frac{1}{(2\pi)^{d/2}} \int_{\mathbb{R}^d} e^{-i\omega^T x} f(x) dx \quad (\omega \in \mathbb{R}^d).$$

Falls  $\tilde{F} \in L_1(\mathbb{R}^d)$ , dann gilt die Umkehrformel

$$f(x) = \frac{1}{(2\pi)^{d/2}} \int_{\mathbb{R}^d} e^{i\omega^T x} \tilde{F}(\omega) d\omega \quad (2.16)$$

fast überall hinsichtlich des Lebesgue-Maßes. Sei  $0 < \Gamma < \infty$ . Betrachtet wird die Klasse von Funktionen  $\mathcal{F}_\Gamma$ , für die (2.16) auf  $\mathbb{R}^d$  und außerdem

$$\int_{\mathbb{R}^d} \|\omega\| F(\omega) d\omega \leq \Gamma \quad (2.17)$$

gilt. Eine Klasse von Funktionen, die (2.17) erfüllen, ist eine Unterklasse der Funktionen, die eine Fourier-Transformierte mit endlichem ersten absoluten Moment besitzen, d. h.  $\int_{\mathbb{R}^d} \|\omega\| F(\omega) d\omega < \infty$  (diese Funktionen sind stetig differenzierbar auf  $\mathbb{R}^d$ ). Das nächste Korollar liefert nun die Konvergenzrate des Neuronale-Netze-Schätzers.

**Korollar 2.13 (Konvergenzgeschwindigkeit)** Sei  $L > 0$ . Sei  $X_0, X_1, \dots, X_T$  ein  $\mathbb{R}^d$ -wertiger Markov-Prozess, und sei  $X_t \in [-A, A]^d$  fast sicher für ein  $A > 0$  und für alle  $t \in \{0, 1, \dots, T\}$ . Des Weiteren sei die diskontierte Auszahlungsfunktion  $f_t$  absolut beschränkt durch  $L$ , d. h.

$$|f_t(x)| \leq L \text{ für } x \in \mathbb{R}^d \text{ und } t \in \{0, 1, \dots, T\}.$$

Sei  $q_t$  der Kontinuierungswert definiert durch (1.12). Des Weiteren sei  $w \in \{0, 1, \dots, T - t - 1\}$  fest gewählt. Nehme an, dass die Fourier-Transformierte  $\tilde{Q}_t$  von  $q_t$  für alle  $x \in \mathbb{R}^d$  und alle  $t \in \{0, \dots, T\}$  (2.16) und (2.17) genügt. Setze  $\beta_n = \text{const} \cdot \log(n)$  sowie  $K_n = \lceil (n/(\log^5(n)))^{1/2} \rceil$ , und definiere den Schätzer  $q_{n,t}^{w,p}$  (mit  $p = p_n = K_n$ ) für alle  $t \in \{0, \dots, T\}$  durch (2.6) und (2.7) mit  $\mathcal{F}_{n,p} = \mathcal{F}_{K_n}(\beta_n)$  sowie mit Daten definiert in (2.5). Dann gilt

$$\int |q_{n,t}^{w,p}(x) - q_t(x)|^2 \mathbf{P}_{X_t}(dx) = O_{\mathbf{P}} \left( \sqrt{\frac{\log^5(n)}{n}} \right)$$

für alle  $t \in \{0, \dots, T\}$ .

**Beweis:** Da im Beweis Satz 2.1 zum Einsatz kommt, setze  $\delta_{n,t} = c_{11} \cdot \beta_n^4 \cdot K_n \cdot \log(n)/n$ . Mit dieser Wahl von  $\delta_{n,t}$  sind die Voraussetzungen im obigen Satz erfüllt, was bereits im Beweis von Korollar 2.9 verifiziert wurde.

Aus Lemma A.9 erhält man für  $n$  genügend groß

$$\max_{s \in \{t, t+1, \dots, T-1\}} \inf_{\substack{f \in \mathcal{F}_{K_n(\beta_n)} \\ \|f\|_\infty \leq \beta_{n+1}}} \int |f(x) - q_s(x)|^2 \mathbf{P}_{X_s}(dx) \leq \frac{(2\sqrt{d}A\Gamma)^2}{K_n}.$$

Satz 2.1 impliziert weiter

$$\begin{aligned} & \int |q_{n,t}^{w,p}(x) - q_t(x)|^2 \mathbf{P}_{X_t}(dx) \\ &= O_{\mathbf{P}} \left( \sum_{s=t}^{T-1} \frac{c_{11} \cdot \beta_n^4 \cdot K_n \cdot \log(n)}{n} + \max_{s \in \{t, t+1, \dots, T-1\}} \inf_{\substack{f \in \mathcal{F}_{K_n(\beta_n)} \\ \|f\|_\infty \leq \beta_{n+1}}} \int |f(x) - q_s(x)|^2 \mathbf{P}_{X_s}(dx) \right) \\ &= O_{\mathbf{P}} \left( \sum_{s=t}^{T-1} \frac{c_{11} \cdot \beta_n^4 \cdot K_n \cdot \log(n)}{n} + \frac{(2\sqrt{d}A\Gamma)^2}{K_n} \right) \\ &= O_{\mathbf{P}} \left( \sqrt{\frac{\log^5(n)}{n}} \right) \end{aligned}$$

für alle  $t \in \{0, 1, \dots, T\}$ . □

### Bemerkungen.

**Bemerkung 2.14** *Es sei hier betont, dass in Korollar 2.9 keine Voraussetzung an die Verteilung von  $X_t$  gestellt wird. In Korollar 2.13 wird - außer dass  $X_t \in [-A, A]^d$  fast sicher gilt - ebenfalls keine Voraussetzung an die Verteilung von  $X_t$  gestellt. Insbesondere wird in beiden Korollaren nicht die Existenz einer Dichte von  $X_t$  bezüglich des Lebesgue-Borel-Maßes gefordert.*

**Bemerkung 2.15** *Als Konvergenzrate erhält man in Korollar 2.13 bis auf einen logarithmischen Faktor  $(\log^5(n))^{1/2}$  eine parametrische Konvergenzrate  $n^{-1/2}$ .*

**Bemerkung 2.16** *Angenommen,  $X_0 = x_0$  fast sicher für ein  $x_0 \in \mathbb{R}^d$ . Der Preis*

$$V_0 = \max\{f_0(x_0), q_0(x_0)\}$$

*einer Amerikanischen Option (vgl. (1.10) und (1.12)) kann durch*

$$\hat{V}_0 = \max\{f_0(x_0), q_{n,0}^{w,p}(x_0)\}$$

*geschätzt werden. Da die Verteilung von  $X_0$  auf  $x_0$  konzentriert ist, gilt unter den Voraussetzungen in Korollar 2.13 die folgende Fehlerschranke:*

$$|\hat{V}_0 - V_0|^2 = |\max\{f_0(x_0), q_{n,0}^{w,p}(x_0)\} - \max\{f_0(x_0), q_0(x_0)\}|^2$$

$$\begin{aligned} &\leq |q_{n,0}^{w,p}(x_0) - q_0(x_0)|^2 \\ &= O_{\mathbf{P}} \left( \left( \frac{\log^5(n)}{n} \right)^{1/2} \right). \end{aligned}$$

# Kapitel 3

## Ein alternativer Zugang zur Analyse des Schätzers bei spezieller Regressionsdarstellung

In diesem Kapitel wird ein Spezialfall von Satz 2.1 formuliert und - mit einem alternativen Zugang zur Analyse des Kleinste-Quadrate-Schätzers - verifiziert. Der alternative Zugang entspricht dabei der Anwendung von Resultaten aus der Theorie der Regressionsschätzung mit Fehlern in der abhängigen Variable (die in Kohler (2006a) primär im Zusammenhang mit der Regressionsschätzung bei zensierten Daten eingeführt wurde (siehe auch Dippon und Winter (2006) für eine entsprechende Anwendung)), die hier bei der Schätzung des Wertes einer Amerikanischen Option angewendet werden. Dieser Zugang führt zu einem einfachen Beweis des Spezialfalls. Eine Beschreibung des alternativen Zugangs zur Analyse des Schätzers sowie die Formulierung des Spezialfalls von Satz 2.1 erfolgen in Abschnitt 3.1. Anschließend wird in Abschnitt 3.2 der Spezialfall aus Abschnitt 3.1 bewiesen.

### 3.1 Ein alternativer Zugang

Im Folgenden wird kurz die Theorie der Regressionsschätzung mit Fehlern in der abhängigen Variable beschrieben. Dazu seien  $(X, Y), (X_1, Y_1), \dots$  unabhängig identisch verteilte  $\mathbb{R}^d \times \mathbb{R}$ -wertige Zufallsvariablen mit  $\mathbf{E}Y^2 < \infty$ . Des Weiteren sei  $m(x) = \mathbf{E}\{Y|X = x\}$  die dazugehörige Regressionsfunktion. Angenommen, man möchte  $m$  ausgehend von beobachteten Daten schätzen. In besagter Theorie wird davon ausgegangen, dass anstatt der Stichprobe

$$\mathcal{D}_n = \{(X_1, Y_1), \dots, (X_n, Y_n)\}$$

von  $(X, Y)$  nur eine Datenmenge

$$\bar{\mathcal{D}}_n = \{(X_1, \bar{Y}_1), \dots, (X_n, \bar{Y}_n)\}$$

verfügbar ist, bei der die einzige Voraussetzung an  $\bar{Y}_1, \dots, \bar{Y}_n$  ist, dass der Messfehler

$$\frac{1}{n} \sum_{i=1}^n |Y_i - \bar{Y}_i|^2$$

klein ist. Insbesondere wird hier nicht vorausgesetzt, dass die Zufallsvariablen in  $\bar{\mathcal{D}}_n$  unabhängig oder identisch verteilt sein müssen. Auch muss  $\mathbf{E}\{\bar{Y}_i|X = x\}$  nicht gleich  $m(x)$  sein.

Es könnte nun die Frage aufkommen, welchen Einfluss der Messfehler  $1/n \sum_{i=1}^n |Y_i - \bar{Y}_i|^2$  auf den  $L_2$ -Fehler eines Regressionsschätzers angewandt auf die Daten  $\bar{\mathcal{D}}_n$  besitzt. Da keine Annahmen an die Differenz zwischen den wahren  $y$ -Werten  $Y_i$  und den beobachteten Daten  $\bar{Y}_i$  gestellt werden, außer dass der Messfehler klein ist, ist es klar, dass keine Möglichkeit besteht, den Messfehler loszuwerden. In Kohler (2006a) wurde aber für Kleinste-Quadrate-Schätzer bewiesen, dass ein kleiner Messfehler  $1/n \sum_{i=1}^n |Y_i - \bar{Y}_i|^2$  nur leicht den  $L_2$ -Fehler eines geeignet definierten Regressionsschätzers beeinflusst.

Die Idee, Regressionsschätzung mit zusätzlichen Messfehlern in der abhängigen Variable zu betrachten, stammt von Kohler (2006a). Dort wurden Resultate hinsichtlich der Konvergenzrate von Kleinste-Quadrate-Schätzern hergeleitet. Des Weiteren wurden diese Resultate verwendet, um die Konvergenzrate von Regressionsschätzern, die auf zensierten Daten basieren, zu analysieren. In diesem Kontext wurden in Kohler, Kul und Mathé (2004) adaptive Kleinste-Quadrate-Schätzer basierend auf der Unterteilung der Stichprobe (mehr zur Unterteilung der Stichprobe siehe Kapitel 4) untersucht. Auch wurde die universelle Konsistenz von diesen Schätzern nachgewiesen.

Der Zusammenhang zwischen der Theorie der Regressionsschätzung mit zusätzlichen Messfehlern in der abhängigen Variable und der Optionsbewertung ist durch folgende Betrachtungsweise gegeben : Interpretiere (2.4), also

$$Y_{i,t}^{(t)} = \theta_{t+1:w}(f, q_n^{w,p})(X_{i,t+1}^{(t)}, \dots, X_{i,t+w+1}^{(t)})$$

mit  $i = 1, \dots, n$ , als die abhängige Variable mit zusätzlichen Messfehlern. Dann kann (2.5), also

$$\{(X_{i,t}^{(t)}, Y_{i,t}^{(t)}) : i = 1, \dots, n\},$$

als die obige Datenmenge  $\bar{\mathcal{D}}_n$  gedeutet werden. Ausgehend von (2.5) wird die Regressionsfunktion  $q_t$  geschätzt.

Mit dieser Betrachtungsweise sind nun Resultate aus der Theorie der Regressionsschätzung mit Fehlern in der abhängigen Variable, im Zusammenhang mit der Schätzung des Wertes einer Amerikanischen Option, anwendbar.

Bei der in Egloff (2005) vorgestellten Regressionsdarstellung (2.1) kann der Parameter  $w$  Werte aus  $\{0, 1, \dots, T - t - 1\}$  annehmen. Während der Entstehung dieser Arbeit hat sich anhand von Simulationen herauskristallisiert, dass zumindest bei dem Kleinste-Quadrate-Splineschätzer - im Zusammenhang mit der Schätzung des Wertes einer Amerikanischen Option - das wählbare  $w$  keine signifikanten Verbesserungen mit sich bringt (siehe Kapitel 5). Aus diesem Grund wird  $w = 0$  gesetzt.

Als nächstes folgt ein Spezialfall von Satz 2.1. Der Unterschied zu Satz 2.1 besteht darin, dass im Spezialfall  $w \equiv 0$  gesetzt wird.

**Satz 3.1** *Sei  $\beta_n \geq L > 0$ . Nehme an, dass  $X_0, X_1, \dots, X_T$  ein  $\mathbb{R}^d$ -wertiger Markov-Prozess ist und dass die diskontierte Auszahlungsfunktion  $f_t$  absolut beschränkt ist durch  $L$ . Sei  $q_t$  der Kontinuierungswert definiert durch (1.12). Des Weiteren sei  $w \equiv 0$ . Der Schätzer  $q_{n,t}^{0,p}$  (mit  $p \in \mathcal{P}_n$  fest) sei definiert durch (2.6) und (2.7) mit Daten definiert in (2.5). Sei  $c_1 > 0$ . Angenommen, es existiert ein  $\delta_{n,t} > 0$  mit den Eigenschaften*

$$\delta_{n,t} \rightarrow 0 \quad (n \rightarrow \infty), \quad \frac{n \cdot \delta_{n,t}}{\beta_n^2} \rightarrow \infty \quad (n \rightarrow \infty)$$

und

$$c_1 \sqrt{n} \delta \geq \int_{\delta/(2^9 \sigma_0)}^{\sqrt{\delta}} \left( \log \mathcal{N}_2 \left( u, \left\{ f - g : f \in \mathcal{F}_{n,p}, \frac{1}{n} \sum_{i=1}^n |f(x_i) - g(x_i)|^2 \leq \delta \right\}, x_1^n \right) \right)^{1/2} du \quad (3.1)$$

für alle  $\delta \geq \delta_{n,t}$  und alle  $g \in \mathcal{F}_{n,p}$ , sowie dass für alle  $x_1, \dots, x_n \in \mathbb{R}^d$  und alle  $\delta \geq 2\beta_n^2 \delta_{n,t}$

$$\frac{0.5 \sqrt{n} \delta}{768 \sqrt{2} \beta_n^2} \geq \int_{0.5\delta/(128\beta_n^2)}^{\sqrt{\delta}} \left( \log \mathcal{N}_2 \left( \frac{u}{4\beta_n}, \left\{ f - q_t : f \in T_{\beta_n} \mathcal{F}_{n,p}, \frac{1}{n} \sum_{i=1}^n |f(x_i) - q_t(x_i)|^2 \leq \frac{\delta}{\beta_n^2} \right\}, x_1^n \right) \right)^{1/2} du. \quad (3.2)$$

Dann gilt

$$\begin{aligned} & \int |q_{n,t}^{0,p}(x) - q_t(x)|^2 \mathbf{P}_{X_t}(dx) \\ &= O_{\mathbf{P}} \left( \sum_{s=t}^{T-1} \delta_{n,s} + \max_{s \in \{t, t+1, \dots, T-1\}} \inf_{\substack{f \in \mathcal{F}_{n,p} \\ \|f\|_{\infty} \leq \beta_{n+1}}} \int |f(x) - q_s(x)|^2 \mathbf{P}_{X_s}(dx) \right) \end{aligned}$$

für alle  $t \in \{0, \dots, T\}$ .

## 3.2 Beweis von Satz 3.1

**Beweis von Satz 3.1:** Bevor auf den Beweis eingegangen wird, ist zu beachten, dass die Beschränktheit der diskontierten Auszahlungsfunktion  $f_t$  durch  $L$   $|q_t(x)| \leq L$  für alle  $x \in \mathbb{R}^d$  impliziert. Des Weiteren impliziert die Definition von  $q_{n,t}^{0,p}$   $|q_{n,t}^{0,p}(x)| \leq \beta_n$  für alle  $x \in \mathbb{R}^d$ .

Im Folgenden wird gezeigt, dass

$$\begin{aligned} & \int |q_{n,s}^{0,p}(x) - q_s(x)|^2 \mathbf{P}_{X_s}(dx) \\ &= O_{\mathbf{P}} \left( \sum_{t=s}^{T-1} \delta_{n,t} + \max_{t \in \{s, s+1, \dots, T-1\}} \inf_{\substack{f \in \mathcal{F}_{n,p} \\ \|f\|_{\infty} \leq \beta_{n+1}}} \int |f(x) - q_t(x)|^2 \mathbf{P}_{X_t}(dx) \right) \quad (3.3) \end{aligned}$$

für alle  $s \in \{0, \dots, T\}$  gilt.

Für  $s = T$  gilt  $q_{n,t}^{0,p}(x) = 0 = q_T(x)$ , und somit ist die Aussage trivial. Sei  $t < T$ , und nehme an, dass die Aussage für jedes  $s \in \{t+1, \dots, T-1\}$  gilt. Durch Induktion genügt es, (3.3) für  $s = t$  zu zeigen.

Sei

$$\mathcal{D}_{n,t} = \{X_{j,s}^{(r)} : r \geq t+1, s \in \{0, \dots, T-1\}, j \in \{1, \dots, n\}\},$$

d. h. in  $\mathcal{D}_{n,t}$  sind alle Daten, die zur Konstruktion von  $q_{n,t+1}^{0,p}$  genutzt werden.

Im ersten Schritt des Beweises wird die Beziehung

$$\int |q_{n,t}^{0,p}(x) - q_t(x)|^2 \mathbf{P}_{X_t}(dx) = O_{\mathbf{P}} \left( \frac{1}{n} \sum_{i=1}^n |q_{n,t}^{0,p}(X_{i,t}^{(t)}) - q_t(X_{i,t}^{(t)})|^2 + \delta_{n,t} \right)$$

gezeigt. Um obige Gleichung zu verifizieren, wird Lemma A.8 verwendet. Zuvor sei auf die folgenden Beziehungen verwiesen:

$$\begin{aligned} & \mathbf{P} \left\{ \int |q_{n,t}^{0,p}(x) - q_t(x)|^2 \mathbf{P}_{X_t}(dx) > 3 \frac{1}{n} \sum_{i=1}^n |q_{n,t}^{0,p}(X_{i,t}^{(t)}) - q_t(X_{i,t}^{(t)})|^2 + \delta_{n,t} \mid \mathcal{D}_{n,t} \right\} \\ &= \mathbf{P} \left\{ \int |q_{n,t}^{0,p}(x) - q_t(x)|^2 \mathbf{P}_{X_t}(dx) - 3 \frac{1}{n} \sum_{i=1}^n |q_{n,t}^{0,p}(X_{i,t}^{(t)}) - q_t(X_{i,t}^{(t)})|^2 > \delta_{n,t} \mid \mathcal{D}_{n,t} \right\} \\ &\leq \mathbf{P} \left\{ \exists f \in T_{\beta_n} \mathcal{F}_{n,p} : \int |f(x) - q_t(x)|^2 \mathbf{P}_{X_t}(dx) - 3 \frac{1}{n} \sum_{i=1}^n |f(X_{i,t}^{(t)}) - q_t(X_{i,t}^{(t)})|^2 > \delta_{n,t} \mid \mathcal{D}_{n,t} \right\} \\ &\leq \mathbf{P} \left\{ \exists f \in T_{\beta_n} \mathcal{F}_{n,p} : \frac{\int |f(x) - q_t(x)|^2 \mathbf{P}_{X_t}(dx) - \frac{1}{n} \sum_{i=1}^n |f(X_{i,t}^{(t)}) - q_t(X_{i,t}^{(t)})|^2}{\delta_{n,t} + \int |f(x) - q_t(x)|^2 \mathbf{P}_{X_t}(dx) + \frac{1}{n} \sum_{i=1}^n |f(X_{i,t}^{(t)}) - q_t(X_{i,t}^{(t)})|^2} > \frac{1}{2} \mid \mathcal{D}_{n,t} \right\} \\ &\leq \mathbf{P} \left\{ \sup_{f \in T_{\beta_n} \mathcal{F}_{n,p}} \frac{\int |f(x) - q_t(x)|^2 \mathbf{P}_{X_t}(dx) - \frac{1}{n} \sum_{i=1}^n |f(X_{i,t}^{(t)}) - q_t(X_{i,t}^{(t)})|^2}{\delta_{n,t} + \int |f(x) - q_t(x)|^2 \mathbf{P}_{X_t}(dx) + \frac{1}{n} \sum_{i=1}^n |f(X_{i,t}^{(t)}) - q_t(X_{i,t}^{(t)})|^2} > \frac{1}{2} \mid \mathcal{D}_{n,t} \right\} \end{aligned}$$

$$\leq \mathbf{P} \left\{ \sup_{f \in T_{\beta_n} \mathcal{F}_{n,p}} \frac{\int |f(x) - q_t(x)|^2 \mathbf{P}_{X_t}(dx) - \frac{1}{n} \sum_{i=1}^n |f(X_{i,t}^{(t)}) - q_t(X_{i,t}^{(t)})|^2}{\delta_{n,t} + \int |f(x) - q_t(x)|^2 \mathbf{P}_{X_t}(dx) + \frac{1}{n} \sum_{i=1}^n |f(X_{i,t}^{(t)}) - q_t(X_{i,t}^{(t)})|^2} > \frac{1}{2} \mid \mathcal{D}_{n,t} \right\}.$$

Um Lemma A.8 auf die letzte Wahrscheinlichkeit anwenden zu können, setzt man

$$\epsilon = 0.5, \quad \alpha = \delta_{n,t}, \quad B = \beta_n$$

und beachtet, dass mit diesen Notationen (A.4) durch (3.2) erfüllt ist, sowie für hinreichend großes  $n$  (A.3) durch die Annahme  $n \cdot \delta_{n,t} / \beta_n^2 \rightarrow \infty$  für  $n \rightarrow \infty$  impliziert wird. Des Weiteren gilt  $|q_t(x)| \leq L$  ( $x \in \mathbb{R}^d$ ) und  $|q_{n,t}^{0,p}(x)| \leq \beta_n$  ( $x \in \mathbb{R}^d$ ) sowie dass der betrachtete Funktionenraum  $T_{\beta_n} \mathcal{F}_{n,p}$  beschränkt ist durch  $\beta_n$ . Da nun alle Voraussetzungen von Lemma A.8 erfüllt sind, erhält man mit genügend großem  $n$  für die letzte Wahrscheinlichkeit eine obere Schranke gemäß

$$15 \exp \left( -\frac{n \cdot \delta_{n,t} \cdot 1/4}{512 \cdot 2304 \beta_n^2} \right) \rightarrow 0 \quad (n \rightarrow \infty),$$

da nach Voraussetzung  $n \delta_{n,t} / \beta_n^2 \rightarrow \infty$  für  $n$  gegen unendlich gilt.

Im zweiten Schritt des Beweises wird die Beziehung

$$\begin{aligned} & \frac{1}{n} \sum_{i=1}^n |q_{n,t}^{0,p}(X_{i,t}^{(t)}) - q_t(X_{i,t}^{(t)})|^2 \\ &= O_{\mathbf{P}} \left( \frac{1}{n} \sum_{i=1}^n |q_{n,t+1}^{0,p}(X_{i,t+1}^{(t)}) - q_{t+1}(X_{i,t+1}^{(t)})|^2 + \delta_{n,t} + \inf_{f \in \mathcal{F}_{n,p}} \frac{1}{n} \sum_{i=1}^n |f(X_{i,t}^{(t)}) - q_t(X_{i,t}^{(t)})|^2 \right) \end{aligned}$$

gezeigt. Um dies zu verifizieren, wird Lemma A.1 angewandt. Um Lemma A.1 anwenden zu können, wird auf  $X_i = X_{i,t}^{(t)}$  konditioniert ( $Y_i$  sowie  $\bar{Y}_i$  siehe unten). Die Voraussetzung (A.1) im Lemma ist erfüllt, da (nach Definition)  $Y_{i,t}^{(t)}$  beschränkt ist. Wird  $\delta_n = \delta_{n,t}$  gesetzt, so wird auch für alle  $\delta \geq \delta_{n,t}$  und alle  $g \in \mathcal{F}_{n,p}$  die Voraussetzung (A.2) durch die Annahme (3.1) erfüllt. Weiter gilt  $\delta_n \rightarrow 0$  und  $n \cdot \delta_n \rightarrow \infty$  für  $n$  gegen unendlich. Im Kontext von Lemma A.1 ist

$$X_i = X_{i,t}^{(t)}, \quad Y_i = \max\{f_{t+1}(X_{i,t+1}^{(t)}), q_{t+1}(X_{i,t+1}^{(t)})\} \text{ und } \bar{Y}_i = \max\{f_{t+1}(X_{i,t+1}^{(t)}), q_{n,t+1}^{0,p}(X_{i,t+1}^{(t)})\}.$$

Mit der Beziehung  $|\max\{a, b\} - \max\{a, c\}| \leq |b - c|$  erhält man

$$\frac{1}{n} \sum_{i=1}^n |Y_i - \bar{Y}_i|^2 \leq \frac{1}{n} \sum_{i=1}^n |q_{t+1}(X_{i,t+1}^{(t)}) - q_{n,t+1}^{0,p}(X_{i,t+1}^{(t)})|^2,$$

und die zu zeigende Beziehung folgt mit Lemma A.1, wenn man auf  $\mathcal{D}_{n,t}$  konditioniert.

Im dritten Schritt des Beweises wird die Beziehung

$$\frac{1}{n} \sum_{i=1}^n |q_{n,t+1}^{0,p}(X_{i,t+1}^{(t)}) - q_{t+1}(X_{i,t+1}^{(t)})|^2 = O_{\mathbf{P}} \left( \int |q_{n,t+1}^{0,p}(x) - q_{t+1}(x)|^2 \mathbf{P}_{X_{t+1}}(dx) + \delta_{n,t+1} \right)$$



gezeigt. Analog wie im ersten Schritt des Beweises wird diese Behauptung für genügend großes  $n$  mit Lemma A.8 – konditioniert auf  $\mathcal{D}_{n,t}$  – verifiziert. Der Vollständigkeit halber sei erwähnt, dass

$$\frac{1}{n} \sum_{i=1}^n |q_{n,t+1}^{0,p}(X_{i,t+1}^{(t)}) - q_{t+1}(X_{i,t+1}^{(t)})|^2 > 3 \int |q_{n,t+1}^{0,p}(x) - q_{t+1}(x)|^2 \mathbf{P}_{X_{t+1}}(dx) + \delta_{n,t+1}$$

äquivalent ist zu

$$\begin{aligned} & 2 \frac{1}{n} \sum_{i=1}^n |q_{n,t+1}^{0,p}(X_{i,t+1}^{(t)}) - q_{t+1}(X_{i,t+1}^{(t)})|^2 - 2 \int |q_{n,t+1}^{0,p}(x) - q_{t+1}(x)|^2 \mathbf{P}_{X_{t+1}}(dx) \\ & > \frac{1}{n} \sum_{i=1}^n |q_{n,t+1}^{0,p}(X_{i,t+1}^{(t)}) - q_{t+1}(X_{i,t+1}^{(t)})|^2 + \int |q_{n,t+1}^{0,p}(x) - q_{t+1}(x)|^2 \mathbf{P}_{X_{t+1}}(dx) + \delta_{n,t+1} \end{aligned}$$

was wiederum äquivalent ist zu

$$\frac{\frac{1}{n} \sum_{i=1}^n |q_{n,t+1}^{0,p}(X_{i,t+1}^{(t)}) - q_{t+1}(X_{i,t+1}^{(t)})|^2 - \int |q_{n,t+1}^{0,p}(x) - q_{t+1}(x)|^2 \mathbf{P}_{X_{t+1}}(dx)}{\delta_{n,t+1} + \frac{1}{n} \sum_{i=1}^n |q_{n,t+1}^{0,p}(X_{i,t+1}^{(t)}) - q_{t+1}(X_{i,t+1}^{(t)})|^2 + \int |q_{n,t+1}^{0,p}(x) - q_{t+1}(x)|^2 \mathbf{P}_{X_{t+1}}(dx)} > \frac{1}{2},$$

und dass die letzte Ungleichung (wie im ersten Schritt) die folgende Aussage impliziert:

$$\frac{|\frac{1}{n} \sum_{i=1}^n |q_{n,t+1}^{0,p}(X_{i,t+1}^{(t)}) - q_{t+1}(X_{i,t+1}^{(t)})|^2 - \int |q_{n,t+1}^{0,p}(x) - q_{t+1}(x)|^2 \mathbf{P}_{X_{t+1}}(dx)|}{\delta_{n,t+1} + \int |q_{n,t+1}^{0,p}(x) - q_{t+1}(x)|^2 \mathbf{P}_{X_{t+1}}(dx) + \frac{1}{n} \sum_{i=1}^n |q_{n,t+1}^{0,p}(X_{i,t+1}^{(t)}) - q_{t+1}(X_{i,t+1}^{(t)})|^2} > \frac{1}{2}.$$

Im vierten Schritt des Beweises wird die Beziehung

$$\begin{aligned} & \int |q_{n,t+1}^{0,p}(x) - q_{t+1}(x)|^2 \mathbf{P}_{X_{t+1}}(dx) \\ & = O_{\mathbf{P}} \left( \sum_{s=t+1}^{T-1} \delta_{n,s} + \max_{s \in \{t+1, \dots, T-1\}} \inf_{\substack{f \in \mathcal{F}_{n,p} \\ \|f\|_{\infty} \leq \beta_{n+1}}} \int |f(x) - q_s(x)|^2 \mathbf{P}_{X_s}(dx) \right) \end{aligned}$$

gezeigt. Dies folgt mit Induktion und (3.3).

Im fünften (und letzten) Schritt des Beweises wird die Beziehung

$$\begin{aligned} & \inf_{f \in \mathcal{F}_{n,p}} \frac{1}{n} \sum_{i=1}^n |f(X_{i,t}^{(t)}) - q_t(X_{i,t}^{(t)})|^2 \\ & = O_{\mathbf{P}} \left( \sum_{s=t}^{T-1} \delta_{n,s} + \max_{s \in \{t, t+1, \dots, T-1\}} \inf_{\substack{f \in \mathcal{F}_{n,p} \\ \|f\|_{\infty} \leq \beta_{n+1}}} \int |f(x) - q_s(x)|^2 \mathbf{P}_{X_s}(dx) \right) \end{aligned}$$

gezeigt. Um das zu zeigen, beachte man

$$\begin{aligned}
 & \inf_{f \in \mathcal{F}_{n,p}} \frac{1}{n} \sum_{i=1}^n |f(X_{i,t}^{(t)}) - q_t(X_{i,t}^{(t)})|^2 \\
 & \leq \inf_{\substack{f \in \mathcal{F}_{n,p} \\ \|f\|_\infty \leq \beta_n + 1}} \frac{1}{n} \sum_{i=1}^n |f(X_{i,t}^{(t)}) - q_t(X_{i,t}^{(t)})|^2 \\
 & = \inf_{\substack{f \in \mathcal{F}_{n,p} \\ \|f\|_\infty \leq \beta_n + 1}} \left( \frac{1}{n} \sum_{i=1}^n |f(X_{i,t}^{(t)}) - q_t(X_{i,t}^{(t)})|^2 - 3 \int |f(x) - q_t(x)|^2 \mathbf{P}_{X_t}(dx) \right. \\
 & \quad \left. + 3 \int |f(x) - q_t(x)|^2 \mathbf{P}_{X_t}(dx) \right) \\
 & \leq \sup_{\substack{f \in \mathcal{F}_{n,p} \\ \|f\|_\infty \leq \beta_n + 1}} \left( \frac{1}{n} \sum_{i=1}^n |f(X_{i,t}^{(t)}) - q_t(X_{i,t}^{(t)})|^2 - 3 \int |f(x) - q_t(x)|^2 \mathbf{P}_{X_t}(dx) \right) \\
 & \quad + 3 \inf_{\substack{f \in \mathcal{F}_{n,p} \\ \|f\|_\infty \leq \beta_n + 1}} \int |f(x) - q_t(x)|^2 \mathbf{P}_{X_t}(dx),
 \end{aligned}$$

wobei für den linken Term in letzter Ungleichung analog wie im ersten und dritten Schritt für genügend großes  $n$  mit Lemma A.8 – konditioniert auf  $\mathcal{D}_{n,t}$  – gezeigt werden kann, dass

$$\sup_{\substack{f \in \mathcal{F}_{n,p} \\ \|f\|_\infty \leq \beta_n + 1}} \left( \frac{1}{n} \sum_{i=1}^n |f(X_{i,t}^{(t)}) - q_t(X_{i,t}^{(t)})|^2 - 3 \int |f(x) - q_t(x)|^2 \mathbf{P}_{X_t}(dx) \right) = O_{\mathbf{P}}(\delta_{n,t})$$

gilt. Um dies zu sehen ist zu beachten, dass nach Voraussetzung  $\beta_n \geq L$  gilt. Des Weiteren folgt mit  $|q_t(x)| \leq \beta_n$  für  $x \in \mathbb{R}^d$  und

$$\{f \in \mathcal{F}_{n,p} : \|f\|_\infty \leq \beta_n + 1\} \subseteq T_{\beta_n + 1} \mathcal{F}_{n,p}$$

die Beziehung

$$\begin{aligned}
 & \mathbf{P} \left\{ \sup_{\substack{f \in \mathcal{F}_{n,p} \\ \|f\|_\infty \leq \beta_n + 1}} \left( \frac{1}{n} \sum_{i=1}^n |f(X_{i,t}^{(t)}) - q_t(X_{i,t}^{(t)})|^2 - 3 \int |f(x) - q_t(x)|^2 \mathbf{P}_{X_t}(dx) \right) > \delta_{n,t} \mid \mathcal{D}_{n,t} \right\} \\
 & \leq \mathbf{P} \left\{ \sup_{f \in T_{\beta_n + 1} \mathcal{F}_{n,p}} \left( \frac{1}{n} \sum_{i=1}^n |f(X_{i,t}^{(t)}) - q_t(X_{i,t}^{(t)})|^2 - 3 \int |f(x) - q_t(x)|^2 \mathbf{P}_{X_t}(dx) \right) \right. \\
 & \quad \left. > \delta_{n,t} \mid \mathcal{D}_{n,t} \right\} \\
 & \leq \mathbf{P} \left\{ \exists f \in T_{\beta_n + 1} \mathcal{F}_{n,p} : \right. \\
 & \quad \left. \frac{1}{n} \sum_{i=1}^n |f(X_{i,t}^{(t)}) - q_t(X_{i,t}^{(t)})|^2 - 3 \int |f(x) - q_t(x)|^2 \mathbf{P}_{X_t}(dx) > \delta_{n,t} \mid \mathcal{D}_{n,t} \right\}.
 \end{aligned}$$

Durch das Erfassen aller Resultate aus den obigen fünf Schritten ist Satz 3.1 vollständig bewiesen.  $\square$

# Kapitel 4

## Adaptive Wahl von Parametern

Die erzielten Resultate der vorangegangenen Kapitel beziehen sich auf Kleinste-Quadrate-Schätzer, deren Definition (siehe Abschnitt 2.1) von der Wahl des Parameters  $p \in \mathcal{P}_n$  abhängt. In diesem Kapitel wird diese Abhängigkeit der Schätzer von  $p \in \mathcal{P}_n$  umgangen, indem eine adaptive Wahl der Parameter erfolgt. Dazu wird in Abschnitt 4.1 ein von  $p \in \mathcal{P}_n$  unabhängiger Schätzer definiert und ein Resultat betreffend der Adaption formuliert. Mit Hilfe dieses Resultats werden anschließend in Abschnitt 4.2 Resultate hinsichtlich Konsistenz und Konvergenzgeschwindigkeit betreffend den in Abschnitt 4.2 neu definierten Schätzern erzielt. Dabei werden die in Abschnitt 4.2 neu definierten Schätzer im Zusammenhang mit der Schätzung des Wertes einer Amerikanischen Option eingesetzt.

### 4.1 Adaption

Wie oben bereits erläutert, beziehen sich die Resultate der vorangegangenen Kapitel auf Kleinste-Quadrate-Schätzer, deren Definition (siehe Abschnitt 2.1) von der Wahl des Parameters  $p \in \mathcal{P}_n$  abhängt. Im weiteren Verlauf dieser Arbeit werden nun Resultate formuliert, bei denen die darin verwendeten Kleinste-Quadrate-Schätzer unabhängig vom Parameter  $p \in \mathcal{P}_n$  sind; beispielsweise durch eine vollautomatische Parameterwahl. Diese Schätzer werden im Folgenden mittels Techniken der Adaption eingeführt. Dazu wird die Methode der Unterteilung der Stichprobe für die vollautomatische Parameterwahl verwendet. Bei dieser Methode wird die Datenmenge, auf der der Schätzer definiert wird, in zwei disjunkte Untermengen geteilt, eine so genannte Lernmenge (“learning sample”) mit Kardinalität  $n_l$  und eine Testmenge (“testing sample”) der Kardinalität  $n - n_l$ .

Des Weiteren wird einerseits angenommen, dass die nichtleere Parametermenge  $\mathcal{P}_n$  endlich ist sowie andererseits, dass für jedes  $p \in \mathcal{P}_n$  ein Kleinste-Quadrate-Schätzer existiert und wie folgt (auf den so genannten Lerndaten) definiert wird:

$$\tilde{q}_{n_l,t}^{w,p}(\cdot) = \arg \min_{f \in \mathcal{F}_{n,p}} \left( \frac{1}{n_l} \sum_{i=1}^{n_l} |f(X_{i,t}^{(t)}) - \hat{Y}_{i,t}^{(t)}|^2 \right) \quad (4.1)$$

und

$$q_{n_l,t}^{w,p}(x) = T_{\beta_n} \tilde{q}_{n_l,t}^{w,p}(x) \quad (x \in \mathbb{R}^d), \quad (4.2)$$

wobei für jedes  $p \in \mathcal{P}_n$   $\mathcal{F}_{n,p}$  ein Raum von Funktionen  $f : \mathbb{R}^d \rightarrow \mathbb{R}$  ist. Anschließend wird das empirische  $L_2$ -Risiko auf den so genannten Testdaten minimiert, um den Wert des Parameters  $p \in \mathcal{P}_n$  zu wählen. Mit

$$\hat{p} = \arg \min_{p \in \mathcal{P}_n} \left( \frac{1}{n - n_l} \sum_{i=n_l+1}^n |q_{n_l,t}^{w,p}(X_{i,t}^{(t)}) - \hat{Y}_{i,t}^{(t)}|^2 \right), \quad (4.3)$$

wobei

$$\hat{Y}_{i,t}^{(t)} = \theta_{t+1:w}(f, \hat{q}_n^w)(X_{i,t+1}^{(t)}, \dots, X_{i,t+w+1}^{(t)}) \quad (4.4)$$

für  $1 \leq i \leq n$  (vgl. (2.2)), ist der (Regressions-)Schätzer  $\hat{q}_{n,t}^w$  mit automatischer Parameterwahl der Parameter  $p \in \mathcal{P}_n$  von  $q_t$  gegeben durch

$$\hat{q}_{n,t}^w(\cdot) = q_{n_l,t}^{w,\hat{p}}(\cdot). \quad (4.5)$$

Für eine detaillierte Beschreibung zur Unterteilung der Stichprobe sei beispielsweise auf Kapitel 7 in Györfi et al. (2002) verwiesen. Als nächstes folgt ein Resultat, welches Auskunft über den integrierten quadratischen Fehler des obigen (Regressions-)Schätzers gibt.

**Satz 4.1 (Adaption)** Sei  $\beta_n \geq L > 0$ . Nehme an, dass  $X_0, X_1, \dots, X_T$  ein  $\mathbb{R}^d$ -wertiger Markov-Prozess ist und dass die diskontierte Auszahlungsfunktion  $f_t$  absolut beschränkt ist durch  $L$ . Sei  $q_t$  der Kontinuierungswert definiert durch (1.12). Des Weiteren sei  $w \in \{0, 1, \dots, T - t - 1\}$  fest gewählt. Der Schätzer  $\hat{q}_{n,t}^w$  sei definiert durch (4.1), (4.2), (4.3), (4.4) und (4.5), wobei  $\mathcal{P}_n$  eine nichtleere endliche Parametermenge ist und für jedes  $p \in \mathcal{P}_n$  ein Funktionenraum  $\mathcal{F}_{n,p}$  gegeben ist. Dann gilt

$$\begin{aligned} & \int |\hat{q}_{n,t}^w(x) - q_t(x)|^2 \mathbf{P}_{X_t}(dx) \\ &= O_{\mathbf{P}} \left( \max_{s \in \{t, t+1, \dots, T-1\}} \min_{p \in \mathcal{P}_n} \int |q_{n_l,s}^{w,p}(x) - q_s(x)|^2 \mathbf{P}_{X_s}(dx) + \frac{\beta_n^2 \cdot \log |\mathcal{P}_n|}{n - n_l} \right) \end{aligned}$$

für alle  $t \in \{0, \dots, T\}$ .

**Bemerkung 4.2** Satz 4.1 dient als “Verknüpfung” zwischen dem Schätzer mit vollautomatischer Parameterwahl und dem von der Wahl des Parameters abhängigen Schätzer. Des Weiteren werden mit Hilfe des Satzes Aussagen zur Konsistenz sowie zur Konvergenzgeschwindigkeit für die in Abschnitt 4.2 verwendeten Schätzer hergeleitet.

**Beweis von Satz 4.1:** Bevor auf den Beweis eingegangen wird, ist zu beachten, dass die Beschränktheit der diskontierten Auszahlungsfunktion  $f_t$  durch  $L$   $|q_t(x)| \leq L$  für alle  $x \in \mathbb{R}^d$  impliziert. Des Weiteren impliziert die Definition von  $\hat{q}_{n,t}^w$   $|\hat{q}_{n,t}^w(x)| \leq \beta_n$  für alle  $x \in \mathbb{R}^d$ .

Im Folgenden wird gezeigt, dass

$$\begin{aligned} & \int |\hat{q}_{n,s}^w(x) - q_s(x)|^2 \mathbf{P}_{X_s}(dx) \\ &= O_{\mathbf{P}} \left( \max_{t \in \{s, s+1, \dots, T-1\}} \min_{p \in \mathcal{P}_n} \int |q_{n_l,t}^{w,p}(x) - q_t(x)|^2 \mathbf{P}_{X_t}(dx) + \frac{\beta_n^2 \cdot \log |\mathcal{P}_n|}{n - n_l} \right) \end{aligned} \quad (4.6)$$

für alle  $s \in \{0, \dots, T\}$  gilt.

Für  $s = T$  gilt  $\hat{q}_{n,T}^w(x) = 0 = q_T(x)$ , und somit ist die Aussage trivial. Sei  $t < T$ , und nehme an, dass die Aussage für jedes  $s \in \{t+1, \dots, T-1\}$  gilt. Durch Induktion genügt es, (4.6) für  $s = t$  zu zeigen.

Sei

$$\mathcal{D}_{n,t} = \{X_{j,s}^{(r)} : r \geq t+1, s \in \{0, \dots, T-1\}, j \in \{1, \dots, n\}\},$$

und setze

$$q_t^w(x) = \mathbf{E}^* \{\hat{Y}_{1,t}^{(t)} | X_{1,t}^{(t)} = x\},$$

wobei in  $\mathbf{E}^* \{ \cdot | X_{1,t}^{(t)} = x \}$  der bedingte Erwartungswert hinsichtlich  $\mathcal{D}_{n,t}$  genommen wird. D. h. alle Daten, die zur Konstruktion von  $\hat{q}_{n,t+1}^w, \dots, \hat{q}_{n,t+w+1}^w$  genutzt wurden, sind fest. Weiter sei

$$\begin{aligned} \bar{\mathcal{D}}_{n,t} = & \left\{ X_{j,s}^{(r)} : r \geq t+1, s \in \{0, \dots, T-1\}, j \in \{1, \dots, n\} \text{ oder} \right. \\ & \left. r = t, s \in \{0, \dots, T-1\}, j \in \{1, \dots, n_l\} \right\}. \end{aligned}$$

Setze

$$p^* = \arg \min_{p \in \mathcal{P}_n} \int |q_{n_l,t}^{w,p}(x) - q_t(x)|^2 \mathbf{P}_{X_t}(dx).$$

Es gilt nun, (4.6) für  $s = t$  zu zeigen. Dazu wird die Fehler-Zerlegung des  $L_2$ -Fehlers betrachtet:

$$\begin{aligned} & \int |\hat{q}_{n,t}^w(x) - q_t(x)|^2 \mathbf{P}_{X_t}(dx) \\ &= \int |\hat{q}_{n,t}^w(x) - q_t(x)|^2 \mathbf{P}_{X_t}(dx) - \frac{2}{n - n_l} \sum_{i=n_l+1}^n |\hat{q}_{n,t}^w(X_{i,t}^{(t)}) - q_t(X_{i,t}^{(t)})|^2 \\ &+ \frac{2}{n - n_l} \sum_{i=n_l+1}^n |\hat{q}_{n,t}^w(X_{i,t}^{(t)}) - q_t(X_{i,t}^{(t)})|^2 - \frac{4}{n - n_l} \sum_{i=n_l+1}^n |\hat{q}_{n,t}^w(X_{i,t}^{(t)}) - q_t^w(X_{i,t}^{(t)})|^2 \\ &+ \frac{4}{n - n_l} \sum_{i=n_l+1}^n |\hat{q}_{n,t}^w(X_{i,t}^{(t)}) - q_t^w(X_{i,t}^{(t)})|^2 - \frac{72}{n - n_l} \sum_{i=n_l+1}^n |q_{n_l,t}^{w,p^*}(X_{i,t}^{(t)}) - q_t^w(X_{i,t}^{(t)})|^2 \\ &+ \frac{72}{n - n_l} \sum_{i=n_l+1}^n |q_{n_l,t}^{w,p^*}(X_{i,t}^{(t)}) - q_t^w(X_{i,t}^{(t)})|^2 - \frac{144}{n - n_l} \sum_{i=n_l+1}^n |q_{n_l,t}^{w,p^*}(X_{i,t}^{(t)}) - q_t(X_{i,t}^{(t)})|^2 \\ &+ \frac{144}{n - n_l} \sum_{i=n_l+1}^n |q_{n_l,t}^{w,p^*}(X_{i,t}^{(t)}) - q_t(X_{i,t}^{(t)})|^2 - 288 \int |q_{n_l,t}^{w,p^*}(x) - q_t(x)|^2 \mathbf{P}_{X_t}(dx) \end{aligned}$$

$$\begin{aligned}
 & + 288 \int |q_{n_l,t}^{w,p^*}(x) - q_t(x)|^2 \mathbf{P}_{X_t}(dx) \\
 & = \sum_{j=1}^6 T_{j,n}.
 \end{aligned}$$

Es sei hier erwähnt, dass der Beweis vollständig ist, sobald einmal gezeigt wurde, dass für  $j \in \{1, 2, \dots, 6\}$  gilt:

$$T_{j,n} = O_{\mathbf{P}} \left( \max_{s \in \{t, t+1, \dots, T-1\}} \min_{p \in \mathcal{P}_n} \int |q_{n_l,s}^{w,p} - q_s(x)|^2 \mathbf{P}_{X_s}(dx) + \frac{\beta_n^2 \cdot \log |\mathcal{P}_n|}{n - n_l} \right). \quad (4.7)$$

Im ersten Schritt des Beweises wird  $T_{1,n}$  betrachtet. Wird die Menge

$$\{q_{n_l,t}^{w,p} : p \in \mathcal{P}_n\}$$

auf  $\bar{\mathcal{D}}_{n,t}$  konditioniert, dann besteht die Menge aus  $|\mathcal{P}_n|$  verschiedenen Funktionen. Des Weiteren gilt wegen der Beschränktheit von  $q_{n_l,t}^{w,p}$  und  $q_t$  durch  $\beta_n$ :

$$\begin{aligned}
 \sigma_p^2 & := \mathbf{Var}\{|q_{n_l,t}^{w,p}(X_{n_l+1,t}^{(t)}) - q_t(X_{n_l+1,t}^{(t)})|^2 \mid \bar{\mathcal{D}}_{n,t}\} \\
 & \leq \mathbf{E}\{|q_{n_l,t}^{w,p}(X_{n_l+1,t}^{(t)}) - q_t(X_{n_l+1,t}^{(t)})|^4 \mid \bar{\mathcal{D}}_{n,t}\} \\
 & \leq 4\beta_n^2 \mathbf{E}\{|q_{n_l,t}^{w,p}(X_{n_l+1,t}^{(t)}) - q_t(X_{n_l+1,t}^{(t)})|^2 \mid \bar{\mathcal{D}}_{n,t}\} \\
 & \leq 4\beta_n^2 \int |q_{n_l,t}^{w,p}(x) - q_t(x)|^2 \mathbf{P}_{X_t}(dx).
 \end{aligned}$$

Mit Hilfe dieses Resultats und der Ungleichung von Bernstein (siehe Lemma A.11) erhält man mit der Notation  $\epsilon_n = c_{16} \cdot (\beta_n^2 \log |\mathcal{P}_n|)/(n - n_l)$ , wobei  $c_{16} > 0$  konstant:

$$\begin{aligned}
 & \mathbf{P}\{T_{1,n} > \epsilon_n \mid \bar{\mathcal{D}}_{n,t}\} \\
 & \leq |\mathcal{P}_n| \cdot \max_{p \in \mathcal{P}_n} \mathbf{P}\left\{ \int |q_{n_l,t}^{w,p}(x) - q_t(x)|^2 \mathbf{P}_{X_t}(dx) - \frac{2}{n - n_l} \sum_{i=n_l+1}^n |q_{n_l,t}^{w,p}(X_{i,t}) - q_t(X_{i,t})|^2 \right. \\
 & \quad \left. > \epsilon_n \mid \bar{\mathcal{D}}_{n,t} \right\} \\
 & = |\mathcal{P}_n| \cdot \max_{p \in \mathcal{P}_n} \mathbf{P}\left\{ \int |q_{n_l,t}^{w,p}(x) - q_t(x)|^2 \mathbf{P}_{X_t}(dx) - \frac{1}{n - n_l} \sum_{i=n_l+1}^n |q_{n_l,t}^{w,p}(X_{i,t}) - q_t(X_{i,t})|^2 \right. \\
 & \quad \left. > \frac{\epsilon_n}{2} + \frac{1}{2} \int |q_{n_l,t}^{w,p}(x) - q_t(x)|^2 \mathbf{P}_{X_t}(dx) \mid \bar{\mathcal{D}}_{n,t} \right\} \\
 & \leq |\mathcal{P}_n| \cdot \max_{p \in \mathcal{P}_n} \mathbf{P}\left\{ \int |q_{n_l,t}^{w,p}(x) - q_t(x)|^2 \mathbf{P}_{X_t}(dx) - \frac{1}{n - n_l} \sum_{i=n_l+1}^n |q_{n_l,t}^{w,p}(X_{i,t}) - q_t(X_{i,t})|^2 \right. \\
 & \quad \left. > \frac{\epsilon_n}{2} + \frac{1}{2} \cdot \frac{\sigma_p^2}{4\beta_n^2} \mid \bar{\mathcal{D}}_{n,t} \right\} \\
 & \leq |\mathcal{P}_n| \cdot \max_{p \in \mathcal{P}_n} \exp\left( -\frac{(n - n_l)\left(\frac{\epsilon_n}{2} + \frac{\sigma_p^2}{8\beta_n^2}\right)^2}{2\sigma_p^2 + 2\left(\frac{\epsilon_n}{2} + \frac{\sigma_p^2}{8\beta_n^2}\right) \cdot \frac{4\beta_n^2}{3}} \right)
 \end{aligned}$$

$$\begin{aligned}
 &= |\mathcal{P}_n| \cdot \max_{p \in \mathcal{P}_n} \exp \left( - \frac{(n - n_l) \left( \frac{4\epsilon_n \beta_n^2 + \sigma_p^2}{8\beta_n^2} \right)^2}{2\sigma_p^2 + 2 \left( \frac{4\epsilon_n \beta_n^2 + \sigma_p^2}{8\beta_n^2} \right) \cdot \frac{4\beta_n^2}{3}} \right) \\
 &= |\mathcal{P}_n| \cdot \max_{p \in \mathcal{P}_n} \exp \left( - \frac{(n - n_l)(4\epsilon_n \beta_n^2 + \sigma_p^2)}{2\sigma_p^2 (8\beta_n^2)^2 \cdot \frac{1}{4\epsilon_n \beta_n^2 + \sigma_p^2} + 2(8\beta_n^2) \cdot \frac{4\beta_n^2}{3}} \right) \\
 &= |\mathcal{P}_n| \cdot \max_{p \in \mathcal{P}_n} \exp \left( - \frac{(n - n_l)(4\epsilon_n \beta_n^2 + \sigma_p^2)}{2(8\beta_n^2)^2 \cdot \frac{\sigma_p^2}{4\epsilon_n \beta_n^2 + \sigma_p^2} + 2(8\beta_n^2) \cdot \frac{4\beta_n^2}{3}} \right) \\
 &\leq |\mathcal{P}_n| \cdot \max_{p \in \mathcal{P}_n} \exp \left( - \frac{(n - n_l)(4\epsilon_n \beta_n^2 + \sigma_p^2)}{2(8\beta_n^2)^2 + 2(8\beta_n^2) \cdot \frac{4\beta_n^2}{3}} \right) \\
 &\leq |\mathcal{P}_n| \cdot \max_{p \in \mathcal{P}_n} \exp \left( - \frac{(n - n_l)(4\epsilon_n \beta_n^2)}{2(8\beta_n^2)^2 + 2(8\beta_n^2) \cdot \frac{4\beta_n^2}{3}} \right) \\
 &\leq |\mathcal{P}_n| \cdot \exp \left( - \frac{c_{16}}{2(8\beta_n^2)^2 + 2(8\beta_n^2) \cdot \frac{4\beta_n^2}{3}} \cdot \beta_n^4 \log |\mathcal{P}_n| \right) \\
 &\rightarrow 0 \quad (n \rightarrow \infty)
 \end{aligned}$$

für  $c_{16}$  genügend groß gewählt. Insgesamt erhält man für  $T_{1,n}$

$$\mathbf{P}\{T_{1,n} > \epsilon_n \mid \bar{\mathcal{D}}_{n,t}\} \rightarrow 0 \quad (n \rightarrow \infty),$$

woraus wiederum  $T_{1,n} = O_{\mathbf{P}}((\beta_n^2 \cdot \log |\mathcal{P}_n|)/(n - n_l))$  folgt. Dies impliziert (4.7) mit  $j = 1$ .

Im zweiten Schritt des Beweises wird  $T_{5,n}$  betrachtet. Wie im ersten Schritt gilt, dass die Menge  $\{q_{n_l,t}^{w,p} : p \in \mathcal{P}_n\}$  aus  $|\mathcal{P}_n|$  verschiedenen Funktionen besteht, falls die Menge auf  $\bar{\mathcal{D}}_{n,t}$  konditioniert wird. Weiter gilt (siehe oben):

$$\begin{aligned}
 \sigma_p^2 &:= \mathbf{Var}\{|q_{n_l,t}^{w,p}(X_{n_l+1,t}) - q_t(X_{n_l+1,t})|^2 \mid \bar{\mathcal{D}}_{n,t}\} \\
 &\leq 4\beta_n^2 \int |q_{n_l,t}^{w,p}(x) - q_t(x)|^2 \mathbf{P}_{X_t}(dx).
 \end{aligned}$$

Analog wie im ersten Schritt erhält man weiter mit der Ungleichung von Bernstein sowie mit  $c_{17} > 0$  konstant und der Notation  $\epsilon_n = c_{17} \cdot (\beta_n^2 \log |\mathcal{P}_n|)/(n - n_l)$ :

$$\begin{aligned}
 &\mathbf{P}\{T_{5,n} > \epsilon_n \mid \bar{\mathcal{D}}_{n,t}\} \\
 &\leq |\mathcal{P}_n| \cdot \max_{p \in \mathcal{P}_n} \mathbf{P}\left\{ \frac{144}{n - n_l} \sum_{i=n_l+1}^n |q_{n_l,t}^{w,p}(X_{i,t}) - q_t(X_{i,t})|^2 - 288 \int |q_{n_l,t}^{w,p}(x) - q_t(x)|^2 \mathbf{P}_{X_t}(dx) \right. \\
 &\quad \left. > \epsilon_n \mid \bar{\mathcal{D}}_{n,t} \right\} \\
 &= |\mathcal{P}_n| \cdot \max_{p \in \mathcal{P}_n} \mathbf{P}\left\{ \frac{144}{n - n_l} \sum_{i=n_l+1}^n |q_{n_l,t}^{w,p}(X_{i,t}) - q_t(X_{i,t})|^2 - 144 \int |q_{n_l,t}^{w,p}(x) - q_t(x)|^2 \mathbf{P}_{X_t}(dx) \right. \\
 &\quad \left. > \epsilon_n + 144 \int |q_{n_l,t}^{w,p}(x) - q_t(x)|^2 \mathbf{P}_{X_t}(dx) \mid \bar{\mathcal{D}}_{n,t} \right\}
 \end{aligned}$$



$$\begin{aligned}
 &\leq |\mathcal{P}_n| \cdot \max_{p \in \mathcal{P}_n} \mathbf{P} \left\{ \frac{1}{n - n_l} \sum_{i=n_l+1}^n |q_{n_l,t}^{w,p}(X_{i,t}) - q_t(X_{i,t})|^2 - \int |q_{n_l,t}^{w,p}(x) - q_t(x)|^2 \mathbf{P}_{X_t}(dx) \right. \\
 &\quad \left. > \frac{\epsilon_n}{144} + \frac{\sigma_p^2}{4\beta_n^2} \mid \bar{\mathcal{D}}_{n,t} \right\} \\
 &\leq |\mathcal{P}_n| \cdot \max_{p \in \mathcal{P}_n} \exp \left( - \frac{(n - n_l) \left( \frac{\epsilon_n}{144} + \frac{\sigma_p^2}{4\beta_n^2} \right)^2}{2\sigma_p^2 + 2 \left( \frac{\epsilon_n}{144} + \frac{\sigma_p^2}{4\beta_n^2} \right) \cdot \frac{4\beta_n^2}{3}} \right) \\
 &= |\mathcal{P}_n| \cdot \max_{p \in \mathcal{P}_n} \exp \left( - \frac{(n - n_l) \left( \frac{\epsilon_n \beta_n^2 + 36\sigma_p^2}{144\beta_n^2} \right)^2}{2\sigma_p^2 + 2 \left( \frac{\epsilon_n \beta_n^2 + 36\sigma_p^2}{144\beta_n^2} \right) \cdot \frac{4\beta_n^2}{3}} \right) \\
 &= |\mathcal{P}_n| \cdot \max_{p \in \mathcal{P}_n} \exp \left( - \frac{(n - n_l) (\epsilon_n \beta_n^2 + 36\sigma_p^2)}{2\sigma_p^2 (144\beta_n^2)^2 \cdot \frac{1}{\epsilon_n \beta_n^2 + 36\sigma_p^2} + 2(144\beta_n^2) \cdot \frac{4\beta_n^2}{3}} \right) \\
 &= |\mathcal{P}_n| \cdot \max_{p \in \mathcal{P}_n} \exp \left( - \frac{(n - n_l) (\epsilon_n \beta_n^2 + 36\sigma_p^2)}{2(144\beta_n^2)^2 \cdot \frac{\sigma_p^2}{\epsilon_n \beta_n^2 + 36\sigma_p^2} + 2(144\beta_n^2) \cdot \frac{4\beta_n^2}{3}} \right) \\
 &\leq |\mathcal{P}_n| \cdot \max_{p \in \mathcal{P}_n} \exp \left( - \frac{(n - n_l) (\epsilon_n \beta_n^2 + 36\sigma_p^2)}{2(144\beta_n^2)^2 + 2(144\beta_n^2) \cdot \frac{4\beta_n^2}{3}} \right) \\
 &\leq |\mathcal{P}_n| \cdot \max_{p \in \mathcal{P}_n} \exp \left( - \frac{(n - n_l) (\epsilon_n \beta_n^2)}{2(144\beta_n^2)^2 + 2(144\beta_n^2) \cdot \frac{4\beta_n^2}{3}} \right) \\
 &\leq |\mathcal{P}_n| \cdot \exp \left( - \frac{c_{17}}{2(144\beta_n^2)^2 + 2(144\beta_n^2) \cdot \frac{4\beta_n^2}{3}} \cdot \beta_n^4 \log |\mathcal{P}_n| \right) \\
 &\rightarrow 0 \quad (n \rightarrow \infty)
 \end{aligned}$$

für  $c_{17}$  genügend groß gewählt. Mit diesem Resultat folgt die Behauptung (4.7) mit  $j = 5$ .

Im dritten Schritt des Beweises wird  $T_{2,n}$  betrachtet. Mit der Beziehung  $a^2 = (a - b + b)^2 \leq 2(a - b)^2 + 2b^2$  folgt für  $T_{2,n}$ :

$$\begin{aligned}
 T_{2,n} &\leq \frac{4}{n - n_l} \sum_{i=n_l+1}^n |q_t^w(X_{i,t}^{(t)}) - q_t(X_{i,t}^{(t)})|^2 \\
 &= \frac{4}{n - n_l} \sum_{i=n_l+1}^n |q_t^w(X_{i,t}^{(t)}) - q_t(X_{i,t}^{(t)})|^2 - 8 \int |q_t^w(x) - q_t(x)|^2 \mathbf{P}_{X_t}(dx) \\
 &\quad + 8 \int |q_t^w(x) - q_t(x)|^2 \mathbf{P}_{X_t}(dx) \\
 &= \sum_{j=7}^8 T_{j,n}.
 \end{aligned}$$

Im vierten Schritt des Beweises wird  $T_{7,n}$  betrachtet. Es ist einerseits zu beachten, dass  $|q_t(x)| \leq L$  für alle  $x \in \mathbb{R}^d$  gilt, sowie andererseits, dass  $\hat{Y}_{i,t}^{(t)}$  beschränkt ist, was (nach

Definition von  $q_t^w$   $|q_t^w(x)| \leq \beta_n$  für alle  $x \in \mathbb{R}^d$  zur Folge hat. Des Weiteren gilt wegen der Beschränktheit von  $q_t^w$  und  $q_t$  durch  $\beta_n$ :

$$\begin{aligned} \sigma_p^2 &:= \mathbf{Var}\{|q_t^w(X_{n_l+1,t}^{(t)}) - q_t(X_{n_l+1,t}^{(t)})|^2 \mid \mathcal{D}_{n,t}\} \\ &\leq \mathbf{E}\{|q_t^w(X_{n_l+1,t}^{(t)}) - q_t(X_{n_l+1,t}^{(t)})|^4 \mid \mathcal{D}_{n,t}\} \\ &\leq 4\beta_n^2 \int |q_t^w(x) - q_t(x)|^2 \mathbf{P}_{X_t}(dx). \end{aligned}$$

Wird  $T_{7,n}$  auf  $\mathcal{D}_{n,t}$  konditioniert - was zur Folge hat, dass  $q_t^w$  fest ist - dann gilt mit der Notation  $\epsilon_n = \text{const} \cdot \beta_n^2/(n - n_l)$ :

$$\begin{aligned} &\mathbf{P}\left\{T_{7,n} > \epsilon_n \mid \mathcal{D}_{n,t}\right\} \\ &= \mathbf{P}\left\{\frac{1}{n - n_l} \sum_{i=n_l+1}^n |q_t^w(X_{i,t}^{(t)}) - q_t(X_{i,t}^{(t)})|^2 - 2 \int |q_t^w(x) - q_t(x)|^2 \mathbf{P}_{X_t}(dx) > \frac{\epsilon_n}{4} \mid \mathcal{D}_{n,t}\right\} \\ &= \mathbf{P}\left\{\frac{1}{n - n_l} \sum_{i=n_l+1}^n |q_t^w(X_{i,t}^{(t)}) - q_t(X_{i,t}^{(t)})|^2 - \int |q_t^w(x) - q_t(x)|^2 \mathbf{P}_{X_t}(dx) \right. \\ &\quad \left. > \frac{\epsilon_n}{4} + \int |q_t^w(x) - q_t(x)|^2 \mathbf{P}_{X_t}(dx) \mid \mathcal{D}_{n,t}\right\} \\ &\leq \mathbf{P}\left\{\frac{1}{n - n_l} \sum_{i=n_l+1}^n |q_t^w(X_{i,t}^{(t)}) - q_t(X_{i,t}^{(t)})|^2 - \int |q_t^w(x) - q_t(x)|^2 \mathbf{P}_{X_t}(dx) \right. \\ &\quad \left. > \frac{\epsilon_n}{4} + \frac{\sigma_p^2}{4\beta_n^2} \mid \mathcal{D}_{n,t}\right\}. \end{aligned}$$

Die letzte Wahrscheinlichkeit kann analog zu oben mit der Ungleichung von Bernstein (siehe Lemma A.11) abgeschätzt werden. Insgesamt ergibt sich so für  $T_{7,n}$

$$T_{7,n} = O_{\mathbf{P}}(\beta_n^2/(n - n_l)).$$

Im fünften Schritt des Beweises wird  $T_{8,n}$  betrachtet. Die Anwendung von (2.12) mit  $\hat{q}_{n,s}^w$  anstelle von  $q_{n,s}^{w,p}$  führt zu

$$\begin{aligned} T_{8,n} &= 8 \int |q_t^w(x) - q_t(x)|^2 \mathbf{P}_{X_t}(dx) \\ &\leq 8 \left( \sum_{s=t+1}^{t+w+1} \left( \int |\hat{q}_{n,s}^w(x) - q_s(x)|^2 \mathbf{P}_{X_s}(dx) \right)^{1/2} \right)^2 \\ &\leq 8(w+1)^2 \cdot \sum_{s=t+1}^{t+w+1} \int |\hat{q}_{n,s}^w(x) - q_s(x)|^2 \mathbf{P}_{X_s}(dx), \end{aligned}$$

und mit (4.6) für  $s \geq t+1$  erhält man

$$T_{8,n} = O_{\mathbf{P}} \left( \max_{s \in \{t+1, \dots, T-1\}} \min_{p \in \mathcal{P}_n} \int |q_{n_i,s}^{w,p} - q_s(x)|^2 \mathbf{P}_{X_s}(dx) + \frac{\beta_n^2 \cdot \log |\mathcal{P}_n|}{n - n_l} \right).$$

Insgesamt implizieren die letzten drei Schritte (4.7) mit  $j = 2$ .

Im sechsten Schritt des Beweises wird  $T_{4,n}$  betrachtet. Mit ähnlicher Vorgehensweise wie bei  $T_{2,n}$  kann (4.7) mit  $j = 4$  gezeigt werden.

Im siebten Schritt des Beweises wird  $T_{3,n}$  betrachtet. Setze

$$q^* = \arg \min_{f \in \{q_{n_l,t}^{w,p} : p \in \mathcal{P}_n\}} \frac{1}{n - n_l} \sum_{i=n_l+1}^n |f(X_{i,t}^{(t)}) - q_t^w(X_{i,t}^{(t)})|^2.$$

Mit Lemma A.7 (oder mit Standard-Resultaten aus dem Buch Van de Geer (2000), siehe dazu Beweis von Theorem 10.11) angewendet mit  $x_i = X_{i,t}^{(t)}$  sowie  $y_i = \bar{y}_i = \hat{Y}_{i,t}^{(t)}$  und der Notation  $\epsilon_n = c_{18} \cdot (\beta_n^2 \log |\mathcal{P}_n|) / (n - n_l)$  erhält man:

$$\begin{aligned} & \mathbf{P} \left\{ T_{3,n} > \epsilon_n \mid \bar{\mathcal{D}}_{n,t} \right\} \\ & \leq \mathbf{P} \left\{ \frac{1}{n - n_l} \sum_{i=n_l+1}^n |\hat{q}_{n,t}^w(X_{i,t}^{(t)}) - q_t^w(X_{i,t}^{(t)})|^2 > \frac{\epsilon_n}{4} \right. \\ & \quad \left. + 18 \cdot \min_{p \in \mathcal{P}_n} \frac{1}{n - n_l} \sum_{i=n_l+1}^n |q_{n_l,t}^{w,p}(X_{i,t}^{(t)}) - q_t^w(X_{i,t}^{(t)})|^2 \mid \bar{\mathcal{D}}_{n,t} \right\} \\ & \leq \mathbf{P} \left\{ \frac{\epsilon_n/4}{2} < \frac{1}{n - n_l} \sum_{i=n_l+1}^n |\hat{q}_{n,t}^w(X_{i,t}^{(t)}) - q^*(X_{i,t}^{(t)})|^2 \right. \\ & \quad \left. \leq \frac{16}{n - n_l} \sum_{i=n_l+1}^n (\hat{q}_{n,t}^w(X_{i,t}^{(t)}) - q^*(X_{i,t}^{(t)})) \cdot (\hat{Y}_{i,t}^{(t)} - q_t^w(X_{i,t}^{(t)})) \mid \bar{\mathcal{D}}_{n,t} \right\} \\ & \leq |\mathcal{P}_n| \cdot \max_{p \in \mathcal{P}_n} \mathbf{P} \left\{ \frac{\epsilon_n/4}{2} < \frac{1}{n - n_l} \sum_{i=n_l+1}^n |q_{n_l,t}^{w,p}(X_{i,t}^{(t)}) - q^*(X_{i,t}^{(t)})|^2 \right. \\ & \quad \left. \leq \frac{16}{n - n_l} \sum_{i=n_l+1}^n (q_{n_l,t}^{w,p}(X_{i,t}^{(t)}) - q^*(X_{i,t}^{(t)})) \cdot (\hat{Y}_{i,t}^{(t)} - q_t^w(X_{i,t}^{(t)})) \mid \bar{\mathcal{D}}_{n,t} \right\} \\ & \leq |\mathcal{P}_n| \cdot \max_{p \in \mathcal{P}_n} \sum_{s=0}^{\infty} \mathbf{P} \left\{ 2^{s-1} \epsilon_n/4 < \frac{1}{n - n_l} \sum_{i=n_l+1}^n |q_{n_l,t}^{w,p}(X_{i,t}^{(t)}) - q^*(X_{i,t}^{(t)})|^2 \leq 2^s \epsilon_n/4, \right. \\ & \quad \left. \frac{1}{n - n_l} \sum_{i=n_l+1}^n |q_{n_l,t}^{w,p}(X_{i,t}^{(t)}) - q^*(X_{i,t}^{(t)})|^2 \right. \\ & \quad \left. \leq \frac{16}{n - n_l} \sum_{i=n_l+1}^n (q_{n_l,t}^{w,p}(X_{i,t}^{(t)}) - q^*(X_{i,t}^{(t)})) \cdot (\hat{Y}_{i,t}^{(t)} - q_t^w(X_{i,t}^{(t)})) \mid \bar{\mathcal{D}}_{n,t} \right\} \\ & \leq |\mathcal{P}_n| \cdot \sum_{s=0}^{\infty} \max_{p \in \mathcal{P}_n: \frac{1}{n-n_l} \sum_{i=n_l+1}^n |q_{n_l,t}^{w,p}(X_{i,t}^{(t)}) - q^*(X_{i,t}^{(t)})|^2 \leq 2^s \epsilon_n/4} \\ & \quad \mathbf{P} \left\{ \frac{1}{n - n_l} \sum_{i=n_l+1}^n (q_{n_l,t}^{w,p}(X_{i,t}^{(t)}) - q^*(X_{i,t}^{(t)})) \cdot (\hat{Y}_{i,t}^{(t)} - q_t^w(X_{i,t}^{(t)})) > \frac{2^s \epsilon_n}{128} \mid \bar{\mathcal{D}}_{n,t} \right\}. \end{aligned}$$

Wird die letzte Wahrscheinlichkeit oben mit der Ungleichung von Hoeffding (siehe Lemma A.10) und unter Berücksichtigung, dass  $\hat{Y}_{i,t}^{(t)}$  sowie  $q_t^w$  absolut beschränkt sind durch  $\beta_n$ , abgeschätzt, dann erhält man für die rechte Seite in letzter Ungleichung eine obere Schranke

$$\begin{aligned}
 & |\mathcal{P}_n| \cdot \sum_{s=0}^{\infty} 2 \exp\left(-\frac{2(n-n_l)(2^s \epsilon_n/128)^2}{2^s \epsilon_n/4 \cdot (2\beta_n)^2}\right) \\
 &= 2|\mathcal{P}_n| \cdot \sum_{s=0}^{\infty} \exp\left(-\frac{8 \cdot 2^s (n-n_l) \epsilon_n}{128^2 \cdot 4\beta_n^2}\right) \\
 &\leq 2|\mathcal{P}_n| \cdot \sum_{s=0}^{\infty} \exp\left(-\frac{c_{18} \cdot 2^{s+1}}{128^2} \cdot \log |\mathcal{P}_n|\right) \\
 &\leq 2|\mathcal{P}_n| \cdot \sum_{s=0}^{\infty} \exp\left(-\frac{c_{18} \cdot (s+1)}{128^2} \cdot \log |\mathcal{P}_n|\right) \\
 &= 2|\mathcal{P}_n| \cdot \exp\left(-\frac{c_{18}}{128^2} \cdot \log |\mathcal{P}_n|\right) \sum_{s=0}^{\infty} \exp\left(-s \cdot \frac{c_{18}}{128^2} \cdot \log |\mathcal{P}_n|\right) \\
 &\rightarrow 0 \quad (n \rightarrow \infty)
 \end{aligned}$$

für  $c_{18}$  genügend groß gewählt und da in der letzten Gleichung die (unendliche) Reihe sich darstellen lässt als eine 1 (für  $s = 0$ ) plus eine geometrische Reihe die konvergiert (da der Betrag von  $e^{-c_{18} \log |\mathcal{P}_n|/128^2}$  kleiner 1 ist). Insgesamt wurde für  $T_{3,n}$

$$T_{3,n} = O_{\mathbf{P}}((\beta_n^2 \cdot \log |\mathcal{P}_n|)/(n - n_l))$$

gezeigt, was wiederum (4.7) mit  $j = 3$  impliziert.

Im achten (und letzten) Schritt des Beweises wird  $T_{6,n}$  betrachtet. Die Wahl von  $p^*$ , also

$$p^* = \arg \min_{p \in \mathcal{P}_n} \int |q_{n_l,t}^{w,p}(x) - q_t(x)|^2 \mathbf{P}_{X_t}(dx),$$

impliziert dabei die Behauptung (4.7) mit  $j = 6$ , und der Beweis ist vollständig.  $\square$

**Bemerkung 4.3** Wird der Spezialfall von Satz 4.1 mit  $w \equiv 0$  betrachtet, dann kann alternativ zum obigen Beweis der Beweis des Spezialfalls durch den Zugang zur Analyse des Kleinste-Quadrate-Schätzers wie in Kapitel 3 erfolgen.

In diesem Abschnitt erfolgte eine Aussage betreffend eines Kleinste-Quadrate-Schätzers mit vollautomatischer Parameterwahl. Im nächsten Abschnitt wird dieser Schätzer dahingehend modifiziert, dass auch der noch von der Regressionsdarstellung (2.1) verbleibende Parameter  $w$  automatisch gewählt wird.

## 4.2 Adaptive Spline- und Neuronale-Netze-Schätzer

In den Abschnitten 2.2 und 2.3 erfolgten Resultate hinsichtlich der Konsistenz und der Konvergenzgeschwindigkeit von Kleinste-Quadrate-Schätzern ohne vollautomatische Wahl der Parameter. Im Unterschied dazu werden in diesem Abschnitt Resultate hinsichtlich Konsistenz sowie Konvergenzgeschwindigkeit für Schätzer mit vollautomatischer Parameterwahl formuliert. Als zusätzliche Eigenschaft der unten verwendeten Schätzer wird neben der automatischen Wahl der Parameter, wie im vorherigen Abschnitt beschrieben, auch der verbleibende Parameter  $w$  aus der Regressionsdarstellung (2.1) automatisch gewählt. Es wird sich zeigen, dass diese Schätzer (unten) die gleiche Konvergenzrate erreichen wie die Schätzer aus den Abschnitten 2.2 und 2.3, wenn man dieselben Regularitätsvoraussetzungen an die zugrunde liegende Verteilung von  $X_t$  stellt.

Bevor die oben angekündigten Resultate formuliert werden können, wird als erstes der vollautomatische (nichtparametrische Kleinste-Quadrate-Regressions-)Schätzer eingeführt. Dazu wird die Datenmenge der Kardinalität  $n$ , auf der der Schätzer definiert ist, in eine Lernmenge der Kardinalität  $n_l$ , eine Testmenge der Kardinalität  $n_t$  und eine so genannte Validierungsmenge der Kardinalität  $n_v$  unterteilt, wobei

$$n_v = n_t = \lfloor n/3 \rfloor \quad \text{und} \quad n_l = n - n_v - n_t$$

gilt. Des Weiteren existiere für jeden Parameter  $p$  aus der nichtleeren endlichen Parametermenge  $\mathcal{P}_n$  ein Raum  $\mathcal{F}_{n,p}$  von Funktionen  $f : \mathbb{R}^d \rightarrow \mathbb{R}$ . In einem ersten Schritt wird der vollautomatische (Regressions-)Schätzer  $\hat{q}_{n,t}$  von  $q_t$  in Abhängigkeit von  $w \in \{0, 1, \dots, T - t - 1\}$  definiert. Anschließend erfolgt in einem zweiten Schritt die (datenabhängige) Wahl von  $w$ , mit der die Definition des Schätzers vervollständigt wird.

Sei  $w \in \{0, 1, \dots, T - t - 1\}$  fest. Der Schätzer  $\hat{q}_{n,t}^w$  mit automatischer Parameterwahl der Parameter  $p \in \mathcal{P}_n$  von  $q_t$  ist definiert durch

$$\tilde{q}_{n,t}^{w,p}(\cdot) = \arg \min_{f \in \mathcal{F}_{n,p}} \left( \frac{1}{n_l} \sum_{i=1}^{n_l} |f(X_{i,t}^{(t)}) - \hat{Y}_{i,t}^{(t)}|^2 \right), \quad (4.8)$$

$$q_{n,t}^{w,p}(x) = T_{\beta_n} \tilde{q}_{n,t}^{w,p}(x) \quad (x \in \mathbb{R}^d), \quad (4.9)$$

$$\hat{p} = \arg \min_{p \in \mathcal{P}_n} \left( \frac{1}{n_t} \sum_{i=n_l+1}^{n_l+n_t} |q_{n,t}^{w,p}(X_{i,t}^{(t)}) - \hat{Y}_{i,t}^{(t)}|^2 \right) \quad \text{und} \quad (4.10)$$

$$\hat{q}_{n,t}^w(\cdot) = q_{n,t}^{w,\hat{p}}(\cdot), \quad (4.11)$$

mit  $\hat{Y}_{i,t}^{(t)}$  wie in (4.4) mit  $\hat{q}_n$  anstelle von  $\hat{q}_n^w$ . Für jedes  $w \in \{0, 1, \dots, T - t - 1\}$  bietet Definition (4.11) einen Schätzer von  $q_t$  an. Die Idee für die Wahl von  $w$  ist nun, ausgehend

von  $\hat{q}_{n,t}^w$  eine approximative optimale Stoppregel zu berechnen, die eine untere Schranke für die Lösung des optimalen Stoppproblems zur Zeit  $t$  angibt. Der optimale Kandidat für  $w$  ist derjenige, der diese untere Schranke maximiert. Deshalb gelte

$$\hat{w}_t = \arg \max_{w \in \{0,1,\dots,T-t-1\}} \frac{1}{n_v} \sum_{i=n_l+n_t+1}^{n_l+n_t+n_v} f_{\hat{\tau}_t^w}^{(X_{i,\hat{\tau}_t^w}^w)}, \quad (4.12)$$

wobei für  $w \in \{0,1,\dots,T-t-1\}$  die optimale Stoppregel  $\hat{\tau}_t^w$  gegeben ist durch

$$\begin{aligned} \hat{\tau}_t^w &= \hat{\tau}_t^w(X_{i,t}^{(t)}, \dots, X_{i,T-1}^{(t)}) = t \cdot I_{\{f_t(X_{i,t}^{(t)}) \geq \hat{q}_{n,t}^w(X_{i,t}^{(t)})\}} + \\ &\left( \sum_{s=t+1}^{T-1} s \cdot I_{\{f_{t+1}(X_{i,t+1}^{(t)}) < \hat{q}_{n,t+1}(X_{i,t+1}^{(t)}), \dots, f_{s-1}(X_{i,s-1}^{(t)}) < \hat{q}_{n,s-1}(X_{i,s-1}^{(t)}), f_s(X_{i,s}^{(t)}) \geq \hat{q}_{n,s}(X_{i,s}^{(t)})\}} \right. \\ &\left. + T \cdot I_{\{f_{t+1}(X_{i,t+1}^{(t)}) < \hat{q}_{n,t+1}(X_{i,t+1}^{(t)}), \dots, f_{T-1}(X_{i,T-1}^{(t)}) < \hat{q}_{n,T-1}(X_{i,T-1}^{(t)})\}} \right) \cdot I_{\{f_t(X_{i,t}^{(t)}) < \hat{q}_{n,t}^w(X_{i,t}^{(t)})\}}. \end{aligned}$$

Mit der Spezifikation (4.12) für  $\hat{w}_t$  ist der vollautomatische (Regressions-)Schätzer  $\hat{q}_{n,t}$  von  $q_t$  definiert durch

$$\hat{q}_{n,t}(\cdot) = \hat{q}_{n,t}^{\hat{w}_t}(\cdot). \quad (4.13)$$

An dieser Stelle sei erwähnt, dass der vollautomatische (Regressions-)Schätzer  $\hat{q}_{n,t}$  im Zusammenhang mit der Schätzung des Wertes einer Amerikanischen Option eingesetzt wird. Auch sei wie in Abschnitt 2.1 bereits erläutert darauf hingewiesen, dass  $\hat{q}_{n,T}(x) = 0$  für  $x \in \mathbb{R}^d$  ist.

**Bemerkung 4.4** In (4.10) und (4.12) wird über eine endliche Menge optimiert, was die Existenz eines Optimierers impliziert.

Als nächstes folgen Resultate hinsichtlich der Konsistenz und der Konvergenzgeschwindigkeit für den oben definierten Schätzer. In einer ersten Betrachtungsweise werden Funktionenräume mit polynomialen Splinefunktionen und in einer zweiten Betrachtungsweise werden neuronale Netze als Funktionenräume, auf die (4.8) definiert ist, gewählt. Für ersteres bedeutet das, dass der Splineraum  $S_{M,\alpha}$  als Funktionenraum gewählt wird, d. h. in (4.8) entspricht  $\mathcal{F}_{n,p}$  Funktionenräumen mit polynomialen Splinefunktionen für jeden Parameter  $p = (M, \alpha) \in \mathcal{P}_n$ . Der Splineraum  $S_{M,\alpha}$ , welcher in Abschnitt 2.2 definiert wurde, ist gegeben durch

$$S_{M,\alpha} = \left\{ \sum_{\mathbf{k} \in \mathbb{Z}^d: \text{supp}(B_{\mathbf{k},M}) \cap [-A,A]^d \neq \emptyset} a_{\mathbf{k}} \cdot B_{\mathbf{k},M} : a_{\mathbf{k}} \in \mathbb{R} \right\}.$$

Wie auch in Abschnitt 2.2 erwähnt, wird des Weiteren angenommen, dass der stochastische Prozess  $X_t$  ( $t \in \{0,1,\dots,T\}$ ) durch  $A > 0$  beschränkt ist. Weiter wird aus Gründen der Einfachheit angenommen, dass der stochastische Prozess stetige Pfade besitzt und somit eine zusätzliche Stützung des Prozesses vernachlässigbar ist.

Im Folgenden wird die Parametermenge  $\mathcal{P}_n$  für den Splineraum eingeführt. Dabei werden für eine Datenmenge mit Stichprobenumfang bzw. Kardinalität  $n$  die Parameter  $M$  sowie  $\alpha$  aus der Menge

$$\mathcal{P}_n = \left\{ (M, \alpha) : M \in \mathbb{N}_0, M \leq \lceil \log(n) \rceil, \alpha = 2^k \text{ für } k \in \mathbb{Z} \text{ mit } |k| \leq \lceil \log(n) \rceil \right\}$$

verwendet. Es sei hier erwähnt, dass durch die Wahl von  $M$  und  $\alpha$  der Splineraum  $S_{M,\alpha}$  eindeutig festgelegt wird. Mit obiger Parametermenge  $\mathcal{P}_n$  liefert das nächste Korollar eine Aussage zur Konsistenz des oben definierten Schätzers  $\hat{q}_{n,t}$ .

**Korollar 4.5 (Konsistenz)** Sei  $L > 0$ . Sei  $X_0, X_1, \dots, X_T$  ein  $\mathbb{R}^d$ -wertiger Markov-Prozess und sei  $X_t \in [-A, A]^d$  fast sicher für ein  $A > 0$  und für alle  $t \in \{0, 1, \dots, T\}$ . Des Weiteren sei die diskontierte Auszahlungsfunktion  $f_t$  absolut beschränkt durch  $L$ , d.h.

$$|f_t(x)| \leq L \quad \text{für } x \in [-A, A]^d \quad \text{und } t \in \{0, 1, \dots, T\}.$$

Sei  $q_t$  der Kontinuierungswert definiert durch (1.12). Definiere den Schätzer  $\hat{q}_{n,t}$  für alle  $t \in \{0, \dots, T\}$  durch (4.8) - (4.13) mit Daten definiert in (2.3) sowie mit  $\mathcal{F}_{n,p} = S_{M,\alpha}$  für  $p = (M, \alpha) \in \mathcal{P}_n$  und  $\mathcal{P}_n$  wie oben, und setze  $\beta_n = L$ . Dann gilt

$$\int |\hat{q}_{n,t}(x) - q_t(x)|^2 \mathbf{P}_{X_t}(dx) \rightarrow 0 \quad \text{nach Wahrscheinlichkeit}$$

für alle  $t \in \{0, \dots, T\}$ .

**Beweis:** Wegen der Beziehung

$$\int |\hat{q}_{n,t}(x) - q_t(x)|^2 \mathbf{P}_{X_t}(dx) \leq \sum_{w=0}^{T-t-1} \int |\hat{q}_{n,t}^w(x) - q_t(x)|^2 \mathbf{P}_{X_t}(dx)$$

genügt es,

$$\int |\hat{q}_{n,t}^w(x) - q_t(x)|^2 \mathbf{P}_{X_t}(dx) \rightarrow 0$$

nach Wahrscheinlichkeit für jedes  $t \in \{0, 1, \dots, T\}$  und jedes  $w \in \{0, 1, \dots, T-t-1\}$  zu zeigen, um Korollar 4.5 zu verifizieren.

Halte  $w \in \{0, 1, \dots, T-t-1\}$  fest. Für alle  $t \in \{0, 1, \dots, T\}$  liefert die Anwendung von Satz 4.1 für den  $L_2$ -Fehler des Schätzers  $\hat{q}_{n,t}^w$ :

$$\begin{aligned} & \int |\hat{q}_{n,t}^w(x) - q_t(x)|^2 \mathbf{P}_{X_t}(dx) \\ &= O_{\mathbf{P}} \left( \max_{s \in \{t, t+1, \dots, T-1\}} \min_{p \in \mathcal{P}_n} \int |q_{n_t, s}^{w,p}(x) - q_s(x)|^2 \mathbf{P}_{X_s}(dx) + \frac{L^2 \cdot \log |\mathcal{P}_n|}{n_t} \right), \end{aligned}$$

mit  $p = (M, \alpha)$ . Da  $M \leq \lceil \log(n) \rceil$ ,  $|k| \leq \lceil \log(n) \rceil$  und  $n_t = \lfloor n/3 \rfloor$  ist, folgt weiter für den letzten Term in obiger Gleichung

$$\frac{L^2 \cdot \log |\mathcal{P}_n|}{n_t} \leq \frac{L^2 \cdot \log(\lceil \log(n) \rceil) \cdot (2\lceil \log(n) \rceil + 1)}{n/3 - 1} \rightarrow 0 \quad (n \rightarrow \infty).$$

Um den Beweis abzuschließen, bleibt noch zu zeigen:

$$\max_{s \in \{t, t+1, \dots, T-1\}} \min_{p \in \mathcal{P}_n} \int |q_{n_l, s}^{w, p}(x) - q_s(x)|^2 \mathbf{P}_{X_s}(dx) \rightarrow 0 \quad (n \rightarrow \infty). \quad (4.14)$$

(4.14) wird mit Hilfe von Korollar 2.3 verifiziert. Dazu halte  $s \in \{t, t+1, \dots, T-1\}$  fest, und beachte, dass für  $p_n \in \mathcal{P}_n$  (mit  $p_n = (M, \alpha_n)$ ) die Beziehung

$$\min_{p \in \mathcal{P}_n} \int |q_{n_l, s}^{w, p}(x) - q_s(x)|^2 \mathbf{P}_{X_s}(dx) \leq \int |q_{n_l, s}^{w, p_n}(x) - q_s(x)|^2 \mathbf{P}_{X_s}(dx)$$

gilt. Weiter kann, aufgrund der Definition der Parametermenge  $\mathcal{P}_n$ , ein  $p_n \in \mathcal{P}_n$  mit  $p_n = (M, \alpha_n)$  gewählt werden, so dass  $\alpha_n$  die Bedingungen

$$\alpha_n \rightarrow 0 \quad (n \rightarrow \infty) \quad \text{und} \quad \frac{\log(n)}{\alpha_n^d \cdot n} \rightarrow 0 \quad (n \rightarrow \infty)$$

erfüllt. Die Anwendung von Korollar 2.3 auf den rechten Term in obiger Ungleichung liefert nun

$$\int |q_{n_l, s}^{w, p_n}(x) - q_s(x)|^2 \mathbf{P}_{X_s}(dx) \rightarrow 0$$

nach Wahrscheinlichkeit für alle  $s \in \{t, t+1, \dots, T-1\}$ . Unter Ausnutzung dieses Resultats kann (4.14) gefolgert werden, indem die Beziehung

$$\max_{s \in \{t, t+1, \dots, T-1\}} \min_{p \in \mathcal{P}_n} \int |q_{n_l, s}^{w, p}(x) - q_s(x)|^2 \mathbf{P}_{X_s}(dx) \leq \sum_{s=t}^{T-1} \int |q_{n_l, s}^{w, p_n}(x) - q_s(x)|^2 \mathbf{P}_{X_s}(dx)$$

berücksichtigt wird. Da (4.14) für festes aber beliebiges  $w \in \{0, 1, \dots, T-t-1\}$  gilt, ist der Beweis vollständig.  $\square$

Das nächste Korollar enthält eine Aussage zur Konvergenzgeschwindigkeit des Schätzers von  $q_t$ , wobei  $q_t$  als  $(p, C)$ -glatt vorausgesetzt wurde, um nicht-triviale Resultate herleiten zu können.

**Korollar 4.6 (Konvergenzgeschwindigkeit)** *Es gelten die Voraussetzungen von Korollar 4.5. Des Weiteren sei  $p, C > 0$  fest und sei  $q_t$   $(p, C)$ -glatt. Dann gilt*

$$\int |\hat{q}_{n, t}(x) - q_t(x)|^2 \mathbf{P}_{X_t}(dx) = O_{\mathbf{P}} \left( C^{\frac{2d}{2p+d}} \cdot \left( \frac{\log(n)}{n} \right)^{\frac{2p}{2p+d}} \right)$$

für alle  $t \in \{0, \dots, T\}$ .

**Beweis:** Analog zum Beweis von Korollar 4.5 genügt es, wegen

$$\int |\hat{q}_{n, t}(x) - q_t(x)|^2 \mathbf{P}_{X_t}(dx) \leq \sum_{w=0}^{T-t-1} \int |\hat{q}_{n, t}^w(x) - q_t(x)|^2 \mathbf{P}_{X_t}(dx)$$



die Beziehung

$$\int |\hat{q}_{n,t}^w(x) - q_t(x)|^2 \mathbf{P}_{X_t}(dx) = O_{\mathbf{P}} \left( C^{\frac{2d}{2p+d}} \cdot \left( \frac{\log(n)}{n} \right)^{\frac{2p}{2p+d}} \right)$$

für jedes  $t \in \{0, 1, \dots, T\}$  und jedes  $w \in \{0, 1, \dots, T-t-1\}$  zu zeigen.

Halte  $w \in \{0, 1, \dots, T-t-1\}$  fest. Die Anwendung von Satz 4.1 liefert, für alle  $t \in \{0, 1, \dots, T\}$ , die folgende Beziehung für den  $L_2$ -Fehler des Schätzers  $\hat{q}_{n,t}^w$ :

$$\begin{aligned} & \int |\hat{q}_{n,t}^w(x) - q_t(x)|^2 \mathbf{P}_{X_t}(dx) \\ &= O_{\mathbf{P}} \left( \max_{s \in \{t, t+1, \dots, T-1\}} \min_{p \in \mathcal{P}_n} \int |q_{n_i, s}^{w,p}(x) - q_s(x)|^2 \mathbf{P}_{X_s}(dx) + \frac{L^2 \cdot \log |\mathcal{P}_n|}{n_t} \right) \end{aligned} \quad (4.15)$$

mit  $p = (M, \alpha)$ . Unter Berücksichtigung folgender Ungleichungen für den linken Term auf der rechten Seite in (4.15),

$$\begin{aligned} & \max_{s \in \{t, t+1, \dots, T-1\}} \min_{p \in \mathcal{P}_n} \int |q_{n_i, s}^{w,p}(x) - q_s(x)|^2 \mathbf{P}_{X_s}(dx) \\ & \leq \sum_{s=t}^{T-1} \min_{p \in \mathcal{P}_n} \int |q_{n_i, s}^{w,p}(x) - q_s(x)|^2 \mathbf{P}_{X_s}(dx) \\ & \leq \sum_{s=t}^{T-1} \int |q_{n_i, s}^{w, p_n}(x) - q_s(x)|^2 \mathbf{P}_{X_s}(dx), \end{aligned}$$

und mit der Wahl des Parameters  $p_n = (M, \alpha_n) \in \mathcal{P}_n$ , wobei

$$\alpha_n \approx C^{-2/(2p+d)} \cdot (\log(n)/n)^{1/(2p+d)}$$

(genauer soll gelten  $|\alpha_n / (C^{-2/(2p+d)} \cdot (\log(n)/n)^{1/(2p+d)})| \leq \text{const}$ ), impliziert Korollar 2.5

$$\int |q_{n_i, s}^{w, p_n}(x) - q_s(x)|^2 \mathbf{P}_{X_s}(dx) = O_{\mathbf{P}} \left( C^{\frac{2d}{2p+d}} \cdot \left( \frac{\log(n)}{n} \right)^{\frac{2p}{2p+d}} \right).$$

Zum einen sei hier erwähnt, dass ein solches  $p_n \in \mathcal{P}_n$  mit  $\alpha_n$  wie oben aufgrund der Definition von  $\mathcal{P}_n$  gefunden bzw. gewählt werden kann. Zum anderen sei darauf hingewiesen, dass letztere Beziehung für festes aber beliebiges  $w \in \{0, 1, \dots, T-t-1\}$  gilt.

Als nächstes wird der letzte Term in (4.15) betrachtet. Für eine hinreichend große Konstante  $c_{19} > 0$  erhält man:

$$\begin{aligned} \frac{L^2 \cdot \log |\mathcal{P}_n|}{n_t} & \leq \frac{L^2 \cdot \log(\lceil \log(n) \rceil \cdot (2\lceil \log(n) \rceil + 1))}{n/3 - 1} \\ & \leq c_{19} \cdot \frac{\log(n)}{n}. \end{aligned}$$

Da weiter der Glättungsparameter  $C$  fest ist, kann die letzte Schranke nach oben durch

$$c_{20} \cdot C^{\frac{2d}{2p+d}} \cdot \left( \frac{\log(n)}{n} \right)^{\frac{2p}{2p+d}},$$

mit  $c_{20} > 0$  konstant, abgeschätzt werden, was den Beweis vervollständigt.  $\square$

**Bemerkung 4.7** *Der Schätzer in Korollar 4.5 und 4.6 entspricht einem adaptiven Kleinste-Quadrate-Splineschätzer.*

Nachdem Resultate hinsichtlich der Konsistenz und der Konvergenzgeschwindigkeit des Schätzers  $\hat{q}_{n,t}$  mit Splineräumen als Funktionenräume erfolgten, werden als nächstes vorwärtsgerichtete neuronale Netze (mit einer verdeckten Schicht) als Funktionenräume betrachtet.

Die Klasse der vorwärtsgerichteten neuronalen Netze, welche bereits in Abschnitt 2.3 definiert wurde, ist gegeben durch

$$\mathcal{F}_K(\beta_n) = \left\{ \sum_{i=1}^K d_i \cdot \sigma(a_i^T x + b_i) + d_0 \quad : \quad a_i \in \mathbb{R}^d, b_i \in \mathbb{R}, \sum_{i=0}^K |d_i| \leq \beta_n \right\},$$

wobei  $\beta_n > 0$  (mit  $\beta_n \rightarrow \infty$  ( $n \rightarrow \infty$ )) und  $\sigma$  die logistische Funktion  $\sigma(x) = 1/(1 + e^{-x})$  für  $x \in \mathbb{R}$  ist. Mit dieser Klasse sowie mit  $p = K \in \mathcal{P}_n$  und  $\mathcal{P}_n$  wie unten definiert, weist das nächste Korollar die Konsistenz des Schätzers  $\hat{q}_{n,t}$  von  $q_t$  nach. Dabei werden keine Beschränktheitsannahmen an  $X_t$  gestellt.

**Korollar 4.8 (Konsistenz)** *Sei  $L > 0$ . Sei  $X_0, X_1, \dots, X_T$  ein  $\mathbb{R}^d$ -wertiger Markov-Prozess und sei die diskontierte Auszahlungsfunktion  $f_t$  absolut beschränkt durch  $L$ , d.h.*

$$|f_t(x)| \leq L \quad \text{für } x \in \mathbb{R}^d \quad \text{und } t \in \{0, 1, \dots, T\}.$$

*Sei  $q_t$  der Kontinuierungswert definiert durch (1.12). Definiere den Schätzer  $\hat{q}_{n,t}$  für alle  $t \in \{0, \dots, T\}$  durch (4.8) - (4.13) mit Daten definiert in (2.3) sowie mit  $\mathcal{F}_{n,p} = \mathcal{F}_K(\beta_n)$  für  $p = K \in \mathcal{P}_n$ .  $\mathcal{P}_n$  sei gegeben durch  $\mathcal{P}_n = \{1, \dots, n\}$ . Sei weiter  $\beta_n > 0$  und erfülle*

$$\beta_n \rightarrow \infty \quad (n \rightarrow \infty) \quad \text{und} \quad \frac{\beta_n^4 \cdot \log(n)}{n} \rightarrow 0 \quad (n \rightarrow \infty).$$

*Dann gilt*

$$\int |\hat{q}_{n,t}(x) - q_t(x)|^2 \mathbf{P}_{X_t}(dx) \rightarrow 0 \quad \text{nach Wahrscheinlichkeit}$$

*für alle  $t \in \{0, \dots, T\}$ .*

**Beweis:** Der Beweis verläuft ähnlich wie der Beweis von Korollar 4.5. Der Vollständigkeit halber wird er explizit angegeben. Aufgrund der Beziehung

$$\int |\hat{q}_{n,t}(x) - q_t(x)|^2 \mathbf{P}_{X_t}(dx) \leq \sum_{w=0}^{T-t-1} \int |\hat{q}_{n,t}^w(x) - q_t(x)|^2 \mathbf{P}_{X_t}(dx)$$

genügt es,

$$\int |\hat{q}_{n,t}^w(x) - q_t(x)|^2 \mathbf{P}_{X_t}(dx) \rightarrow 0$$

nach Wahrscheinlichkeit für jedes  $t \in \{0, 1, \dots, T\}$  und jedes  $w \in \{0, 1, \dots, T-t-1\}$  zu zeigen, um Korollar 4.8 zu verifizieren.

Halte  $w \in \{0, 1, \dots, T-t-1\}$  fest. Mit der Anwendung von Satz 4.1 auf den  $L_2$ -Fehler des Schätzers  $\hat{q}_{n,t}^w$  erhält man für alle  $t \in \{0, 1, \dots, T\}$ :

$$\begin{aligned} & \int |\hat{q}_{n,t}^w(x) - q_t(x)|^2 \mathbf{P}_{X_t}(dx) \\ &= O_{\mathbf{P}} \left( \max_{s \in \{t, t+1, \dots, T-1\}} \min_{p \in \mathcal{P}_n} \int |q_{n_t, s}^{w,p}(x) - q_s(x)|^2 \mathbf{P}_{X_s}(dx) + \frac{\beta_n^2 \cdot \log |\mathcal{P}_n|}{n_t} \right), \end{aligned}$$

wobei  $p = K$  ist. Mit  $\mathcal{P}_n = \{1, \dots, n\}$  und  $n_t = \lfloor n/3 \rfloor$  folgt weiter für den letzten Term in obiger Gleichung

$$\frac{\beta_n^2 \cdot \log |\mathcal{P}_n|}{n_t} \leq \frac{\beta_n^2 \cdot \log(n)}{n/3 - 1} \rightarrow 0 \quad (n \rightarrow \infty),$$

da nach Voraussetzung  $\beta_n^4 \cdot \log(n)/n$  gegen Null konvergiert für  $n$  gegen unendlich.

Um den Beweis abzuschließen, bleibt noch

$$\max_{s \in \{t, t+1, \dots, T-1\}} \min_{p \in \mathcal{P}_n} \int |q_{n_t, s}^{w,p}(x) - q_s(x)|^2 \mathbf{P}_{X_s}(dx) \rightarrow 0 \quad (n \rightarrow \infty) \quad (4.16)$$

zu zeigen, was mit Hilfe von Korollar 2.9 verifiziert wird. Dazu halte  $s \in \{t, t+1, \dots, T-1\}$  fest und beachte, dass für  $p_n = K_n \in \mathcal{P}_n$  die Beziehung

$$\min_{p \in \mathcal{P}_n} \int |q_{n_t, s}^{w,p}(x) - q_s(x)|^2 \mathbf{P}_{X_s}(dx) \leq \int |q_{n_t, s}^{w, p_n}(x) - q_s(x)|^2 \mathbf{P}_{X_s}(dx)$$

gilt. Wegen der Bedingungen in Korollar 4.8 kann weiter ein  $p_n = K_n \in \mathcal{P}_n$  so gewählt werden, dass  $K_n$  nachstehenden Bedingungen genüge:

$$K_n \rightarrow \infty \quad (n \rightarrow \infty) \quad \text{und} \quad \frac{\beta_n^4 \cdot K_n \cdot \log(n)}{n} \rightarrow 0 \quad (n \rightarrow \infty).$$

Zum Beispiel erfüllt

$$K_n = \left\lceil (n / \log^5(n))^{1/2} \right\rceil$$

diese Bedingungen. Die Anwendung von Korollar 2.9 auf den rechten Term in obiger Ungleichung liefert nun

$$\int |q_{n_t, s}^{w, p_n}(x) - q_s(x)|^2 \mathbf{P}_{X_s}(dx) \rightarrow 0$$

nach Wahrscheinlichkeit für alle  $s \in \{t, t+1, \dots, T-1\}$ . Unter Ausnutzung dieses Resultats und mit der Beziehung

$$\max_{s \in \{t, t+1, \dots, T-1\}} \min_{p \in \mathcal{P}_n} \int |q_{n_i, s}^{w, p}(x) - q_s(x)|^2 \mathbf{P}_{X_s}(dx) \leq \sum_{s=t}^{T-1} \int |q_{n_i, s}^{w, p_n}(x) - q_s(x)|^2 \mathbf{P}_{X_s}(dx)$$

kann (4.16) gefolgert werden. Da (4.16) für festes aber beliebiges  $w \in \{0, 1, \dots, T-t-1\}$  gilt, ist der Beweis vollständig.  $\square$

**Bemerkung 4.9** *Der Schätzer im obigen sowie im nächsten Korollar entspricht einem adaptiven Neuronale-Netze-Schätzer.*

Korollar 4.8 liefert einen Konsistenznachweis des Schätzers  $\hat{q}_{n,t}$ . Um eine Aussage hinsichtlich der Konvergenzgeschwindigkeit des Schätzers zu treffen, wird die Beschränktheitsannahme von oben (oder aus Abschnitt 2.3) an den stochastischen Prozess  $X_t$  ( $t \in \{0, 1, \dots, T\}$ ) durch  $A > 0$  gestellt. Aus Gründen der Einfachheit wird auch angenommen, dass der stochastische Prozess stetige Pfade besitzt, womit eine zusätzliche Stützung des Prozesses vernachlässigbar ist. Des Weiteren werden, um nicht-triviale Resultate herleiten zu können, Forderungen an die Fourier-Transformierte von  $q_t$  gemäß (2.16) und (2.17) gestellt, was eine Einschränkung der Glattheit von  $q_t$  zur Folge hat. Das nächste Korollar liefert eine Aussage zur Konvergenzgeschwindigkeit des Schätzers  $\hat{q}_{n,t}$  von  $q_t$ .

**Korollar 4.10 (Konvergenzgeschwindigkeit)** *Es gelten die Voraussetzungen von Korollar 4.8. Setze  $\beta_n = \text{const} \cdot \log(n)$ . Des Weiteren sei  $X_t \in [-A, A]^d$  fast sicher für ein  $A > 0$  und für alle  $t \in \{0, 1, \dots, T\}$ . Erfüllt die Fourier-Transformierte  $\tilde{Q}_t$  von  $q_t$  für alle  $x \in \mathbb{R}^d$  und alle  $t \in \{0, \dots, T\}$  (2.16) und (2.17), dann gilt*

$$\int |\hat{q}_{n,t}(x) - q_t(x)|^2 \mathbf{P}_{X_t}(dx) = O_{\mathbf{P}} \left( \sqrt{\frac{\log^5(n)}{n}} \right)$$

für alle  $t \in \{0, \dots, T\}$ .

**Beweis:** Die Vorgehensweise ist ähnlich wie die Vorgehensweise im Beweis von Korollar 4.6. Analog zum Beweis von Korollar 4.8 genügt es, wegen

$$\int |\hat{q}_{n,t}(x) - q_t(x)|^2 \mathbf{P}_{X_t}(dx) \leq \sum_{w=0}^{T-t-1} \int |\hat{q}_{n,t}^w(x) - q_t(x)|^2 \mathbf{P}_{X_t}(dx)$$

die Beziehung

$$\int |\hat{q}_{n,t}^w(x) - q_t(x)|^2 \mathbf{P}_{X_t}(dx) = O_{\mathbf{P}} \left( \sqrt{\frac{\log^5(n)}{n}} \right)$$

für jedes  $t \in \{0, 1, \dots, T\}$  und jedes  $w \in \{0, 1, \dots, T-t-1\}$  zu zeigen.

Halte  $w \in \{0, 1, \dots, T-t-1\}$  fest. Mit der Anwendung von Satz 4.1 auf den  $L_2$ -Fehler des Schätzers  $\hat{q}_{n,t}^w$  wird für alle  $t \in \{0, 1, \dots, T\}$  die folgende Beziehung erhalten:

$$\begin{aligned} & \int |\hat{q}_{n,t}^w(x) - q_t(x)|^2 \mathbf{P}_{X_t}(dx) \\ &= O_{\mathbf{P}} \left( \max_{s \in \{t, t+1, \dots, T-1\}} \min_{p \in \mathcal{P}_n} \int |q_{n_i, s}^{w,p}(x) - q_s(x)|^2 \mathbf{P}_{X_s}(dx) + \frac{\beta_n^2 \cdot \log |\mathcal{P}_n|}{n_t} \right) \end{aligned} \quad (4.17)$$

mit  $p = K$ . Werden weiter für den linken Term auf der rechten Seite in (4.17) die Ungleichungen

$$\begin{aligned} & \max_{s \in \{t, t+1, \dots, T-1\}} \min_{p \in \mathcal{P}_n} \int |q_{n_i, s}^{w,p}(x) - q_s(x)|^2 \mathbf{P}_{X_s}(dx) \\ & \leq \sum_{s=t}^{T-1} \min_{p \in \mathcal{P}_n} \int |q_{n_i, s}^{w,p}(x) - q_s(x)|^2 \mathbf{P}_{X_s}(dx) \\ & \leq \sum_{s=t}^{T-1} \int |q_{n_i, s}^{w, p_n}(x) - q_s(x)|^2 \mathbf{P}_{X_s}(dx) \end{aligned}$$

berücksichtigt, und wird  $p_n = K_n \in \mathcal{P}_n$  gemäß

$$K_n = \left\lceil (n / \log^5(n))^{1/2} \right\rceil$$

gewählt, dann impliziert Korollar 2.9 (wegen  $\beta_n = \text{const} \cdot \log(n)$ )

$$\int |q_{n_i, s}^{w, p_n}(x) - q_s(x)|^2 \mathbf{P}_{X_s}(dx) = O_{\mathbf{P}} \left( \sqrt{\frac{\log^5(n)}{n}} \right).$$

Zum einen sei hier erwähnt, dass ein solches  $p_n = K_n \in \mathcal{P}_n$  mit  $K_n$  wie oben aufgrund der Definition von  $\mathcal{P}_n$  gefunden bzw. gewählt werden kann. Zum anderen sei darauf hingewiesen, dass letztere Beziehung für festes aber beliebiges  $w \in \{0, 1, \dots, T-t-1\}$  gilt.

Als nächstes wird der letzte Term in (4.17) betrachtet. Da nach Voraussetzung  $\beta_n = \text{const} \cdot \log(n)$  gilt, erhält man für hinreichend große Konstanten  $c_{21}, c_{22} > 0$  sowie für  $n$  genügend groß

$$\begin{aligned} \frac{\beta_n^2 \cdot \log |\mathcal{P}_n|}{n_t} & \leq \frac{\beta_n^2 \cdot \log(n)}{n/3 - 1} \\ & \leq c_{21} \cdot \frac{\log^3(n)}{n} \\ & \leq c_{22} \cdot \left( \frac{\log^5(n)}{n} \right)^{1/2}, \end{aligned}$$

was den Beweis vervollständigt. □

**Bemerkungen.**

**Bemerkung 4.11** *Es sei hier betont, dass in Korollar 4.8 keine Voraussetzung an die Verteilung von  $X_t$  gestellt wird. In Korollar 4.5, 4.6 und 4.10 wird ebenfalls keine Voraussetzung an die Verteilung von  $X_t$  gestellt, außer dass  $X_t \in [-A, A]^d$  fast sicher gilt. Insbesondere wird nicht die Existenz einer Dichte von  $X_t$  bezüglich des Lebesgue–Borel–Maßes gefordert.*

**Bemerkung 4.12** *Der Schätzer in Korollar 4.6 bzw. 4.10 erreicht die gleiche Konvergenzrate wie der Schätzer in Abschnitt 2.2 bzw. 2.3. D. h. die Konvergenzrate in Korollar 4.6 stimmt bis auf den logarithmischen Faktor  $(\log(n))^{2p/(2p+d)}$  mit der üblichen optimalen Konvergenzrate aus der Regressionsschätzung überein, und in Korollar 4.10 erhält man bis auf einen logarithmischen Faktor  $(\log^5(n))^{1/2}$  die parametrische Konvergenzrate  $n^{-1/2}$ .*

**Bemerkung 4.13** *Die Definition des Schätzers in Korollar 4.6 bzw. 4.10 hängt nicht vom Grad der Glattheit von  $q_t$  ab, welche durch  $(p, C)$  im ersten bzw. durch die Bedingungen (2.16) und (2.17) an die Fourier-Transformierte (von  $q_t$ ) im zweiten Korollar repräsentiert wird. In Korollar 4.6 erreicht der Schätzer dennoch die (bis auf den logarithmischen Faktor  $(\log(n))^{2p/(2p+d)}$ ) “optimale” Konvergenzrate für eine bestimmte Glattheit des Kontinuierungswertes. In diesem Sinne ist der Schätzer in der Lage, sich automatisch an die Glattheit des Kontinuierungswertes anzupassen.*

**Bemerkung 4.14** *Angenommen,  $X_0 = x_0$  fast sicher für ein  $x_0 \in [-A, A]^d$  bzw. für ein  $x_0 \in \mathbb{R}^d$ . Der Preis*

$$V_0 = \max\{f_0(x_0), q_0(x_0)\}$$

*einer Amerikanischen Option (vgl. (1.10) und (1.12)) kann durch*

$$\hat{V}_0 = \max\{f_0(x_0), \hat{q}_{n,0}(x_0)\}$$

*geschätzt werden. Da die Verteilung von  $X_0$  auf  $x_0$  konzentriert ist, gilt unter den Voraussetzungen in Korollar 4.6 (mit  $x_0 \in [-A, A]^d$ ) die folgende Fehlerschranke:*

$$\begin{aligned} |\hat{V}_0 - V_0|^2 &= |\max\{f_0(x_0), \hat{q}_{n,0}(x_0)\} - \max\{f_0(x_0), q_0(x_0)\}|^2 \\ &\leq |\hat{q}_{n,0}(x_0) - q_0(x_0)|^2 \\ &= O_{\mathbf{P}} \left( C^{\frac{2d}{2p+d}} \cdot \left( \frac{\log(n)}{n} \right)^{\frac{2p}{2p+d}} \right). \end{aligned}$$

*Mit  $x_0 \in \mathbb{R}^d$  und unter den Voraussetzungen in Korollar 4.10 erhält man die Fehlerschranke*

$$|\hat{V}_0 - V_0|^2 \leq O_{\mathbf{P}} \left( \left( \frac{\log^5(n)}{n} \right)^{1/2} \right).$$

**Bemerkung 4.15** *Es ist zu beachten, dass in den Korollaren 4.5, 4.6, 4.8 und 4.10 für  $q_t$  die Regressionsdarstellung (2.1) verwendet wurde. Dies impliziert für  $w = 0$  die Regressionsdarstellung von Tsitsiklis und Van Roy (1999) (vgl. (1.15)) bzw. für  $w = T - t - 1$  die*

Darstellung von Longstaff und Schwartz (2001) (vgl. (1.16)). Da die simulationsbasierten Methoden zur Bewertung von Amerikanischen Optionen in Tsitsiklis und Van Roy (1999) sowie in Longstaff und Schwartz (2001) auf der parametrischen Regressionsschätzung beruhen, wurden mit den oben genannten Korollaren die Methoden in Tsitsiklis und Van Roy (1999) sowie in Longstaff und Schwartz (2001) auf die nichtparametrische Regressionsschätzung erweitert.

**Bemerkung 4.16** *Es sei hier betont, dass das in dieser Arbeit vorgestellte Verfahren zur Schätzung des Wertes einer Amerikanischen Option bzw. der Kleinste-Quadrate-Schätzer  $\hat{q}_{n,t}$  auf alle Finanzprozesse  $X_t$  angewendet werden kann, sofern  $X_t$  die Markov-Eigenschaft erfüllt. Somit ist das Verfahren universell einsetzbar in dem Sinne, dass durch Verwendung unterschiedlicher Ansätze zur Modellierung des Finanzprozesses  $X_t$  das Verfahren nicht modifiziert werden muss und deshalb sofort anwendbar ist (sofern  $X_t$  die Markov-Eigenschaft besitzt). Beispielsweise können GARCH-Modelle (siehe Kapitel 1) zur Modellierung von Finanzprozessen verwendet werden, wenn man berücksichtigt, dass sie durch eine geeignete Darstellung die Markov-Eigenschaft erfüllen. Werden des Weiteren Voraussetzungen an die Verteilung von  $X_t$  gemäß Bemerkung 4.11 gestellt, dann gelten die in diesem Abschnitt erzielten Resultate.*

**Bemerkung 4.17** *Adaptive nichtparametrische Smoothing-Spline-Regressionsschätzer zur Bewertung Amerikanischer Optionen wurden in diesem Kontext auch in Kohler (2006b) untersucht. Dort erfolgten Aussagen zur Konsistenz und zur Konvergenzgeschwindigkeit.*

Zusammenfassend lässt sich festhalten, dass die Schätzer in diesem Abschnitt - im Zusammenhang mit der Schätzung des Wertes einer Amerikanischen Option - durch Techniken der Adaption für die vollautomatische Wahl der Parameter die gleichen Resultate bezüglich Konsistenz und Konvergenzgeschwindigkeit erzielen wie die Schätzer aus den vorherigen Kapiteln. Als Technik der Adaption für die vollautomatische Parameterwahl wurde dabei die Unterteilung der Stichprobe gewählt. Des Weiteren wurden durch eine datenabhängige Wahl des Parameters  $w$  vollautomatische Kleinste-Quadrate-Schätzer  $\hat{q}_{n,t}$  definiert. Diese Schätzer sind einerseits in der Lage,  $w$  sowie alle weiteren Parameter vollautomatisch zu bestimmen, um eine "gute" Schätzung der unbekanntenen Regressionsfunktion  $q_t$  in dem Sinne zu liefern, dass das  $L_2$ -Risiko der Schätzfunktion nahe beim  $L_2$ -Risiko der zu schätzenden Regressionsfunktion liegt. Andererseits berechnen sie eine approximative optimale Stoppregel, die eine untere Schranke für die Lösung des optimalen Stoppproblems zur Zeit  $t$  liefert und maximieren diese Schranke mit der Spezifikation (4.12) für  $\hat{w}_t$ . In diesem Sinne ist auch der Algorithmus, der rekursiv Schätzer von  $q_{t-1}, \dots, q_0$  berechnet (vgl. Kapitel 2), adaptiv.

Mit den Schätzern  $\hat{q}_{n,t}$  ( $t \in \{0, 1, \dots, T\}$ ), die als adaptive Kleinste-Quadrate-Splineschätzer sowie adaptive Neuronale-Netze-Schätzer eingeführt wurden, kann weiterhin der Preis einer Amerikanischen Option mit Hilfe von  $\hat{q}_{n,0}$  geschätzt werden. Hiermit und mit den vorangehenden Kapiteln sind die theoretischen Grundlagen bereit, die es erlauben, eine praktische Regressionsschätzung bzw. eine Schätzung des Wertes einer Amerikanischen Option im Ein- oder Hochdimensionalen auf simulierten Daten zu betrachten.

Das nächste Kapitel wird sich mit simulierten Anwendungsbeispielen beschäftigen.



# Kapitel 5

## Anwendung auf simulierte Daten

In diesem Kapitel soll anhand simulierter Anwendungsbeispiele die praktische Anwendbarkeit des in dieser Arbeit vorgestellten rekursiven Algorithmus zur Schätzung der Kontinuierungswerte mit Anwendung in der Bewertung Amerikanischer Optionen demonstriert werden. Für dieses Vorhaben werden einerseits Kleinste-Quadrate-Splineschätzer sowie andererseits Neuronale-Netze-Schätzer als nichtparametrische Schätzer gewählt, mit denen der Wert einer Amerikanischen Option auf simulierten Daten geschätzt wird. Bei beiden Schätzern erfolgt die automatische Wahl der Parameter mit der Unterteilung der Stichprobe. Der Kleinste-Quadrate-Splineschätzer ist - wie bereits angeführt - zum einen numerisch gut zu handhaben, und zum anderen besitzt er gute Approximationseigenschaften im Deterministischen. Ein weiterer Grund für die Wahl dieses Schätzers ist seine Glattheit. Die hier betrachteten Amerikanischen Optionen basieren auf einem oder auch auf mehreren Underlyings, die alle samt Aktien entsprechen. Bei Simulationen stellte sich jedoch heraus, dass der zur Verfügung stehende Speicherplatz in Matlab<sup>®</sup> für die Kleinste-Quadrate-Splineschätzer im Hochdimensionalen nicht ausreichte (aufgrund der verwendeten Tensorprodukträume). Daher wurde zusätzlich ein Neuronale-Netze-Schätzer gewählt, da dieser eines von wenigen Standardverfahren im Hochdimensionalen darstellt. Der Neuronale-Netze-Schätzer wird hier approximativ mit "Backpropagation" (d. h. durch Gradientenabstieg) berechnet.

Bevor man auf die praktische Anwendbarkeit der aus dieser Arbeit resultierenden Verfahren eingeht, wird als erstes die Regressionsschätzung mit nichtparametrischen Methoden, angewandt auf die Regressionsdarstellung von Tsitsiklis und Van Roy (1999) sowie von Longstaff und Schwartz (2001), betrachtet. Genauer bedeutet dies, dass der Wert einer Amerikanischen Option (auf simulierten Daten) mittels des Kleinste-Quadrate-Splineschätzers mit  $w = 0$  und  $w = T - t - 1$  geschätzt wird. Zusätzlich wird auch der Fall  $w \in \{0, 1, \dots, T - t - 1\}$  betrachtet. Mit dieser Betrachtungsweise der nichtparametrischen Regressionsschätzung wird das Verhalten des (nichtparametrischen) Schätzers in Abhängigkeit von der Wahl von  $w$  geprüft. Da es sich bei den Schätzern herausgestellt hat, dass die datenabhängige Wahl des Parameters  $w$  keine besseren (aber sehr wohl viel aufwendig zu berechnenden) Resultate liefern, basieren alle Simulationen mit dem Neuronale-Netze-Schätzer auf  $w = 0$ .

Weiter wird das endliche Stichproben-Verhalten des in dieser Arbeit vorgestellten Algorithmus durch einen Vergleich mit dem Algorithmus von Tsitsiklis und Van Roy (1999) und dem Algorithmus von Longstaff und Schwartz (2001) illustriert werden.

Dazu wurden einerseits die oben erwähnten Algorithmen und die daraus resultierenden Schätzer in Matlab<sup>®</sup>, Version 7.0.4.365 (R14) Service Pack 2, implementiert sowie alle Simulationen mit der selben Software durchgeführt.

### Daten-Erzeugung

Die Pfade der zugrunde liegenden Aktien werden mittels eines einfachen Black-Scholes-Modells erzeugt. Ein Grund für die Wahl des Black-Scholes-Modells ist seine häufige Anwendung in der Praxis (in Bezug auf Optionsbewertung). Ein weiterer Grund ist die (relativ) leichte Verständlichkeit des Modells. An dieser Stelle sei erwähnt, dass beispielsweise auch ohne Weiteres ein GARCH-Modell oder alle anderen Modelle, bei denen durch geeignete Darstellung der Preisprozess ein Markov-Prozess ist, zur Modellierung des Pfades verwendet werden können. Damit sind auch modernere Methoden zur Modellierung des Kurses einer Aktie einsetzbar.

Die Verfallszeit  $T$  der Amerikanischen Optionen, sowohl basierend auf einer Aktie als auch basierend auf mehreren Aktien (Basket-Optionen), wird als ein Jahr angenommen. Dementsprechend ist die maximale Laufzeit der Optionen ein Jahr. Das Zeitintervall  $[0, 1]$  wird in  $m$  äquidistante Zeitschritte mit

$$t_0 = 0 < t_1 < \dots < t_m = T = 1$$

diskretisiert. Sowohl in den ersten beiden als auch im vierten und fünften Beispiel wird eine Option auf nur eine Aktie betrachtet. Die Kurse der zugrunde liegenden Aktie sind zu den Zeitpunkten  $t_j$  ( $j = 0, \dots, m$ ) gegeben durch

$$X_{i,t_j} = x_0 \cdot \exp\left((r - 1/2 \cdot \sigma^2) \cdot t_j + \sigma \cdot W_{t_j}\right) \quad (i = 1, \dots, n, j = 0, \dots, m).$$

Die jeweiligen Werte von  $x_0$ ,  $r$  und  $\sigma$  in obiger Gleichung sowie  $n$  und  $m$  sind in den jeweiligen Beispielen (siehe unten) angegeben. Weiter werden in allen Beispielen als Diskontierungsfaktoren  $e^{-rt_j}$  für  $j = 0, \dots, m$  gewählt. Um mit den bisher verwendeten Notationen für  $w$  weiterzuarbeiten, sollen die folgenden Abkürzungen für die Werte von  $w$  gelten:  $T - t - 1 := T - t_j - \frac{1}{m}$  und  $4 := \frac{4}{m}$ .  $w \in \{0, 4, T - t - 1\}$  beispielsweise bedeutet demnach  $w \in \{0, \frac{4}{m}, T - t_j - \frac{1}{m}\}$ .

### Betrachtete Optionstypen

Wie oben bereits erwähnt wurde, wird der Wert einer Amerikanischen Option mittels des Algorithmus von Tsitsiklis und Van Roy (1999), des Algorithmus von Longstaff und Schwartz (2001) und des aus dieser Arbeit vorgestellten Algorithmus geschätzt. Dabei werden basierend auf einer Aktie eine Put-Option mit Ausübungspreis 90 sowie eine Strangle-Spread-Option mit Ausübungspreisen 50, 90, 110 und 150 verwendet. Die

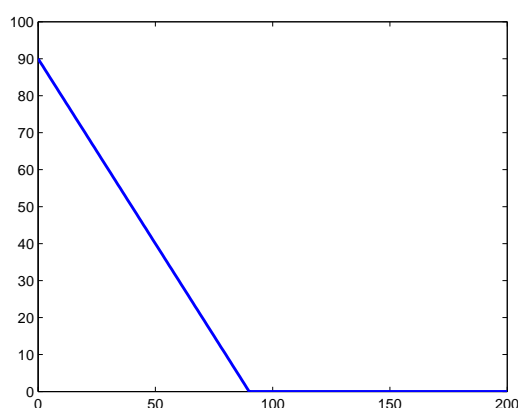


Abbildung 5.1: *Put-Auszahlung der Put-Option mit Ausübungspreis 90.*

Auszahlung ersterer Option ist in Abbildung 5.1 und die Strangle-Spread-Auszahlung (der Strangle-Spread-Option) ist in Abbildung 5.2 illustriert.

Im hochdimensionalen Fall werden Basket-Optionen verwendet. Dabei basieren diese einmal auf drei und einmal auf fünf Aktien. Im Falle von drei Aktien sind die Optionspreise normalisiert, so dass sie bei 1 starten. Dies rührt daher, dass bei Simulationen mit dem Kleinste-Quadrate-Splineschätzer der interne Speicher in Matlab<sup>®</sup> nicht ausreichte, was wiederum zurückzuführen ist auf die enorme Anzahl der zu berechnenden Koeffizienten des zu lösenden linearen Gleichungssystems des Schätzers. (Die enorme Anzahl der Koeffizienten wird erreicht, falls beispielsweise  $A$  sehr groß gewählt wird, was bei  $x_0 = 100$  sinnvoll wäre, und der Knotenabstand  $\alpha$  zu klein gewählt wird. Damit erhält man schnell ein sehr großes lineares Gleichungssystem, das es zu lösen gilt, bzw. macht Operationen mit Matrizen der Größe 10000x10000 oder mehr). Weiter wird, im Falle von drei Aktien, bei der Basket-Option die Strangle-Spread-Auszahlung mit Ausübungspreisen 0.85, 0.95, 1.05 und 1.15 für den Durchschnittswert der drei zugrunde liegenden korrelierten Aktienkurse verwendet. Im Falle von fünf Aktien wird ebenfalls bei der Basket-Option die Strangle-Spread-Auszahlung, nun aber mit Ausübungspreisen 75, 90, 110 und 125, für den Durchschnittswert der fünf zugrunde liegenden korrelierten Aktienkurse verwendet.

### Vergleichskriterien

Im Folgenden wird die Vorgehensweise zum Vergleichen der drei Algorithmen erläutert, was mit Hilfe von so genannten Boxplots bewerkstelligt wird. Auf die selbe Art und Weise werden auch die Resultate der Regressionsschätzung mit nichtparametrischen Methoden, angewandt auf die Regressionsdarstellung von Tsitsiklis und Van Roy (1999), von Longstaff und Schwartz (2001) und jene in (2.1), verglichen. Wie oben bereits erwähnt, bedeutet dies, dass bei ersterer Regressionsdarstellung ein Kleinste-Quadrate-Splineschätzer mit  $w = 0$ , bei zweiterer mit  $w = T - t - 1$  und bei letzterer

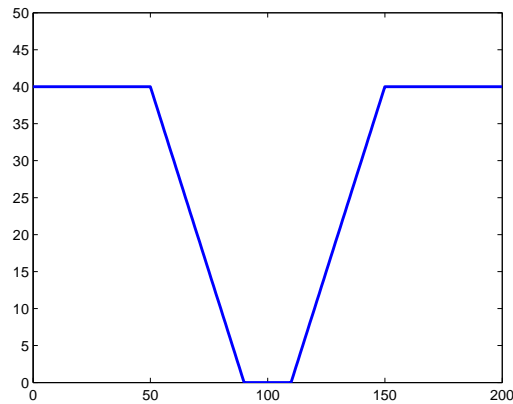


Abbildung 5.2: *Strangle-Spread-Auszahlung mit Ausübungspreisen 50, 90, 110 und 150.*

Regressionsdarstellung ein Kleinste-Quadrate-Splineschätzer mit  $w \in \{0, 4, T - t - 1\}$  die Werte einer Amerikanischen Option (auf simulierten Daten) schätzt. In diesem Sinne wird auf den Einfluss von  $w$  bei der Regressionsschätzung mit nichtparametrischen Methoden eingegangen. Bevor nun die Vorgehensweise zum Vergleichen erläutert wird, sei noch erwähnt, dass der Algorithmus aus dieser Arbeit genutzt wird, um die daraus resultierenden Verfahren praktisch umzusetzen. Als Schätzer erhält man einen Kleinste-Quadrate-Splineschätzer sowie einen Neuronale-Netze-Schätzer, mit denen der Wert einer Amerikanischen Option geschätzt wird. Mit den verbleibenden beiden Algorithmen erhält man parametrische Schätzer, die ebenfalls den Wert einer Amerikanischen Option schätzen. Alle vier Schätzer sind Monte-Carlo-Schätzer und liefern Monte-Carlo-Schätzungen.

Um alle drei Algorithmen zu vergleichen, wird wie folgt vorgegangen:

1. Es werden  $n$  Pfade bzw.  $n$  Aktienkursverläufe  $X_{i,t_j}$  ( $i = 1, \dots, n$ ,  $j = 1, \dots, m$ ) gemäß der oben angegebenen Gleichung simuliert (die Werte von  $n$  und  $m$  sind unten angegeben).
2. Anhand dieser simulierten Pfade werden die jeweiligen Monte-Carlo-Schätzer berechnet.
3. Jeder der berechneten Monte-Carlo-Schätzer wird auf 4000 neu simulierten Pfaden mit der jeweiligen Stoppregel ausgewertet (und somit erhält man für jeden Monte-Carlo-Schätzer 4000 geschätzte Optionspreise).
4. Das arithmetische Mittel wird über die jeweiligen 4000 geschätzten Optionspreise gebildet, und der daraus resultierende Optionspreis wird als endgültig geschätzter Optionspreis des jeweiligen Monte-Carlo-Schätzers verwendet.
5. Die Schritte 1. bis 4. werden 100 mal (auf neuen Daten) wiederholt.

6. Die 100 geschätzten Optionspreise der jeweiligen Monte-Carlo-Schätzer werden in Boxplots dargestellt.
7. Diese Boxplots werden miteinander verglichen.

Somit erhält man für jeden Schätzer 100 geschätzte Optionspreise, die sich in den entsprechenden Boxplots widerspiegeln. Es sei hier betont, dass alle drei Algorithmen eine untere Schranke des optimalen Stoppwertes anbieten. Weiter gilt, dass die approximative optimale Stoppregel auf neu generierten Daten ausgewertet wird. Aufgrund dessen kennzeichnet eine höhere Monte-Carlo-Schätzung eine bessere Leistung des Algorithmus. Da die geschätzten Optionspreise der jeweiligen Monte-Carlo-Schätzer mittels Boxplots dargestellt werden und die Algorithmen mit Hilfe dieser Boxplots verglichen werden, gilt, dass der Algorithmus umso besser ist, je höher der Boxplot der Monte-Carlo-Schätzungen liegt.

### **Beeinflusst die Wahl von $w$ die Regressionsschätzung mit nichtparametrischen Methoden?**

Hier wird der Einfluss von  $w$  auf die Regressionsschätzung mit nichtparametrischen Methoden, angewandt auf die Regressionsdarstellung von Tsitsiklis und Van Roy (1999), Longstaff und Schwartz (2001) und (2.1), analysiert. Um den Einfluss von  $w$  bei der Regressionsschätzung mit nichtparametrischen Methoden zu prüfen, wird der Wert einer Amerikanischen Option mittels des Kleinste-Quadrate-Splineschätzers für  $w = 0$ ,  $w = T - t - 1$  und  $w \in \{0, 4, T - t - 1\}$  100 mal geschätzt und mittels Boxplots verglichen (wie oben beschrieben). Dabei werden im eindimensionalen Fall die Put-Option sowie die Strangle-Spread-Option und im dreidimensionalen Fall die Basket-Option verwendet. Im Weiteren wird der Einfachheit halber der Kleinste-Quadrate-Splineschätzer als KQS bezeichnet.

In einem ersten Beispiel wird die Put-Auszahlung mit Ausübungspreis 90, wie in Abbildung 5.1 gegeben, analysiert. Die Kursverläufe der zugrunde liegenden Aktie werden, wie oben beschrieben, mit

$$x_0 = 100, r = 0.05, \sigma = 0.25, n = 10000 \quad \text{und} \quad m = 12$$

simuliert. Für den KQS ist dabei die Anzahl der Lerndaten, Testdaten und Validierungsdaten durch  $n_l = 6000$  und  $n_t = n_v = 2000$  gegeben. Um die Implementierung zu vereinfachen, wird weiter der Grad  $M$  aus der Menge  $\{0, 1, 2\}$  sowie der Knotenabstand  $\alpha$  aus der Menge  $\{\frac{100}{2}, \frac{100}{2^2}, \frac{100}{2^3}, \frac{100}{2^4}\}$ , datenabhängig wie in Abschnitt 4.1 beschrieben, gewählt. Wie in Abbildung 5.3 zu sehen ist, unterscheiden sich die Resultate bei den verwendeten Regressionsdarstellungen kaum, da keiner der Boxplots signifikant höher ist. Daraus lässt sich folgern, dass in diesem Beispiel die Wahl von  $w$  kaum einen Einfluss auf die Schätzungen der Optionspreise besitzt.

In einem zweiten Beispiel wird die kompliziertere Strangle-Spread-Auszahlung mit Ausübungspreisen 50, 90, 110 und 150, wie in Abbildung 5.2 illustriert, analysiert. Im Gegensatz zum ersten Beispiel werden nun die Volatilität auf  $\sigma = 0.5$  sowie die Zeitschritte auf  $m = 48$  erhöht. Alle anderen Werte bleiben gleich. Ebenfalls ist in Abbildung 5.4

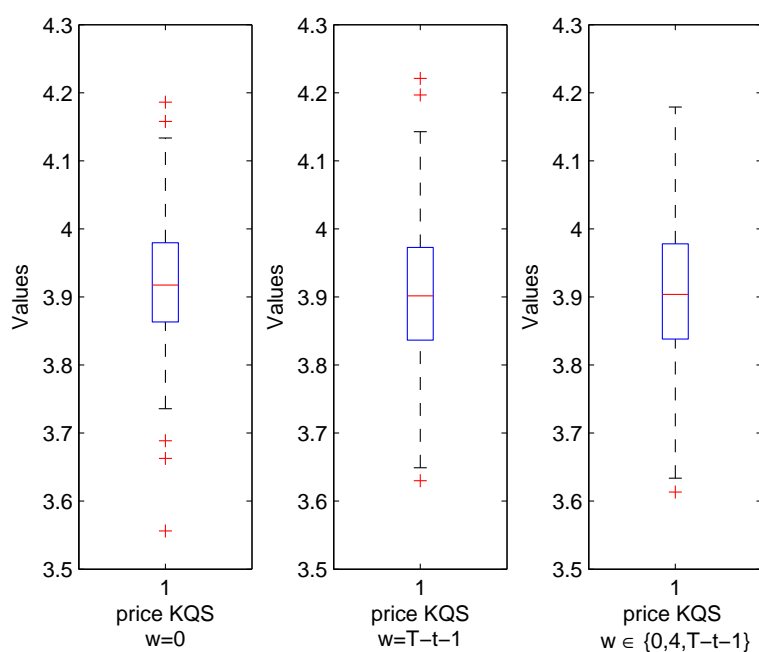


Abbildung 5.3: Boxplots der Optionspreise (price KQS) für die Put-Auszahlung, die unter Verwendung des Kleinste-Quadrate-Splineschätzers (KQS) mit fest gewählten Werten  $w = 0$  (links) und  $w = T - t - 1$  (Mitte) sowie mit der datenabhängigen Wahl von  $w \in \{0, 4, T - t - 1\}$  (rechts) realisiert wurden. In dem Boxplot beschreibt die mittlere waagerechte Linie die Lage des Medians, die untere Kante des Rechtecks stellt das 25. Perzentil dar während die obere Kante das 75. Perzentil darstellt.

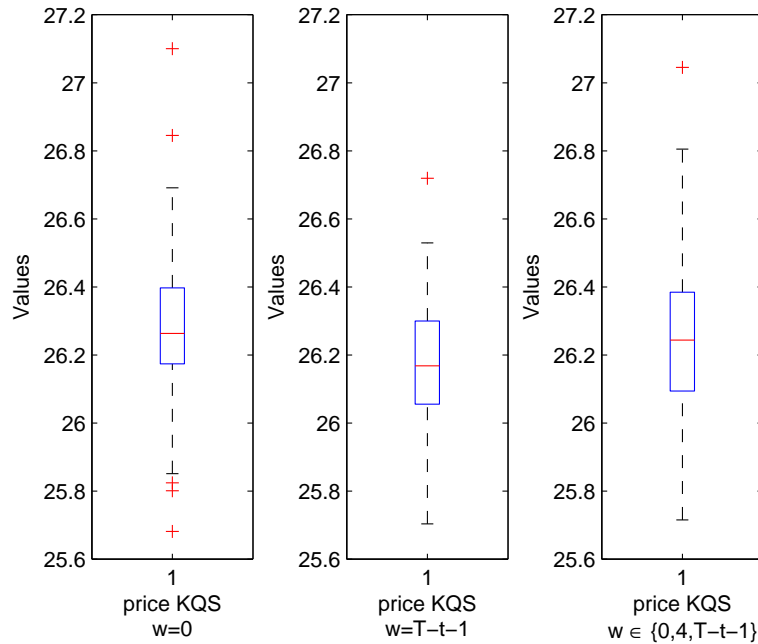


Abbildung 5.4: Optionspreise (*price KQS*) für die *Strangle-Spread-Auszahlung*, die unter Verwendung des *Kleinste-Quadrate-Splineschätzers (KQS)* mit fest gewählten Werten  $w = 0$  (links) und  $w = T - t - 1$  (Mitte) sowie mit der datenabhängigen Wahl von  $w \in \{0, 4, T - t - 1\}$  (rechts) im 1-dimensionalen Fall realisiert wurden.

zu sehen, dass die Wahl von  $w$  die Schätzungen der Werte der *Strangle-Spread-Option* kaum beeinflusst.

In einem dritten (und letzten) Beispiel wird eine *Basket-Option* mit einer *Strangle-Spread-Auszahlung* mit Ausübungspreisen 0.85, 0.95, 1.05 und 1.15 für den Durchschnittswert von drei korrelierten Aktienkursen analysiert. Die Aktien sind ADECCO R, BALOISE R und CIBA. Die Aktienkurse wurden vom 10. November 2000 bis zum 3. Oktober 2003 an Wochentagen, an denen die Börse geöffnet hatte, also an 756 Tagen, beobachtet. Als Volatilität wurde nun die aus obigen historischen Daten geschätzte historische Volatilität

$$\sigma = \begin{pmatrix} 0.3024 & 0.1354 & 0.0722 \\ 0.1354 & 0.2270 & 0.0613 \\ 0.0722 & 0.0613 & 0.0717 \end{pmatrix} \quad (5.1)$$

verwendet. Weiter werden die Kursverläufe der korrelierten Aktien mit

$$x_0 = 1, \quad r = 0.05, \quad n = 4000 \quad \text{und} \quad m = 48$$

simuliert. Dabei ist die Anzahl der Lern-, Test- und Validierungsdaten wie folgt:

$$n_l = 2400 \quad \text{und} \quad n_t = n_v = 800 . \quad (5.2)$$

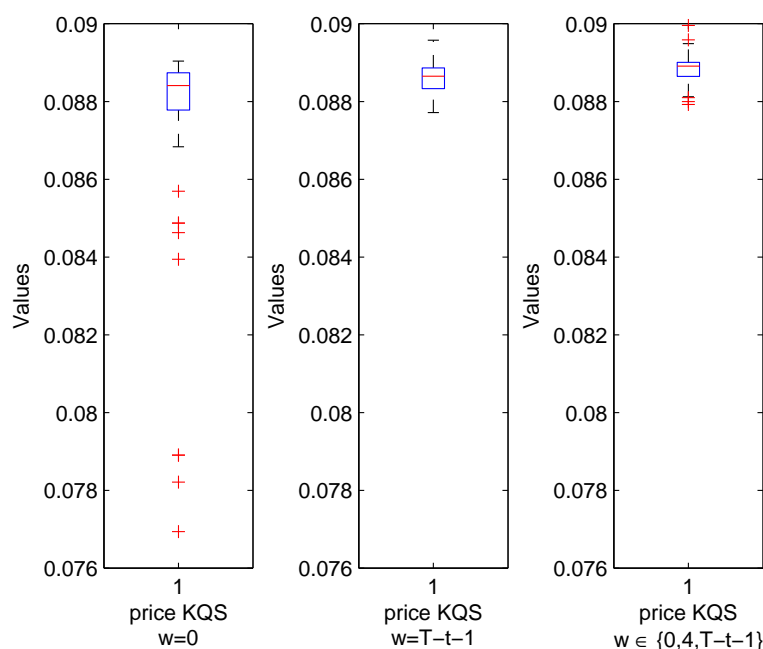


Abbildung 5.5: Optionspreise (*price KQS*) für die Strangle-Spread-Auszahlung, die unter Verwendung des Kleinst-Quadrate-Splineschätzers (*KQS*) mit fest gewählten Werten  $w = 0$  (links) und  $w = T - t - 1$  (Mitte) sowie mit der datenabhängigen Wahl von  $w \in \{0, 4, T - t - 1\}$  (rechts) im 3-dimensionalen Fall realisiert wurden.

Der Knotenabstand  $\alpha$  wird in diesem Beispiel aus der Menge  $\{1, 1.5, 2, 4\}$  gewählt ( $M \in \{0, 1, 2\}$ ). Wie auch in den vorherigen Beispielen beeinflusst hier die Wahl von  $w$  die Schätzungen der Werte der Basket-Option kaum, was in Abbildung 5.5 zu sehen ist.

Zusammenfassend lässt sich festhalten, dass zumindest beim KQS sowohl im Eindimensionalen als auch im Hochdimensionalen - im Zusammenhang mit der Schätzung des Wertes einer Amerikanischen Option - die Wahl von  $w$  keine signifikanten Verbesserungen mit sich bringt. Weiter hat sich herausgestellt, dass sich im Ein- sowie im Hochdimensionalen die Resultate bei den Regressionsdarstellungen von Tsitsiklis und Van Roy (1999) sowie Longstaff und Schwartz (2001) nicht unterscheiden, sofern die Regressionsschätzung mit nichtparametrischen Methoden durchgeführt wird.

### Vergleich der Algorithmen

Im Folgenden werden die Algorithmen von Tsitsiklis und Van Roy (1999), Longstaff und Schwartz (2001) und der Algorithmus aus dieser Arbeit mittels Boxplots verglichen. Wie oben bereits erwähnt, werden die Algorithmen auf 100 unabhängig generierte Mengen von Pfaden (Aktienkursverläufe) angewandt. Der Einfachheit halber wird im Weiteren der Algorithmus von Tsitsiklis und Van Roy (1999) als TR-Algorithmus und der Algorithmus von Longstaff und Schwartz (2001) als LS-Algorithmus bezeichnet.



In einem ersten Beispiel wird obige Put-Auszahlung mit Ausübungspreis 90 (siehe Abbildung 5.1) analysiert. Die Kursverläufe der zugrunde liegenden Aktie werden mit

$$x_0 = 100, r = 0.05, \sigma = 0.25 \quad \text{und} \quad m = 12$$

simuliert. Der Stichprobenumfang  $n$  wird bei dem TR-Algorithmus sowie beim LS-Algorithmus auf  $n = 10000$  gesetzt. Des Weiteren werden für diese beiden Algorithmen Polynome vom Grad 3 gewählt, da mit dieser Wahl die besten Resultate erzielt wurden. Für den in dieser Arbeit vorgestellten Algorithmus werden für die jeweiligen Schätzer verschiedene Werte von  $n$  angenommen. Für den Kleinste-Quadrate-Splineschätzer (KQS) ist  $n = 10000$ , wobei die Anzahl der Lern-, Test- sowie Validierungsdaten  $n_l = 6000$  und  $n_t = n_v = 2000$  beträgt. Die Parameter  $M$  und  $\alpha$  werden aus den Mengen  $\{0, 1, 2\}$  und  $\{\frac{100}{2}, \frac{100}{2^2}, \frac{100}{2^3}, \frac{100}{2^4}\}$  durch Unterteilung der Stichprobe (siehe Abschnitt 4.1) gewählt. Weiter wird  $w$ , wie in Abschnitt 4.2 beschrieben, aus der Menge  $\{0, 4, T - t - 1\}$  datenabhängig gewählt. Für den Neuronale-Netze-Schätzer (NNS) ist  $n = 2000$ . Wie sich zeigen wird, reicht dieser geringe Stichprobenumfang schon aus, um gute Resultate zu erzielen. Da beim NNS  $w = 0$  ist (siehe oben), wird hier die Stichprobe nur in Lern- und Testdaten unterteilt. Dabei beträgt die Anzahl der Lern- sowie Testdaten  $n_l = n_t = 1000$ . Die  $K$  versteckten Neuronen werden aus der Menge  $\{2^0, 2^1, \dots, 2^5\}$  ebenfalls mittels Unterteilung der Stichprobe gewählt. Wie in Abbildung 5.6 zu sehen ist, ist der in dieser Arbeit vorgestellte Algorithmus geringfügig besser als der LS-Algorithmus und vergleichbar mit dem TR-Algorithmus. Dies erkennt man, indem man die Lage der Boxplots der jeweiligen Monte-Carlo-Schätzer betrachtet. Dieses Resultat ist nicht weiter erstaunlich, denn es ist wohlbekannt, dass für einfache Auszahlungsfunktionen der TR- sowie LS-Algorithmus gute Ergebnisse liefert.

In einem zweiten Beispiel wird das Bewertungsproblem schwieriger. Anstatt der einfachen Put-Auszahlung wird nun die kompliziertere Strangle-Spread-Auszahlung mit Ausübungspreisen 50, 90, 110 und 150, wie in Abbildung 5.2 gegeben, analysiert. Weiter wird im Unterschied zum ersten Beispiel eine hohe Volatilität von  $\sigma = 0.5$  gewählt, und die Zeitschritte werden auf  $m = 48$  erhöht. Des Weiteren werden für den TR-Algorithmus Polynome vom Grad 1 gewählt, da nun mit dieser Wahl die besten Resultate erzielt wurden. Beim LS-Algorithmus werden weiterhin Polynome vom Grad 3 verwendet. Wie nun in Abbildung 5.7 zu erkennen ist, sind die Monte-Carlo-Schätzungen des in dieser Arbeit vorgestellten Algorithmus höher als die der LS- und TR-Algorithmen. Weiter ist in Abbildung 5.7 erkennbar, dass der KQS gleichwertig mit dem NNS ist.

In einem dritten Beispiel wird das Bewertungsproblem noch komplizierter, da der hochdimensionale Fall betrachtet wird. Genauer: Es wird die Basket-Option mit einer Strangle-Spread-Auszahlung mit Ausübungspreisen 0.85, 0.95, 1.05 und 1.15 für den Durchschnittswert von drei korrelierten Aktienkursen verwendet. Die Aktien sind ADECCO R, BALOISE R und CIBA (siehe oben). Aus diesen historischen Daten wurde wie oben die historische Volatilität (5.1) verwendet. Die Kursverläufe der korrelierten Aktien werden mit

$$x_0 = 1, r = 0.05 \quad \text{und} \quad m = 48$$

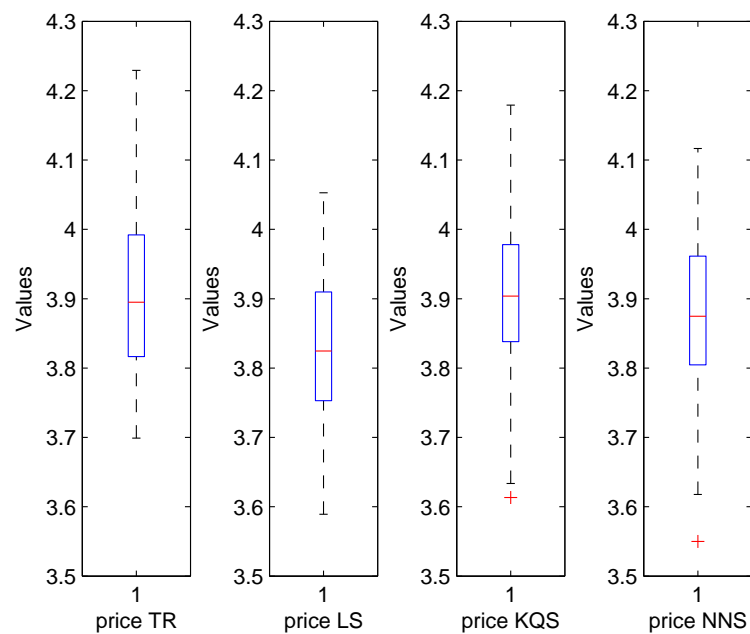


Abbildung 5.6: Optionspreise für die Put-Auszahlung, die unter Verwendung des TR-Algorithmus (*price TR*), des LS-Algorithmus (*price LS*) und des in dieser Arbeit vorgestellten Algorithmus mit dem Kleinsten-Quadrate-Splineschätzer (*price KQS*) und dem Neuronale-Netze-Schätzer (*price NNS*) realisiert wurden.

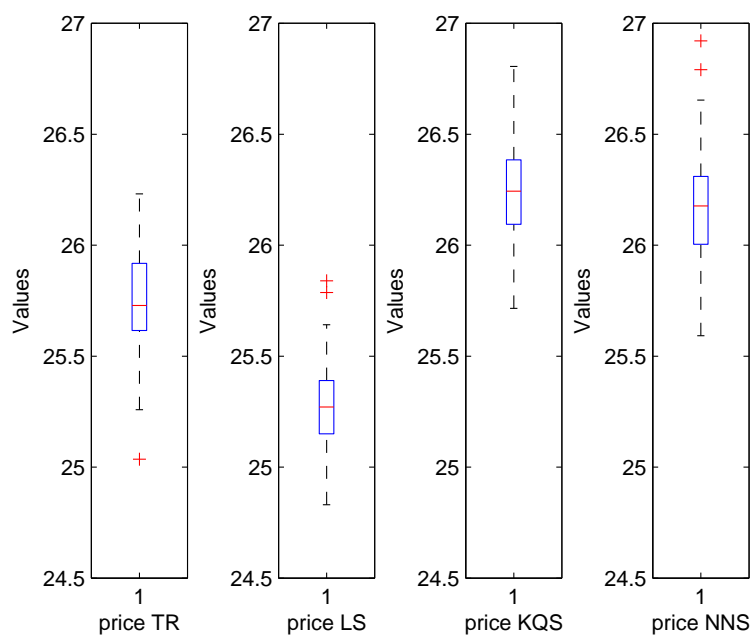


Abbildung 5.7: Optionspreise für die Strangle-Spread-Auszahlung, die unter Verwendung des TR-Algorithmus (price TR), des LS-Algorithmus (price LS) und des in dieser Arbeit vorgestellten Algorithmus mit dem Kleinsten-Quadrate-Splineschätzer (price KQS) und dem Neuronale-Netze-Schätzer (price NNS) im 1-dimensionalen Fall realisiert wurden.

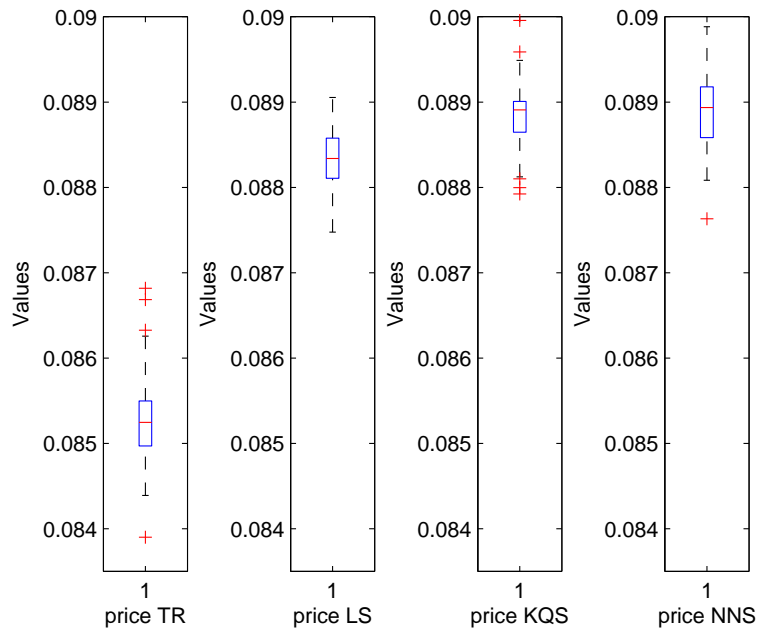


Abbildung 5.8: Optionspreise für die Strangle-Spread-Auszahlung, die unter Verwendung des TR-Algorithmus (*price TR*), des LS-Algorithmus (*price LS*) und des in dieser Arbeit vorgestellten Algorithmus mit dem Kleinsten-Quadrate-Splineschätzer (*price KQS*) und dem Neuronale-Netze-Schätzer (*price NNS*) im 3-dimensionalen Fall realisiert wurden.

simuliert. Der Stichprobenumfang  $n$  ist weiterhin beim TR- sowie beim LS-Algorithmus gegeben durch  $n = 10000$ . Im Gegensatz zu den vorherigen Beispielen werden nun für beide Algorithmen Polynome vom Grad 2 gewählt, da sich dadurch die höheren Monte-Carlo-Schätzungen ergeben haben. Für den in dieser Arbeit vorgestellten Algorithmus wird für den KQS  $n = 4000$  gewählt, und die Anzahl der Lern-, Test- sowie Validierungsdaten beträgt  $n_l = 2400$  und  $n_t = n_v = 800$ . Dieses Mal basiert der KQS auf  $M \in \{0, 1, 2\}$  sowie auf  $\alpha \in \{1, 1.5, 2, 4\}$  ( $w \in \{0, 4, T - t - 1\}$ ). Für den NNS bleiben alle Größen gleich, d. h.  $n = 2000$ ,  $n_l = n_t = 1000$ ,  $K \in \{2^0, 2^1, \dots, 2^5\}$  und  $w = 0$ . Wieder liefert der in dieser Arbeit vorgestellte Algorithmus die höchsten Monte-Carlo-Schätzungen, wie in Abbildung 5.8 zu sehen ist.

Abschließend wird in einem vierten Beispiel noch einmal der hochdimensionale Fall betrachtet. Dieses Mal wird für das Bewertungsproblem die Basket-Option mit einer Strangle-Spread-Auszahlung mit Ausübungspreisen 75, 90, 110 und 125 (anstatt 0.85, 0.95, 1.05 und 1.15) für den Durchschnittswert von fünf (anstatt drei) korrelierten Aktienkursen analysiert. Die Aktien sind ADECCO R, BALOISE R, CIBA, CLARIANT und CREDIT SUISSE R. Die Aktien wurden wie im vorherigen Beispiel vom 10. November 2000 bis zum 3. Oktober 2003 an den 756 Wochentagen, an denen die Börse während dieses Zeitraumes geöffnet hatte, beobachtet. Die Volatilität wurde aus den historischen

Daten geschätzt durch die historische Volatilität

$$\sigma = \begin{pmatrix} 0.3024 & 0.1354 & 0.0722 & 0.1367 & 0.1641 \\ 0.1354 & 0.2270 & 0.0613 & 0.1264 & 0.1610 \\ 0.0722 & 0.0613 & 0.0717 & 0.0884 & 0.0699 \\ 0.1367 & 0.1264 & 0.0884 & 0.2937 & 0.1394 \\ 0.1641 & 0.1610 & 0.0699 & 0.1394 & 0.2535 \end{pmatrix}.$$

Für die Simulationen der Kursverläufe wurde

$$x_0 = 100, \quad r = 0.05 \quad \text{und} \quad m = 48$$

gesetzt. Bei dem TR- sowie LS-Algorithmus gilt weiterhin  $n = 10000$ . Aber dieses Mal werden Polynome vom Grad 1 gewählt, da diese Wahl zu den besten Resultaten führte. Wie oben bereits erwähnt, reichte der interne Speicher in Matlab<sup>®</sup> beim KQS (aufgrund der verwendeten Tensorprodukträume) für dieses Beispiel nicht aus. Demzufolge wurden die beiden Algorithmen bzw. deren Monte-Carlo-Schätzer mit dem NNS, und somit mit dem Algorithmus aus dieser Arbeit, verglichen. Beim NNS blieben wieder alle Größen gleich (also  $n = 2000$ ,  $n_l = n_t = 1000$ ,  $K \in \{2^0, 2^1, \dots, 2^5\}$  und  $w \equiv 0$ ). Wie in Abbildung 5.9 zu sehen, ist der in dieser Arbeit vorgestellte Algorithmus wieder besser als die TR- und LS-Algorithmen, da der Boxplot der Optionspreise, die durch den NNS ermittelt wurden, am höchsten liegt.

Zusammenfassend lässt sich festhalten, dass die aus dieser Arbeit resultierenden Verfahren bei handelsüblichen Put-Optionen (basierend auf Black-Scholes-Modellen) in Simulationen keine schlechteren Resultate liefern als die Verfahren in Tsitsiklis und Van Roy (1999) und Longstaff und Schwartz (2001). Anders verhält es sich bei einer etwas komplizierteren Auszahlungsfunktion wie beispielsweise bei einer Strangle-Spread-Option. Hier liefern die Verfahren aus dieser Arbeit in Simulationen bessere Resultate als die Verfahren in Tsitsiklis und Van Roy (1999) und Longstaff und Schwartz (2001). Das gleiche Verhalten zeigt sich auch bei einer Basket-Option (basierend auf Black-Scholes-Modellen), bei der ebenfalls die Verfahren aus dieser Arbeit in Simulationen bessere Resultate liefern als die Verfahren in Tsitsiklis und Van Roy (1999) und Longstaff und Schwartz (2001).

Schlussfolgernd ist festzuhalten, dass der Einsatz von Regressionsschätzung mit nichtparametrischen Methoden bei der Bewertung Amerikanischer Optionen deutliche Vorteile bringt.

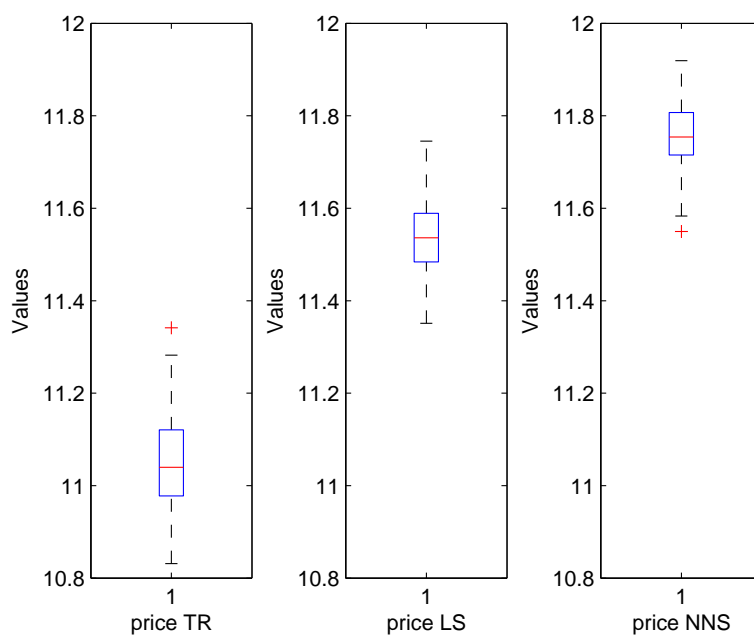


Abbildung 5.9: Optionspreise für die Strangle-Spread-Auszahlung, die unter Verwendung des TR-Algorithmus (price TR), des LS-Algorithmus (price LS) und des in dieser Arbeit vorgestellten Algorithmus mit dem Neuronale-Netze-Schätzer (price NNS) im 5-dimensionalen Fall realisiert wurden.

# Anhang A

## Hilfsresultate

In diesem Anhang werden Definitionen und Lemmata aufgeführt, die Anwendungen in den Beweisen der Resultate dieser Arbeit finden.

### Eine Standardabschätzung für Kleinste-Quadrate-Schätzer

**Lemma A.1** (*Lemma 2 in Kohler (2006a)*)

Sei  $x_1, \dots, x_n \in \mathbb{R}^d$ , sei  $m : \mathbb{R}^d \rightarrow \mathbb{R}$  eine fest gewählte Funktion, und seien  $W_1, \dots, W_n$  unabhängige reelle Zufallsvariablen mit Erwartungswert Null. Sei

$$Y_i = m(x_i) + W_i \quad (i = 1, \dots, n),$$

und es gelte

$$\max_{i=1, \dots, n} V^2 \mathbf{E}\{e^{(Y_i - m(x_i))^2/V^2} - 1\} \leq \sigma_0^2 \quad (\text{A.1})$$

für ein  $V > 0$  sowie für ein  $\sigma_0 > 0$ . Es sei  $\mathcal{F}_n$  eine Klasse von Funktionen  $f : \mathbb{R}^d \rightarrow \mathbb{R}$ , und die Schätzfunktion  $\bar{m}_n$  sei definiert durch

$$\bar{m}_n(\cdot) = \arg \min_{f \in \mathcal{F}_n} \frac{1}{n} \sum_{i=1}^n |f(x_i) - \bar{Y}_i|^2,$$

wobei  $\bar{Y}_1, \dots, \bar{Y}_n$  beliebige Zufallsvariablen sind mit der Eigenschaft, dass

$$\frac{1}{n} \sum_{i=1}^n |Y_i - \bar{Y}_i|^2$$

klein ist. Es existieren Konstanten  $c_1, c_2 > 0$ , die nur von  $\sigma_0$  und  $V$  abhängen, so dass für jedes  $\delta_n > 0$  mit

$$\delta_n \rightarrow 0 \quad (n \rightarrow \infty) \quad \text{und} \quad n \cdot \delta_n \rightarrow \infty \quad (n \rightarrow \infty)$$

und für alle  $\delta \geq \delta_n$ , alle  $g \in \mathcal{F}_n$ , für die

$$c_1 \cdot \sqrt{n} \cdot \delta \geq \int_{\delta/(2^9 \sigma_0)}^{\sqrt{\delta}} \left( \log \mathcal{N}_2 \left( u, \left\{ f - g : f \in \mathcal{F}_n, \frac{1}{n} \sum_{i=1}^n |f(x_i) - g(x_i)|^2 \leq \delta \right\}, x_1^n \right) \right)^{1/2} du \quad (\text{A.2})$$

für alle  $x_1^n \in \mathbb{R}^{d \cdot n}$  gilt, die Beziehung

$$\begin{aligned} & \mathbf{P} \left\{ \frac{1}{n} \sum_{i=1}^n |\bar{m}_n(x_i) - m(x_i)|^2 \right. \\ & \quad \left. > c_2 \left( \frac{1}{n} \sum_{i=1}^n |Y_i - \bar{Y}_i|^2 + \delta_n + \min_{f \in \mathcal{F}_n} \frac{1}{n} \sum_{i=1}^n |f(x_i) - m(x_i)|^2 \right) \right\} \\ & \rightarrow 0 \quad (n \rightarrow \infty) \end{aligned}$$

erfüllt wird.

### Lemmata und Definitionen hinsichtlich Überdeckungszahlen

**Definition A.2** Seien  $z_1, \dots, z_n \in \mathbb{R}^d$ , und setze  $z_1^n = (z_1, \dots, z_n)$ . Definiere den Abstand  $d_2(f, g)$  zwischen  $f, g : \mathbb{R}^d \rightarrow \mathbb{R}$  durch

$$d_2(f, g) = \left( \frac{1}{n} \sum_{i=1}^n |f(z_i) - g(z_i)|^2 \right)^{1/2}.$$

Eine  $\epsilon$ -Füllung von  $\mathcal{G}$  (hinsichtlich des Abstandes  $d_2$ ) ist eine Menge von Funktionen  $g_1, \dots, g_\kappa : \mathbb{R}^d \rightarrow \mathbb{R}$  mit der Eigenschaft

$$d_2(g_i, g_j) \geq \epsilon \quad \text{für alle } i, j \in \{1, \dots, \kappa\}, i \neq j.$$

Sei  $\mathcal{M}_2(\epsilon, \mathcal{G}, z_1^n)$  die Größe  $\kappa$  der größten  $\epsilon$ -Füllung von  $\mathcal{G}$  hinsichtlich des Abstandes  $d_2$ , und setze  $\mathcal{M}_2(\epsilon, \mathcal{G}, z_1^n) = \infty$ , falls solche Funktionen für jedes  $\kappa \in \mathbb{N}$  existieren.  $\mathcal{M}_2(\epsilon, \mathcal{G}, z_1^n)$  ist die so genannte  $L_2$ - $\epsilon$ -Füllzahl von  $\mathcal{G}$  auf  $z_1^n$ .

**Lemma A.3** (Spezialfall von Lemma 9.2 in Györfi et al. (2002))

Sei  $\mathcal{G}$  eine Klasse von Funktionen auf  $\mathbb{R}^d$ , und sei  $\epsilon > 0$ . Des Weiteren seien  $z_1, \dots, z_n \in \mathbb{R}^d$  und  $z_1^n := (z_1, \dots, z_n)$ . Dann gilt:

$$\mathcal{M}_2(2\epsilon, \mathcal{G}, z_1^n) \leq \mathcal{N}_2(\epsilon, \mathcal{G}, z_1^n) \leq \mathcal{M}_2(\epsilon, \mathcal{G}, z_1^n)$$

für alle  $z_1, \dots, z_n \in \mathbb{R}^d$ .

**Lemma A.4** (Lemma 9.3 in Györfi et al. (2002))

Sei  $\mathcal{F}$  ein linearer Vektorraum der Dimension  $D$  bestehend aus Funktionen  $f : \mathbb{R}^d \rightarrow \mathbb{R}$ . Dann gilt für beliebiges  $R > 0$ ,  $\epsilon > 0$  und  $z_1, \dots, z_n \in \mathbb{R}^d$ :

$$\mathcal{N}_2 \left( \epsilon, \left\{ f \in \mathcal{F} : \frac{1}{n} \sum_{i=1}^n |f(z_i)|^2 \leq R^2 \right\}, z_1^n \right) \leq \left( \frac{4R + \epsilon}{\epsilon} \right)^D.$$

**Definition A.5** Sei  $\mathcal{A}$  ein System von Teilmengen von  $\mathbb{R}^d$  mit  $\mathcal{A} \neq \emptyset$ . Des Weiteren sei  $F$  eine Teilmenge von  $\mathbb{R}^d$ . Man sagt,  $\mathcal{A}$  zerlegt  $F$ , falls jede Teilmenge von  $F$  sich in der Form  $A \cap F$  für ein  $A \in \mathcal{A}$  darstellen lässt.

Die Vapnik-Chervonenkis-Dimension (kurz: VC-Dimension)  $V_{\mathcal{A}}$  von  $\mathcal{A}$  wird definiert als die größte Zahl  $n \in \mathbb{N}$ , für die eine Menge bestehend aus  $n$  Punkten in  $\mathbb{R}^d$  existiert, die von  $\mathcal{A}$  zerlegt wird. Setze  $V_{\mathcal{A}} = \infty$ , falls  $\mathcal{A}$  Mengen beliebiger Kardinalität zerlegt.



Als nächstes wird die  $L_2$ - $\epsilon$ -Füllzahl nach oben abgeschätzt. Dabei verwendet das folgende Lemma die VC-Dimension  $V_{\mathcal{G}^+}$  der Menge

$$\mathcal{G}^+ := \{ \{ (z, t) \in \mathbb{R}^d \times \mathbb{R} : t \leq g(z) \} : g \in \mathcal{G} \}$$

aller Graphen von Funktionen aus  $\mathcal{G}$ , wobei  $\mathcal{G}$  eine Menge von Funktionen  $g : \mathbb{R}^d \rightarrow [0, B]$  ist.

**Lemma A.6** (Theorem 9.4 in Györfi et al. (2002))

Seien  $B > 0$ ,  $\mathcal{G}$  eine Menge von Funktionen  $g : \mathbb{R}^d \rightarrow [0, B]$  mit  $V_{\mathcal{G}^+} \geq 2$ ,  $z_1, \dots, z_n \in \mathbb{R}^d$  und  $0 < \epsilon < \frac{B}{4}$ . Dann gilt:

$$\mathcal{M}_2(\epsilon, \mathcal{G}, z_1^n) \leq 3 \left( \frac{2eB^2}{\epsilon^2} \log \frac{3eB^2}{\epsilon^2} \right)^{V_{\mathcal{G}^+}}.$$

### Lemmata hinsichtlich Ungleichungen

**Lemma A.7** (Lemma 1 in Kohler (2006a))

Seien  $t > 0$ ,  $x_1, \dots, x_n \in \mathbb{R}^d$  und  $y_1, \bar{y}_1, \dots, y_n, \bar{y}_n \in \mathbb{R}$ . Des Weiteren sei  $m$  eine Funktion mit  $m : \mathbb{R}^d \rightarrow \mathbb{R}$  und sei  $\mathcal{F}$  eine Menge von Funktionen  $f : \mathbb{R}^d \rightarrow \mathbb{R}$ . Setze

$$\bar{m}_n = \arg \min_{f \in \mathcal{F}} \frac{1}{n} \sum_{i=1}^n |f(x_i) - \bar{y}_i|^2$$

und

$$m_n^* = \arg \min_{f \in \mathcal{F}} \frac{1}{n} \sum_{i=1}^n |f(x_i) - m(x_i)|^2,$$

und nehme an, dass beide Minima existieren. Dann impliziert

$$\begin{aligned} \frac{1}{n} \sum_{i=1}^n |\bar{m}_n(x_i) - m(x_i)|^2 &> t + 512 \frac{1}{n} \sum_{i=1}^n |y_i - \bar{y}_i|^2 \\ &\quad + 18 \min_{f \in \mathcal{F}} \frac{1}{n} \sum_{i=1}^n |f(x_i) - m(x_i)|^2 \end{aligned}$$

die Ungleichungen

$$\frac{t}{2} < \frac{1}{n} \sum_{i=1}^n |\bar{m}_n(x_i) - m_n^*(x_i)|^2 \leq \frac{16}{n} \sum_{i=1}^n (\bar{m}_n(x_i) - m_n^*(x_i)) \cdot (y_i - m(x_i)).$$

**Lemma A.8** (Lemma 5 in Kohler (2006a))

Sei  $B \geq 1$ , sei  $\mathcal{F}$  eine Menge von Funktionen  $f : \mathbb{R}^d \rightarrow [-B, B]$  und sei  $m : \mathbb{R}^d \rightarrow [-B, B]$  eine fest gewählte Funktion. Des Weiteren sei  $0 < \epsilon < 1$  sowie  $\alpha > 0$ . Es gelte

$$\sqrt{n\epsilon}\sqrt{\alpha} \geq 1152B, \tag{A.3}$$

und für alle  $x_1, \dots, x_n \in \mathbb{R}^d$  und alle  $\delta \geq 2B^2\alpha$  sei

$$\frac{\sqrt{n}\epsilon\delta}{768\sqrt{2}B^2} \geq \int_{\frac{\epsilon\delta}{128B^2}}^{\sqrt{\delta}} \left( \log \mathcal{N}_2 \left( \frac{u}{4B}, \left\{ f - m : f \in \mathcal{F} \right. \right. \right. \\ \left. \left. \left. \frac{1}{n} \sum_{i=1}^n |f(x_i) - m(x_i)|^2 \leq \frac{\delta}{B^2} \right\}, x_1^n \right) \right)^{1/2} du \quad (\text{A.4})$$

erfüllt. Dann gilt

$$\mathbf{P} \left\{ \sup_{f \in \mathcal{F}} \frac{|\mathbf{E}\{|f(X) - m(X)|^2\} - \frac{1}{n} \sum_{i=1}^n |f(X_i) - m(X_i)|^2|}{\alpha + \mathbf{E}\{|f(X) - m(X)|^2\} + \frac{1}{n} \sum_{i=1}^n |f(X_i) - m(X_i)|^2} > \epsilon \right\} \\ \leq 15 \exp \left( -\frac{n\alpha\epsilon^2}{512 \cdot 2304B^2} \right).$$

**Lemma A.9** (Lemma 16.8 in Györfi et al. (2002))

Sei  $\sigma$  eine Squashing-Funktion, und sei  $\mathcal{F}_\Gamma$  die Klasse von Funktionen, für die (2.16) auf  $\mathbb{R}^d$  und (2.17) (siehe Seite 49) gilt. Für jedes Wahrscheinlichkeitsmaß  $\mu$  auf  $\mathbb{R}^d$ , jede messbare Funktion  $f \in \mathcal{F}_\Gamma$  und jedes  $K \geq 1$  existiert dann ein neuronales Netz  $f_K$  in

$$\mathcal{F}_K = \left\{ \sum_{i=1}^K d_i \sigma(a_i^T x + b_i) + d_0 : K \in \mathbb{N}, a_i \in \mathbb{R}^d, b_i, d_i \in \mathbb{R} \right\},$$

so dass gilt:

$$\int_{S_r} (f(x) - f_K(x))^2 \mu(dx) \leq \frac{(2r\Gamma)^2}{K},$$

wobei  $S_r = \{y \in \mathbb{R}^d : \|y - 0\| \leq r\}$  mit Radius  $r > 0$ . Die Koeffizienten der Linearkombination in der Menge  $\mathcal{F}_K$  können so gewählt werden, dass  $\sum_{i=0}^K |d_i| \leq 3r\Gamma + f(0)$ .

**Lemma A.10** (Ungleichung von Hoeffding, Hoeffding (1963))

Seien  $X_1, \dots, X_n$  unabhängige reelle Zufallsvariablen mit  $X_i \in [a_i, b_i]$  fast sicher ( $i = 1, \dots, n$ ), wobei  $a_1, b_1, \dots, a_n, b_n \in \mathbb{R}$ . Dann gilt für beliebiges  $\epsilon > 0$

$$\mathbf{P} \left\{ \left| \frac{1}{n} \sum_{i=1}^n (X_i - \mathbf{E}\{X_i\}) \right| > \epsilon \right\} \leq 2 \exp \left( -\frac{2n\epsilon^2}{\frac{1}{n} \sum_{i=1}^n |b_i - a_i|^2} \right).$$

**Lemma A.11** (Ungleichung von Bernstein, Bernstein (1946))

Seien  $X_1, \dots, X_n$  unabhängige reelle Zufallsvariablen mit  $X_i \in [a, b]$  fast sicher ( $i = 1, \dots, n$ ), wobei  $a, b \in \mathbb{R}$  und  $a < b$ . Sei

$$\sigma^2 = \frac{1}{n} \sum_{i=1}^n \mathbf{Var}(X_i) > 0.$$

Dann gilt für beliebiges  $\epsilon > 0$

$$\mathbf{P} \left\{ \left| \frac{1}{n} \sum_{i=1}^n (X_i - \mathbf{E}\{X_i\}) \right| > \epsilon \right\} \leq \exp \left( -\frac{n\epsilon^2}{2\sigma^2 + 2\epsilon \cdot \frac{b-a}{3}} \right).$$

**Lemma A.12** (Theorem 9.5 in Györfi et al. (2002))

Sei  $\mathcal{G}$  ein  $D$ -dimensionaler linearer Vektorraum von Funktionen auf  $\mathbb{R}^d$ . Dann gilt für die VC-Dimension von

$$\mathcal{A} = \left\{ \{z : g(z) \geq 0\} : g \in \mathcal{G} \right\}$$

die Abschätzung

$$V_{\mathcal{A}} \leq D .$$

**Lemma A.13** (Lemma 16.3 in Györfi et al. (2002))

Sei  $\mathcal{F}$  ein Familie von reellen Funktionen auf  $\mathbb{R}^m$ . Weiter sei  $g : \mathbb{R} \rightarrow \mathbb{R}$  eine feste nichtfallende Funktion. Sei die Klasse  $\mathcal{G}$  definiert durch  $\mathcal{G} = \{g \circ f : f \in \mathcal{F}\}$ . Dann gilt für die VC-Dimension

$$V_{\mathcal{G}^+} \leq V_{\mathcal{F}^+} .$$

### Lemmata hinsichtlich Approximationsresultaten

**Lemma A.14** (Theorem A.1 in Györfi et al. (2002))

Für beliebiges  $p \geq 1$  und beliebiges Wahrscheinlichkeitsmaß  $\mu$  ist die Menge der stetigen Funktionen mit beschränktem Support dicht in  $L_p(\mu)$ , d. h. für beliebiges  $f \in L_p(\mu)$  und  $\epsilon > 0$  existiert eine stetige Funktion  $g$  mit kompaktem Support, so dass gilt:

$$\int |f(x) - g(x)|^p \mu(dx) \leq \epsilon .$$

**Lemma A.15** (Lemma 16.2 in Györfi et al. (2002))

Sei  $\sigma$  eine Squashing-Funktion. Für jedes Wahrscheinlichkeitsmaß  $\mu$  auf  $\mathbb{R}^d$ , jede messbare Funktion  $f : \mathbb{R}^d \rightarrow \mathbb{R}$  mit  $\int |f(x)|^2 \mu(dx) < \infty$  und jedes  $\epsilon > 0$  existiert dann ein neuronales Netz

$$h(x) = \sum_{i=1}^K d_i \sigma(a_i^T x + b_i) + d_0 \quad (K \in \mathbb{N}, a_i \in \mathbb{R}^d, b_i, d_i \in \mathbb{R}),$$

so dass gilt:

$$\int |f(x) - h(x)|^2 \mu(dx) < \epsilon .$$

# Literaturverzeichnis

- [1] Andersen, L. und Broadie, M. (2004). “Primal-dual simulation algorithm for pricing multidimensional American options”, *Management Sci.* 50(9), pp. 1222-1234.
- [2] Barraquand, J. und Martineau, D. (1995). “Numerical Valuation of High Dimensional Multivariate American Securities”, *Journal of Financial and Quantitative Analysis* 30, pp. 383-405.
- [3] Bernstein, S. N. (1946). *The Theory of Probabilities*. Gastehizdat Publishing House, Moskau.
- [4] Bingham, N. H. und Kiesel, R. (2004). *Risk-Neutral Valuation*. Springer Finance, Springer Verlag.
- [5] Black, F. und Scholes, M. (1973). “The pricing of options and corporate liabilities”, *Journal of Political Economy* 81, pp. 637-659.
- [6] Boessarts, P. (1989). “Simulation Estimators of Optimal Early Exercise”, Working paper, Carnegie-Mellon University.
- [7] Bollerslev, T. P. (1986). “Generalized autoregressive conditional heteroscedasticity”, *Journal of Econometrics* 31, pp. 307-327.
- [8] Boyle, P. (1977). “Options: A Monte Carlo Approach”, *Journal of Financial Economics* 4, pp. 323-338.
- [9] Broadie, M. und Glasserman, P. (1997). “Pricing American-Style Securities using Simulation”, *Journal of Economic Dynamics and Control* 21, pp. 1323-1352.
- [10] Broadie, M. und Glasserman, P. (1998). “Monte Carlo Methods for Pricing High-dimensional American Options: An Overview”, in: *Monte Carlo Methodologies and Applications for Pricing and Risk Management*, Risk Books, pp. 149-161.
- [11] Broadie, M., Glasserman, P. und Jain, G. (1997). “Enhanced Monte Carlo estimates for American option prices”, *J. Derivatives* 5, pp. 25-44.
- [12] Carriere, J. F. (1996). “Valuation of the early-exercise price for options using simulations and nonparametric regression”, *Insurance: Math. Economics* 19, pp. 19-30.
- [13] Chow, Y. S., Robbins, H. und Siegmund, D. (1971). *Great Expectations: The Theory of Optimal Stopping*. Houghton Mifflin, Boston.

- [14] Clément, E., Lamberton, D. und Protter, P. (2002). “An analysis of the Longstaff-Schwartz algorithm for American option pricing”, *Finance and Stochastics* 6(4), pp. 449-471.
- [15] Cont, R. und Tankov, P. (2004). *Financial Modelling with Jump Processes*. Chapman & Hall.
- [16] Cover, T. (1968). “Rates of convergence for nearest neighbor procedures”, in: *Proceedings of the Hawaii International Conference on System Sciences*, pp. 413-415, Honolulu, HI.
- [17] Cox, J. C. und Ross, S. (1976). “The Valuation of Options for Alternative Stochastic Processes”, *Journal of Financial Economics* 3, pp. 145-166.
- [18] Cox, J. C., Ross, S. und Rubinstein, M. (1979). “Option Pricing: A Simplified Approach”, *Journal of Financial Economics* 7, pp. 229-264.
- [19] de Boor, C. (1978). *A Practical Guide to Splines*. Springer, New York.
- [20] Derman, E. und Kani, I. (1994). “Riding on a Smile”, *RISK* 7, pp. 32-39.
- [21] Deutsch, H. und Eller, R. (1999). *Derivates and Internal Models*. Macmillan Press.
- [22] Devroye, L. (1982). “Necessary and sufficient conditions for the almost everywhere convergence of nearest neighbor regression function estimates”, *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete* 61, pp. 467-481.
- [23] Dippon, J. und Winter, S. (2006). “Smoothing spline regressions estimates for randomly right censored data”, Preprint, Universität Stuttgart.
- [24] Duan, J. C. (1995). “The GARCH Option Pricing Model”, *Mathematical Finance* 5(1), pp. 13-32.
- [25] Duan, J. C., Gauthier, G., Sasservielle, C. und Simonate, J. G. (2003). “Approximating American Option Prices in teh GARCH Framework”, *Journal of Futures Markets* 23(10), pp. 915-929.
- [26] Duan, J. C., Ritchken, P. und Sun, Z. (2005). *Approximating GARCH-Jump Models, Jump-Diffusion Processes, and Option Pricing*. Forthcoming, Mathematical Finance.
- [27] Duffie, D. (2001). *Dynamic Asset Pricing Theory*. (Third ed.). Princeton, New Jersey: Princeton University Press.
- [28] Dupire, B. (1994). “Pricing with a smile”, *RISK* 7, pp. 18-20.
- [29] Egloff, D. (2005). “Monte Carlo Algorithms for Optimal Stopping and Statistical Learning”, *Annals of Applied Probability* 15, pp. 1-37.
- [30] Egloff, D., Kohler, M. und Todorovic, N. (2006). “A dynamic look-ahead Monte Carlo algorithm for pricing American options”, erscheint in: *Annals of Applied Probability*.

- [31] El Karoui, N. (1981). *Les aspects probabilistes du contrôle stochastique*. Springer.
- [32] Elliott, R. J. und Kopp, P. E. (1999). *Mathematics of Financial Markets*. Springer.
- [33] Engle, R. (1982). "Autoregressive conditional heteroscedasticity with estimates of the variance of U.K. inflation", *Econometrica* 50, pp. 345-360.
- [34] Föllmer, H. und Schied, A. (2002). *Stochastic Finance*. De Gruyter.
- [35] Franke, J., Härdle, W. und Hafner, C. (2004). *Einführung in die Statistik der Finanzmärkte*. Springer.
- [36] Glasserman, P. (2004). *Monte Carlo methods in financial engineering*. Springer.
- [37] Györfi, L., Kohler, M., Krzyżak, A. und Walk, H. (2002). *A Distribution-Free Theory of Nonparametric Regression*. Springer Series in Statistics, Springer.
- [38] Haugh, M. und Kogan, L. (2004). "Pricing American Options: A Duality Approach", *Operations Research* 52(2), pp. 258-270.
- [39] Heston, S. (1993). "A closed-form solution for options with stochastic volatility with applications to bond and currency options", *Rev. Fin. Studies* 6, pp. 327-343.
- [40] Hoeffding, W. (1963). "Probability inequalities for sums of bounded random variables", *Journal of the American Statistical Association* 58, pp. 13-30.
- [41] Hull, J. (2006). *Optionen, Futures und andere Derivate*. Pearson Studium.
- [42] Hull, J. und White, A. (1987). "The pricing of options on assets with stochastic volatilities", *Journal of Finance* 42, pp. 281-300.
- [43] Jondeau, E., Poon, S.-H. und Rockinger, M. (2007). *Financial Modeling Under Non-Gaussian Distributions*. Springer.
- [44] Karatzas, I. und Shreve, S. E. (1998). *Methods of Mathematical Finance*. Springer.
- [45] Kohler, M. (2000). *Analyse von nichtparametrischen Regressionsschätzern unter minimalen Voraussetzungen*. Habilitationsschrift, Shaker Verlag, Aachen.
- [46] Kohler, M. (2006a). "Nonparametric regression with additional measurement errors in the dependent variable", *Journal of Statistical Planning and Inference* 136, pp. 3339-3361.
- [47] Kohler, M. (2006b). "A regression based smoothing spline Monte Carlo algorithm for pricing American options", Preprint, Universität des Saarlandes.
- [48] Kohler, M., Krzyżak, A. und Todorovic, N. (2006). "Pricing of high-dimensional American options by neural networks", Preprint, Universität des Saarlandes.
- [49] Kohler, M., Kul, S. und Mathé, K. (2004). "Least squares estimates for random design regression", Preprint, Universität Stuttgart.

- [50] Korn, R. und Korn, E. (2001). *Optionsbewertung und Portfolio-Optimierung*. Vieweg Verlag.
- [51] Lamberton, D. und Pagès, G. (1990). “Sur l’approximation des réduites”, *Ann. Inst. Henri Poincarés, Probab. Statist.* 26(2), pp. 331-355.
- [52] Laprise, S. B., Su, Y., Wu, R., Fu, M. C. und Madan, D. B. (2001). “Pricing American options: A comparison of Monte Carlo Simulation Approaches”, *Journal Comput. Finance* 4(3), pp. 39-88.
- [53] Longstaff, F. A. und Schwartz, E. S. (2001). “Valuing American options by simulation: a simple least-squares approach”, *Review of Financial Studies* 14, pp. 113-147.
- [54] Merton, R. C. (1973). “Theory of Rational Option Pricing”, *Bell Journal of Economics and Management Science* 4, pp. 141-183.
- [55] Merton, R. C. (1976). “Option Pricing When Underlying Stock Returns Are Discontinuous”, *Journal of Financial Economics* 3, pp. 125-144.
- [56] Nelson, D. B. (1990). “Arch models as diffusion approximations”, *Journal of Econometrics* 45, pp. 7-38.
- [57] Nelson, R. B. (1999). *An Introduction to Copulas*. Springer, New York.
- [58] Neveu, J. (1975). *Discrete-Parameter Martingales*. North-Holland Publ. Comp.
- [59] Rogers, L. C. G. (2002). “Monte Carlo Valuing of American Options”, *Mathematical Finance* 12, pp. 271-286.
- [60] Sandmann, K. (2001). *Einführung in die Stochastik der Finanzmärkte*. Springer.
- [61] Schumaker, L. (1981). *Spline functions: Basic Theory*. Wiley, New York.
- [62] Schweitzer, M. (2002). “On bermudan options”. *Advances in Finance and Stochastics. Essays in Honour of Dieter Sondermann* (K. Sandmann und P. J. Schönbucher, eds.), Springer 2002, pp. 257-269.
- [63] Seydel, R. (2006). *Tools for Computational Finance*. Springer.
- [64] Shiriyayev, A. N. (1978). *Optimal Stopping Rules*. Applications of Mathematics, Springer Verlag.
- [65] Steele, J. M. (2001). *Stochastic Calculus and Financial Applications*. Applications of Mathematics, Springer Verlag.
- [66] Stentoft, L. (2005). “Pricing American Options when the Underlying Asset follows GARCH processes”, *Journal of Empirical Finance* 12(4), pp. 576-611.
- [67] Stone, C. J. (1982). “Optimal global Rates of Convergence for Nonparametric Regression”, *Annals of Statist.* 10(4), pp. 1040-1053.

- [68] Tilley, J. A. (1993). “Valuing American Options in a Path Simulation Model”, Transactions of the Society of Actuaries 45, pp. 83-104.
- [69] Tsitsiklis, J. N. und Van Roy, B. (1999). “Optimal stopping of Markov processes: Hilbert space theory, approximation algorithms, and an application to pricing high-dimensional financial derivatives”, IEEE Transactions On Automatic Control 44, pp. 1840-1851.
- [70] Tsitsiklis, J. N. und Van Roy, B. (2001). “Regression Methods for Pricing Complex American-Style Options”, IEEE Transactions On Neural Networks 12, pp. 694-703.
- [71] Van de Geer, S. (2000). *Empirical Processes in M-estimation*. Cambridge University Press.



## Lebenslauf

|                    |  |
|--------------------|--|
| 23. Juni 1974      | Geboren in Schwäbisch Hall   |
| 1980 bis 1984      | Grundschule in Gailenkirchen bzw. in Gottwollshausen   |
| 1984 bis 1990      | Leonhard-Kern-Realschule in Schwäbisch Hall  |
| 1990 bis 1994      | Ausbildung zum Industriemechaniker,<br>Fachrichtung Maschinen- und Systemtechnik,<br>bei Fa. Breitner in Schwäbisch Hall |
| 1994 bis 1995      | Industriemechaniker bei Fa. Breitner in<br>Schwäbisch Hall   |
| 1995 bis 1997      | Schlosser bei Fa. Klenk in Oberrot   |
| 1997 bis Juli 1999 | Technische Oberschule in Stuttgart   |
| Oktober 1999       | Beginn des Mathematikstudiums an der Universität<br>Stuttgart  |
| März 2005          | Beendigung des Studiums mit dem Abschluss<br>Diplom-Mathematiker   |
| Seit Mai 2005      | Wissenschaftlicher Mitarbeiter an der Universität<br>des Saarlandes  |