

Diplomarbeit

Parameterschätzung
stochastischer Prozesse
aus der Finanzwelt
mittels
(Dünngitter-)Histogramm-
Matching-Verfahren

Angefertigt am
Institut für Numerische Simulation

Vorgelegt der
Mathematisch-Naturwissenschaftlichen Fakultät der
Rheinischen Friedrich-Wilhelms-Universität Bonn

Mai 2006
Von
Vera Gerig
Aus
Köln

Inhaltsverzeichnis

1. Einleitung	1
2. Wertpapiere und ihre Risiken	7
2.1. Aktien	9
2.2. Schuldverschreibungen / Festverzinsliche Wertpapiere	10
3. Stochastische Prozesse	13
3.1. Die Zufallskomponente	14
3.1.1. Ein Random-Walk-Modell für Kursänderungen	16
3.1.2. Einsatz des Wiener-Prozesses	18
3.2. Die deterministische Komponente	19
3.3. Itô-Prozesse	19
3.3.1. Das Kursmodell nach dem Itô-Lemma	21
3.4. Spezielle stochastische Prozesse	23
3.5. Die Übergangsdichte eines stochastischen Prozesses	27
4. Stochastische Modellierung von Risikofaktoren an den Finanzmärkten	29
4.1. Das Kursrisiko bei Aktien	30
4.2. Das Zinsänderungsrisiko festverzinslicher Wertpapiere	32
4.2.1. Die Bestimmung der Zinsstruktur	34
4.3. Short-Rate-Modelle	44
4.3.1. Das Vasicek-Modell	48
4.3.2. Das Cox-Ingersoll-Ross-Modell	52
4.4. Mehr-Faktor-Modelle	55
4.4.1. Der (unkorrelierte) Mehr-Faktor-Fall des Vasicek-Modells	56
4.4.2. Der (unkorrelierte) Mehr-Faktor-Fall des CIR-Modells	58
5. Diskretisierung der Modelle	59
5.1. Das Euler-Maruyama-Schema für einen einzelnen stochastischen Prozess . .	60
5.2. Das Euler-Maruyama-Schema für mehrere korrelierte stochastische Prozesse	60
6. Die Simulation der Modelle	65
6.1. Aktienkurse	65
6.2. Short Rate	68
6.3. Ein korreliertes Finanzmodell (Aktienkurs und Short Rate)	68
7. Dichte-Approximationen	69
7.1. Der Tensorprodukt-Ansatz	69
7.1.1. Die Wahl der Basisfunktionen	70
7.1.2. Die Basis der charakteristischen Funktionen	71
7.1.3. Die Basis der stückweise linearen Hutfunktionen	72

7.1.4.	Die Bestimmung der Gewichtskoeffizienten	75
7.2.	Der Dünngitter-Ansatz	79
7.2.1.	Herleitung der Dünnen Gitter über die Differenzen der eindimensio- nodalen Basisfunktionen	79
7.2.2.	Die Kombinationstechnik zur Erstellung von Dünnen Gittern	82
7.2.3.	Die Gewichtskoeffizienten im Dünngitter-Ansatz	82
8.	Die Parameterschätzverfahren	87
8.1.	Die Repräsentativität der historischen Stichprobe	88
8.2.	Parameterspezifische Parameterschätzverfahren	90
8.3.	Allgemeinere Parameterschätzverfahren	92
8.4.	Das allgemeine Histogramm-Matching-Verfahren	93
8.4.1.	Skalierungen für das Histogramm-Matching-Verfahren	98
8.4.2.	Das iterative Histogramm-Matching-Verfahren	102
8.4.3.	Paralleles (Dünngitter-)Histogramm-Matching-Verfahren	103
8.5.	Das (diskretisierte) Maximum-Likelihood-Verfahren	104
8.6.	Die Schätzproblematik des Zwei-Faktor-Falls des CIR-Modells	109
8.7.	Schätzung eines korrelierten Finanzmodells	112
9.	Numerische Ergebnisse	113
9.1.	Schätzung des Mean-Reversion-Niveaus	113
9.2.	Skalierungen	114
9.3.	Verfahrensparameter des eindimensionalen Histogramm-Matching-Verfahrens	115
9.3.1.	Der diskretisierte Zeitzuwachs Δt	116
9.3.2.	Der Datensatzumfang N	117
9.3.3.	Die Simulationsanzahl M	118
9.3.4.	Die Intervallgrenzen	119
9.3.5.	Die Anzahl k der Teilintervalle	120
9.3.6.	Die Basisfunktionen	123
9.3.7.	Die Iterationsanzahl und die Startwertwahl des iterativen Histogramm- Matching-Verfahrens basierend auf eindimensionalen Histogrammen	124
9.4.	Verfahrensparameter des zweidimensionalen Histogramm-Matching-Verfah- rens	126
9.4.1.	Die Simulationsanzahl M	126
9.4.2.	Die Anzahl k_q der Teilquader	127
9.4.3.	Die Levelanzahl beim Einsatz des Dünngitter-Ansatzes entsprechend der Kombinationstechnik	129
9.5.	Vergleich der Verfahren	131
9.6.	Schätzung eines korrelierten Finanzmodells anhand realer Daten	135
10.	Ausblick	139
A.	Wichtiges aus der Wahrscheinlichkeitstheorie und Statistik	141
	Literaturverzeichnis	147
	Index	151

Bezeichnungen

t	Laufende Zeitvariable
dt	Infinitesimaler Zeitzuwachs
Δt	Zur Diskretisierung verwendeter endlicher Zeitzuwachs
T	Fälligkeit(-sdatum)
$\tau = T - t$	(Rest-)Laufzeit
$X(t)$	Stetig stochastischer Prozess
$dX(t)$	Änderung des durch $X(t)$ beschriebenen stochastischen Prozesses über einen infinitesimal kleinen Zeitraum dt
$\Delta X(t)$ $= X(t + \Delta t) - X(t)$	Änderung des durch $X(t)$ beschriebenen stochastischen Prozesses über einen endlichen Zeitraum Δt
$W(t)$	Wiener Prozess
$S(t)$	Prozess zur Beschreibung eines Aktienkurses
$r(t)$	Prozess zur Beschreibung einer Short Rate
$r_i(t)$	Prozess zur Beschreibung der Teilrate einer Short Rate in einem Mehr-Faktor-Modell
$m(r, t)$	Driftkoeffizient eines Short-Rate-Modells
$v(r, t)$	Diffusionskoeffizient eines Short-Rate-Modells
μ, σ, a, b, m	Parameter zur Beschreibung von Drift- und Diffusionskoeffizienten
$F(tT, T')$	Diskreter Terminzins
$f(tT, T')$	Stetiger Terminzins (engl. Forward Rate)
$f(t, T)$	Instantaner Terminzins
r	Konstanter Zinssatz
$y(t, T)$	Diskreter Kassazins
$R(t, T)$	Stetiger Kassazins (engl. Spot Rate)
$r(t)$	Instantaner Kassazins (engl. Short Rate)
k_i	Kuponauszahlungen an diskreten Zeitpunkten t_i
$P^k(t, T)$	Preis einer Kupon-Anleihe zur Zeit t mit Fälligkeitsdatum T und Kuponzahlungsstrom $k(t)$
$P(t, T), P(r, tT)$	Preis einer Nullkupon-Anleihe zur Zeit t mit Fälligkeitsdatum T
\mathbb{P}	Reales Wahrscheinlichkeitsmaß
Q	Äquivalentes (risikoneutrales) Martingalmaß
$\lambda(r, t)$	Marktpreis des Risikos
$E[X]$	Erwartungswert der Zufallsvariable X
$\text{var}[X]$	Varianz der Zufallsvariable X
$\text{cov}[X, Y]$	Kovarianz der Zufallsvariablen X und Y
$\rho(X, Y)$	Korrelationskoeffizient der Zufallsvariablen X und Y
$X \sim N(p_1, p_2)$	X ist normalverteilt mit Mittelwert p_1 und Varianz p_2
$X \sim LN(p_1, p_2)$	X ist lognormalverteilt mit den Parametern p_1 und p_2
N	Anzahl der Zeitschritte
$\theta \in \Theta$	Parametersatz und Parameterraum
M	Anzahl simulierter Datensätze pro Parametersatz
$\underline{x} \in \mathbb{R}^d$	Datenpunkt der Dimension d
$x \in \mathbb{R}$	Eindimensionaler Datenpunkt
$\rho(\underline{x})$	Dichte-Approximation der Dimension d
$\rho(x)$	Eindimensionale Dichte-Approximation

$\rho_{\underline{l}}^{PA}(\underline{x})$	Mittels Tensorprodukt-Ansatz approximierte Dichtefunktion
$\rho_{\underline{l}}^{DG}(\underline{x})$	Mittels Dünngitter-Ansatz approximierte Dichtefunktion
$\Omega \subset \mathbb{R}^d$	Trägerbereich
$r \in \{1, \dots, d\}$	Dimensionsindex, „Richtung“
$\underline{l} = (l_1, \dots, l_d)^T \in \mathbb{N}^d$	(Multi-)Index der Verfeinerungslevel bezüglich der einzelnen Richtungen
$\underline{i} = (i_1, \dots, i_d)^T$	(Multi-)Index ($1 \leq i_r \leq n_{l_r}$)
$\varphi_{\underline{l}, \underline{i}}(\underline{x}) = \prod_{r=1}^d \varphi_{l_r, i_r}(x_r)$	Basisfunktionen der Dimension d
$\varphi_{l_r, i_r}(x_r)$	Eindimensionale Basisfunktionen für $r \in \{1, \dots, d\}$
$w_{\underline{l}, \underline{i}}$	Gewichtskoeffizienten
n_{l_r}	Anzahl der eindimensionalen Basisfunktionen und Gewichtskoeffizienten pro Richtung $r \in \{1, \dots, d\}$
b_r	Breite des Träger-Intervalls in Richtung $r \in \{1, \dots, d\}$
h_{l_r}	Schrittweite in Richtung $r \in \{1, \dots, d\}$
I_{l_r, i_r}	Teilintervall in Richtung $r \in \{1, \dots, d\}$
\hat{N}	Anzahl der tatsächlich gewerteten historischen Daten

Kapitel 1.

Einleitung

Die Verwendung stochastischer Differentialgleichungen zur Modellierung der zukünftigen Wertentwicklungen verschiedener Anlageformen hat in den letzten Jahren im Banken- und Versicherungsgewerbe mehr und mehr an Bedeutung gewonnen. Sie dient unter anderem als Grundlage für eine adäquate Bewertung von Optionen sowie für Projektionen und Analysen im Bereich des Risikomanagements. So ist es zum Beispiel im Rahmen des Asset-Liability-Managements (der Aktiv-Passiv-Steuerung) eines Versicherungsprodukts notwendig, die zukünftige Entwicklung des Kapitalmarktes (der „Aktivseite“) zu modellieren. Zu den wesentlichen Faktoren, die dabei von Interesse sind, zählen die Kursentwicklung von Aktien, etwa die des Deutschen Aktienindex DAX, und die Entwicklung der Zinssätze festverzinslicher Wertpapiere, zum Beispiel in Form des Tagesgeldsatzes.

Für Aktienkurse wird bei der stochastischen Modellierung in der Regel eine Geometrische Brownsche Bewegung verwendet. Diese ist ein stochastischer Prozess, der über zwei Parameter, die Drift und die Volatilität, spezifiziert wird. Die Modellierung von Zinssätzen gestaltet sich dagegen komplexer. Im Unterschied zu Aktienkursen, deren Wert lediglich vom Betrachtungszeitpunkt abhängt, besteht bei Zinssätzen eine zusätzliche Abhängigkeit von der Laufzeit des zugrunde liegenden festverzinslichen Wertpapiers. Diese wird in der Zinsstruktur ausgedrückt, die die Zinssätze $R(t, T)$ zu jedem Zeitpunkt t in Abhängigkeit von ihrer Laufzeit $T - t$ angibt. Bei der Darstellung der Zinsstruktur unterscheidet man Ein- und Mehr-Faktor-Modelle (vergleiche etwa [Gla04], [FMW03]). Ein-Faktor-Modelle stellen die gesamte Zinsstruktur mit Hilfe eines einzelnen zugrunde liegenden Faktors, der so genannten Short Rate $r(t)$, dar. Zu den Ein-Faktor-Modellen zählt zum Beispiel das Cox-Ingersoll-Ross-Modell (CIR-Modell, [CIR85]), welches insgesamt drei Parameter beinhaltet.

Ist ein Modell gefunden, durch das die Realität angemessen beschrieben werden kann, ergibt sich im nächsten Schritt das Problem der passenden Parameterwahl, die das Modell möglichst gut an die aktuelle Marktsituation anpasst. Die Parameterwerte haben erheblichen Einfluss auf die Güte der Modellierung und alle abgeleiteten Kenngrößen, die von Banken und Versicherungen als Entscheidungsstütze im Managementbereich herangezogen werden. Da die konkreten Parameterwerte der Modelle im Allgemeinen nicht direkt am Markt beobachtbar sind, werden die benötigten Parameter üblicherweise auf Grundlage historischer Daten geschätzt. Aufgrund des endlichen Umfangs einer Stichprobe ist hierbei nur ein „best fit“ und keine genaue Bestimmung der gesuchten Parameterwerte möglich. Die Güte einer solchen „best fit“-Schätzung wird maßgeblich vom Grad der Repräsentativität einer Stichprobe bestimmt. Ausschlaggebend dafür sind sowohl Quantität als auch Qualität der Stichprobe.

Bei den Parameterschätzverfahren ist zwischen den sehr speziellen Parameter-spezifischen, den etwas allgemeineren Prozess-spezifischen sowie gänzlich allgemein gehaltenen Verfahren zu unterscheiden. Parameter-spezifische Verfahren nutzen spezielle Eigenschaften eines einzelnen Parameters zu dessen Schätzung. Solche Verfahren existieren allerdings nur für

ausgewählte Parameter in wenigen Modellen. Prozess-spezifische Verfahren verwenden a-priori-Wissen über den modellierten Prozess, zum Beispiel die Kenntnis der Verteilung, die dem Prozess zugrunde liegt. Zu den Prozess-spezifischen Verfahren zählt unter anderem die Maximum-Likelihood-Methode (siehe zum Beispiel [Lo86], [Sø97], [FMW03]). Diese verwendet eine Likelihood-Funktion, die in Abhängigkeit von den gewählten Parametern angibt, wie gut ein Modell gegebene historische Daten erklären kann. Zu ihrer Definition wird die Dichtefunktion oder eine geeignete Approximation des betrachteten stochastischen Prozesses herangezogen. Durch Minimierung der Likelihoodfunktion über die Parameter wird schließlich der optimale Parametersatz bestimmt. Für das CIR-Modell kann zum Beispiel der Ansatz von Andrew W. Lo [Lo86] zur Anwendung kommen, der die stochastische Differentialgleichung des CIR-Modells mit Hilfe des Euler-Maruyama-Schemas diskretisiert. Die approximative Likelihood-Funktion ist dabei durch die normalverteilten Übergangsdichten der diskretisierten Zuwächse gegeben. Aufgegriffen wurde dieser Ansatz von Lo unter anderem von Michael Sørensen [Sø97], der sich mit der Spezifizierung verschiedener Schätzfunktionen bezüglich der diskret beobachteten Daten beschäftigt hat, sowie der Arbeitsgruppe von Angelika May [FMW03].

Gegenstand dieser Arbeit ist, ein allgemeines Parameterschätzverfahren – das Histogramm-Matching-Verfahren – vorzustellen und seine Eigenschaften systematisch zu untersuchen. Dieses Verfahren beruht auf dem Vergleich von Dichte-Approximationen und kommt, im Gegensatz zur Maximum-Likelihood-Methode, ohne spezielle Kenntnisse der Übergangsdichten beziehungsweise entsprechender Approximationen aus. Damit ist es prinzipiell für beliebige stochastische Prozesse einsetzbar.

Das Verfahren basiert zum Einen auf gegebenen historischen und zum Anderen auf numerisch simulierten Datenwerten des zugrunde liegenden Prozesses. Die historischen Daten bezeichnen wir mit $\tilde{r}_0, \dots, \tilde{r}_{N-1}$, wobei \tilde{r}_i die Realisierung des Prozesses zum Zeitpunkt t_i angibt. Die Zeitpunkte t_0, \dots, t_{N-1} sind äquidistant im Zeitintervall $[0, T]$ verteilt mit Abstand Δt . In seiner Funktionsweise basiert das Histogramm-Matching-Verfahren auf drei Komponenten:

1. Bestimmung einer Dichte-Approximation der historischen Zuwächse des Prozesses,
2. numerische Generierung einer simulierten Dichte-Approximation,
3. Vergleich der simulierten mit der historischen Dichte-Approximation,
4. Parameteroptimierung, so dass simulierte und historische Dichte-Approximation möglichst gut übereinstimmen.

Im ersten Schritt des Verfahrens berechnen wir aus den historischen Datenwerten die Dichte der empirischen Verteilung der Zuwächse $\Delta\tilde{r}_i = \tilde{r}_{i+1} - \tilde{r}_i$. Im nächsten Schritt generieren wir zu gegebenen Parameterwerten mit Hilfe der Euler-Maruyama-Diskretisierung (siehe zum Beispiel [KP99]) der zugrunde liegenden stochastischen Differentialgleichung simulierte Zuwächse Δr_i für $i = 0, \dots, N - 1$. Um die zufälligen Einflüsse herauszumitteln, wiederholen wir diesen zweiten Schritt M mal. Aus der Gesamtheit aller simulierten Zuwächse bilden wir, wie unten dargestellt, eine Dichte-Approximation.

Die Dimension der Dichte-Approximationen wird durch die Anzahl der zu schätzenden Parameter bestimmt. Die einfachste Möglichkeit einer Dichte-Approximation besteht in der Verwendung von Histogrammen. Darüber hinaus existieren zahlreiche Kerndichteschätzer [Sil86]. Histogramme können als Kerndichteschätzer, basierend auf stückweise

konstanten Funktionen, aufgefasst werden. In dieser Arbeit untersuchen wir den Einsatz von Histogrammen sowie von Kerndichteschätzern basierend auf stückweise linearen Funktionen. Hierbei wird sowohl ein klassischer Tensorprodukt-Ansatz als auch ein Dünngitter-Ansatz (zum Beispiel [BG04]) vorgestellt. Beim Dünngitter-Ansatz wird, im Gegensatz zum Tensorprodukt-Ansatz, nicht mit der feinsten äquidistanten Intervallaufteilung des Bereichs gearbeitet. Stattdessen werden Intervallaufteilungen mit Teilintervallen verschiedener Größe ausgewertet und mittels der so genannten Kombinationstechnik (vergleiche [GSZ92], [Gar04]) zusammengesetzt. Damit ergibt sich die Dichte-Approximation als Kombination verschiedener Dichte-Approximationen auf Grundlage des Tensorprodukt-Ansatzes, die aber auf erheblich größeren Intervallaufteilungen basieren als die entsprechende Dichte-Approximation des klassischen Tensorprodukt-Ansatzes.

Beim klassischen Tensorprodukt-Ansatz hat eine Erhöhung der Dimension einen exponentiellen Anstieg der Anzahl der Basisfunktionen zur Folge. Die Anzahl der Basisfunktionen bestimmt die Anzahl der Auswertungen, die für die Ermittlung einer einzelnen Histogramm-Abweichung durchzuführen sind, und damit den Aufwand des Verfahrens. Diese mit der Zahl der Dimensionen exponentiell wachsende Komplexität wird auch als „Fluch der Dimension“ bezeichnet. Im Gegensatz zum Tensorprodukt-Ansatz kann mit Hilfe des Dünngitter-Ansatzes der „Fluch der Dimension“ in gewisser Weise gebrochen werden, da nur mit einer Auswahl von Basisfunktionen gearbeitet wird. Eine zusätzliche Dimension bedeutet keinen exponentiellen Anstieg der Anzahl der benötigten Basisfunktionen mehr.

Ein weiterer Vorteil des Dünngitter-Ansatzes ist, dass die geringere Anzahl der Basisfunktionen zu einer geringeren Zerstreung der Daten führen und somit einem Overfitting-Effekt stärker entgegenwirkt als es beim Tensorprodukt-Ansatz der Fall ist. Besonders die Kombinationstechnik bietet sich im Hinblick auf eine möglichst große Ausschöpfung der Dateninformationen an, da sie verschieden feine Gitterauswertungen miteinander kombiniert. Ein vergleichbarer Einsatz von dünnen Gittern wird zum Beispiel in [CGY04] beschrieben.

Nachdem auf diese Weise historische und simulierte Dichte-Approximationen bestimmt worden sind, liefert die Abweichung der simulierten von der historischen Dichte-Approximation im dritten Schritt des Verfahrens ein Maß für die Güte der getroffenen Parameterwahl. Die Abweichungen zweier gegebener Dichte-Approximationen werden hierbei in der diskreten L_2 -Norm gemessen.

Durch Minimierung der Abweichung werden im letzten Schritt des Histogramm-Matching-Verfahrens die optimalen Parameterwerte bestimmt. Im einfachsten Fall wird dazu der Parameterbereich systematisch durchsucht. Diese Methode ist zwar – im Gegensatz zu gradientenbasierten Suchverfahren wie dem Newton-Verfahren – robust; jedoch ist der Aufwand wesentlich höher und steigt exponentiell mit der Anzahl der zu schätzenden Parameter.

Um mehrere Parameter gleichzeitig zu schätzen, unterscheiden wir iterative und parallele Histogramm-Matching-Verfahren. Bei den iterativen Verfahren werden alle Parameter – bis auf einen einzelnen – abwechselnd festgehalten. Dieser einzelne Parameter wird, basierend auf den zuvor ermittelten Schätzwerten der anderen Parameter, mittels des (eindimensionalen) Histogramm-Matching-Verfahrens geschätzt. Im Gegensatz dazu müssen beim parallelen Histogramm-Matching-Verfahren Optimierungsprobleme gelöst werden, deren Dimension der Anzahl der Parameter entspricht.

Insgesamt werden in dieser Arbeit folgende Histogramm-Matching-Verfahren für die Parameterschätzung stochastischer Prozesse vorgestellt und untersucht:

- ein iteratives Histogramm-Matching-Verfahren basierend auf eindimensionalen Dichte-Approximationen,
- ein paralleles Histogramm-Matching-Verfahren basierend auf zweidimensionalen Dichte-Approximationen mittels des Tensorprodukt-Ansatzes,
- ein paralleles Histogramm-Matching-Verfahren basierend auf zweidimensionalen Dichte-Approximationen mittels des Dünngitter-Ansatzes.

Bei numerischen Tests hat sich gezeigt, dass eine vorausgehende Skalierung der Daten sinnvoll ist. Aus diesem Grund wurden zwei mögliche Vorgehensweisen untersucht: eine Skalierung der Daten bezüglich des Erwartungswertes und eine Skalierung bezüglich der Varianz.

Im numerischen Teil dieser Arbeit werden zunächst die Verfahrensparameter der verschiedenen Histogramm-Matching-Verfahren am Beispiel des CIR-Modells näher untersucht. Dazu zählen unter anderem der Umfang der historischen Daten, die Anzahl simulierter Datensätze pro betrachtetem Parametersatz, die Teilintervallbreite, die durch die Intervallgrenzen und die Anzahl von Teilintervallen festgelegt wird, die Wahl der Startwerte und die Anzahl der Iterationen bei den iterativen Verfahren sowie die Wahl der Basisfunktionen bei den Dichte-Approximationen.

Ein Vergleich der Histogramm-Matching-Verfahren findet sowohl untereinander als auch mit der Maximum-Likelihood-Methode statt. Dabei stellt sich heraus, dass die auf der Kombinationstechnik beruhenden Histogramm-Matching-Verfahren für kleine und sehr große Datensatzumfänge im Mittel die besten Schätzergebnisse liefern. Dafür benötigen sie allerdings einen erheblich höheren Rechen- und Speicheraufwand als die anderen Verfahren. Dies ist einerseits darauf zurückzuführen, dass jeder Datenpunkt bezüglich der einzelnen zu kombinierenden Gitter mehrfach ausgewertet werden muss. Andererseits muss beim Einsatz der Kombinationstechnik eine vergleichsweise hohe Auflösung des Gitters verwendet werden, um die selbe Genauigkeit wie beim Tensorprodukt-Ansatz zu erreichen.

Nachdem die Parameterschätzverfahren zunächst speziell für das CIR-Modell untersucht wurden, wenden wir es auf ein komplizierteres Modell an, welches zwei gekoppelte stochastische Differentialgleichungen jeweils für den Aktienkurs beziehungsweise für die Short Rate beinhaltet. Die Schätzung der fünf Parameter dieses Modells wird auf Grundlage realer Marktdaten, den DAX-Werten und der Tagesgeldsätze der letzten 27 Jahre (entnommen aus der Zeitreihendatenbank der Deutschen Bundesbank [Ze]), durchgeführt.

Die Arbeit ist folgendermaßen gegliedert: In Kapitel 2 werden verschiedene Arten von Wertpapieren und die mit ihnen verbundenen Kapitalmarktrisiken erläutert. In Kapitel 3 werden die benötigten theoretischen Grundlagen der stochastischen Analysis eingeführt. Kapitel 4 bietet einen Überblick über unterschiedliche stochastische Zins- und Aktienkursmodelle. Verschiedene Diskretisierungs- und Simulationsmöglichkeiten der stochastischen Modelle werden in den Kapiteln 5 und 6 diskutiert. Kapitel 7 stellt die Dichte-Approximationen vor, die in Kapitel 8 die Grundlage der Histogramm-Matching-Verfahren darstellen. Dabei wird in Kapitel 8 neben einer ausführlichen Erörterung der Histogramm-Matching-Verfahren die Maximum-Likelihood-Methode nach [Lo86] für das CIR-Modell erklärt. In Kapitel 9 werden numerische Analysen hinsichtlich der Verfahrensparameter sowie der Effizienz der einzelnen Verfahren durchgeführt. Zum Abschluss werden die Parameter eines komplizierten Finanzmarktmodells geschätzt, welches die in der Realität beobachtbare negative Korrelation zwischen Aktien und Zinsen berücksichtigt. Schlussbe-

merkungen zum Einsatz der Histogramm-Matching-Verfahren und ein Ausblick werden in Kapitel 10 gemacht.

An dieser Stelle möchte ich meinen Dank all jenen aussprechen, die durch ihre fachliche und persönliche Unterstützung zum Gelingen meiner Diplomarbeit beigetragen haben. Vor allem danke ich Prof. Dr. Michael Griebel für die Ermöglichung und Förderung dieser Diplomarbeit sowie für das in diesem Zusammenhang entgegenbrachte Interesse. Darüber hinaus gilt mein Dank Dr. Thomas Gerstner, dessen intensive Betreuung stets hilfreich und für die Entwicklung dieser Arbeit sehr wichtig war. Für die persönliche Unterstützung bedanke ich mich bei meinen Eltern, meinen Freunden und meinen Mitdiplomandinnen Margrit Klitz, Melanie Reiferscheid und Claudia Warawko.

Kapitel 2.

Wertpapiere und ihre Risiken

In diesem Kapitel wird – in Anlehnung an [GP99] – ein Einblick in die Welt der Wertpapiere an den Finanzmärkten gegeben. Damit soll der wirtschaftliche Hintergrund dieser Diplomarbeit umrissen und eine Einführung in die später von uns betrachteten Risikofaktoren in Form des Aktien innewohnenden Kursrisikos (2.1) und des Zinsänderungsrisikos festverzinslicher Wertpapiere (2.2) gegeben werden.

Definition 2.0.1 [WERTPAPIERE; GELD-, KAPITAL- UND WARENWERTPAPIERE]

Ein Wertpapier ist eine Urkunde¹, in der ein privates Vermögensrecht derart verbrieft ist, dass zur Ausübung des Rechts der Besitz an der Urkunde erforderlich ist. Damit stellt ein Wertpapier die Verkörperung eines Vermögensrechts dar. Je nach Art des verbrieften Vermögenswertes kann unterschieden werden zwischen Geldwertpapieren, die eine Verkörperung kurzfristiger Forderungen darstellen, Kapitalwertpapieren, in denen langfristige Forderungen oder Teilhaberrechte verbrieft sind, und Warenwertpapieren, bei denen Rechte an schwimmenden oder lagernden Waren Gegenstand der Urkunde sind.

Zu den Kapitalwertpapieren zählen unter anderem *Aktien* und *Schuldverschreibungen*, die insbesondere *festverzinsliche Wertpapiere* umfassen. Da in Aktien die Teilhaberrechte an Unternehmen verbrieft sind, werden sie auch als *Mitgliedschaftspapiere* bezeichnet. Schuldverschreibungen hingegen sind so genannte *Forderungspapiere*. Nähere Ausführungen zu Aktien und Schuldverschreibungen finden sich in 2.1 und 2.2.

Definition 2.0.2 [BÖRSENFÄHIGE WERTPAPIERE; EFFEKTE]

Um als börsenfähig eingestuft zu werden, müssen Wertpapiere Voraussetzungen erfüllen, die es ermöglichen, dass in ihnen verbrieft Vermögenwerte an Börsen und anderen geregelten Wertpapiermärkten schnell, kostengünstig und zuverlässig bewertet, erworben und verkauft werden können. Im Wesentlichen sollten die zu handelnden Wertpapiere beweglich sowie von gleicher Beschaffenheit sein und sich nach Zahl, Maß oder Gewicht bestimmen lassen (vergleiche [GP99]). Börsenfähige Kapitalwertpapiere werden auch als Effekte bezeichnet.

Aktien und Schuldverschreibungen verkörpern je nach Art der Emission bei gleichem Nominalwert (siehe dazu Definition 2.0.6) beziehungsweise gleicher Stückelung die gleichen Rechte. Aufgrund dieser Eigenschaft sind die einzelnen Stücke untereinander austauschbar, ohne dass der Inhaber dadurch eine Minderung oder Mehrung seiner dem Wertpapier entspringenden Rechte erfährt. Damit erfüllen sowohl Aktien als auch Schuldverschreibungen die wichtigsten Voraussetzungen, um als börsenfähig eingestuft zu werden. Sie zählen somit zur Gruppe der Effekte.

¹Als *Urkunden* werden im Allgemeinen schriftliche Niederlegungen von rechtlich oder wirtschaftlich bedeutsamen Erklärungen bezeichnet.

Bemerkung 2.0.3 [WERTPAPIERE IM SINNE DES WERTPAPIERHANDELSGESETZES]

Im Sinne des Wertpapierhandelsgesetzes zählen Aktien, Zertifikate, die Aktien vertreten, Schuldverschreibungen, Genussscheine, Optionsscheine sowie andere vergleichbare Wertpapiere, insoweit sie an einem Markt handelbar sind, zur Gruppe der Wertpapiere. Dabei muss keine explizite Urkunde vorliegen.

Gerade der letzte Satz der Bemerkung 2.0.3 spielt heutzutage im Umgang mit Wertpapieren eine große Rolle, denn da für den Handel eines Wertpapiers (entgegen früherer Zeiten) keine explizite Urkunde vorliegen muss, entfällt die Notwendigkeit eines realen Transports. Stattdessen kann auf zentrale Verwahrungsorte (etwa das so genannte "Fort Knox" in Olten, der größte Wertpapier-Tresor der Schweiz) zurück gegriffen werden und Transaktionen werden in der Regel auf elektronischem Wege gegen Zahlung von Transaktionsgebühren in Echtzeit ausgeführt. Als Nachweis für die sich in ihrem Besitz befindlichen Wertpapiere erhalten die Inhaber Depotauszüge von den eingelagerten Papieren. Inhaber einer Aktie haben damit etwa die Möglichkeit, das Recht auf die Teilnahme an der Hauptversammlung des entsprechenden Unternehmens geltend zu machen.

Bemerkung 2.0.4 [WERTPAPIERHANDEL ALS RISIKOTRANSFER]

Sowohl Emittenten, die die Wertpapiere in Emissionen ausstellen, als auch Anleger verfolgen mit dem von ihnen eingegangenen Handel von Wertpapieren einen bestimmten Zweck. Den Emittenten dienen Wertpapiere der Beschaffung von Eigen- oder Fremdkapital. Für Anleger stellen Wertpapiere eine meist flexible Geld- und Vermögensanlage dar. Benötigt ein Anleger liquide Mittel, so kann er die Wertpapiere wieder veräußern oder beleihen. Da Wertpapiere nicht risikofrei sind, kann ihr Handel auch als Risiko-Transfer betrachtet werden. Es geht um das bewußte Abgeben beziehungsweise Eingehen finanzieller Risiken. Dabei wird zum Ausgleich ein gewisser Preis, oder zumindest die Erwartung auf diesen Preis, vereinbart.

Definition 2.0.5 [FINANZMÄRKTE; GELD-, KAPITAL- UND DEVISENMARKT]

Die Märkte, auf denen mit Finanzinstrumenten wie Geld, Devisen, Krediten, Wertpapieren und Finanzderivaten tatsächlich gehandelt wird, werden als Finanzmärkte bezeichnet. Diese müssen die Aufgabe erfüllen, Angebot und Nachfrage bezüglich der Finanzinstrumente über die Bildung aktueller und fairer Preise ins Gleichgewicht zu bringen. Der Handel kann dabei einerseits von den Marktteilnehmern persönlich an konkreten Märkten, andererseits über elektronische Kommunikationsnetze vollzogen werden (vergleiche die Erläuterungen vor Bemerkung 2.0.4). Die Finanzmärkte teilen sich auf in:

- **Geldmarkt (auch Bankengeldmarkt):** Dies ist der Markt für kurz-² und mittelfristige³ Geldaufnahmen (Kredite) und Geldanlagen. Der Geldzins bezeichnet den Zinssatz, mit dem am Geldmarkt kurzfristiger Kredite, besonders im Verkehr von Kreditinstituten untereinander oder zwischen Kreditinstituten und Zentralbank, gehandelt werden.
- **Kapitalmarkt:** An diesem werden langfristige⁴ Kapitalanlagen gehandelt. Im engeren Sinne ist dies der Markt für Wertpapiere und Finanzderivate, im weiteren Sinne ist es der Markt für alle langfristigen Kredite und Vermögensanlagen. Zum Kapitalmarkt, der sich in viele Teilmärkte aufteilt, gehört im weitesten Sinne auch der

²Laufzeit bis zu 4 Jahren, meist weniger als 365 Tage.

³Laufzeit von 4 bis 8 Jahren.

⁴Laufzeit von mehr als 8 Jahren, oft 30 Jahre oder noch länger.

Immobilienmarkt. Der Kapitalzins bezeichnet den (langfristigen) Zinssatz am Kapitalmarkt.

- *Devisenmarkt: Hier werden Auszahlungen in fremden Währungen gehandelt.*

Geld- und Kapitalmarkt werden auch als Kassamarkt zusammengefasst.

Definition 2.0.6 [ABZINSUNGSPAPIERE, ZINSTRAGENDE PAPIERE, NOMINALWERT]

An den Finanzmärkten wird zwischen Abzinsungspapieren und zinstragenden Papieren unterschieden. Während zinstragende Papiere, bei denen es nach den getroffenen Handelsvereinbarungen zu einer oder mehreren laufenden Zinszahlungen kommt, zum ausgehandelten Nominalwert emittiert werden, werden Abzinsungspapiere, die über keinerlei laufende Zinszahlungen verfügen, mit einem vom Zinssatz abhängigen Abschlag des Nominalwertes gehandelt. Der Anleger zahlt dabei als Kaufpreis den um den Zinsertrag reduzierten Nominalwert und erhält am Ende der Laufzeit den gesamten Nominalwert zurück.

2.1. Aktien

Definition 2.1.1 [AKTIE, AKTIONÄR]

Aktien sind börsenfähige Wertpapiere, die Teilhaberrechte an einer Aktiengesellschaft verbriefen. Der Inhaber einer Aktie, der Aktionär, ist am Grundkapital der Aktiengesellschaft beteiligt.

Zu den verbrieften Grundrechten der Aktionäre zählen vor allem das Recht auf Anteil am Unternehmensgewinn (*Dividendenrecht*), *Teilnahmerecht* an und *Stimmrecht* bei der Hauptversammlung, *Auskunftsrecht* durch den Vorstand, das *Bezugsrecht* bei der Ausgabe junger Aktien sowie das *Recht auf Anteil am Liquidationserlös*. Diese Rechte sind im Aktiengesetz und in der Satzung der Gesellschaft festgelegt.

Je nach Art der Zerlegung des Grundkapitals unterscheidet man zwischen *Nominalwertaktien* und *Stückaktien*. Bei Nominalwertaktien, die einen Mindestwert von 1 € haben müssen, ergibt sich der Anteil am Grundkapital durch das Verhältnis des Aktiennominalwertes zum selbigen. Bei Stückaktien wird der Anteil am Grundkapital durch die Gesamtanzahl der Stückaktien des Unternehmens bestimmt.

Tabelle 2.1 zeigt die für Anlagen in Aktien spezifischen Risiken auf. Von besonderem Interesse für die weiteren Betrachtungen innerhalb dieser Diplomarbeit ist das Kursrisiko. Dieses Risiko wird später mit Hilfe stochastischer Prozesse modelliert werden.

SPEZIFISCHE RISIKEN BEI ANLAGEN IN AKTIEN	
RISIKOART	CHARAKTERISIERUNG
Kursrisiko	Gefahr einer negativen Kursentwicklung der Aktie...
- unsystematisch (allgemeines Marktrisiko)	...bei allgemein rückläufiger Entwicklung des gesamten Aktienmarktes (beziehungsweise des betreffenden Teilmarktes) - unabhängig von der besonderen Situation des Unternehmens.
- systematisch	...aufgrund unternehmensspezifischer Entwicklungen der Aktiengesellschaft.
Unternehmerisches Risiko	Gefahr, dass sich die wirtschaftliche Lage des Unternehmens anders entwickelt als erwartet und Verluste drohen.
Dividendenrisiko	Gefahr, dass die Dividende in gewinnschwachen oder in Verlustjahren gekürzt wird oder ausfallen kann.
Risiko der Kursprognose	Gefahr aufgrund von Entscheidungen, die auf Kursprognosen beruhen, die nicht mit der tatsächlichen Marktsituation übereinstimmen.

Tabelle 2.1.: Spezifische Risiken bei Anlagen in Aktien.

2.2. Schuldverschreibungen / Festverzinsliche Wertpapiere

Definition 2.2.1 [SCHULDVERSCHREIBUNGEN]

Schuldverschreibungen sind börsenfähige Wertpapiere, die Forderungsrechte verkörpern. Es handelt sich um verbriefte Kredite, mit denen sich die Emittenten, in diesem Zusammenhang auch Schuldner genannt, Fremdkapital auf dem Kapitalmarkt beschaffen wollen. Die Anleger, auch Gläubiger genannt, erwerben eine verbriefte, verzinsliche Darlehensforderung. Sie haben grundsätzlich den Anspruch auf Rückzahlung des Nominalwerts und der anfallenden (Nominal-)Zinsen.

Tabelle 2.2 gibt eine Übersicht der Einteilung von Schuldverschreibungen.

EINTEILUNG VON SCHULDVERSCHREIBUNGEN NACH					
Emittenten	Verzinsung	Laufzeit	Rückzahlung	Sitz des Emittenten	Währung
▶ Öffentliche Anleihen	▶ festverzinsliche Anleihen	▶ kurzfristige Anleihen	▶ gesamtfällige Anleihen (am Ende der Laufzeit)	▶ Inlandsanleihen	▶ Euro-Anleihen
▶ Bankschuldverschreibungen	▶ variabel verzinsliche Anleihen	▶ mittelfristige Anleihen	▶ Tilgungsanleihen (über die Laufzeit verteilt)	▶ Auslandsanleihen	▶ Währungsanleihen
▶ Industrieobligationen	▶ unverzinsliche Anleihen	▶ langfristige Anleihen	▶ ewige Anleihen (keine Rückzahlungspflicht) ⁵		

Tabelle 2.2.: Einteilung von Schuldverschreibungen.

Um einen kleinen Einblick in tatsächlich am Markt vorhandene Schuldverschreibungen zu geben, findet sich in Tabelle 2.3 eine Übersicht börsennotierter Bundeswertpapiere, die im Sinne von Tabelle 2.2 gerade die börsenfähigen öffentlichen Anleihen verkörpern.

BÖRSENNOTIERTE BUNDESWERTPAPIERE		
WERTPAPIERART	CHARAKTERISIERUNG	LAUFZEIT
Bundesanleihen	Einmalemissionen, zur Beschaffung langfristiger Mittel	meistens 10 Jahre, aber auch 12,15,30 Jahre
Bundesobligationen	Daueremissionen ⁶ , zur Beschaffung mittelfristiger Mittel	5 Jahre
Schatzanweisungen des Bundes, seiner Sondervermögen und der Treuhandanstalt	Einmalemission, zur Beschaffung kurzfristiger Mittel	Anfangslaufzeit von 2 Jahren

Tabelle 2.3.: Börsennotierte Bundeswertpapiere.

Definition 2.2.2 [KUPON-ANLEIHEN (ENGL. BONDS)]

Bei einer Kupon-Anleihe (engl. Bond) handelt es sich um eine Schuldverschreibung, die das Recht auf Rückzahlung eines Nominalwerts zuzüglich einer Verzinsung verbrieft. Die im Zuge der Verzinsung meist periodisch anfallenden Zahlungsströme werden auch als Kupon-Auszahlungen bezeichnet. Diese Bezeichnung rührt daher, dass früher so genannte (Zins-)Kupon-Bücher ausgegeben wurden. Für jedes Formular (Kupon) aus diesem Buch erhielten die Besitzer eine entsprechende Zinsauszahlung.

Definition 2.2.3 [FESTVERZINSLICHE ANLEIHEN (ENGL. STRAIGHT BONDS), VARIABEL VERZINSLICHE ANLEIHEN (ENGL. FLOATING RATE NOTES)]

Je nach zugrunde liegender Verzinsungsart unterscheidet man zwischen festverzinslichen Anleihen (engl. Straight Bonds) bei festem Nominalzins und variabel verzinslichen Anleihen (engl. Floating Rate Notes) bei entsprechend variabler Verzinsung. Eine festverzinsliche Anleihe ist meistens über die gesamte Laufzeit konstant in der Höhe ihrer Verzinsung. Es kann höchstens zu abgestuften Veränderungen des Zinssatzes kommen. Dagegen wird bei variabler Verzinsung der für jede Zinsperiode gültige Zinssatz⁷ auf Grundlage eines Referenzzinssatzes (Leitzinses), wie etwa des EURIBORs⁸ oder des LIBORs⁹, jeweils aktualisiert.

Definition 2.2.4 [NULL-KUPON-ANLEIHEN (ENGL. ZERO-BONDS)]

Null-Kupon-Anleihen (engl. Zero-Bonds) bezeichnen Schuldverschreibungen ohne laufende Zinszahlung. Bei diesen werden während der gesamten Laufzeit keine Zinszahlungen geleistet. Der Zinsertrag ergibt sich aus der Differenz des Emissions- und des Rückzahlungskurses. Je nachdem, ob das Wertpapier zum abgezinsten oder zum einfachen Nominalwert emittiert wird, kommt es am Ende der Laufzeit zur Rückzahlung des Nominalwerts inklusive oder exklusive Zinsauszahlung.

⁵Bei ewigen Anleihen ist die Laufzeit nicht von vorneherein festgelegt.

⁶Neue Serie wird aufgelegt, wenn Marktlage den Übergang zu neuem Nominalzins erfordert.

⁷Wird der Zins in Prozent angegeben, so spricht man von Zinssatz.

⁸EUropean InterBank Offered Rate.

⁹London InterBank Offered Rate.

Anmerkung 2.2.5

In der Praxis werden Schuldverschreibungen häufig als festverzinsliche Wertpapiere bezeichnet, obwohl es, wie erläutert, neben festverzinslichen auch variabel verzinsliche sowie gänzlich unverzinsliche Anleihen gibt, die bei dieser Begriffswahl übergangen werden.

Tabelle 2.4 zeigt die spezifischen Risiken von Schuldverschreibungen. Dabei hat das Zinsänderungsrisiko in Bezug auf diese Arbeit eine besondere Stellung. Für dieses Risiko wird später, ebenso wie für das Kursrisiko von Aktien, ein stochastisches Modell mit geeigneten Parametern gesucht.

SPEZIFISCHE RISIKEN BEI ANLAGEN IN SCHULDVERSCHREIBUNGEN	
RISIKOART	CHARAKTERISIERUNG
Bonitätsrisiko	Gefahr der Zahlungsunfähigkeit des Schuldners.
Zinsänderungsrisiko	ergibt sich aus der Ungewissheit über die zukünftige Zinsentwicklung. Steigt der Marktzins, so wirkt sich das beim Besitzer eines festverzinslichen Wertpapiers in Form von Kursverlusten aus. Die Zinsänderungsempfindlichkeit hängt unter anderem auch vom <i>Nominalzins</i> ¹⁰ , mit dem die Anleihe ausgestattet ist, und von der Restlaufzeit der Anleihe ab.
Kündigungsrisiko	Gefahr, dass der Schuldner bei sinkenden Marktzinsen mit einer vorzeitigen Kündigung reagiert, um seine Zinslast zu senken.
Auslosungsrisiko	Gefahr bei Auslosungsanleihen, dass eine frühzeitige Auslosung zu einem Renditeverlust gegenüber der auf der Grundlage der durchschnittlichen Laufzeit errechneten Emissionen führen kann.
Besondere Risiken einzelner Anleiheformen	bestehen bei Anleihen mit Sonderrechten oder besonderen Ausstattungsmerkmalen (zum Beispiel Währungsrisiko bei Fremdwährungsanleihen, zusätzliches Kursrisiko der Wandelanleihen durch Kursverlauf der Wandelaktie).

Tabelle 2.4.: Spezifische Risiken bei Anlagen in Schuldverschreibungen.

¹⁰Bezeichnung für den für eine Anleihe vereinbarten oder bezahlten Zinssatz.

Kapitel 3.

Stochastische Prozesse

In diesem Kapitel wird die für diese Diplomarbeit relevante Theorie stochastischer Prozesse eingeführt. Dabei folgt die Herangehensweise in großen Teilen derjenigen, die auch in [Deu04] verfolgt wird. Zunächst wird eine Herleitung der in einen stochastischen Prozess einfließenden Zufallskomponente über Random Walks gegeben (3.1), bei der unter anderem bereits ein erstes Modell für Kursänderungen (3.1.1) vorgestellt und der Einsatz des Wiener-Prozesses erläutert wird (3.1.2). Im Anschluss findet sich eine formale Definition allgemeiner stochastischer Prozesse mittels Itô-Prozessen (3.3), die neben der Zufallskomponente in Form eines Wiener-Prozesses zusätzlich auf einer deterministischen Komponente (3.2) basieren. Damit steht ein allgemeines stochastisches Modell zur Verfügung, dessen Koeffizienten für spezielle Prozesse (3.4) noch spezifiziert werden müssen.

Anmerkung 3.0.6

In Anhang A wird ohne Anspruch auf Vollständigkeit ein knapper Überblick über die wichtigsten im Rahmen dieser Diplomarbeit verwendeten Definitionen und Resultate der Wahrscheinlichkeitstheorie gegeben. Für eine Vertiefung der Zusammenhänge wird auf die gängige Literatur ([Bau02]) verwiesen.

Ein *stochastischer Prozess* ist eine Zufallsgröße $X(t)$, die dadurch charakterisiert ist, dass ihr Wert über die Zeit t zufälligen Änderungen ausgesetzt ist. Genauer handelt es sich bei einem stochastischen Prozess um eine Familie von Zufallsvariablen:

Definition 3.0.7 [STETIG STOCHASTISCHER PROZESS, FÄLLIGKEIT, PFAD]

Für eine stetige, kontinuierlich variierende Zeitvariable $t \in I = [0, T] \subset \mathbb{R}$ bezeichnet $\{X(t) | t \in I\}$ (kurz: $X(t)$, X_t oder auch einfach nur X , falls die Abhängigkeit von t aus dem Zusammenhang klar ersichtlich ist.) einen stetig stochastischen Prozess. Das Zeitintervall I wird dabei insbesondere durch den Fälligkeitszeitpunkt (kurz: Fälligkeit) T bestimmt. Unter einem Pfad wird eine konkret beobachtete Realisierung eines solchen Prozesses für ein festes Zeitintervall I verstanden.

Für den aktuellen Wert eines Risikofaktors (etwa eines Aktienkurses oder eines Zinssatzes) wird angenommen, dass er bereits alle aus der Vergangenheit – also den historische Kursentwicklungen – ableitbaren Informationen enthält. Diese Annahme lässt darauf schließen, dass für das zukünftige Kursverhalten nur der augenblickliche Wert zählt.¹ Keinesfalls sind frühere Kurse ausschlaggebend. Stochastische Prozesse, die diese Eigenschaft der so genannten *schwachen Markteffizienz* berücksichtigen, werden *Markov-Prozesse* genannt:

Definition 3.0.8 [MARKOV-PROZESS, GEDÄCHTNISLOSIGKEIT EINES PROZESSES]

Ein stochastischer Prozess $X(t)$ ist genau dann ein Markov-Prozess, wenn er die Eigenschaft besitzt, dass lediglich der aktuelle Wert – nicht aber der Weg, auf dem dieser aktuelle

¹Abgesehen von eventuellen äußeren, etwa politischen Einflüssen.

Wert erreicht wurde – Einfluss auf den nächsten Wert hat. In diesem Zusammenhang wird auch von der Gedächtnislosigkeit eines Prozesses gesprochen.

Zur Modellierung eines Markov-Prozesses $X(t)$ wird ein Prozess verwendet, der sich aus einer deterministischen Komponente und einer Zufallskomponente zusammensetzt. Im nächsten Kapitel (3.1) wird zunächst auf die Modellierung der Zufallskomponente eingegangen und eine mögliche Herleitung über das Prinzip der Random Walks aufgewiesen. Im Anschluss daran wird auch die deterministische Komponente in die Betrachtungen einbezogen (3.2) und schließlich in Kapitel 3.3 die gesamte Modellierung auf Grundlage von Itô-Prozessen beschrieben.

3.1. Die Zufallskomponente

Definition 3.1.1 [RANDOM WALK, SCHRITTVEKTOREN, RW-VEKTOR]

Ein Random Walk (auch: Irrfahrt) ist eine Verkettung zufälliger Schrittvektoren v_i , die für jeden Schritt $i \in \{1, \dots, n\}$ eindeutig seine Richtung und Länge ausgehend vom zuletzt erreichten (Orts-)Punkt festlegen. Die Summe

$$V = \sum_{i=1}^n v_i \tag{3.1}$$

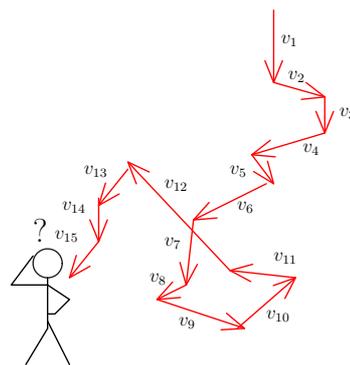
der n Schrittvektoren wird als Random-Walk-Vektor (kurz: RW-Vektor) bezeichnet.

Beispiel 3.1.2 [DESORIENTIERTER FUSSGÄNGER]

Für einen Random Walk im realen Leben kann das Beispiel eines desorientierten Fußgängers herangezogen werden. Abbildung 3.1 zeigt einen möglichen Random Walk eines (nach einem Kneipenbesuch) verwirrten Fußgängers auf dem Bonner Marktplatz.



(a) Der zurückgelegte Weg auf dem Bonner Marktplatz (Mit freundlicher Genehmigung des Landesvermessungsamts NRW [LVA06]).



(b) Die Schrittvektoren des zugehörigen Random Walks.

Abbildung 3.1.: Ein Random Walk am Beispiel eines desorientierten Fußgängers.

Satz 3.1.3 [DIE NORMALVERTEILUNG EINES RANDOM WALKS]

Der RW-Vektor eines Random Walks ist normalverteilt. Diese Normalverteilung besteht für

jede beliebige Dimensionsanzahl d des Random Walks. Dabei beträgt der Erwartungswert Null, also

$$E[V] = 0, \quad (3.2)$$

und die Varianz ist proportional zur Schrittzahl n :

$$\text{var}[V] = n \cdot \underbrace{c^2}_{\text{Proportionalitätskonstante}} \quad (3.3)$$

(Dementsprechend ist die Standardabweichung² proportional zu \sqrt{n} .)

Beweis:

- Da jeder einzelne Schritt (wie schließlich auch der gesamte RW-Vektor) nach Definition 3.1.1 rein zufällig in jede beliebige Richtung mit jeder beliebigen Länge führen kann, heben sich die Schrittvektoren im Mittel auf. Für eine ausreichend große Anzahl von Schritten ist demnach jede Schrittbewegung gleich wahrscheinlich. Es gilt

$$E[v_i] = 0 \quad \text{für alle } i \in \{1, \dots, n\}.$$

Aufgrund der Linearität von Erwartungswerten (Satz A.0.7) folgt

$$E[V] \stackrel{(3.1)}{=} E \left[\sum_{i=1}^n v_i \right] = \sum_{i=1}^n \underbrace{E[v_i]}_{=0} = 0.$$

- Aus der Unabhängigkeit der einzelnen Schritte lässt sich zudem auf die Unkorreliertheit der Schrittvektoren schließen. Für die Korrelationskoeffizienten gilt demzufolge nach Satz A.0.13

$$\rho_{ij} = \rho(v_i, v_j) = 0 \quad \text{für alle } i \neq j \in \{1, \dots, n\}$$

und damit für die Kovarianzen

$$\text{cov}[v_i, v_j] = \underbrace{\rho_{ij}}_{=0} \cdot \sqrt{\text{var}[v_i]\text{var}[v_j]} = 0 \quad \text{für alle } i \neq j \in \{1, \dots, n\}$$

entsprechend der Definition der Korrelationskoeffizienten (Definition A.0.11). Mit dem Verschiebungssatz für Kovarianzen (Satz A.0.14) ergibt sich

$$\begin{aligned} \underbrace{\text{cov}[v_i, v_j]}_{=0} &= E[v_i \cdot v_j] - \underbrace{E[v_i]}_{=0} \cdot \underbrace{E[v_j]}_{=0} \quad \text{für alle } i \neq j \in \{1, \dots, n\} \\ &\Leftrightarrow E[v_i \cdot v_j] = 0 \quad \text{für alle } i \neq j \in \{1, \dots, n\}. \end{aligned} \quad (3.4)$$

Daraus folgt

$$\begin{aligned} E[V^2] &\stackrel{(3.1)}{=} E \left[\sum_{i=1}^n v_i \sum_{j=1}^n v_j \right] = E \left[\sum_{i,j=1}^n v_i \cdot v_j \right] \stackrel{A.0.7}{=} \sum_{i,j=1}^n E[v_i \cdot v_j] \\ &= \sum_{i=1}^n E[v_i \cdot v_i] + \sum_{\substack{i,j=1 \\ i \neq j}}^n \underbrace{E[v_i \cdot v_j]}_{\stackrel{(3.4)}{=} 0} = \sum_{i=1}^n \underbrace{E[v_i^2]}_{\stackrel{(*)}{=} c^2} = n \cdot c^2 \end{aligned} \quad (3.5)$$

²Entspricht der Wurzel aus der Varianz (siehe Definition A.0.10).

unter der Annahme, dass die mittlere Länge des Quadrats eines Einzelschrittes durch die Konstante c^2 beschrieben wird (*). Doch damit wächst nicht nur das mittlere Quadrat des Gesamtabstands vom Ausgangspunkt proportional zur Schrittzahl; vielmehr ergibt sich für die Varianz mittels Definition A.0.9:

$$\text{var}[V] = \underbrace{\text{E}[(V - \text{E}[V])^2]}_{\stackrel{(3.2)}{=} 0} = \text{E}[V^2] \stackrel{(3.5)}{=} n \cdot c^2.$$

□

Satz 3.1.4 [DIE NORMALVERTEILUNG EINES EINDIMENSIONALEN RANDOM WALKS]
Die Wahrscheinlichkeitsverteilung eines eindimensionalen Random Walks ist nach Satz 3.1.3 gerade gegeben durch

$$p(V) = \frac{1}{\sqrt{\text{var}[V] \cdot 2\pi}} \cdot \exp\left(-\frac{(V - \text{E}[V])^2}{2 \cdot \text{var}[V]}\right)$$

mit $\text{E}[V] = 0$ und $\text{var}[V] = n \cdot c^2$.

Bemerkung 3.1.5

Durch den hier verfolgten Herleitungsweg über das Prinzip der Random Walks bedarf die Normalverteilung keiner ausdrücklichen Annahme. Sie ergibt sich automatisch aus dem Aufeinanderfolgen völlig zufälliger Schritte als Folge des Zentralen Grenzwertsatzes (Satz A.0.23).

Bemerkung 3.1.6 [SELBSTÄHNLICHKEIT DER RANDOM WALKS]

Die beschriebenen statistischen Eigenschaften lassen sich ohne Veränderung auf jede beliebige Verfeinerungsstufe übertragen. So lässt sich jeder Schrittvektor eines Random Walks als RW-Vektor eines Random Walks mit kleineren Schritten auffassen. Und umgekehrt lässt sich jeder RW-Vektor als Schrittvektor eines Random Walks mit größeren Schrittlängen verstehen. Diese Eigenschaft wird als Selbstähnlichkeit der Random Walks bezeichnet.

3.1.1. Ein Random-Walk-Modell für Kursänderungen

Um der ursprünglichen Absicht der Modellierung von Kursentwicklungen zu entsprechen, wird nun ein Random Walk gesucht, der für die Modellierung der benötigten Zufallskomponente eingesetzt werden kann. Dabei wird nicht der Kurs selbst, sondern der Logarithmus des Kurses **(3)** als eindimensionaler Random Walk **(1)** aufgefasst. Die Anzahl der Random Walk Schritte wird proportional zur während des Random Walks vergangenen Zeit gewählt **(2)**. Aus dieser Spezifizierung folgt ohne explizite Normalverteilungsannahme (vergleiche Bemerkung 3.1.5), dass die modellierten Kursänderungen normalverteilt sind mit Mittelwert Null und zur Zeit t proportionaler Varianz **(4)**.

Herleitung:

Zu (1): Je nachdem, ob der zu beschreibende Kurs fällt oder steigt, geht der nächste Schritt des Random Walks nach „oben“ oder „unten“. Zur Modellierung von Kursänderungen werden daher exakt zwei Richtungen benötigt. Da sich ein Random Walk für d Dimensionen gerade in $2d$ Richtungen bewegen kann, wird demzufolge ein eindimensionaler Random Walk verwendet.

Zu (2): Basierend auf der Annahme, dass bei jedem der n Schritte eines Random Walks

ein festes Zeitintervall dt vergeht, gilt für die gesamt vom aktuellen (Start-)Zeitpunkt t bis zum Endzeitpunkt T verbrauchte Zeit

$$T - t = n \cdot dt, \quad (3.6)$$

Die Anzahl der Schritte n ist daher proportional zur insgesamt während des Random Walks vergangenen Zeit $T - t$.

Zu (3): Schließlich bleibt es noch aus, die Größe festzulegen, die tatsächlich mit Hilfe des Random Walks modelliert werden soll. Dies kann, wie im Folgenden erläutert wird, nicht einfach der real zu beschreibende Kurs sein.

Zwei wesentliche Ansprüche entscheiden über die Wahl der zu modellierenden Größe:

- Von mathematischer Seite her besteht entsprechend (3.1) der Anspruch, dass ein Random Walk gerade durch die *Summe seiner einzelnen Schritt-Vektoren* definiert ist.
- Aus ökonomischer Sicht spricht alles dafür, die *Kursänderungen nicht absolut sondern relativ* darzustellen. Das heißt, ein Wertpapier sollte in Abhängigkeit von der tatsächlichen Höhe seines Geldwertes eine Kursänderung erfahren. Es wäre nicht sinnvoll, davon auszugehen, dass zwei Wertpapiere verschiedenen Geldwertes (zum Beispiel eines in der Höhe von 10€, ein weiteres in der Höhe von 1000€) eine Kursänderung desselben Geldwertes (zum Beispiel um -2€) erfahren.

Diese Ansprüche führen zu folgenden Erkenntnissen:

1. Würde der absolute Kurs des Risikofaktors modelliert werden, so käme es zu einer Zusammensetzung der Kursentwicklung von S_0 (Startkurs zum aktuellen Zeitpunkt $t = 0$) bis S_n (Schlusskurs zum Endzeitpunkt T) als Summe der n Einzelschritte $dS_i = S_i - S_{i-1}$ für $i = 1, \dots, n$:

$$S_n - S_0 = \sum_{i=1}^n dS_i = \sum_{i=1}^n (S_i - S_{i-1})$$

Damit wäre zwar der erste Anspruch erfüllt, allerdings würde es sich in diesem Modell neben dem gesamten RW-Vektor auch bei den einzelnen Schrittvektoren um absolute Geldwerte handeln. Dies würde dem ökonomischen Anspruch widersprechen.

2. Auch der Ansatz, die Kurs-Relationen S_i/S_{i-1} als Random Walk zu modellieren, würde ins Leere führen. Damit wäre zwar der zweite Anspruch einer relativen Größe erfüllt, der gesamte RW-Vektor ergäbe sich allerdings nicht wie von mathematischer Seite gefordert als Summe der durch S_i/S_{i-1} definierten einzelnen Schritte. Zudem würde es sich weder bei den S_i/S_{i-1} selbst, noch bei der Relation des letzten zum ersten Kurswert

$$\frac{S_n}{S_0} = \prod_{i=1}^n \frac{S_i}{S_{i-1}} \quad (3.7)$$

um Vektoren handeln.

3. An dieser Stelle kommt die Logarithmus-Operation zum Einsatz. Gemäß der Funktionalgleichung des Logarithmus gilt $\ln(a \cdot b) = \ln(a) + \ln(b)$ sowie $\ln(a/b) = \ln(a) -$

$\ln(b)$. Wird nun der Random Walk mittels der Logarithmen der Kurs-Relationen $\ln(S_i/S_{i-1})$ modelliert, so ergibt sich aus (3.7)

$$\ln\left(\frac{S_n}{S_0}\right) = \ln\left(\prod_{i=1}^n \frac{S_i}{S_{i-1}}\right) = \sum_{i=1}^n \ln\left(\frac{S_i}{S_{i-1}}\right)$$

und daher

$$\ln(S_n) - \ln(S_0) = \sum_{i=1}^n d\ln(S_i) \quad \text{mit} \quad d\ln(S_i) = \ln(S_i) - \ln(S_{i-1}). \quad (3.8)$$

Damit beschreibt die modellierte Größe wie auch schon im letzten Versuch relative Kursänderungen. Darüber hinaus ergibt sich, wie gefordert, der gesamte RW-Vektor als Summe der einzelnen Schrittvektoren $d\ln(S_i)$.

Zu (4): Entsprechend Satz 3.1.3 beziehungsweise Satz 3.1.4 gilt auch für den hier vorliegenden RW-Vektor

$$E[V] = E\left[\ln\left(\frac{S(T)}{S(t)}\right)\right] = 0 \quad \text{und} \quad \text{var}[V] = \text{var}\left[\ln\left(\frac{S(T)}{S(t)}\right)\right] = n \cdot c^2$$

Da im vorgestellten Modell die Anzahl der Schritte und die Zeit über (3.6) zusammenhängen, ist mit der Proportionalitätskonstante

$$\sigma^2 = \frac{c^2}{dt} \quad (3.9)$$

nun

$$\text{var}\left[\ln\left(\frac{S(T)}{S(t)}\right)\right] = n \cdot c^2 \stackrel{(3.9)}{=} \sigma^2 \cdot ndt \stackrel{(3.6)}{=} \sigma^2(T-t).$$

Die modellierten logarithmischen Kursänderungen sind demnach normalverteilt mit Mittelwert Null und zur Zeit t proportionaler Varianz.

Satz 3.1.7 [DAS ZEITSTETIGE RANDOM-WALK-MODELL]

Aus der Selbstähnlichkeit der Random Walks (Bemerkung 3.1.6) folgt, dass die Änderungen $d\ln(S(t))$ auch während einer infinitesimal kleinen Zeit $dt = \frac{T-t}{n}$ ($n \rightarrow \infty$) normalverteilt sind mit Mittelwert Null und Varianz

$$\text{var}[d\ln(S(t))] = \text{var}[\ln(S(t+dt)) - \ln(S(t))] = \text{var}\left[\ln\left(\frac{S(t+dt)}{S(t)}\right)\right] = \sigma^2 dt.$$

3.1.2. Einsatz des Wiener-Prozesses

Um den beschriebenen Random Walk als Zufallskomponente für die Modellierung eines Markov-Prozesses tatsächlich nutzen zu können, wird auf den so genannten Wiener-Prozess zurückgegriffen. Ein *Wiener-Prozess* (Definition A.0.27) $W(t)$ ist ein zeitstetiger stochastischer Prozess, dessen Zuwächse $dW(t)$, mit denen sich dieser Prozess über einen Zeitraum dt ändert, unabhängig und normalverteilt sind mit Mittelwert 0 und einer dem Zeitintervall dt entsprechenden Varianz:

$$dW(t) \sim N(0, dt) \quad (3.10)$$

Die Verteilung der Zuwächse lässt sich auch mittels einer standardnormalverteilten Zufallsvariable ausdrücken:

$$dW(t) \sim \sqrt{dt}Y(t) \quad \text{mit} \quad Y(t) \sim N(0, 1) \quad (3.11)$$

Demnach sind die Zuwächse genauso verteilt wie \sqrt{dt} -mal eine standardnormalverteilte Zufallsvariable. Diese Eigenschaft wird vor allem bei der numerischen Generierung von Prozessen ausgenutzt (vergleiche hierzu etwa den Algorithmus zur Generierung von Aktienkursen in Kapitel 6.1, Tabelle 6.3).

Da der oben definierte Random Walk für die logarithmischen Kursänderungen $d \ln S(t)$ normalverteilt sein soll mit Mittelwert 0 und Varianz $\sigma^2 dt$, kann er mit Hilfe der Zuwächse des Wiener-Prozesses ausgedrückt werden:

$$d \ln(S(t)) = \sigma dW(t) \sim N(0, \sigma^2 dt)$$

Damit ist ein stochastischer Prozess gegeben, der lediglich aus einer zufälligen Komponente besteht und der über den Parameter σ gesteuert wird.

3.2. Die deterministische Komponente

Bisher stützt sich die Modellierung der Kursänderungen lediglich auf eine Zufallskomponente. Allerdings kann das Modell damit nicht der so genannten *Martingaleigenschaft*³ handelbarer Risikofaktoren gerecht werden. Im bisherigen Modell war stets

$$\mathbb{E} \left[\ln \left(\frac{S(T)}{S(t)} \right) \right] = 0.$$

Damit ist es nicht möglich, einem Kurs gerecht zu werden, der eine mittlere Rendite abwerfen soll. Zur Abhilfe wird nun eine deterministische Komponente (μdt) hinzugenommen:

$$d \ln(S(t)) = \mu dt + \sigma dW(t) \sim N(\mu dt, \sigma^2 dt) \quad (3.12)$$

Auf Grundlage von (3.12) wird jetzt für einen endlichen Zeitzuwachs Δt unter stetiger Verzinsung eine mittlere Rendite μ erwartet (siehe auch Kapitel 3.3.1, Bemerkung 3.3.6).

3.3. Itô-Prozesse

In der stochastischen Analysis wird die allgemeine Form eines Diffusionsprozesses durch die stochastische Differentialgleichung von Itô⁴ beschrieben:

Definition 3.3.1 [ITÔ-PROZESS, STOCHASTISCHE DIFFERENTIALGLEICHUNG VON ITÔ]
 Ein Itô-Prozess $X(t)$ ist definiert durch die stochastische Differentialgleichung

$$dX(t) = a(X, t)dt + b(X, t)dW(t) \quad \text{mit} \quad dW(t) \sim Y(t)\sqrt{dt} \quad \text{und} \quad Y(t) \sim N(0, 1). \quad (3.13)$$

Dabei stellt $W(t)$ einen Wiener-Prozess und $Y(t)$ eine standardnormalverteilte Zufallsvariable dar.

³Entsprechend dieser Martingaleigenschaft ergibt sich der heutige Wert eines Risikofaktors als der abgezinste zukünftige Erwartungswert.

⁴Itô Kiyoshi (*7. September 1915) gilt als Schöpfer der heutigen Stochastischen Analysis. Mit seinen Arbeiten „On stochastic processes (Infinitely divisible laws of probability)“ (1942) sowie „On a stochastic integral equation“ (1946) legte er wichtige Grundsteine für die Theorie der stochastischen Differentialgleichungen und stochastischen Integration (vergleiche etwa [IWFK96]).

In der stochastischen Differentialgleichung von Itô (3.13) werden die infinitesimalen Änderungen von $X(t)$ beschrieben. Dafür wird neben einem deterministischen Teil auch eine stochastische Komponente, die durch den Wiener-Prozess $W(t)$ charakterisiert ist, verwendet. Die Parameter $a(X, t)$, auch *Drift* genannt, und $b(X, t)$, die *Volatilität*, können sowohl von der Zeit t als auch von der stochastischen Variable $X(t)$ abhängen.

Je nach Anwendung wird der durch (3.13) beschriebene Prozess $X(t)$ durch die Wahl der Parameter $a(X, t)$ und $b(X, t)$ spezifiziert. So ist auch der reine Wiener-Prozess $dW(t)$ ein Itô-Prozess, bei dem die Parameter konstant gewählt sind:

$$a(X, t) = 0 \quad \text{und} \quad b(X, t) = 1 \quad (3.14)$$

Für die Modellierung der Kursänderungen ist mit (3.12) ein weiterer Itô-Prozess gegeben. Die Parameter sind dabei definiert als

$$a(X, t) = \mu \quad \text{und} \quad b(X, t) = \sigma. \quad (3.15)$$

Bevor weitere spezielle stochastische Prozesse vorgestellt werden (Kapitel 3.4), die auf der allgemeinen stochastischen Differentialgleichung von Itô (3.13) basieren, werden nun zunächst die Eigenschaften der Itô-Gleichung näher betrachtet.

Satz 3.3.2 [ERWARTUNGSWERT UND VARIANZ DER ITÔ-GLEICHUNG]

Als Erwartungswert und Varianz der stochastischen Differentialgleichung von Itô ergeben sich

$$\begin{aligned} E[dX(t)] &= a(X, t)dt & (3.16) \\ \text{var}[dX(t)] &\stackrel{(*)}{=} E[(dX(t) - E[dX(t)])^2] \stackrel{(3.16)}{=} E[(b(X, t)dW(t))^2] \stackrel{(3.10)}{=} b(X, t)^2 dt \end{aligned}$$

(*): Definition der Varianz (Definition A.0.9)

Das folgende *Lemma von Itô* beschreibt, inwiefern ein auf einem Itô-Prozess $X(t)$ beruhender Prozess (gegeben durch eine beliebige Funktion $f(X, t)$) durch die stochastische Änderung von $X(t)$ beeinflusst wird.

Satz 3.3.3 [ITÔ-LEMMA]

Gegeben sei ein Itô-Prozess $X(t)$, der entsprechend Definition 3.3.1 der stochastischen Differentialgleichung

$$dX(t) = a(X, t)dt + b(X, t)dW(t)$$

genügt. $f(X, t)$ sei eine Funktion mit stetigen partiellen Ableitungstermen $\frac{\partial f}{\partial X}$, $\frac{\partial^2 f}{\partial X^2}$ und $\frac{\partial f}{\partial t}$. Dann folgt auch $f(X, t)$ einer stochastischen Differentialgleichung von Itô mit dem gleichen Wiener-Prozess $W(t)$

$$df(X, t) = \underbrace{\left[\frac{\partial f}{\partial X} a(X, t) + \frac{\partial f}{\partial t} + \frac{1}{2} \frac{\partial^2 f}{\partial X^2} b(X, t)^2 \right]}_{\text{Drift-Rate}} dt + \underbrace{\frac{\partial f}{\partial X} b(X, t)}_{\text{Volatilität}} dW(t) \quad (3.17)$$

mit $f = f(X, t)$ auf der rechten Seite.

Beweis: (Grobe Skizzierung nach [Deu04])

Mittels Taylorentwicklung ergibt sich

$$\begin{aligned} df(X, t) &= \frac{\partial f}{\partial X} dX(t) + \frac{\partial f}{\partial t} dt + \frac{1}{2} \frac{\partial^2 f}{\partial X^2} (dX(t))^2 + \frac{1}{2} \frac{\partial^2 f}{\partial t^2} (dt)^2 + \dots \\ &= \frac{\partial f}{\partial X} \underbrace{[a(X, t)dt + b(X, t)dW(t)]}_{dX(t)} + \frac{\partial f}{\partial t} dt \\ &\quad + \frac{1}{2} \frac{\partial^2 f}{\partial X^2} [2a(X, t)b(X, t)dW(t) + b(X, t)^2 dW(t)^2] + O(dt^2). \end{aligned}$$

Da nach (3.11) $dW(t) \sim \sqrt{dt}$ gilt, sind die nicht-linearen $dW(t)^2$ -Terme verglichen mit dt nicht vernachlässigbar klein – vielmehr kann

$$dW(t)^2 \approx dt \tag{3.18}$$

gesetzt werden, denn es gilt

$$E[dW(t)^2] = dt \quad \text{sowie} \quad \text{var}[dW(t)^2] \sim dt^2.$$

Der Varianzterm verschwindet also in linearer Ordnung in dt . $dW(t)^2$ ist demnach in linearer Ordnung nicht mehr stochastisch, sondern besteht nur noch aus seinem Erwartungswert dt . Unter Verwendung der Approximation (3.18) und entsprechender Vernachlässigung aller Terme, die kleiner sind als dt (unter der Beachtung, dass $dW(t) \sim \sqrt{dt}$), wird (3.17) erhalten. \square

3.3.1. Das Kursmodell nach dem Itô-Lemma

Wie bereits in (3.15) erwähnt, handelt es sich bei dem in (3.12) gegebenem Modell für die logarithmischen Kursänderungen

$$d \ln(S(t)) = \mu dt + \sigma dW(t) \tag{3.19}$$

um einen Itô-Prozess mit den Koeffizienten

$$a(S, t) = \mu \quad \text{und} \quad b(S, t) = \sigma. \tag{3.20}$$

Mittels des Itô-Lemmas lässt sich daher auf den eigentlichen Kursprozess $S(t)$ schließen. Mit $y(t) = \ln(S(t))$ und der Funktion $f(y, t) = e^{y(t)}$ gilt für die partiellen Ableitungen

$$\frac{\partial f}{\partial y} = \frac{\partial^2 f}{\partial y^2} = f \quad \text{und} \quad \frac{\partial f}{\partial t} = 0 \quad \text{für} \quad f = f(y, t).$$

Aus dem Itô-Lemma resultiert damit

$$df(y, t) = \left(f(y, t)\mu + \frac{1}{2} f(y, t)\sigma^2 \right) dt + f(y, t)\sigma dW(t).$$

Da $f(y, t) = S(t)$ ist, gilt für die absoluten Kursänderungen

$$dS(t) = S(t) \left(\mu + \frac{1}{2} \sigma^2 \right) dt + S(t)\sigma dW(t). \tag{3.21}$$

Satz 3.3.4 [LÖSUNG DES KURSPROZESSES (3.21) FÜR ENDLICHE ZEITZUWÄCHSE]
 Δt bezeichne einen endlichen Zeitzuwachs. Dann ist die Lösung der stochastischen Differentialgleichung in (3.21) gegeben durch die Funktion

$$S(t + \Delta t) = S(t) \cdot \exp(\mu\Delta t + \sigma\Delta W(t)) \quad (3.22)$$

mit $\Delta W(t) \sim N(0, \Delta t)$ $\left(\Delta W(t) \sim Y(t)\sqrt{\Delta t} \text{ für } Y(t) \sim N(0, 1) \right)$

Mit (3.22) werden demnach die prozentualen Kursänderungen von $S(t)$ über endliche Zeitzuwächse Δt definiert:

$$\frac{S(t + \Delta t)}{S(t)} = \exp(\mu\Delta t + \sigma\Delta W(t))$$

Beweis: (In Anlehnung an [Deu04])

Es sei $y(t) = W(t)$ ein beliebiger Wiener-Prozess und

$$S(y, t) = S_0 \cdot \exp(\mu t + \sigma y(t))$$

eine Funktion der stochastischen Variable $y(t)$ mit beliebigem Faktor S_0 . Für Wiener-Prozesse gilt entsprechend (3.14), dass $a(y, t) = 0$ und $b(y, t) = 1$. Daher ergibt sich aus dem Itô-Lemma für die Funktion $S(y, t)$:

$$\begin{aligned} dS(y, t) &= \left[\underbrace{\frac{\partial S}{\partial y}}_{\sigma S(y,t)} \underbrace{a(y, t)}_0 + \underbrace{\frac{\partial S}{\partial t}}_{\mu S(y,t)} + \frac{1}{2} \underbrace{\frac{\partial^2 S}{\partial y^2}}_{\sigma^2 S(y,t)} \underbrace{b(y, t)^2}_1 \right] dt + \underbrace{\frac{\partial S}{\partial y}}_{\sigma S(y,t)} \underbrace{b(y, t)}_1 dW(t) \\ &= S(y, t) \left(\mu + \frac{1}{2} \sigma^2 \right) dt + S(y, t) \sigma dW(t) \end{aligned}$$

Ersetzen von t durch $t + \Delta t$ führt zu

$$\begin{aligned} S(y, t + \Delta t) &= S_0 \cdot \exp(\mu(t + \Delta t) + \sigma y(t + \Delta t)) \\ &\stackrel{S1}{=} S_0 \cdot \exp(\mu\Delta t + \sigma W(t + \Delta t)) \\ &= S_0 \cdot \exp \left(\mu\Delta t + \sigma W(t) + \underbrace{\sigma(W(t + \Delta t) - W(t))}_{=\Delta W(t) \sim N(0, \Delta t)} \right) \\ &\stackrel{S2}{=} S_0 \cdot \exp(\mu\Delta t + \sigma\Delta W(t)) \end{aligned}$$

Dabei wird in den Schritten $S1$ und $S2$ angenommen, dass die sich lediglich auf die Zeit t beziehenden Faktoren $\exp(\mu t)$ und $\exp(\sigma W(t))$ vom (bisher beliebigen) Faktor S_0 absorbiert werden. Anschließend wird S_0 so gewählt, dass $S(y, t + \Delta t) \xrightarrow{\Delta t \rightarrow 0} S(y, t)$. Das entspricht der Anfangsbedingung für die Lösung der stochastischen Differentialgleichung. Damit ergibt sich (3.22) als Lösung des Kursprozesses für endliche Zeitzuwächse Δt . \square

Lemma 3.3.5 [LOGNORMALVERTEILUNG DER KURSÄNDERUNGEN]

Die logarithmischen Kursänderungen sind entsprechend (3.19) normalverteilt mit

$$\ln \left(\frac{S(t + dt)}{S(t)} \right) = d \ln(S(t)) \sim N(\mu dt, \sigma^2 dt).$$

Deshalb unterliegen die prozentualen Kursänderungen $\frac{S(t+dt)}{S(t)}$ beziehungsweise die absoluten Kursänderungen $S(t)$ einer Lognormalverteilung.

Bemerkung 3.3.6 [DIE MITTLERE RENDITE UNTER STETIGER VERZINSUNG]

Gehen wir von stetiger Verzinsung aus, so ist die Rendite R über einen endlichen Zeitraum Δt gegeben durch $S(t + \Delta t) = S(t)e^{R\Delta t}$ beziehungsweise

$$R = \frac{1}{\Delta t} \ln \left(\frac{S(t + \Delta t)}{S(t)} \right) \stackrel{(3.22)}{=} \frac{1}{\Delta t} \ln \left(\frac{S(t) \cdot \exp(\mu\Delta t + \sigma\Delta W(t))}{S(t)} \right) = \mu + \frac{1}{\Delta t} \sigma \Delta W(t).$$

Die mittlere Rendite entspricht damit gerade

$$\mathbb{E}[R] = \mathbb{E} \left[\mu + \frac{1}{\Delta t} \sigma \Delta W(t) \right] = \mu,$$

da $\mathbb{E}[\Delta W(t)] = 0$ (vergleiche (3.10)).

3.4. Spezielle stochastische Prozesse

Die Erläuterungen in diesem Kapitel orientieren sich an [Nef96]. Die Variablen μ , σ sowie b und \mathbf{m} werden als konstant angenommen. Für die Abbildungen wird auf Diskretisierungen mittels des Euler-Maruyama-Schemas zurückgegriffen. Näheres zur Diskretisierung und den darauf beruhenden Simulationsmöglichkeiten findet sich in den Kapiteln 5 und 9.

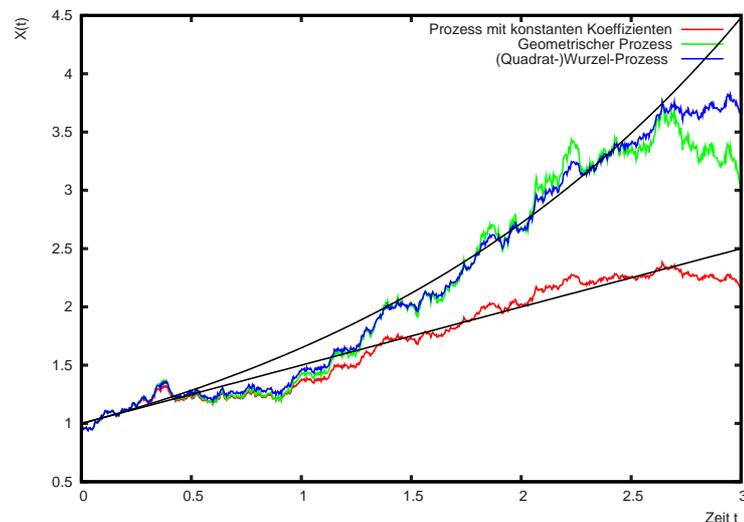


Abbildung 3.2.: Dargestellt sind mögliche Pfade der vorgestellten Prozessstypen mit Anfangswert $X(0) = 1.0$, den Parametern $\mu = 0.5$, $\sigma = 0.3$ und dem diskretisierten konstanten Zeitzuwachs $\Delta t = 0.003$ (Diskretisierung entsprechend des Euler-Maruyama-Schemas). Alle drei Pfade basieren hinsichtlich ihres stochastischen Anteils auf dem gleichen Wiener-Prozess.

Prozesse, die auf SDGL mit konstanten Koeffizienten beruhen

Beim einfachsten Fall stochastischer Differentialgleichungen sind Drift- und Diffusionskoeffizient konstant und somit vollkommen unabhängig von den über die Zeit erhaltenen Informationen:

$$a(X, t) = \mu, \quad b(X, t) = \sigma \quad \Longrightarrow \quad dX(t) = \mu dt + \sigma dW(t)$$

Nach Satz 3.3.2 sind Erwartungswert und Varianz stochastischer Differentialgleichungen dieser Art gegeben durch

$$E[dX(t)] = \mu dt \quad \text{und} \quad \text{var}[dX(t)] = \sigma^2 dt. \quad (3.23)$$

Der in (3.19) beschriebene Prozess für die logarithmischen Kursänderungen ist durch eine stochastische Differentialgleichung mit konstanten Koeffizienten gegeben (siehe (3.20)). In Abbildung 3.2 ist unter anderem ein Pfad eines auf einer stochastischen Differentialgleichung mit konstanten Koeffizienten beruhenden Prozesses mit Anfangswert $X(0) = 1.0$, $\mu = 0.5$, $\sigma = 0.3$ und diskretisiertem konstanten Zeitzuwachs $\Delta t = 0.003$ dargestellt. Entsprechend des in (3.23) gegebenen Erwartungswertes seiner Änderungen entwickelt sich dieser Pfad um die Trendgerade $g(t) = \mu \cdot t + X(0)$. Die Schwankungen um diese Trendgerade bleiben aufgrund des konstanten Diffusionsparameters $b(X, t) = \sigma$ über die Zeit t gleich stark.

Der Einsatz dieses Prozesstyps ist also sinnvoll, wenn die zu modellierende Größe im Großen und Ganzen einem konstant linearen Trend folgt und ihre Schwankungen gleichbleibend stark sind.

Prozesse mit geometrischer SDGL

Entgegen einer stochastischen Differentialgleichung mit konstanten Koeffizienten sind Drift- und Diffusionskoeffizient einer geometrischen stochastischen Differentialgleichung von den zur Zeit t verfügbaren Informationen abhängig:

$$a(X, t) = \mu X(t), \quad b(X, t) = \sigma X(t) \quad \Longrightarrow \quad dX(t) = \mu X(t) dt + \sigma X(t) dW(t)$$

Erwartungswert und Standardabweichung sind proportional zu $X(t)$: Es gilt

$$E[dX(t)] = \mu X(t) dt \quad \text{sowie} \quad \text{var}[dX(t)] = \sigma^2 X(t)^2 dt$$

nach Satz 3.3.2. Mit (3.21) sind die absoluten Kursänderungen mittels eines Prozesses mit geometrischer stochastischer Differentialgleichung beschrieben.

Der in Abbildung 3.2 dargestellte Pfad eines geometrischen Prozesses mit Anfangswert $X(0) = 1.0$, $\mu = 0.5$, $\sigma = 0.3$ und diskretisiertem konstanten Zeitzuwachs $\Delta t = 0.003$ folgt einem exponentiellen Trend. Die Schwankungen des Prozesses um die durch $f(t) = \exp(\mu t)$ gegebene Trendfunktion nehmen entsprechend der Proportionalität der Varianz zu $X(t)^2$ mit der Zeit t zu, da auch der Prozess $X(t)$ steigt.

Prozesse geometrischer Art sind aufgrund ihrer proportionalen Eigenschaften zu Standardmodellen für Aktienkurse geworden, da für die meisten Aktienkurse ein exponentieller Trend weitaus realistischer als ein linearer Trend ist und auch der Effekt der zur $X(t)^2$ proportionalen Varianz durchaus als sinnvoll zu erachten ist.

Anhand einer Umformung der stochastischen Differentialgleichung zu

$$\frac{dX(t)}{X(t)} = \mu dt + \sigma dW(t)$$

lässt sich erkennen, dass Drift und Diffusion der relativen Zuwächse bei diesem Prozesstyp nach wie vor konstant sind.

(Quadrat-)Wurzel-Prozesse

Eng verwandt zu den zuletzt diskutierten Prozessen mit geometrischen stochastischen Differentialgleichungen sind die (Quadrat-)Wurzel-Prozesse. Auch bei diesen sind Drift- und Diffusionskoeffizient von den zur Zeit t verfügbaren Informationen abhängig:

$$a(X, t) = \mu X(t), \quad b(X, t) = \sigma \sqrt{X(t)} \quad \implies \quad dX(t) = \mu X(t)dt + \sigma \sqrt{X(t)}dW(t)$$

Für den Erwartungswert und die Varianz gilt nach Satz 3.3.2

$$E[dX(t)] = \mu X(t)dt \quad \text{und} \quad \text{var}[dX(t)] = \sigma^2 X(t)dt.$$

Damit unterscheidet sich lediglich die Varianz von der des Prozesses mit geometrischer stochastischer Differentialgleichung. Die vom Diffusionsterm gesteuerten Schwankungen fallen im Vergleich schwächer aus, da die Varianz hier „nur“ proportional zu $X(t)$ und nicht zu $X(t)^2$ ist.

In Abbildung 3.2 ist auch ein Pfad eines (Quadrat-)Wurzel-Prozesses mit Anfangswert $X(0) = 1.0$, $\mu = 0.5$, $\sigma = 0.3$ und diskretisiertem konstanten Zeitzuwachs $\Delta t = 0.003$ dargestellt. Dieser Pfad folgt wiederum der exponentiellen Trendfunktion $f(t) = \exp(\mu t)$. Die Schwankungen des Pfades um diesen Trend fallen allerdings, verglichen mit denen des Pfades des Prozesses mit geometrischer stochastischer Differentialgleichung, besonders für höhere Werte von $X(t)$ (fortgeschrittener Zeitpunkt t) deutlich schwächer aus.

Mean-Reversion-Prozesse

Bei den Mean-Reversion-Prozessen sind, ebenso wie bei den vorangegangenen Prozessen, der Drift- als auch der Diffusionskoeffizient der zugrunde liegenden stochastischen Differentialgleichung von den zur Zeit t verfügbaren Informationen abhängig. Dabei ist der Driftkoeffizient von ganz spezieller Form:

$$a(X, t) = b(m - X(t)), \quad b(X, t) = \sigma X(t) \quad \implies \quad dX(t) = b(m - X(t))dt + \sigma X(t)dW(t)$$

Für den Erwartungswert und die Varianz gilt nach Satz 3.3.2

$$E[dX(t)] = b(m - X(t))dt \quad \text{und} \quad \text{var}[dX(t)] = \sigma^2 X(t)^2 dt.$$

Als *Mean-Reversion-Effekt* wird die im Driftterm verankerte Eigenschaft bezeichnet, nach der der Prozess genau dann eine positive (negative) Driftrate besitzt, wenn $X(t)$ unter (über) das *Mean-Reversion-Niveau* m rutscht (also $X(t) < m$ ($X(t) > m$)). Der Prozess $X(t)$ wird dann – in Abhängigkeit von der Entfernung ($m - X(t)$) – mit der *Mean-Reversion-Stärke* b zurück zum Mean-Reversion-Niveau m gezogen. Dieser tendenzielle Zug wird allerdings von der Zufallskomponente der stochastischen Differentialgleichung überlagert. So kann zum Beispiel aufgrund eines hohen positiven Diffusionseinflusses trotz negativer Driftrate ein positiver Zuwachs resultieren. Entsprechend des relativen Faktors ($m - X(t)$) wird die Wahrscheinlichkeit, dass der Prozess im nächsten Schritt in Richtung m zurückkehrt, umso größer, je weiter sich der Prozess $X(t)$ von m entfernt hat. Der Zufallsterm muss daher bei größerer Entfernung auch entsprechend größer ausfallen, um die Rückzugstendenz zum Mean-Reversion-Niveau im nächsten Schritt zu unterbinden.

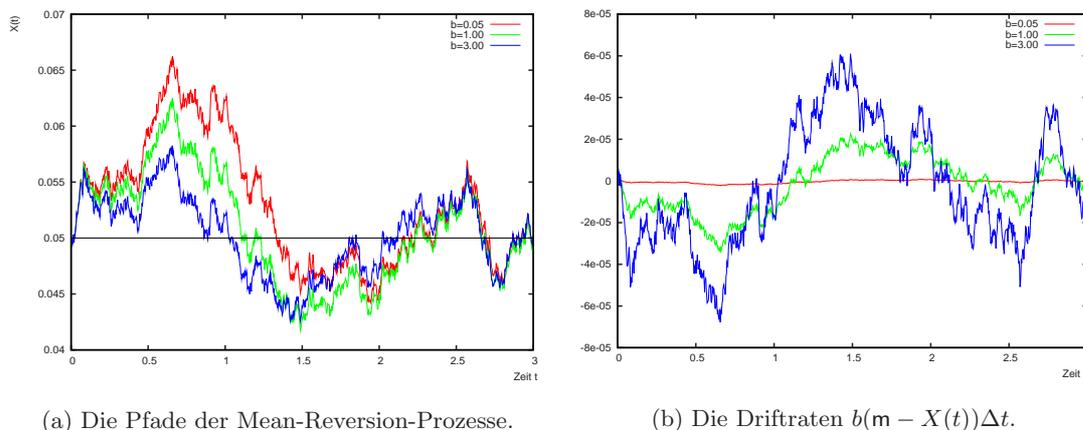


Abbildung 3.3.: Darstellung diskretisierter Pfade (basierend auf denselben Zufallszahlen) dreier Mean-Reversion-Prozesse mit identischem Mean-Reversion-Niveau $m = 0,05$, Diffusionsparameter $\sigma = 0,3$ und Anfangswert $X(0) = 0,05$ auf Grundlage verschiedener Mean-Reversion-Stärken b .

Abbildung 3.3(a) zeigt Pfade dreier Mean-Reversion-Prozesse mit demselben Mean-Reversion-Niveau $m = 0,05$, Diffusionsparameter $\sigma = 0,3$ und Anfangswert $X(0) = 0,05$ und diskretisierten konstanten Zeitzuwachs $\Delta t = 0,003$ – jedoch unterschiedlicher Mean-Reversion-Stärke b . Alle drei Pfade basieren auf dem gleichen Wiener-Prozess. In Abbildung 3.3(b) sind die zugehörigen Driftraten der Zuwächse dargestellt. Zusätzlich ist in Abbildung 3.3(a) das Mean-Reversion-Niveau $m = 0,05$ eingezeichnet, das den langfristigen Trend der Prozesse darstellt. Im Gegensatz zu den anderen vorgestellten Prozessen sind die Abweichungen vom Trend nicht rein zufälliger Natur, sondern sie werden auch vom Driftterm beeinflusst beziehungsweise kontrolliert.

Dem Prozess $X(t)$ ist es sozusagen möglich, „Ausflüge“ von seinem Mean-Reversion-Niveau m zu unternehmen. Wie auch in Abbildung 3.3(a) beobachtbar ist, ist das Ausmaß dieser Exkursionen vom Parameter b , der Mean-Reversion-Stärke, abhängig. Je größer b ist, desto kleiner fallen die Ausflüge im Mittel aus.⁵ Ferner veranschaulicht Abbildung 3.3(b) die Stärke des Rückzugs. Der Driftanteil eines Pfades ist entsprechend des Verhältnisses vom augenblicklichen Wert des Pfades zum Mean-Reversion-Niveau positiv beziehungsweise negativ (vergleiche die beiden Abbildungen). Er fällt umso höher aus, je weiter sich $X(t)$ vom Mean-Reversion-Niveau m entfernt hat. Gleichzeitig fällt er proportional zur Mean-Reversion-Stärke b aus. Vergleiche dazu die Driftraten für die verschiedenen Werte des Parameters b . Zum Beispiel entspricht die Driftrate zu festem Zeitpunkt $t = 1,5$ für $b = 1,0$ in etwa dem Wert $2e^{-5}$, während sie für $b = 3,0$ ungefähr den dreifachen Wert ($6e^{-5}$) annimmt.

Mean-Reversion-Prozesse werden vor allem bei der Modellierung von Zinssätzen eingesetzt, die immer wieder zu einem mittleren Wert zurückpendeln (vergleiche dazu Kapitel 4.3).

⁵Für einen sinnvollen Einsatz des Mean-Reversion-Prozesses darf b nicht zu groß gewählt werden, denn sonst wird $X(t)$ schließlich direkt auf die andere Seite des Mean-Reversion-Niveaus gezogen.

3.5. Die Übergangsdichte eines stochastischen Prozesses

Mit Hilfe der Übergangsdichte $\rho(x)$ lässt sich angeben, mit welcher Wahrscheinlichkeit der stochastische Prozess $X(t)$ zur Zeit t_{i+1} einen Wert zwischen a und b annimmt – unter der Voraussetzung, dass er zur Zeit t_i gleich X war. Bezeichnet $\rho(X', t_{i+1}|X, t_i)$ die Übergangsdichte, die die Wahrscheinlichkeiten beschreibt, mit der der stochastische Prozess unter der Bedingung, dass er zur Zeit t_i genau gleich X war, zur Zeit t_{i+1} genau gleich X' ist, so gilt

$$P(a \leq X(t_{i+1}) \leq b | X(t_i) = X) = \int_a^b \rho(X', t_{i+1}|X, t_i) dX'.$$

Die Übergangsdichte enthält damit sämtliche den Prozess betreffende Informationen.

Kapitel 4.

Stochastische Modellierung von Risikofaktoren an den Finanzmärkten

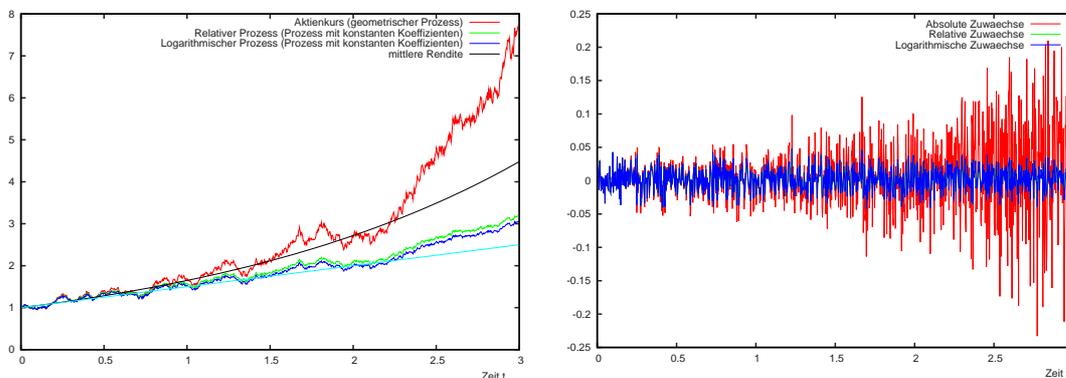
Nachdem im letzten Kapitel stochastische Prozesse vorgestellt wurden, wird nun dargestellt, wie mit ihrer Hilfe das Kurs- (4.1) sowie das Zinsänderungsrisiko (4.2) (vorgestellt in Kapitel 2) beschrieben werden können. Bei Aktienkursen wird dafür standardmäßig auf einen Prozess mit geometrischer stochastischer Differentialgleichung zurückgegriffen (vergleiche Kapitel 3.3.1). In Bezug auf das Zinsänderungsrisiko ist es dagegen notwendig, sich zunächst die zusätzliche Abhängigkeit von der (Rest-)Laufzeit des zugrunde liegenden festverzinslichen Wertpapiers zu verdeutlichen, die neben der Abhängigkeit vom Betrachtungszeitpunkt besteht. Ihren Ausdruck findet diese Abhängigkeit in der Zinsstruktur (grafisch: Zinsstrukturkurve). Da Zinssätze selbst am Markt nicht beobachtbar sind, wird in 4.2.1 beschrieben, inwieweit auf festverzinsliche Wertpapiere in Form von Nullkupon-Anleihen zurückgegriffen werden kann, um die Zinsstruktur zu bestimmen. Die Modellierungsaufgabe verlagert sich damit auf die der Preise von Nullkupon-Anleihen. Dabei ist es möglich, diese Preise durch einen besonderen Zinssatz, die Short Rate, zu beschreiben. Unter der Voraussetzung, dass diese Short Rate durch eine stochastische Differentialgleichung beschrieben wird, können die Preise der Nullkupon-Anleihen als Erwartungswerte der auf der Short Rate basierenden Diskontierungsfaktoren bestimmt werden. Dies lässt sich einerseits über den Ansatz der partiellen Differentialgleichungen andererseits über den Martingal-Ansatz herleiten. Im Rahmen dieser Herleitungen wird auch der Marktpreis des Risikos vorgestellt, der das Bindeglied zwischen realem und risikoneutralem Wahrscheinlichkeitsraum darstellt. Damit ist schließlich spezifiziert, inwieweit die Zinsstruktur mittels eines einzelnen treibenden Faktors – in Form der durch eine stochastische Differentialgleichung gegebenen Short Rate – beschrieben werden kann. Spezielle Modelle für die Short Rate werden in Kapitel 4.3 vorgestellt. Dabei werden das Vasicek- und das CIR-Modell näher betrachtet. Die Analysen hinsichtlich der Parameterschätzung werden sich vor allem auf das CIR-Modell beziehen. An dieser Stelle muss erwähnt werden, dass die Zinsstruktur auch durch mehrere Faktoren (in Form stochastischer Prozesse) angetrieben werden kann. Es wird dann von so genannten Mehr-Faktor-Modellen gesprochen. Diese können zum Beispiel auf dem Prinzip der Modellierung der Short Rate als Summe mehrerer Teilraten basieren. Wie sich herausgestellt hat (vergleiche etwa [Kwo98], [Gla04], [Hau05], [FMW03]), können mit Mehr-Faktor-Modellen die Dynamiken der Zinsstruktur meist besser beschrieben werden (siehe Kapitel 4.4). Solche Mehr-Faktor-Modelle werden in Kapitel 4.4 (besonders im Hinblick auf eine Mehr-Faktor-Verallgemeinerung des Vasicek- sowie CIR-Modells) vorgestellt.

4.1. Das Kursrisiko bei Aktien

Bei der stochastischen Modellierung von Aktienkursen wird standardmäßig auf das in Kapitel 3.3.1 erläuterte Kursmodell zurückgegriffen. Wir fassen noch einmal zusammen:

Kursänderungen	Stochastische Differentialgleichung	Verteilung
absolut	$dS(t) = S(t) (\mu + \frac{1}{2}\sigma^2) dt + S(t)\sigma dW(t)$	$\sim LN (\mu dt, \sigma^2 dt)$
relativ	$\frac{dS(t)}{S(t)} = (\mu + \frac{1}{2}\sigma^2) dt + \sigma dW(t)$	$\sim N ((\mu + \frac{1}{2}\sigma^2) dt, \sigma^2 dt)$
logarithmisch	$d \ln(S(t)) = \mu dt + \sigma dW(t)$	$\sim N (\mu dt, \sigma^2 dt)$
Lösung für endliche Zeitzuwächse Δt	$S(t + \Delta t) = S(t) \cdot e^{\mu\Delta t + \sigma\Delta W(t)}$ mit $\underbrace{\Delta W(t)}_{=W(t+\Delta t)-W(t)} \sim N(0, \Delta t)$	

Unter Beachtung der stochastischen Differentialgleichung für die absoluten Kursänderungen lässt sich erkennen, dass der Aktienkurs $S(t)$ mittels eines Prozesses mit geometrischer stochastischer Differentialgleichung¹ beschrieben wird. Die entsprechenden stochastischen Differentialgleichungen für die relativen und logarithmischen Kursänderungen hingegen definieren Prozesse, denen eine stochastische Differentialgleichung mit konstanten Koeffizienten zugrundeliegt.



(a) Die Pfade der Aktienprozesse.

(b) Die Zuwächse der Pfade.

Abbildung 4.1.: Darstellung diskretisierter Pfade (basierend auf denselben Zufallszahlen) der verschiedenen stochastischen Differentialgleichungen mit identischen Parameter $\mu = 0.5$ $\sigma = 0.3$ und Anfangswert 1.

Abbildung 4.1(a) veranschaulicht für jeden der drei Prozesse einen Pfad, der den anderen beiden in Bezug auf den Anfangswert (vom Wert 1), die Parameter ($\mu = 0.5$, $\sigma = 0.3$) sowie dem zugrunde liegenden Wiener-Prozess entspricht. Außerdem ist die erwartete (mittlere) Rendite eingezeichnet. Die Schwankungen lassen den stochastischen Einfluß erkennen. In Abbildung 4.1(b) sind die zugehörigen Zuwächse der Pfade dargestellt. Dabei ist

¹Für die Eigenschaften eines solchen Prozesses vergleiche Kapitel 3.4.

(im Vergleich mit Abbildung 4.1(a)) deutlich zu erkennen, dass beim geometrischen Prozess, der den eigentlichen Aktienkurs beschreibt, die Schwankungsbreite der Zuwächse zusammen mit dem Level des Aktienkurses größer wird. Die Schwankungen der Prozesse mit den konstanten Koeffizienten hingegen bleiben gleichmäßig.

Die logarithmischen und die relativen Kursänderungen sind entsprechend ihren stochastischen Differentialgleichungen normalverteilt. Die absoluten Kursänderungen hingegen sind nach Bemerkung 3.3.5 lognormalverteilt.

Bei der Analyse von Aktienkursen wird in der Regel mit den logarithmischen Kursänderungen gearbeitet. Sie lassen sich als Logarithmus des Verhältnisses zwischen neuem und altem Kurs bestimmen:

$$\ln\left(\frac{S(t+\Delta t)}{S(t)}\right) = \ln(S(t+\Delta t)) - \ln(S(t)) = d\ln(S(t)).$$

Die Verwendung dieser logarithmischen Größe hat unter anderem den Vorteil, dass die Bestimmung der Modellparameter auf Grundlage historischer Daten sehr einfach ist. So lassen sich die Modellparameter μ und σ als historische mittlere Rendite (vergleiche Bemerkung 3.3.6) und historische Volatilität dieser Größe abschätzen. Doch dazu später, in Kapitel 8 zur Parameterschätzung, mehr.

Um anschließend entsprechende Rückschlüsse auf die absoluten Kursänderungen zu ziehen, ist es notwendig, eine Brücke zwischen logarithmischen und absoluten beziehungsweise logarithmischen und relativen Kursänderungen zu schlagen. In Abbildung 4.1 konnten wir bereits erahnen, dass die logarithmischen und die relativen Kursänderungen stets sehr nahe beieinander liegen. In der Tat gilt:

Satz 4.1.1

In linearer Näherung ist

$$\Delta \ln(S(t)) \approx \frac{\Delta S(t)}{S(t)} \quad \Rightarrow \quad \Delta S(t) \approx S(t) \Delta \ln(S(t)).$$

Für Zeitwüchse $\Delta t \rightarrow 0$ entsprechen die logarithmischen Kursänderungen damit den relativen Kursänderungen.

Beweis:

$$\begin{aligned} \Delta \ln(S(t)) &= \ln(S(t+\Delta t)) - \ln(S(t)) = \ln\left(\frac{S(t+\Delta t)}{S(t)}\right) \\ &= \ln\left(\frac{S(t) + (S(t+\Delta t) - S(t))}{S(t)}\right) = \ln\left(1 + \frac{\Delta S(t)}{S(t)}\right) \\ &\stackrel{2}{=} \frac{\Delta S(t)}{S(t)} - \frac{1}{2} \left(\frac{\Delta S(t)}{S(t)}\right)^2 + \dots \approx \frac{\Delta S(t)}{S(t)} \end{aligned}$$

□

²Logarithmusentwicklung: $\ln(1+x) = x - \frac{x^2}{2} + \frac{x^3}{3} - \frac{x^4}{4} + \dots$.

4.2. Das Zinsänderungsrisiko festverzinslicher Wertpapiere

Werden die Vorgänge am Finanzmarkt näher betrachtet, um einen Eindruck der zeitlichen Abhängigkeiten von Zinsen zu erhalten, so wird schnell offensichtlich, dass ein Anleger in der Regel einen höheren Zinssatz für sein Geld erhält, wenn er es für eine längere Laufzeit anlegt. Der Zinssatz steigt also mit der Laufzeit der zugrunde liegenden Anlage.

Definition 4.2.1 [(REST-)LAUFZEIT, GEAMTLAUFZEIT]

Die (Rest-)Laufzeit $\tau = T - t$ einer Anlage wird durch den Betrachtungszeitpunkt t sowie das Fälligkeitsdatum T der Anlage bestimmt. Ist $t = t_0$, so handelt es sich um die Gesamtlaufzeit einer Anlage.

Im Unterschied zu Aktienkursen, die in ihrer Höhe lediglich vom Betrachtungszeitpunkt t abhängen (vergleiche Kapitel 4.1), besteht also bei Zinssätzen eine zusätzliche zeitliche Abhängigkeit von der (Rest-)Laufzeit des jeweils zugrunde liegenden festverzinslichen Wertpapiers. Diese Abhängigkeit findet ihren Ausdruck in der Zinsstruktur. Um die Zinsstruktur definieren zu können, ist es notwendig, vorher die Begriffe Kassazins (engl. Spot Rate) und Terminzins (engl. Forward Rate) einzuführen.

Definition 4.2.2 [KASSAZINS (ENGL. SPOT RATE)]

Unter dem Kassazins (engl. Spot Rate) $R(t, T)$ wird der Zinssatz verstanden, welcher mit aktuell am Markt vorhandenen Anleihen über eine gewisse Laufzeit von heute an risikolos abgesichert werden kann. Diese Anleihen sind in der Regel für verschiedene Laufzeiten unterschiedlich.

Definition 4.2.3 [TERMINZINS (ENGL. FORWARD RATE)]

Das Gegenstück zum Kassazins ist der Terminzins (engl. Forward Rate) $f(t, T)$. Dabei handelt es sich um den Zinssatz, dessen Laufzeit nicht unmittelbar (wie es beim Kassazins der Fall ist) – sondern erst ab einem bestimmten Zeitpunkt in der Zukunft – beginnt.

Bemerkung 4.2.4 [VERHÄLTNISS VON KASSAZINS UND TERMINZINS]

Ausgehend vom Prinzip der Arbitragefreiheit³ lässt sich der Terminzins aus den Kassazinsen zu verschiedenen Laufzeiten eindeutig bestimmen.

Zur Verdeutlichung können zwei Anleger betrachtet werden, die beide planen, ihr Geld für die nächsten zwölf Jahre zu investieren. Der erste Anleger möchte sein Geld zu einem bestimmten Zinssatz (Kassazins) für den gesamten Zeitraum bereits heute fest anlegen. Der andere Anleger hingegen plant, sein Geld erst für fünf Jahre festanzulegen (Kassazins) und für die noch ausstehenden sieben Jahre bereits heute mit der Bank einen Zinssatz (Terminzins) zu vereinbaren, zu welchem er das Geld nach den ersten fünf Jahren anlegen kann. Entsprechend des Prinzips der Arbitragefreiheit müssen beide Anleger unabhängig von ihrer Anlagestrategie denselben Ertrag erhalten. Der Terminzins kann demzufolge aus den beiden Kassazinsen abgeleitet werden.

Mit Hilfe der Kassazinsen ist es jetzt möglich, eine formale Definition der die zeitlichen Abhängigkeiten der Zinsen reflektierenden Zinsstruktur anzugeben:

Definition 4.2.5 [ZINSSTRUKTUR (ENGL. TERM STRUCTURE)]

Bei der Zinsstruktur (engl. Term structure) handelt es sich um eine Abfolge von Kassazinsen verschiedener Laufzeiten, die ab dem aktuellen Zeitpunkt t über ihre Laufzeit $\tau = T - t$ bis zu ihrem Fälligkeitsdatum T Gültigkeit besitzen.

³Nach dem Prinzip der Arbitragefreiheit ist die Möglichkeit risikoloser Gewinne etwa durch den gleichzeitigen Kauf und Verkauf von Wertpapieren ausgeschlossen.

Damit liegt der Zinsstruktur zum jeweiligen Zeitpunkt t eine Abbildung der Art

$$\varphi_t : [0, \hat{\tau}] \rightarrow \mathbb{R} \quad \text{mit} \quad \tau \rightarrow \varphi_t(\tau) = R(t, t + \tau) \quad (4.1)$$

zugrunde, wobei $\hat{\tau}$ eine obere Grenze für die (Rest-)Laufzeit angibt.

Definition 4.2.6 [ZINSSTRUKTURKURVEN (ENGL. TERM STRUCTURE CURVES)]

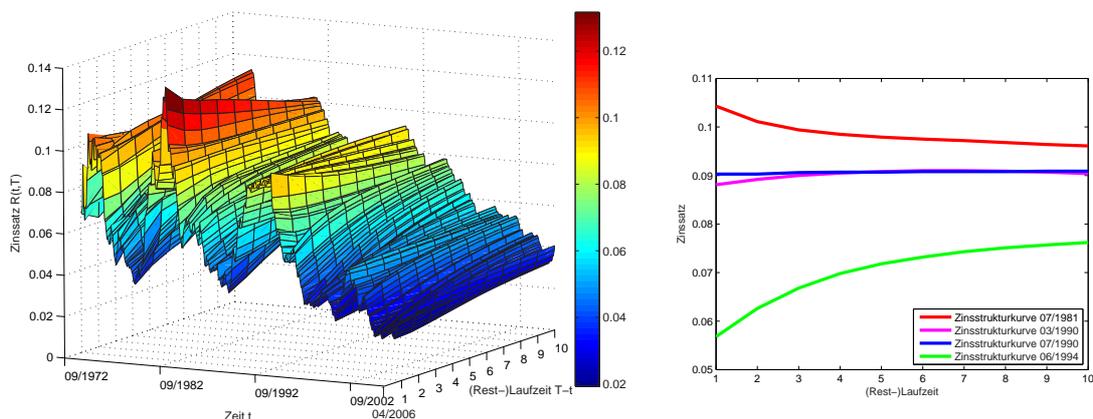
Ihren graphischen Ausdruck findet die Zinsstruktur in den Zinsstrukturkurven. Diese sind abhängig von der Laufzeit und können verschiedene Ausformungen haben.

Rechnet der Markt etwa mit höheren Zinsen in der Zukunft, so werden auch für längere Laufzeiten höhere Zinsen erwartet. Dieser Effekt wird durch die Auszahlung von so genannten Liquiditäts- und Risikoprämien für längerfristige Anlagen unterstützt. Es kommt zu einer steigenden Form der Zinsstrukturkurve. Da dies dem Regelfall entspricht, wird die zugehörige steigende Form auch als normal bezeichnet. Sollte der Markt allerdings für die Zukunft stark fallende Zinsen erwarten (etwa durch Senkungen der Leitzinsen), so entsteht eine größere Nachfrage an längerfristigeren Anlagen. Diese Nachfrage wiederum drückt die Zinssätze an diesem so genannten langen Ende der Zinsstrukturkurve. Die Zinsstrukturkurve ist schließlich fallend und wird auch als invers bezeichnet. Hält die Zinsstrukturkurve einen konstanten Wert, so wird von einer flachen Zinsstrukturkurve gesprochen. Die Höhe der Zinsen sind dann unabhängig von der Laufzeit der Anlage. Darüber hinaus gibt es auch unregelmäßige Zinsstrukturkurven. Unter ihnen ist die buckelige die häufigste.

Abbildung 4.2(b) zeigt die wichtigsten Ausformungen von Zinsstrukturkurven. In Abbildung 4.2(a) ist die gesamte Zinsstruktur der letzten 33 Jahre und 5 Monate abgebildet. Die Zinsstrukturkurven der einzelnen Zeitpunkte t entsprechen in dieser Abbildung den Querschnitten parallel zur Laufzeitachse.

Bemerkung 4.2.7 [ZINSSTRUKTURMODELLE (ENGL. TERM STRUCTURE MODELS)]

Da sich der Zinssatz für jede (Rest-)Laufzeit τ als Zufallsvariable in Abhängigkeit von der Zeit t auffassen lässt, besteht die Zinsstruktur im stochastischen Sinne aus einer unendlichen Anzahl von Zufallsvariablen. Wie allerdings auch in Abbildung 4.2(a) zu erkennen ist, besteht eine starke Abhängigkeit zwischen diesen parallel zur Zeitachse angeordneten Zufallsvariablen. Auf dieser Beobachtung fußen die so genannten Zinsstrukturmodelle, mit denen die stochastische Veränderung der Zinsstruktur mit Hilfe möglichst weniger Faktoren (Zufallsvariablen) modelliert werden soll. In der Praxis genügen aufgrund der starken Korrelation sogar nur zwei bis drei stochastische Prozesse als treibende Faktoren der Zinsstrukturentwicklung, die im Idealfall gänzlich voneinander unabhängig (demnach auch unkorreliert) sind. ([]) Das bedeutet, dass die mittels dieser Faktoren modellierten Zinssätze bezüglich ihrer Laufzeiten weit genug auseinander liegen. Oft werden sogar Modelle verwendet, die lediglich von einem einzigen stochastischen Prozess ausgehen. Im Sprachgebrauch wird je nach Anzahl der eingesetzten Faktoren zwischen Ein-Faktor-Modellen und Mehr-Faktor-Modellen unterschieden.



(a) Zinsstruktur von September 1972 bis April 2006. Datenquelle ist die Zeitreihendatenbank der Deutschen Bundesbank [Zei]: Es handelt sich um nach der Svensson-Methode bestimmte Zinsstrukturkurven entsprechend der Monatsendstände der börsennotierten Bundeswertpapiere mit (Rest-)Laufzeiten $\tau = T - t$ bis zu 10 Jahren.

(b) Dargestellt sind vier der Zinsstruktur entnommene Zinsstrukturkurven zur Visualisierung der möglicher Ausformungen (grün - steigend, rot - fallend, blau - flach, pink - buckelig).

Abbildung 4.2.: Die Zinsstruktur von September 1972 bis April 2006 und mögliche Ausformungen der Zinsstrukturkurven.

4.2.1. Die Bestimmung der Zinsstruktur

Zur Bestimmung der Zinsstruktur wird auf die Beobachtung festverzinslicher Wertpapiere (siehe Kapitel 2.2) am Finanzmarkt zurückgegriffen, denn Zinssätze selbst stellen keine am Markt handelbaren Objekte dar. Mit Hilfe von Nullkupon-Anleihen (Kapitel 2.2, Definition 2.2.4) lässt sich, wie im Folgenden dargestellt, die Zinsstruktur vollständig ermitteln.

Bemerkung 4.2.8 [SPEZIFIZIERUNG DER BETRACHTETEN NULLKUPON-ANLEIHEN]

Ab jetzt wird von Nullkupon-Anleihen ausgegangen, die zu ihrem abgezinsten Nominalwert emittiert werden. Entsprechend dieser Annahme ist die Rückzahlung des Nominalwerts ebenso wie bei Kupon-Anleihen am Ende der Laufzeit verbrieft. Der Nominalwert wird als auf 1 normiert angenommen.

Replikationen mittels Nullkupon-Anleihen

Einleitend wird die Replikation von Kupon-Anleihen mittels Nullkupon-Anleihen erläutert. Dabei wird – analog zu Bemerkung 4.2.8 – jeweils von auf den Wert 1 normierten Nominalwerten der Anleihen ausgegangen. In diesem Zusammenhang wird der Preisbegriff für Kupon- und Nullkupon-Anleihen abgesteckt, mit dem im Anschluss gearbeitet wird. Anschließend wird die Möglichkeit der Replikation diskreter Terminzinsen (siehe Definition 4.2.3) durch Nullkupon-Anleihen aufgewiesen. Aufgrund dieser Replikationsmöglichkeit wird offensichtlich, inwieweit die gesamte Zinsstruktur durch die Beobachtung von Nullkupon-Anleihen bestimmt werden kann.

Bemerkung 4.2.9 [REPLIKATION EINER KUPON-ANLEIHE MIT NOMINALWERT 1(1)]

Der Zahlungstrom einer Kupon-Anleihe mit n Kupon-Auszahlungen der Höhe k_i zu den Zeitpunkten T_i für $1 \leq i \leq n$ und Rückzahlung des Nominalwerts 1 zum Fälligkeitsdatum

$T = T_n$ kann durch ein Portfolio von n Nullkupon-Anleihen mit entsprechenden Fälligkeitsdaten T_i für $i \in \{1, 2, \dots, n\}$ und Rückzahlungsbeträgen (entspricht der Annahme nach den Nominalwerten (Bemerkung 4.2.8)) der Höhe 1 repliziert werden .

Definition 4.2.10 [DIE PREISE VON KUPON- UND NULLKUPON-ANLEIHEN]

Der Preis einer Anleihe entspricht der Prämie, die der Inhaber der Anleihe, der Gläubiger, dem Schuldner zu Beginn zahlen muss, damit es sich um ein faires Geschäft für beide handelt. Der Preis einer Kupon-Anleihe mit Fälligkeitsdatum T sei zum Zeitpunkt $t < T$ im Folgenden beschrieben durch $P^k(t, T)$, während der entsprechende Preis einer Nullkupon-Anleihe durch $P(t, T)$ gegeben sei. Insbesondere ist zu beachten, dass der faire Preis einer sofort fälligen Nullkupon-Anleihe mit Nominalwert 1 gerade $P(t, t) = 1$ sein muss.

Bemerkung 4.2.11 [REPLIKATION EINER KUPON-ANLEIHE MIT NOMINALWERT 1(2)]

Bezeichnet $P^k(t, T)$ nun den Preis der in Bemerkung 4.2.9 eingeführten Kupon-Anleihe und $P(t, T_i)$ den Preis der jeweiligen Nullkupon-Anleihe, so ergibt sich

$$P^k(t, T) = k_1 \cdot P(t, T_1) + k_2 \cdot P(t, T_2) + \dots + (1 + k_n) \cdot P(t, T_n)$$

als Replikationsformel für die Preise der Kupon-Anleihe. Abbildung 4.3 veranschaulicht die Replikation und die in ihrem Rahmen stattfindenden Auszahlungen.

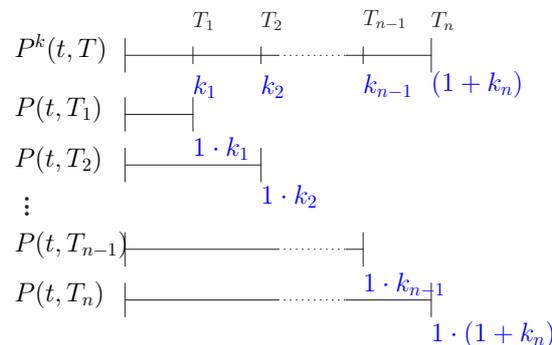


Abbildung 4.3.: Replikation einer Kupon-Anleihe mit Nominalwert 1 mittels Nullkupon-Anleihen mit Nominalwert 1.

Bemerkung 4.2.12 [REPLIKATION EINES DISKRETEN TERMINZINSES DURCH NULLKUPON-ANLEIHEN]

Falls ein Anleger bereits heute, zur Zeit t , die Investition eines gewissen Betrages in einem zukünftigen (diskreten) Zeitintervall $[T, T']$ (mit $T' = T + \Delta T$ und $\Delta T > 0$ sowie $T > t$) plant, bedient er sich eines fest vereinbarten Terminzinses entsprechend Definition 4.2.3. Aufgrund des Prinzips der Arbitragefreiheit ist es möglich, einen solchen (diskreten) Terminzins durch die Emission und den Kauf bestimmter Nullkupon-Anleihen zum heutigen Zeitpunkt t nachzubilden: Dazu wird heute, zum Zeitpunkt t , eine Nullkupon-Anleihe mit Fälligkeitsdatum T emittiert. Für den erhaltenen Preis $P(t, T)$ dieser Nullkupon-Anleihe werden $\frac{P(t, T)}{P(t, T')}$ Nullkupon-Anleihen mit Fälligkeitsdatum T' gekauft. Am Fälligkeitsdatum T wird dem Inhaber der ersten Nullkupon-Anleihe der Nominalwert 1 ausgezahlt. Zur Fälligkeit T' wird der Betrag $\frac{P(t, T)}{P(t, T')}$ als Rückzahlung aus den erworbenen Nullkupon-Anleihen mit Fälligkeitsdatum T' erhalten. In Abbildung 4.4 sind die Replikation und die in ihrem Rahmen stattfindenden Auszahlungen visualisiert.

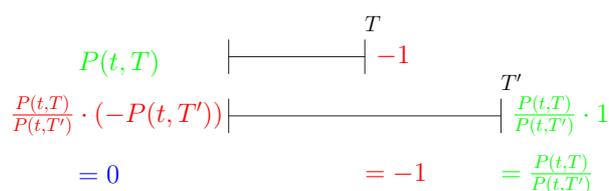


Abbildung 4.4.: Replikation eines Terminzinses mittels Nullkupon-Anleihen mit Nominalwert 1.

Bestimmung der Zinsstruktur über die Preise von Nullkupon-Anleihen

Basierend auf der letzten Replikationsmöglichkeit lässt sich eine formale Definition des diskreten Terminzinses basierend auf den Preisen von Nullkupon-Anleihen angeben:

Definition 4.2.13 [DER DISKRETE TERMINZINS, DER DISKRETE KASSAZINS]

Der diskrete Terminzins $F(t; T, T')$ für den Zeitraum $[T, T']$ ist zur Zeit t bestimmt durch

$$\begin{aligned} 1 + (T' - T)F(t; T, T') &= \frac{P(t, T)}{P(t, T')} \\ \Leftrightarrow F(t; T, T') &= \frac{1}{T' - T} \left(\frac{P(t, T)}{P(t, T')} - 1 \right). \end{aligned} \quad (4.2)$$

Demnach ist der durch $y(t, T) = F(t; t, T)$ definierte diskrete Kassazins gegeben durch

$$y(t, T) = \frac{1}{T - t} \left(\frac{1}{P(t, T)} - 1 \right).$$

Dafür ist zu beachten, dass der faire Preis einer sofort fälligen Nullkupon-Anleihe mit Nominalwert 1 als $P(t, t) = 1$ festgelegt ist (vergleiche Definition 4.2.10).

Definition 4.2.14 [DER STETIGE TERMINZINS, DER STETIGE KASSAZINS]

Der stetige Terminzins wird analog zu Definition 4.2.13 mittels stetiger Verzinsung über

$$\begin{aligned} e^{f(t; T, T')(T' - T)} &= \frac{P(t, T)}{P(t, T')} \\ \Leftrightarrow f(t; T, T') &= -\frac{1}{T' - T} (\ln P(t, T') - \ln P(t, T)) \end{aligned} \quad (4.3)$$

erhalten, so dass für den stetigen Kassazins $R(t, T)$ mit $R(t, T) = f(t; t, T)$

$$R(t, T) = -\frac{1}{T - t} \ln P(t, T) \quad (4.4)$$

gilt. Somit entspricht der stetige Kassazins der Zuwachsrate, unter der die Preise den Wert 1 zum Fälligkeitsdatum T erreichen:

$$P(t, T) \cdot e^{R(t, T)(T - t)} = 1.$$

Folgerung 4.2.15

An dieser Stelle ist mittels (4.4) ersichtlich, dass sich die durch (4.1) definierte Zinsstruktur über die Preisbestimmung von Nullkupon-Anleihen vollständig ermitteln lässt.

Die Preisbestimmung von Nullkupon-Anleihen mittels instantaner Zinssätze, die im risikoneutralen Wahrscheinlichkeitsraum gegeben sind

Werden die stetigen Zinssätze aus Definition 4.2.14 für $T' \rightarrow T$ betrachtet, so ergeben sich der instantane Termin- sowie Kassazins:

Definition 4.2.16 [DER INSTANTANE TERMINZINS, DER INSTANTANE KASSAZINS (ENGL. SHORT RATE)]

Der instantane Terminzins $f(t, T)$ mit Fälligkeitsdatum T ist gegeben durch

$$\begin{aligned} f(t, T) = \lim_{T' \rightarrow T} f(t; T, T') &= \lim_{T' \rightarrow T} \left(-\frac{1}{T' - T} (\ln P(t, T') - \ln P(t, T)) \right) \\ &= -\frac{\partial}{\partial T} \ln P(t, T). \end{aligned} \quad (4.5)$$

Der instantane Kassazins (engl. Short Rate) $r(t)$ ist gegeben durch

$$r(t) (= f(t, t)) = \lim_{T \rightarrow t} R(t, T) = \lim_{T \rightarrow t} \left(-\frac{1}{T - t} \ln P(t, T) \right). \quad (4.6)$$

Damit entspricht der instantane Kassazins dem Kauf einer unmittelbar zum aktuellen Zeitpunkt t fälligen Nullkupon-Anleihe. Hier erklärt sich auch $P(t, t) = 1$ für alle t , denn dies entspricht dem fairen Preis einer unmittelbar erfolgenden Auszahlung vom Wert 1.

Bemerkung 4.2.17

In Anlehnung an die gängige Literatur zu diesem Thema werden auch in dieser Arbeit im Weiteren die aus dem Englischen entlehnten Begriffe Spot Rate für den Kassazins, Short Rate für den instantanen Kassazins und Forward Rate für den Terminzins verwendet.

Bemerkung 4.2.18 [PREISBESTIMMUNG EINER NULLKUPON-ANLEIHE MITTELS DES INSTANTANEN TERMINZINSES]

Aus Gleichung (4.5) ergibt sich durch Integration:

$$\begin{aligned} \int_t^T f(t, s) ds &= - \int_t^T \frac{\partial}{\partial s} \ln P(t, s) ds = -(\ln P(t, T) - \underbrace{\ln P(t, t)}_{=1}) = -\ln P(t, T) \\ \Leftrightarrow P(t, T) &= \exp \left(- \int_t^T f(t, s) ds \right). \end{aligned}$$

Bemerkung 4.2.19 [PREISBESTIMMUNG EINER NULLKUPON-ANLEIHE MITTELS KONSTANT ANGENOMMENER SHORT RATE]

Es sei r der als konstant angenommene Nominalzins, mit dem eine Nullkupon-Anleihe kontinuierlich verzinst sei. Der Rückzahlungsbetrag am Ende der Laufzeit entspreche wiederum dem Wert 1. Dann ist der Preis der Nullkupon-Anleihe zum Zeitpunkt t gegeben durch

$$P(t, T) = e^{-r(T-t)}. \quad (4.7)$$

Dies ist aufgrund des Prinzips der Arbitragefreiheit durch eine mögliche Replizierung der Nullkupon-Anleihe durch eine Investition von $e^{-r(T-t)}$ in eine mit dem konstanten Zinssatz r verzinsten Anlage einzusehen. Denn auch auf diesem Wege würde die Investition entsprechend der konstant kontinuierlicher Verzinsung zum Fälligkeitsdatum T gerade dem Wert 1 ($= e^{-r(T-t)} \cdot e^{r(T-t)}$) entsprechen.⁴ Der Preis (4.7) entspricht somit gerade dem Diskontierungsfaktor konstant kontinuierlicher Verzinsung, mit dem der Wert eines zukünftigen

⁴Unter konstant kontinuierlicher Verzinsung wächst eine Geldeinheit über eine Laufzeit $T - t$ gerade auf den Betrag $e^{r(T-t)}$ an.

(zum Fälligkeitsdatum T vorliegenden) Geldbetrags zum aktuellen Zeitpunkt $t < T$ erhalten wird.

Da die meisten Anleihen, wie etwa die Bundesanleihen, allerdings eine Laufzeit von zehn Jahren oder mehr haben, ist es nicht sinnvoll anzunehmen, dass sich der Nominalzins bis zum Fälligkeitsdatum T nicht ändert und konstant bleiben wird. Daher wird nun der Fall betrachtet, in dem sich der Zinssatz – die Short Rate – deterministisch in Abhängigkeit der Zeit durch eine bekannte Funktion ermitteln lässt.

Bemerkung 4.2.20 [PREISBESTIMMUNG EINER KUPON-ANLEIHE MITTELS NICHTKONSTANTER DETERMINISTISCHER SHORT RATE]

Die Short Rate sei durch eine zeitabhängige Funktion $r(t)$ für $t \in [0, T]$ gegeben. $P^k(t, T)$ bezeichne weiterhin den Preis einer Kupon-Anleihe mit Kuponzahlungsstrom $k(t)$ für $t \in [0, T]$. Zunächst werden Anleihen mit allgemeinem Rückzahlungswert $Z (= P^k(T, T))$ betrachtet. Die Änderung des Preises der Anleihe sei für einen beliebigen Zeitzuwachs dt durch $dP^k(t, T)$ bezeichnet und die Höhe der Kuponauszahlungen während dieser Zeit sei mit $k(t)dt$ gekennzeichnet. Nach dem Prinzip der Arbitragefreiheit muss die Summe dieser beiden Beträge gerade mit dem Ertrag bezüglich der Short Rate im Zeitintervall dt übereinstimmen, das heißt, es muss gelten

$$dP^k(t, T) + k(t)dt = r(t)P^k(t, T)dt \quad \Rightarrow \quad \frac{dP^k(t, T)}{dt} + k(t) = r(t)P^k(t, T). \quad (4.8)$$

Zusammen mit der Endbedingung $P^k(T, T) = Z$ (siehe oben) kann diese lineare gewöhnliche Differentialgleichung erster Ordnung analytisch gelöst werden:

$$P^k(t, T) = e^{-\int_t^T r(s)ds} \left[Z + \int_t^T k(u)e^{\int_u^T r(s)ds} du \right]. \quad (4.9)$$

Satz 4.2.21 [PREISBESTIMMUNG EINER NULLKUPON-ANLEIHE MITTELS NICHTKONSTANTER DETERMINISTISCHER SHORT RATE]

Da bei einer Nullkupon-Anleihe keine Kuponzahlungen existieren, ergibt sich aus (4.9) für den Preis einer Nullkupon-Anleihe

$$P(t, T) = e^{-\int_t^T r(s)ds} Z. \quad (4.10)$$

Folgerung 4.2.22 [DER PREIS EINER NULLKUPON-ANLEIHE MIT NOMINALWERT 1]

Aus (4.10) folgt für den Preis einer Nullkupon-Anleihe mit Nominalwert 1:

$$P(t, T) = e^{-\int_t^T r(s)ds}. \quad (4.11)$$

Bemerkung 4.2.23 [DISKONT- UND PRÄMIEN-ANLEIHE, PULL-TO-PAR-PHÄNOMEN]

In Abhängigkeit von der Short Rate $r(t)$ sowie dem Kuponzahlungsstrom $k(t)$ kann die Preisfunktion über die Zeit t fallen oder steigen. Fällt der Preis einer Anleihe unter den entsprechenden Nominalwert, so wird von einer Diskont-Anleihe gesprochen; ansonsten handelt es sich um eine Prämien-Anleihe. Als Pull-to-par-Phänomen wird der Effekt bezeichnet, der die Eigenschaft beschreibt, dass der Marktpreis einer Anleihe zum Fälligkeitsdatum T stets seinen Nominalwert (engl. *par value*) erreicht.

Die Preisbestimmung von Nullkupon-Anleihen mittels der im realen Wahrscheinlichkeitsraum gegebenen Short Rate

Im letzten Abschnitt wurde davon ausgegangen, dass die instantanen Zinssätze, mit denen die Preise der Nullkupon-Anleihen bestimmbar sind, im risikoneutralen Wahrscheinlich-

keitsraum gegeben sind. Somit stellt sich die Frage, wie die Beziehung zwischen den Preisen der Nullkupon-Anleihen (beziehungsweise der zu bestimmenden Zinsstruktur) und der Short Rate aussieht, wenn diese nicht direkt im risikoneutralen, sondern im realen Wahrscheinlichkeitsraum gegeben ist, und was dies an der Preisformel (4.11) für Nullkupon-Anleihen mit Nominalwert 1 ändert.

Dazu werden zwei Ansätze näher betrachtet (vergleiche [Hau05]): *Der Ansatz der partiellen Differentialgleichungen* (vergleiche auch [Kwo98]), bei dem der Marktpreis des Risikos (in Hinblick auf die sich ergebende Preisformel (4.19)) eine Schlüsselrolle spielt, und *der Martingal-Ansatz*. Beim Martingal-Ansatz wird der Preis der Nullkupon-Anleihen über den Erwartungswert unter dem zugrunde gelegten äquivalenten Martingalmaß erhalten. Dabei hängt das Maß von dem zur Normierung gebrauchten Finanzinstrument ab. Im vorliegenden Fall wird der Diskontierungsfaktor zur Normierung eingesetzt. Damit wird gerade das risikoneutrale Wahrscheinlichkeitsmaß erhalten.

Der Ansatz der partiellen Differentialgleichungen

Die folgenden Betrachtungen finden im realen Wahrscheinlichkeitsraum (Ω, \mathbb{P}) mit zugrunde liegendem realen Wahrscheinlichkeitsmaß \mathbb{P} statt und basieren nach dem Prinzip der Arbitragefreiheit auf einem Markt, auf dem Nullkupon-Anleihen mit beliebiger Laufzeit gehandelt werden. Die Short Rate wird durch einen Itô-Prozess modelliert, der durch die stochastische Differentialgleichung

$$dr(t) = m_{\mathbb{P}}(r, t)dt + v_{\mathbb{P}}(r, t)dW_{\mathbb{P}}(t) \quad (4.12)$$

mit einem unter dem realen Wahrscheinlichkeitsmaß \mathbb{P} gegebenen Wiener-Prozess $W_{\mathbb{P}}(t)$ gegeben ist. Zur Hervorhebung der stetigen Abhängigkeit des Preises einer Nullkupon-Anleihe vom Beobachtungszeitpunkt t , der Short Rate $r(t)$ und dem Fälligkeitsdatum T wird im Folgenden auch die Bezeichnung $P(r, t; T)$ verwendet.

Über die Anwendung des Itô-Lemmas auf den von der Short Rate $r(t)$ stetig abhängigen Preis $P(r, t; T)$ (1.), das anschließende Hedging zweier Nullkupon-Anleihen mit unterschiedlichen Fälligkeitsdaten inklusive Herleitung des Marktpreises des Risikos (2.) und schließlich die Lösung der resultierenden partiellen Differentialgleichung entsprechend Vasicek (1977) (3.) wird eine Preisformel für Nullkupon-Anleihen (4.19) abgeleitet:

1. Eine Anwendung des Itô-Lemmas unter Beachtung von (4.12) ergibt

$$dP(r, t; T) = \left[\frac{\partial P}{\partial r} m_{\mathbb{P}}(r, t) + \frac{\partial P}{\partial t} + \frac{1}{2} \frac{\partial^2 P}{\partial r^2} v_{\mathbb{P}}(r, t)^2 \right] dt + \frac{\partial P}{\partial r} v_{\mathbb{P}}(r, t) dW_{\mathbb{P}}(t). \quad (4.13)$$

Für die relative Änderung des Preises der Nullkupon-Anleihen folgt daher

$$\frac{dP}{P} = \underbrace{\frac{1}{P} \left[\frac{\partial P}{\partial r} m_{\mathbb{P}}(r, t) + \frac{\partial P}{\partial t} + \frac{1}{2} \frac{\partial^2 P}{\partial r^2} v_{\mathbb{P}}(r, t)^2 \right]}_{= \mu(r, t; T)} dt + \underbrace{\frac{1}{P} \frac{\partial P}{\partial r} v_{\mathbb{P}}(r, t)}_{= \sigma(r, t; T)} dW_{\mathbb{P}}(t). \quad (4.14)$$

2. Es sei $t < T_1 < T_2$. Durch den Ankauf einer ersten Nullkupon-Anleihe mit Wert V_1 und Fälligkeitsdatum T_1 und den Verkauf einer zweiten Nullkupon-Anleihe mit Wert V_2 und Fälligkeitsdatum T_2 wird ein Portfolio konstruiert, dessen Wert $\Pi = V_1 - V_2$

entspricht (siehe [Kwo98]). Dann ist

$$\begin{aligned}
 d\Pi &= \underbrace{V_1 \frac{dP(r, t; T_1)}{P(r, t; T_1)}}_{=dV_1} - \underbrace{V_2 \frac{dP(r, t; T_2)}{P(r, t; T_2)}}_{=dV_2} \\
 &\stackrel{(4.14)}{=} [V_1 \mu(r, t, T_1) - V_2 \mu(r, t, T_2)] dt + [V_1 \sigma(r, t, T_1) - V_2 \sigma(r, t, T_2)] dW_{\mathbb{P}}(t).
 \end{aligned} \tag{4.15}$$

Da aufgrund des Prinzips der Arbitragefreiheit davon ausgegangen wird, dass das Portfolio risikolos ist, müssen V_1 und V_2 so gewählt werden, dass der stochastische Term in Gleichung (4.15) verschwindet. Dies erreichen wir mit

$$V_1 = \frac{\sigma(r, t, T_2)}{\sigma(r, t, T_2) - \sigma(r, t, T_1)} \Pi \quad \text{und} \quad V_2 = \frac{\sigma(r, t, T_1)}{\sigma(r, t, T_2) - \sigma(r, t, T_1)} \Pi; \tag{4.16}$$

denn dann haben wir gerade

$$V_1 \sigma(r, t, T_1) - V_2 \sigma(r, t, T_2) = 0.$$

Desweiteren muss ein risikoloses Portfolio denselben Ertrag erbringen wie der risikolose Zinssatz, so dass die Möglichkeit eines risikolosen Gewinns ausgeschlossen ist (Prinzip der Arbitragefreiheit). Daher muss

$$d\Pi = r(t) \Pi dt$$

gelten. Mit (4.16) ergibt sich daraus

$$\frac{\mu(r, t, T_1) - r(t)}{\sigma(r, t, T_1)} = \frac{\mu(r, t, T_2) - r(t)}{\sigma(r, t, T_2)}.$$

Da in dieser Gleichung die linke Seite unabhängig vom Fälligkeitsdatum T_2 , die rechte unabhängig von T_1 ist, kann der obige Ausdruck mit einem vom Fälligkeitsdatum unabhängigen stochastischen Prozess gleichgesetzt werden. Es ergibt sich der so genannte *Marktpreis des Risikos*:

Definition 4.2.24 [DER MARKTPREIS DES RISIKOS]

$$\lambda(r, t) = \frac{\mu(r, t) - r(t)}{\sigma(r, t)} \tag{4.17}$$

Nach Definition ist der Marktpreis des Risikos somit unabhängig vom Fälligkeitsdatum für alle gleichartigen⁵ Nullkupon-Anleihen identisch. Deswegen ist durch die Festlegung des Marktpreises des Risikos eine interne Konsistenzbedingung gegeben, die von den Nullkupon-Anleihen erfüllt sein muss.

3. Durch Rückeinsetzen von $\mu(r, t, T)$ und $\sigma(r, t, T)$ gemäß ihrer Definition (4.14) wird aus der Definition des Marktpreises des Risikos (4.17) die zu lösende partielle Differentialgleichung

$$\frac{\partial P}{\partial r} (m_{\mathbb{P}}(r, t) - \lambda(r, t) v_{\mathbb{P}}(r, t)) + \frac{\partial P}{\partial t} + \frac{1}{2} \frac{\partial^2 P}{\partial r^2} v_{\mathbb{P}}(r, t)^2 - r(t) P = 0 \tag{4.18}$$

für $t < T$ erhalten. Dabei haben wir die Schlussbedingung $P(r, T; T) = 1$ für die entsprechend Bemerkung 4.2.8 betrachteten Nullkupon-Anleihen mit Nominalwert 1 und Rückzahlung des Nominalwertes am Ende der Laufzeit.

⁵Damit ist vor allem ein Nominalwert derselben Höhe und mit demselben Verbriefungszeitpunkt (zum Beispiel am Ende der Laufzeit) gemeint.

Satz 4.2.25

Als Integraldarstellung der Lösung fand Vasicek 1977

$$P(r, t; T) = E \left[\exp \left(- \int_t^T r(s) ds - \frac{1}{2} \int_t^T \lambda^2(r, s) ds - \int_t^T \lambda(r, s) dW_{\mathbb{P}}(s) \right) \right] \quad (4.19)$$

für $t < T$ ([Vas77]).

Beweis: (siehe [Kwo98])

Als Hilfsfunktion sei für $\xi \geq t$

$$V(r, t; \xi) = \exp \left(- \int_t^\xi r(s) ds - \frac{1}{2} \int_t^\xi \lambda^2(r, s) ds - \int_t^\xi \lambda(r, s) dW_{\mathbb{P}}(s) \right) \quad (4.20)$$

gewählt. Die Anwendung der Produktregel auf PV (mit $P = P(r, t; T)$ sowie $V = V(r, t; \xi)$) führt mit (4.13) auf

$$\begin{aligned} d(PV) &= VdP + PdV + dPdV \\ &= \overbrace{V \left[\frac{\partial P}{\partial r} m_{\mathbb{P}} + \frac{\partial P}{\partial \xi} + \frac{1}{2} \frac{\partial^2 P}{\partial r^2} v_{\mathbb{P}}^2 \right] d\xi + V \frac{\partial P}{\partial r} v_{\mathbb{P}} dW_{\mathbb{P}}}^{VdP} \\ &\quad + \overbrace{PV \left(-r - \frac{1}{2} \lambda^2 \right) d\xi - PV \lambda dW_{\mathbb{P}} + \frac{1}{2} \lambda^2 PV d\xi - V \lambda v_{\mathbb{P}} \frac{\partial P}{\partial r} d\xi}^{PdV} \\ &= V \left[\frac{\partial P}{\partial r} (m_{\mathbb{P}} - \lambda v_{\mathbb{P}}) + \frac{\partial P}{\partial \xi} + \frac{\partial^2 P}{\partial r^2} \frac{1}{2} v_{\mathbb{P}}^2 - r(t)P \right] d\xi + V \left[-P\lambda + \frac{\partial P}{\partial r} v_{\mathbb{P}} \right] dW_{\mathbb{P}} \\ &\stackrel{(4.18)}{=} V \left[-P\lambda + \frac{\partial P}{\partial r} v_{\mathbb{P}} \right] dW_{\mathbb{P}}. \end{aligned}$$

Die Integration dieser Gleichung von t bis T und anschließende Erwartungswertbildung ergibt

$$E[P(r, T; T)V(r, t; T) - P(r, t; T)V(r, t; t)] = 0.$$

Da den Annahmen entsprechend $P(r, T; T) = 1$ sowie nach Definition $V(r, t; t) = 1$ ist, gilt demnach

$$P(r, t; T) = E[V(r, t; T)].$$

□

Folgerung 4.2.26

Wenn also der Diffusionsprozess für $r(t)$ (4.12) und der Marktpreis des Risikos $\lambda(r, t)$ mittels einer am Markt vorhandenen Nullkupon-Anleihe über (4.17) spezifiziert sind, so kann mit (4.19) der Preis jeder gleichartigen Nullkupon-Anleihe mit beliebigem Fälligkeitsdatum bestimmt werden.

Der Martingal-Ansatz

In diesem Ansatz wird – im Gegensatz zum Ansatz der partiellen Differentialgleichungen – nicht vom realen Wahrscheinlichkeitsraum ausgegangen, sondern ein äquivalentes

Martingalmaß Q verwendet, das durch die eingesetzte Normierung mittels des Diskontierungsfaktors D gerade dem risikoneutralen Wahrscheinlichkeitsmaß entspricht.⁶ Das Wahrscheinlichkeitsmaß ist also als das Maß festgelegt, unter dem alle diskontierten Preise Martingale sind: Für die diskontierten Preise $Z(= D \cdot P)$ gilt demzufolge

$$Z(t) = \mathbb{E}_t^Q [Z(u)] \quad \text{für alle } u \geq t \quad (4.21)$$

mit dem bezüglich des Wahrscheinlichkeitsmaßes Q , beziehungsweise der zugrunde liegenden Normierung, für den zum Zeitpunkt t gebildeten Erwartungswert. Für die Zuwächse der Short Rate $r(t)$ wird nun angenommen, dass sie unter Q gegeben sind durch

$$dr(t) = m(r, t)dt + v(r, t)dW(t) \quad (4.22)$$

mit $W(t)$ einem Wiener-Prozess unter Q . Dann ist der Preis einer Nullkupon-Anleihe entsprechend (4.21) über den Erwartungswert unter dem äquivalenten Martingalmaß Q definiert als

$$P(r, t; T) = \exp\left(\int_0^t r(s)ds\right) \mathbb{E}^Q \left[\exp\left(-\int_0^T r(s)ds\right) P(r, T; T) \right]. \quad (4.23)$$

Da von $P(r, T; T) = 1$ ausgegangen wird, gilt demnach

$$P(r, t; T) = \mathbb{E}^Q \left[\exp\left(-\int_t^T r(s)ds\right) \right]. \quad (4.24)$$

Verbindung von PDGL- und Martingal-Ansatz / Maßwechsel von \mathbb{P} zu Q

Ausgehend von einer Nullkupon-Anleihe, die (4.23) erfüllt, lässt sich aus der Feynman-Kac-Formel (siehe [Nef96]) schließen, dass $P(r, t; T)$ ebenso der partiellen Differentialgleichung

$$\frac{\partial P}{\partial r} m(r, t) + \frac{\partial P}{\partial t} + \frac{1}{2} \frac{\partial^2 P}{\partial r^2} v(r, t)^2 - r(t)P = 0 \quad (4.25)$$

folgt. Da $P(r, t; T)$ somit sowohl (4.25) als auch (4.18) erfüllt, muss gelten

$$v_{\mathbb{P}}(r, t) = v(r, t) \quad \text{sowie} \quad \lambda(r, t) = \frac{m_{\mathbb{P}}(r, t) - m(r, t)}{v_{\mathbb{P}}(r, t)} \quad (4.26)$$

Aus dem Vergleich der beiden stochastischen Differentialgleichungen (4.22) und (4.12) resultiert

$$dW_{\mathbb{P}}(t) = dW(t) - \lambda(r, t)dt.$$

Damit lässt sich mittels des Girsanov-Theorems (siehe [Nef96]) für die Spezifikation des Wechsels vom realen Wahrscheinlichkeitsmaß \mathbb{P} zum äquivalenten Martingalmaß Q schließen:

$$\frac{dQ}{d\mathbb{P}} = \exp\left(-\int_0^t \lambda(r, s)dW_{\mathbb{P}}(s) - \frac{1}{2} \int_0^t \lambda^2(r, s)ds\right)$$

⁶Zur Normierung kann im Grunde jedes beliebige Finanzinstrument gewählt werden. Allerdings muss beachtet werden, dass diese Wahl entscheidend beeinflusst, wie elegant ein Problem zu lösen ist.

Bemerkung 4.2.27

Bei der Parameterschätzung auf Grundlage einer Stichprobe realer Daten (vergleiche Kapitel 8) wird im realen Maßraum gearbeitet. Falls die Parameter unter dem risikoneutralen Wahrscheinlichkeitsmaß bestimmt worden sind, muss also anschließend noch ein Maßwechsel vollzogen werden. Dafür kann einerseits der Marktpreis des Risikos explizit bestimmt werden; andererseits kann dies implizit durch eine die Kalibrierung der mittels der geschätzten Parameter generierten Zinsstruktur an die aktuelle reale Zinsstruktur erreicht werden (siehe Kapitel 8.5).

Solange die numerischen Analysen allerdings nur auf Daten basieren, die unter dem risikoneutralen Wahrscheinlichkeitsmaß generiert worden sind, entfällt die Notwendigkeit eines Maßwechsels.

4.3. Short-Rate-Modelle

Ebenso wie die Modellierung der Aktienkurse (Kapitel 4.1) basiert auch die Modellierung der Short Rate auf der stochastischen Differentialgleichung von Itô (siehe (3.13)), um den zugrunde liegenden Diffusionsprozess zu beschreiben. Damit wird die Short Rate $r(t)$ durch einen Itô-Prozess modelliert, der durch die allgemeine stochastische Differentialgleichung

$$dr(t) = m(r, t)dt + v(r, t)dW(t)$$

mit einem den Zufall modellierenden Wiener Prozess $W(t)$ gegeben ist.

Die Koeffizienten $m(r, t)$ und $v(r, t)$, Drift und Volatilität des Prozesses, können je nach Prozesstyp von den bereits gewonnenen Informationen r und/oder der Zeit t abhängen. Im folgenden Abschnitt werden zunächst Short-Rate-Modelle mit zeitunabhängigen Parametern $m(r)$ und $v(r)$ betrachtet. Danach wird kurz auf mögliche Short-Rate-Modelle mit zeitabhängigen Parametern eingegangen ([Hau05], [Hul01]). Die meisten der in diesem Kapitel betrachteten Modelle basieren auf den in Kapitel 3.4 spezifizierten stochastischen Prozessen.

In 4.3.1 wird das auf einem Gauss-Prozess beruhende Vasicek-Modell zur Modellierung der Short Rate untersucht, um anschließend das kompliziertere CIR-Modell, das im Rahmen der numerischen Analyse der Parameterschätzverfahren verwendet wird, zu betrachten (vergleiche [Kwo98], [Gla04], [BM01]).

Short-Rate-Modelle mit zeitunabhängigen Parametern

Short-Rate-Modelle, deren Drift- und Volatilitätsparameter unabhängig von der Zeit t sind, beruhen auf Itô-Prozessen mit einer stochastischen Differentialgleichung der Form

$$dr(t) = m(r)dt + v(r)dW(t).$$

Tabelle 4.1 gibt eine Übersicht der wichtigsten Short-Rate-Modelle mit zeitunabhängigen Parametern. Diese traditionell in der Finanzmathematik zur Modellierung der Zinsstruktur eingesetzten Modelle sind auch als *Gleichgewichtsmodelle* (engl. *Equilibrium models*) bekannt (vergleiche [Hul01]). Allerdings ist es aufgrund der Zeitunabhängigkeit ihrer endlichen Parameteranzahl nicht möglich, mit ihnen die aktuelle Zinsstruktur arbitragefrei zu reproduzieren. Das heißt, die Parameterwerte können in der Regel nicht so gewählt werden, dass die durch das Modell festgelegten Preise vollständig mit den am Markt tatsächlich beobachtbaren übereinstimmen. Stattdessen kann lediglich versucht werden, einen „best fit“ ausfindig zu machen.

Die Mean-Reversion-Eigenschaft bei Zinsen

Bereits in Kapitel 3.4 sind Mean-Reversion-Prozesse und ihre Eigenschaften vorgestellt worden. Daher sei hier nur zur Erinnerung erwähnt, dass ein Mean-Reversion-Prozess $r(t)$ entsprechend seinem Driftterm $b(m - r(t))dt$ stets mit der Mean-Reversion-Stärke b zum Mean-Reversion-Niveau m zurückgezogen wird. Allerdings ist zu beachten, dass dieser tendenzielle Driftanteil vom stochastischen Teil, der Diffusion, überlagert wird und sich der Prozess daher trotz Rückzugstendenz weiter vom Mean-Reversion-Niveau entfernen kann.

Im Zusammenhang mit Zinsen beschreibt die *Mean Reversion* die ökonomische Eigenschaft

SHORT-RATE-MODELLE: ZEITUNABHÄNGIGE PARAMETER		
MODELL	SDGL	PROZESS, VERTEILUNG
Dothan (1978) [Dot78]	$dr(t) = \mu r(t)dt + \sigma r(t)dW(t)$	Prozess mit geometrischer SDGL $\sim LN((\mu - \frac{1}{2}\sigma^2)dt, \sigma^2 dt)$
Vasicek (1977) [Vas77]	$dr(t) = (a - br(t)) dt + \sigma dW(t)$ $\Leftrightarrow dr(t) = b(m - r(t)) dt + \sigma dW(t)$	Mean-Reversion-Prozess $\sim N(b(m - r(t)) dt, \sigma^2 dt)$
CIR (1985) [CIR85] (Cox- Ingersoll- Ross)	$dr(t) = (a - br(t)) dt + \sigma \sqrt{r(t)}dW(t)$ $\Leftrightarrow dr(t) = b(m - r(t)) dt + \sigma \sqrt{r(t)}dW(t)$	Mean-Reversion-Prozess & (Quadrat-)Wurzel-Prozess $\sim \chi_{nz}^2$ (Nichtzentrale Chi-Quadrat-Verteilung)
$a, b, m (= a/b)$ und σ positive Konstante, $W(t)$ Wiener Prozess		

Tabelle 4.1.: Short-Rate-Modelle mit zeitunabhängigen Parametern.

von Zinsen, über die Zeit hinweg nicht beliebig fallen oder wachsen zu können, sondern tendenziell um ein langfristiges Mittel, das Mean-Reversion-Niveau, zu pendeln. Für den Einsatz der Mean Reversion sprechen wesentliche wirtschaftliche Überlegungen (vergleiche [Hul01]): So führen höhere Zinssätze zu einem verlangsamten Wirtschaftswachstum. Von den Kreditnehmern werden immer weniger Geldmittel angefordert. Die Zinssätze sinken wieder ab. Umgekehrt führen niedrigere Zinssätze zu einer erhöhten Kredit-Nachfrage. Infolgedessen steigen auch die Zinssätze erneut.

Sowohl das Vasicek- als auch das CIR-Modell greift in seinem Driftterm die Idee der Mean Reversion auf.

Der Level-Effekt

Ist der Volatilitätsterm von der allgemeinen Form $v(r, t) = \sigma r^\kappa$ mit $\kappa > 0$, dann steigt die Volatilität zusammen mit dem Level des Zinssatzes. In diesem Zusammenhang wird vom *Level-Effekt* gesprochen. Es hat sich gezeigt, dass Modelle, die diesen Effekt aufweisen, etwa das CIR-Modell, wesentlich erfolgreicher sind als solche, die lediglich über einen konstanten Volatilitätsterm verfügen (zum Beispiel das Vasicek-Modell) (vergleiche [Kwo98]). Dieser Effekt wurde unter anderem in Kapitel 3.4 bei den Prozessen mit geometrischer stochastischer Differentialgleichung und den (Quadrat-)Wurzel-Prozessen vorgestellt. Je nachdem wie groß das κ gewählt ist, steigt die Proportionalität der Volatilität zur Höhe des Zinssatzes. Dabei scheint im Vergleich mit realen Zinsdaten eine schwächere Proportionalität mittels $\kappa = \frac{1}{2}$ (wie im CIR-Modell) sinnvoller als $\kappa = 1$ (wie im Dothan-Modell) zu sein.

Short-Rate-Modelle mit zeitabhängigen Parametern

Wie im vorhergehenden Abschnitt erklärt, kann die durch den Markt gegebene Zinsstruktur mit den zeitunabhängigen Gleichgewichtsmodellen nicht arbitragefrei dargestellt werden. Im Gegensatz dazu können die Modelle, die jetzt beschrieben werden, aufgrund ihrer zeitabhängigen Parameter konsistent mit der heutigen Zinsstruktur spezifiziert sein. Daher wird bei ihnen auch von *Arbitragefreien Modellen* (engl. *No-arbitrage models*) gesprochen (vergleiche [Hul01]). Während eine Kalibrierung bezüglich der Zinsstruktur bei den Gleichgewichtsmodellen nur über eine entsprechende Parameterwahl mittels eines „best fit“ im Nachhinein möglich ist, kann und wird bei den Arbitragefreien Modellen die heutige Zinsstruktur wegen der Zeitabhängigkeit bereits als Eingabe für die Modellspezifizierung verwendet. Die zukünftige Entwicklung der Short Rate im Modell orientiert sich daher an der heutigen Zinsstruktur. Eine steigende Zinsstrukturkurve führt so allerdings zum Beispiel dazu, dass auch die Short Rate in der risikoneutralen Welt durchschnittlich steigt. Hier stellt sich das Problem, dass es zum „Overfitting“ bezüglich der aktuellen Zinsstruktur kommen kann. Die Vor- und Nachteile der beiden Ansätze bleiben im Einzelfall gegeneinander abzuwägen.

SHORT-RATE-MODELLE: ZEITABHÄNGIGE PARAMETER		
MODELL	SDGL	MODIFIKATIONEN
Ho-Lee (1986) [HL86]	$dr(t) = \mu(t)dt + \sigma dW(t)$	$\sigma(t)$
Black-Derman-Toy (1990) [BDT90]	$r(t) = e^{Y(t)}$ mit $dY(t) = \left(a(t) - \frac{\sigma'(t)}{\sigma(t)} Y(t) \right) dt + \sigma(t) dW(t)$	
Black-Karasinski (1991) [BK91]	$r(t) = e^{Y(t)}$ mit $dY(t) = (a(t) - b(t)Y(t)) dt + \sigma(t) dW(t)$	
Hull-White I (1990) [HW90] [Erweitertes Vasicek-Modell]	$dr(t) = (a(t) - br(t)) dt + \sigma dW(t)$ $\Leftrightarrow dr(t) = b(m(t) - r(t)) dt + \sigma dW(t)$	$b(t), \sigma(t)$
Hull-White II (1990) [Erweitertes CIR-Modell]	$dr(t) = (a(t) - br(t)) dt + \sigma\sqrt{r(t)}dW(t)$ $\Leftrightarrow dr(t) = b(m(t) - r(t)) dt + \sigma\sqrt{r(t)}dW(t)$	$b(t), \sigma(t)$
$\mu(t), a(t), b(t), m(t), \sigma(t)$ deterministische zeitabhängige Funktionen, b und σ konstant		

Tabelle 4.2.: Short-Rate-Modelle mit zeitabhängigen Parametern.

Der Vorteil der so genannten *Lognormalen Modelle*, zu denen das Black-Derman-Toy-

Modell sowie dessen Verallgemeinerung, das Black-Karasinski-Modell, zählen, liegt darin, dass negative Short Rates von vorneherein über den Exponentialterm ausgeschlossen sind. Allerdings gestaltet sich der Umgang mit ihnen sehr viel komplizierter, als es bei den „normalen“ Modellen der Fall ist.

In den Hull-White-Modellen kann $m(t) = \frac{a(t)}{b}$ als „lokales“ Mean-Reversion-Niveau interpretiert werden.

4.3.1. Das Vasicek-Modell

Der zugrunde liegende Prozess, die Verteilung der Zuwächse der Short Rate

Das *Vasicek-Modell* [Vas77] ist eines der frühesten und bekanntesten zeitstetigen Short-Rate-Modelle. Ihm liegt ein *Mean-Reversion-Prozess* mit zeitunabhängigen Parametern zugrunde. Die Short Rate ist im Vasicek-Modell als die eindeutige Lösung der stochastischen Differentialgleichung

$$dr(t) = b(m - r(t)) dt + \sigma dW(t) \quad (4.27)$$

mit Wiener-Prozess $W(t)$ und positiven Konstanten b , m und σ definiert. Entsprechend des Prinzips der Mean Reversion ist m dabei das Mean-Reversion-Niveau, zu dem die Short Rate $r(t)$ mit der Mean-Reversion-Stärke b gezogen wird. (Für nähere Erläuterungen siehe Kapitel 3.4 und Kapitel 4.3.) Nach (4.27) sind die Änderungen der Short Rate aufgrund der Normalverteilung der Zuwächse des Wiener Prozesses normalverteilt. Genauer folgt

$$dr(t) \sim N(b(m - r(t)) dt, \sigma^2 dt),$$

da $dW(t) \sim N(0, dt)$ (siehe (3.10)). Dabei ist insbesondere der Erwartungswert vom aktuellen Level $r(t)$ der Short Rate abhängig.

Die geschlossene Lösung, die Verteilung der Short Rate

Anwendung des Itô-Lemma auf die Transformation $Y(t) = e^{bt}r(t)$ führt zu

$$Y(t) = Y(0) + bm \int_0^t e^{bs} ds + \sigma \int_0^t e^{bs} dW(s),$$

so dass nach entsprechender Rücksubstitution

$$r(t) = e^{-bt}r(0) + m(1 - e^{-bt}) + \sigma e^{-bt} \int_0^t e^{bs} dW(s) \quad (4.28)$$

ist (vergleiche [BM01]). Daraus ist direkt ersichtlich, dass die Zufallsvariable $r(t)$ zum Zeitpunkt t normalverteilt ist mit Erwartungswert

$$E[r(t)] = e^{-bt}r(0) + m(1 - e^{-bt}) \xrightarrow{t \rightarrow \infty} m \quad (4.29)$$

und Varianz

$$\text{var}[r(t)] = \frac{\sigma^2}{2b} (1 - e^{-2bt}) \xrightarrow{t \rightarrow \infty} \frac{\sigma^2}{2b}.$$

Die Short Rate konvergiert somit im Hinblick auf ihre Verteilung für $t \rightarrow \infty$ gegen die Normalverteilung $N(m, \sigma^2/2b)$. Diese Verteilung ist insoweit stationär, als dass sie bei gegebenem Anfangswert $r(0)$ für alle $r(t)$ mit $t > 0$ gleichermaßen gilt. In der Normalverteilung der Short Rate $r(t)$ liegt der Schwachpunkt des Vasicek-Modells. Denn aufgrund ihrer Normalverteilung kann die Short Rate mit positiver Wahrscheinlichkeit negative Werte annehmen.

Die Bestimmung der Zinsstruktur

Wie in Kapitel 4.2.1 gezeigt, gilt für die Preise von Nullkupon-Anleihen mit Nominalwert 1 und Rückzahlung des Nominalwerts am Ende der Laufzeit

$$P(r, t; T) = \mathbb{E}^Q \left[\exp \left(- \int_t^T r(s) ds \right) \right] \quad (4.30)$$

(siehe (4.24)). Die Preise entsprechen demnach dem Erwartungswert zum Zeitpunkt t unter zugrunde gelegtem risikoneutralem Wahrscheinlichkeitsmaß Q . Zur Vereinfachung wird der Erwartungswert unter dem risikoneutralen Wahrscheinlichkeitsmaß Q von jetzt an einfach mit \mathbb{E} bezeichnet. Aus (4.30) folgt für die Preise zum aktuellen Zeitpunkt $t = 0$

$$P(r, 0; T) = \mathbb{E} \left[\exp \left(- \int_0^T r(s) ds \right) \right].$$

Die im letzten Abschnitt festgestellte Normalverteilung der Short Rate $r(t)$ lässt sich auf die transformierte Zufallsvariable $X(t) = \int_0^T r(s) ds$ übertragen, denn mit $X(t) \sim N(m, v^2)$ gilt $\mathbb{E}[e^{X(t)}] = e^{m + \frac{1}{2}v^2}$. Der Preis einer Nullkupon-Anleihe entspricht daher dem Erwartungswert des Exponentialen einer normalverteilten Zufallsvariable (vergleiche [LL96]):

$$\begin{aligned} P(r, 0; T) &= \mathbb{E} \left[\exp \left(- \int_0^T r(s) ds \right) \right] \\ &= \exp \left(-\mathbb{E} \left[\int_0^T r(s) ds \right] - \frac{1}{2} \text{var} \left[\int_0^T r(s) ds \right] \right). \end{aligned} \quad (4.31)$$

Zur Preisberechnung ist es also notwendig, den Erwartungswert und die Varianz des Integrals der Short Rate $r(t)$ zu bestimmen. Der Erwartungswert entspricht

$$\begin{aligned} \mathbb{E} \left[\int_0^T r(s) ds \right] &= \int_0^T \mathbb{E}[r(s)] ds \stackrel{(4.29)}{=} \int_0^T e^{-bs} r(0) + m (1 - e^{-bs}) ds \\ &= \int_0^T e^{-bs} r(0) ds + \int_0^T m (1 - e^{-bs}) ds \\ &= \frac{1}{b} (1 - e^{-bT}) r(0) + mT - m \frac{1}{b} (1 - e^{-bT}) \\ &= \frac{r(0) - m}{b} (1 - e^{-bT}) + mT. \end{aligned}$$

Für die Varianz ergibt sich mit

$$\text{var} \left[\int_0^T r(s) ds \right] = \text{cov} \left[\int_0^T r(s) ds, \int_0^T r(t) dt \right] = 2 \int_0^T \int_0^t \text{cov}[r(s), r(u)] du ds$$

und

$$\begin{aligned} \text{cov}[r(s), r(u)] &\stackrel{(4.28)}{=} \sigma^2 \int_0^u e^{-b(s-t)} e^{-b(u-t)} dt = \sigma^2 \int_0^u e^{b(2t-s-u)} dt \\ &= \frac{\sigma^2}{2b} (e^{b(u-s)} - e^{-b(u+s)}) \end{aligned}$$

für $u \leq s$ schließlich

$$\begin{aligned}
 \text{var} \left[\int_0^T r(s) ds \right] &= 2 \int_0^T \int_0^s \frac{\sigma^2}{2b} \left(e^{b(u-s)} - e^{-b(u+s)} \right) dud s \\
 &= 2 \int_0^T \frac{\sigma^2}{2b^2} \left(1 + e^{-2bs} - 2e^{-bs} \right) ds \\
 &= \frac{\sigma^2}{b^2} \left[s - \frac{1}{2b} e^{-2bs} + \frac{2}{b} e^{-bs} \right]_0^T \\
 &= \frac{\sigma^2}{b^2} \left[T + \frac{1}{2b} \left(1 - e^{-2bT} \right) - \frac{2}{b} \left(1 - e^{-bT} \right) \right]. \quad (4.32)
 \end{aligned}$$

Wird

$$B(0, T) = \frac{1}{b} \left(1 - e^{-bT} \right)$$

festgelegt, so gilt

$$\mathbb{E} \left[\int_0^T r(t) dt \right] = (r(0) - m) B(0, T) + mT$$

sowie

$$\begin{aligned}
 \text{var} \left[\int_0^T r(t) dt \right] &\stackrel{(4.32)}{=} \frac{\sigma^2}{b^2} \left[T + \frac{1}{2b} \left(1 - e^{-bT} \right) \left(1 + e^{-bT} \right) - \frac{2}{b} \left(1 - e^{-bT} \right) \right] \\
 &= \frac{\sigma^2}{b^2} \left[T + \frac{1}{2} B(0, T) \left(1 + e^{-bT} \right) - 2B(0, T) \right] \\
 &= \frac{\sigma^2}{b^2} T + \frac{\sigma^2}{b^2} \left[-B(0, T) \left(\frac{3}{2} - \frac{1}{2} e^{-bT} \right) \right] \\
 &= \frac{\sigma^2}{b^2} T + \frac{\sigma^2}{b^2} \left[-B(0, T) \left(\frac{b}{2} \frac{1}{b} \left(1 - e^{-bT} \right) + 1 \right) \right] \\
 &= \frac{\sigma^2}{b^2} \left[T - \frac{b}{2} B(0, T)^2 - B(0, T) \right].
 \end{aligned}$$

Für den Preis der Nullkupon-Anleihe ergibt sich daher über die Preisgleichung (4.31)

$$\begin{aligned}
 P(r, 0; T) &= e^{A(0, T) - B(0, T)r(0)} \\
 \text{mit} \quad A(0, T) &= \frac{(B(0, T) - T)(b^2 m - \frac{\sigma^2}{2})}{b^2} - \frac{\sigma^2 B(0, T)^2}{4b} \\
 \text{und} \quad B(0, T) &= \frac{1}{b} (1 - e^{-bT}).
 \end{aligned}$$

Auf demselben Weg kann die Preisgleichung auch für allgemeines t hergeleitet werden⁷:

$$P(r, t; T) = e^{A(t, T) - B(t, T)r(t)} \quad (4.33)$$

$$\text{mit} \quad A(t, T) = \frac{(B(t, T) - (T - t))(b^2 m - \frac{\sigma^2}{2})}{b^2} - \frac{\sigma^2 B(t, T)^2}{4b} \quad (4.34)$$

$$\text{und} \quad B(t, T) = \frac{1}{b} (1 - e^{-b(T-t)}). \quad (4.35)$$

⁷Dies ist auch aus der Eigenschaft von $r(t)$ als Markov-Prozess ableitbar.

Folgerung 4.3.1

Liegt für die Short Rate ein Vasicek-Modell entsprechend (4.27) zugrunde, so lassen sich die Preise der Nullkupon-Anleihen mit Nominalwert 1 und Rückzahlung des Nominalwerts zum Ende der Laufzeit über das Gleichungssystem (4.33)-(4.35) ermitteln (siehe auch [LL96], [BM01], [MR97] und [Reb96]).

Bemerkung 4.3.2

Die Preise $P(r, t; T)$ hängen nur über die (Rest-)Laufzeit $\tau = T - t$ von t und T ab.

Folgerung 4.3.3 [DIE ZINSSTRUKTUR]

Nach Kapitel 4.2.1 Gleichung (4.4) lässt sich mit (4.33) für die stetigen Kassazinsen beziehungsweise die Zinsstruktur folgern:

$$R(t, T) = -\frac{1}{T-t} \ln P(r, t; T) = -\frac{A(t, T) - B(t, T)r(t)}{T-t} \quad (4.36)$$

Sind die Parameter des Vasicek-Modells b , m und σ fest gewählt, kann damit die gesamte Zinsstruktur $\tau \rightarrow R(t, t + \tau)$ als lineare Funktion der Short Rate $r(t)$ bestimmt werden.

Bemerkung 4.3.4 [VASICEK-MODELL ALS BEISPIEL EINES AFFINEN MODELLS]

Die Eigenschaft, dass die Logarithmen der Preise $\ln P(r, t; T)$ beziehungsweise die stetigen Kassazinsen $R(t, T)$ linear abhängig von der Short Rate $r(t)$ sind, zeichnet die von Darrell Duffie and Rui Kan [DK96] eingeführte affine Klasse von Zinsstrukturmodellen aus. Zu dieser gehört unter anderem das Vasicek-Modell sowie das CIR-Modell (siehe Kapitel 4.3.2).

Folgerung 4.3.5 [VERERBUNG DER NORMALVERTEILUNG]

Ist die Short Rate $r(t)$, wie hier im Vasicek-Modell, normalverteilt, so sind die Preise der entsprechenden Nullkupon-Anleihen $P(r, t; T)$ lognormalverteilt (vergleiche (4.31) sowie Definition A.0.25) und die Normalverteilung überträgt sich über (4.36) auf die stetigen Kassazinsen $R(t, T)$, die die Zinsstruktur bilden. Durch diese vererbte Normalverteilung kann es zu negativen Werten sowohl der Short Rate als auch der Zinsstruktur kommen.

Anmerkung 4.3.6

An (4.36) lässt sich sehen, dass die Short Rate $r(t)$ zwar das Niveau der Zinsstrukturkurve zur Zeit t bestimmt, die Form der Zinsstrukturkurve allerdings nur von t selbst abhängig ist.

4.3.2. Das Cox-Ingersoll-Ross-Modell

Der zugrunde liegende Prozess, die Verteilung der Zuwächse der Short Rate

Die dem *Cox-Ingersoll-Ross-Modell* [CIR85] (kurz: *CIR-Modell*) zugrunde liegende stochastische Differentialgleichung lautet

$$dr(t) = b(m - r(t)) dt + \sigma\sqrt{r(t)}dW(t). \quad (4.37)$$

Dabei bezeichnet $W(t)$ einen Wiener-Prozess und b , m und σ sind positive Konstanten. Im Driftterm des CIR-Modells findet also ebenso wie beim Vasicek-Modell die Idee der *Mean Reversion* Anwendung. Der die Short Rate beschreibende Prozess $r(t)$ wird mit der Mean-Reversion-Stärke b zum Mean-Reversion-Niveau m gezogen. (Vergleiche Kapitel 3.4 und Kapitel 4.3.) Dieser Mean-Reversion-Zug wird wiederum durch den stochastischen Diffusionsanteil überlagert. Hier unterscheidet sich das CIR-Modell vom Vasicek-Modell, dessen Diffusionskoeffizient unabhängig vom Level der Short Rate ist. Der Prozess des CIR-Modells ist nicht nur ein Mean-Reversion-Prozess, sondern darüberhinaus ein (*Quadrat-*)*Wurzel-Prozess*. (Für die Eigenschaften dieser Prozesse siehe Kapitel 3.4.) Damit überwindet das CIR-Modell die große Schwäche des Vasicek-Modells, in dem aufgrund der (vererbten) Normalverteilung (vergleiche 4.3.5) negative Werte in Bezug auf die Short Rate als auch auf die gesamte (resultierende) Zinsstruktur auftreten können. Beim CIR-Modell kann es durch die Proportionalität der Volatilität zu $\sqrt{r(t)}$ fast sicher (vergleiche Definition A.0.1) zu keinen negativen Werte mehr kommen – vorausgesetzt es gilt $r(0) > 0$ für den Anfangswert sowie $2bm \geq \sigma^2$ für die Parameter des Modells. Sobald die Short Rate die Nulllinie erreicht ($r(t) = 0$), wird auch der gesamte Diffusionsterm $\sigma\sqrt{r(t)}$ auf Null gesetzt und es können infolge des Driftterms nach dem Mean-Reversion-Prinzip nur noch Werte auftreten, die sich näher am Mean-Reversion-Niveau befinden.

Das CIR-Modell weist außerdem den bereits in Kapitel 4.3 erläuterten *Level-Effekt* auf, der die Proportionalität der Volatilität zur Short Rate beschreibt. So führt eine steigende Short Rate im CIR-Modell gleichzeitig zu einer Erhöhung der Volatilität und im umgekehrten Fall zieht eine fallende Short Rate eine immer kleiner werdende Volatilität nach sich. Gerade bei Werten der Short Rate nahe Null spielt dieser Effekt eine große Rolle. Denn durch die klein gehaltene Diffusion hat die Driftkomponente mehr Gewicht und bekommt eine entsprechend größere Chance, die Short Rate wieder nach „oben“ in Richtung des Mean-Reversion-Niveaus $m > 0$ zu ziehen.

Im Gegenzug für die zu jedem Zeitpunkt fast sicher positive Short Rate büßt das CIR-Modell jedoch an analytischer Umgänglichkeit ein. So existiert für die stochastische Differentialgleichung (4.37) keine geschlossene Lösung, wie es für die stochastische Differentialgleichung des Vasicek-Modells der Fall ist.

Aufgrund der stochastischen Differentialgleichung (4.37) sind die Zuwächse $dr(t)$ normalverteilt:

$$dr(t) \sim N(b(m - r(t)) dt, \sigma^2 r(t) dt).$$

Dabei sind Erwartungswert und Varianz vom aktuellen Wert $r(t)$ der Short Rate abhängig.

Die Übergangsdichte, die Verteilung der Short Rate

Wie bereits erwähnt, ist für die dem CIR-Modell zugrunde liegende stochastische Differentialgleichung (4.37) wegen ihrer Komplexität keine geschlossene Lösung mehr bekannt ([Gla04]). Allerdings kann die Übergangsdichte des Prozesses bestimmt werden, denn die

Verteilung von $r(t)$ bei gegebenen $r(u)$ mit $u < t$ entspricht bis auf einen Skalierungsfaktor einer Nichtzentralen Chi-Quadrat-Verteilung (siehe [Gla04]) mit einem zu $r(u)$ proportionalen Nichtzentralitätsparameter. Für $r(u)$ mit $u < t$ gilt

$$r(t) \sim c_1 \chi_{nz,f}^2(c_2 r(u))$$

mit

$$c_1 = \frac{\sigma^2 (1 - e^{-b(t-u)})}{4b}, \quad c_2 = \frac{4be^{-b(t-u)}}{\sigma^2 (1 - e^{-b(t-u)})}$$

und $f = \frac{4bm}{\sigma^2}$ Freiheitsgraden. Demnach ist $r(t)$ bei gegebenen $r(u)$ verteilt wie c_1 -mal eine Nichtzentrale Chi-Quadrat-Zufallsvariable mit f Freiheitsgraden und Nichtzentralitätsparameter $c_2 r(u)$. Für $t \rightarrow \infty$ konvergieren die $r(t)$ schließlich dagegen, genauso wie $\frac{\sigma^2}{2b}$ -mal eine Nichtzentrale Chi-Quadrat-Zufallsvariable mit f Freiheitsgraden und Nichtzentralitätsparameter 0 (also wie eine gewöhnliche Chi-Quadrat-Zufallsvariable mit f Freiheitsgraden) verteilt zu sein. Diese Verteilung ist in dem Sinne stationär, als dass für $r(0)$, generiert aus dieser Verteilung, sämtliche folgende $r(t)$ mit $t > 0$ dieselbe Verteilung haben.

Der Erwartungswert der Zufallsvariable $r(t)$ zum Zeitpunkt t entspricht dem des Vasicek-Modells (siehe [BM01]):

$$\mathbb{E}[r(t)] = e^{-bt} r(0) + m \left(1 - e^{-bt}\right) \xrightarrow{t \rightarrow \infty} m. \quad (4.38)$$

Die Varianz ist gegeben durch

$$\text{var}[r(t)] = r(0) \frac{\sigma^2}{b} \left(e^{-bt} - e^{-2bt}\right) + m \frac{\sigma^2}{2b} \left(1 - e^{-2bt}\right)^2 \xrightarrow{t \rightarrow \infty} m \frac{\sigma^2}{2b}.$$

Die asymptotische Varianz des CIR-Modells stimmt also genau für $m = 1$ mit der des Vasicek-Modells überein. Dem Level-Effekt entsprechend fällt die Volatilität für $m < 1$ kleiner und für $m > 1$ verhältnismäßig größer aus.

Die Bestimmung der Zinsstruktur

Wie das Vasicek-Modell gehört auch das CIR-Modell zur Klasse der affinen Zinsstrukturmodelle, für die der Preis einer Nullkupon-Anleihe

$$P(r, t; T) = \mathbb{E} \left[\exp \left(- \int_t^T r(s) ds \right) \right]$$

gerade in der allgemeinen Form

$$P(r, t; T) = e^{A(t,T) - B(t,T)r(t)}$$

wiedergegeben werden kann. Explizit ergibt sich für das CIR-Modell (vergleiche etwa [Kwo98], [BM01],[LL96], [MR97] oder [Reb96]):

$$P(r, t; T) = e^{A(t,T) - B(t,T)r(t)} \quad (4.39)$$

$$\text{mit} \quad A(t, T) = \frac{2bm}{\sigma^2} \ln \left(\frac{2\gamma e^{\frac{1}{2}(b+\gamma)(T-t)}}{(\gamma + b)(e^{\gamma(T-t)} - 1) + 2\gamma} \right) \quad (4.40)$$

$$\text{und} \quad B(t, T) = \frac{2(e^{\gamma(T-t)} - 1)}{(\gamma + b)(e^{\gamma(T-t)} - 1) + 2\gamma} \quad (4.41)$$

$$\text{für} \quad \gamma = \sqrt{b^2 + 2\sigma^2}. \quad (4.42)$$

Die analytische Herleitung kann analog zu der des Gleichungssystems des Vasicek-Modells ((4.33) bis (4.35)) erfolgen.

Lemma 4.3.7 [DIE ZINSSTRUKTUR]

Da mit Kapitel 4.2.1 Gleichung (4.4) folgt, dass

$$R(t, T) = -\frac{1}{T-t} \ln P(r, t; T) = -\frac{A(t, T) - B(t, T)r(t)}{T-t}$$

gilt, kann die gesamte Zinsstruktur $\tau \rightarrow R(t, t + \tau)$ wieder als lineare Funktion der Short Rate $r(t)$ bestimmt werden, sofern die Parameter b , m und σ des CIR-Modells fest gewählt sind.

Bemerkung 4.3.8

Wie beim Vasicek-Modell hängen die Preise $P(r, t; T)$ erneut bloß über die (Rest-)Laufzeit $\tau = T - t$ von t und T ab.

Anmerkung 4.3.9

Außerdem bestimmt die Short Rate $r(t)$ zwar wiederum das Niveau der Zinsstrukturkurve zur Zeit t . Die Form der Zinsstrukturkurve ist aber auch hier nur von t selbst abhängig.

4.4. Mehr-Faktor-Modelle

Bei den bislang vorgestellten Short-Rate-Modellen (Kapitel 4.3) handelt es sich um Ein-Faktor-Modelle. Ihnen liegt die Annahme zugrunde, dass die gesamte Zinsstruktur bereits durch die Spezifizierung eines einzigen stochastischen Faktors festgelegt ist. Begründet werden kann diese Annahme durch die hohe Korrelation der Zufallsvariablen, die die Zinsstrukturkurven beschreiben. (Siehe dazu auch Kapitel 4.2 Bemerkung 4.2.7.) Dieser Ansatz tendiert allerdings dazu, das wahre Verhalten der Zinsstruktur zu sehr zu vereinfachen ([Kwo98]). So mangelt es zum Beispiel an der nötigen Flexibilität, Dynamiken von Zinsstrukturen angemessen zu beschreiben, die sich an ihren (bezüglich der (Rest-)Laufzeit) entgegengesetzten Enden konträr entwickeln ([Hau05]). Für das Vasicek-Modell kann festgestellt werden (siehe [Gla04]):

Bemerkung 4.4.1

Wie in Kapitel 4.3.1 hergeleitet, ist der Preis einer Nullkupon-Anleihe im Vasicek-Modell gegeben durch das Gleichungssystem (4.33) bis (4.35). Da $A(t, T)$ sowie $B(t, T)$ lediglich von der (Rest-)Laufzeit $\tau = T - t$ abhängen (vergleiche Bemerkung 4.3.2), lässt sich auch schreiben:

$$\begin{aligned} P(r, t; T) &= e^{A(\tau) - B(\tau)r(t)} \\ \text{mit} \quad A(\tau) &= \frac{(B(\tau) - \tau)(b^2 m - \frac{\sigma^2}{2})}{b^2} - \frac{\sigma^2 B(\tau)^2}{4b} \\ \text{und} \quad B(\tau) &= \frac{1}{b}(1 - e^{-b\tau}) \end{aligned}$$

mit $A(\tau) = A(t, T)$ und $B(\tau) = B(t, T)$ für $\tau = T - t$. Mit dem Itô-Lemma folgt dann

$$\frac{dP(r, t; T)}{P(r, t; T)} = r(t)dt - B(\tau)\sigma dW(t).$$

Für zwei Fälligkeitsdaten T_1 und T_2 ergibt sich daher (mittels (A.7) und (A.6)) der instantane Korrelationskoeffizient

$$\rho \left(\frac{dP(r, t; T_1)}{P(r, t; T_1)}, \frac{dP(r, t; T_2)}{P(r, t; T_2)} \right) = \frac{B(\tau_1)\sigma B(\tau_2)\sigma}{\sqrt{B^2(\tau_1)\sigma^2} \sqrt{B^2(\tau_2)\sigma^2}} = 1$$

mit $\tau_i = T_i - t$ für $i \in \{1, 2\}$. Die Preise der Nullkupon-Anleihen sind somit instantan „perfekt“ korreliert.⁸ Dies widerspricht allerdings dem Wunsch, auch Zinsstrukturen modellieren zu können, die sich an ihren entgegengesetzten Enden konträr entwickeln.

Hier setzen die Mehr-Faktor-Modelle an, mit denen es möglich ist, dieses Verhalten der Zinsstruktur mittels ihrer verschiedenen stochastisch modellierten Faktoren zu beschreiben. So ist es für Mehr-Faktor-Modelle möglich, eine instantane Korrelation vom Wert kleiner 1 zwischen den Preisen der Nullkupon-Anleihen mit verschiedenen Fälligkeitsdaten zu haben (siehe zum Beispiel die Ausführungen in Kapitel 4.4.1). Der Nachteil der Hinzunahme weiterer Faktoren besteht allerdings darin, dass gleichzeitig die Möglichkeiten der analytischen Handhabung und der Parameterschätzung beeinträchtigt werden. So muss im Mehr-Faktor-Fall etwa für die Bewertung von Nullkupon-Anleihen unter Umständen auf numerische Methoden zurückgegriffen werden, da im Gegensatz zum Ein-Faktor-Fall keine Preisformeln mehr existieren. Bei der Parameterschätzung ist es eine entsprechend

⁸Entsprechend Satz A.0.13 gilt $\rho \in [-1, 1]$.

größere Anzahl an Parametern, die mittels der am Markt vorliegenden Daten zu schätzen sind. Es hat sich herausgestellt (vergleiche etwa [FMW03]), dass Modelle mit zwei bis drei stochastisch modellierten Faktoren genau die Grenze in Bezug auf die Qualität der auf ihnen basierenden Modellierung der Zinsstruktur sowie ihrer Handhabungs-Komplexität darstellen.

Im Folgenden sollen spezielle Mehr-Faktor-Modelle vorgestellt werden. Es handelt sich dabei um die Übertragung des Vasicek- sowie des CIR-Modells auf den Mehr-Faktor-Fall. Dabei liegt jeweils die Annahme zugrunde, dass die Short Rate $r(t)$ von einer festen Anzahl k von Teilraten $r_i(t)$ mit $i \in \{1, \dots, k\}$ bestimmt wird (vergleiche auch [Gla04]). Der Vektor dieser Teilraten beschreibt am Markt beobachtbare ökonomische Vorgänge. Die Short Rate hängt somit nicht mehr von einer einzelnen sondern von einem ganzen System stochastischer Differentialgleichungen ab. Im einfachsten Fall, von dem hier ausgegangen wird, entspricht die Short Rate der Summe der einzelnen Teilraten

$$r(t) = \sum_{i=1}^k r_i(t). \quad (4.43)$$

Dabei wird angenommen, dass diese Teilraten unabhängig voneinander und damit unkorreliert sind.

4.4.1. Der (unkorrelierte) Mehr-Faktor-Fall des Vasicek-Modells

Wird für die Modellierung der einzelnen unkorrelierten Teilraten eines durch (4.43) spezifizierten Mehr-Faktor-Modells das Vasicek-Modell genutzt, so ergibt sich folgendes System stochastischer Differentialgleichungen:

$$dr_i(t) = b_i(m_i - r_i(t))dt + \sigma_i dW_i(t) \quad \text{für } i = 1, \dots, k. \quad (4.44)$$

Bei den $W_i(t)$ handelt es sich dabei um k unabhängige, damit unkorrelierte, Wiener-Prozesse. Die Parameter b_i , m_i und σ_i sind positiv und konstant. Wie im eindimensionalen Fall gilt hier für die einzelnen Teilraten $r_i(t) > 0$ fast sicher, wenn $2b_i m_i > \sigma_i^2$ für $i \in \{1, \dots, k\}$ ist. In Matrix-Schreibweise kann dieses System dargestellt werden als:

$$dr(t) = B(m - r(t))dt + \Sigma dW(t) \quad (4.45)$$

mit einem k -dimensionalen Wiener Prozess $W(t)$, zwei $(k \times k)$ -Diagonalmatrizen

$$B = \begin{pmatrix} b_1 & & & 0 \\ & b_2 & & \\ & & \ddots & \\ 0 & & & b_k \end{pmatrix} \quad \text{und} \quad \Sigma = \begin{pmatrix} \sigma_1 & & & 0 \\ & \sigma_2 & & \\ & & \ddots & \\ 0 & & & \sigma_k \end{pmatrix},$$

sowie den k -dimensionalen Vektoren m und $r(t)$.

Die Lösung von (4.45) ist ein Gauss-Prozess, der durch

$$r(t) = e^{-Bt}r(0) + \int_0^t e^{-B(t-s)}m ds + \int_0^t e^{-B(t-s)}\Sigma dW(s)$$

gegeben ist (vergleiche [Gla04], [Hau05]). Ähnlich zum Ein-Faktor-Fall des Vasicek-Modells gibt es für den Mehr-Faktor-Fall Preisformeln, in denen die Preise der Nullkupon-Anleihen Exponentiale affiner Funktionen der Short Rate sind. Da die einzelnen Teilraten durch

das einfache Vasicek-Modell beschriebene Prozesse sind, existieren Funktionen $A_i(\tau)$ und $B_i(\tau)$, so dass

$$\mathbb{E}^Q \left[\exp \left(- \int_t^T r_i(s) ds \right) \right] = e^{A_i(\tau) - B_i(\tau)r_i(t)}.$$

Zur näheren Einsicht können die Erläuterungen zur Zinsstrukturbestimmung mit zugrunde gelegtem Vasicek-Modell herangezogen werden (Kapitel 4.3.1 Gleichung (4.30) und (4.33)). Mit $r(t) = \sum_{i=1}^k r_i(t)$ entsprechend 4.43 sind ausreichend Informationen zur Bestimmung der Zinsstruktur vorhanden, denn es lässt sich schlussfolgern (siehe [Hau05])

$$\begin{aligned} P(r, t; T) &= \mathbb{E}^Q \left[\exp \left(- \int_t^T r(s) ds \right) \right] = \mathbb{E}^Q \left[\exp \left(- \int_t^T \sum_{i=1}^k r_i(t) ds \right) \right] \\ &= \mathbb{E}^Q \left[\exp \left(- \sum_{i=1}^k \int_t^T r_i(t) ds \right) \right] = \mathbb{E}^Q \left[\prod_{i=1}^k \exp \left(- \int_t^T r_i(t) ds \right) \right] \\ &= \prod_{i=1}^k \mathbb{E}^Q \left[\exp \left(- \int_t^T r_i(t) ds \right) \right] = \prod_{i=1}^k e^{A_i(\tau) - B_i(\tau)r_i(t)} \\ &= e^{\sum_{i=1}^k (A_i(\tau) - B_i(\tau)r_i(t))}. \end{aligned}$$

Der Preis einer Nullkupon-Anleihe lässt sich dann in Abhängigkeit von t , T und den stochastischen Faktoren $r_i(t)$ mit $r(t) = \sum_{i=1}^k r_i(t)$ ermitteln über

$$\begin{aligned} P(r, t; T) &= e^{\sum_{i=1}^k (A_i(\tau) - B_i(\tau)r_i(t))} = e^{\tilde{A}(\tau) - \tilde{B}(\tau)^\top r(t)} \\ \text{mit} \quad \tilde{A}(\tau) &= \sum_{i=1}^k A_i(\tau) \quad \text{und} \quad \tilde{B}(\tau) = (B_1(\tau), B_2(\tau), \dots, B_k(\tau))^\top \\ \text{sowie} \quad A_i(\tau) &= \frac{(B_i(\tau) - \tau)(b_i^2 m_i - \frac{1}{2} \sigma_i^2)}{b_i^2} - \frac{\sigma_i^2 B_i(\tau)^2}{4b_i} \\ \text{und} \quad B_i(\tau) &= \frac{1}{b_i} (1 - e^{-b_i \tau}) \\ \text{für} \quad &i = 1, \dots, k. \end{aligned}$$

Daraus resultiert für die Zinsstruktur

$$R(t, T) = -\frac{1}{T-t} \ln P(r, t; T) = \sum_{i=1}^k \left(-\frac{A_i(\tau)}{T-t} + \frac{B_i(\tau)}{T-t} r_i(t) \right). \quad (4.46)$$

Die instantane Korrelation der Preise zweier Nullkupon-Anleihen mit unterschiedlichen Fälligkeitsdaten T_1 und T_2 ergibt sich durch Anwendung des Itô-Lemmas (Kapitel 3.3, Satz 3.3.3) auf

$$\frac{dP(r, t; T)}{P(r, t; T)} = r(t)dt - \tilde{B}(\tau)^\top \Sigma dW(t).$$

Es folgt (siehe [Gla04])

$$\frac{\tilde{B}(\tau_1)^\top \Sigma \Sigma^\top \tilde{B}(\tau_2)}{\|\tilde{B}(\tau_1)^\top \Sigma\| \|\tilde{B}(\tau_2)^\top \Sigma\|}.$$

Damit kann die instantane Korrelation (im Gegensatz zum Ein-Faktor-Modell) Werte < 1 annehmen.

4.4.2. Der (unkorrelierte) Mehr-Faktor-Fall des CIR-Modells

Bereits in ihrer Arbeit zum (Ein-Faktor-)CIR-Modell 1985 [CIR85] haben Cox, Ingersoll und Ross eine Übertragung ihres Modells auf den Mehr-Faktor-Fall umrissen. Realisiert wurde diese Idee schließlich 1992 von Chen und Scott [CS92] und [CS95].

Wie im bereits vorgestellten Mehr-Faktor Vasicek-Modell ist $r(t)$ durch eine Summe von k Teilraten $r(t) = \sum_{i=1}^k r_i(t)$ gegeben. Die einzelnen Teilraten entspringen einer dem (Ein-Faktor-)CIR-Modell entsprechenden stochastischen Differentialgleichung, so dass sich folgendes System von stochastischen Differentialgleichungen ergibt (siehe [Gla04], [Hau05]):

$$dr_i(t) = b_i(m_i - r_i(t))dt + \sigma_i \sqrt{r_i(t)} dW_i(t) \quad \text{für } i = 1, \dots, k. \quad (4.47)$$

Die $W_i(t)$ sind dabei wieder k unabhängige (unkorrelierte) Wiener-Prozesse und die b_i , m_i sowie σ_i positive Konstanten. Wie im eindimensionalen Fall gilt auch hier $r_i > 0$ fast sicher für $2b_i m_i > \sigma_i^2$.

Auf dem gleichen Weg wie beim Mehr-Faktor-Fall des Vasicek-Modells lässt sich mit $r(t) = \sum_{i=1}^k r_i(t)$ zeigen, dass sich die Preise von Nullkupon-Anleihen über folgendes Gleichungssystem in Abhängigkeit von t , T und den Teilraten $r_i(t)$ ermitteln lassen:

$$\begin{aligned} P(r, t; T) &= e^{\sum_{i=1}^k (A_i(\tau) - B_i(\tau)r_i(t))} = e^{\tilde{A}(\tau) - \tilde{B}(\tau)^\top r(t)} \\ \text{mit} \quad \tilde{A}(\tau) &= \sum_{i=1}^k A_i(\tau) \quad \text{und} \quad \tilde{B}(\tau) = \begin{pmatrix} B_1(\tau) \\ B_2(\tau) \\ \vdots \\ B_k(\tau) \end{pmatrix} \\ \text{sowie} \quad A_i(\tau) &= \frac{2b_i m_i}{\sigma_i^2} \ln \left[\frac{2h_i e^{(b_i+h_i)\tau/2}}{(h_i + b_i)(e^{h_i\tau} - 1) + 2h_i} \right] \\ \text{und} \quad B_i(\tau) &= \frac{2(e^{h_i\tau} - 1)}{(h_i + b_i)(e^{h_i\tau} - 1) + 2h_i} \\ \text{mit} \quad h_i &= \sqrt{b_i^2 + 2\sigma_i^2} \\ \text{für} \quad &i = 1, \dots, k. \end{aligned}$$

Daraus resultiert für die Zinsstruktur

$$R(t, T) = -\frac{1}{T-t} \ln P(r, t; T) = \sum_{i=1}^k \left(-\frac{A_i(\tau)}{T-t} + \frac{B_i(\tau)}{T-t} r_i(t) \right). \quad (4.48)$$

Kapitel 5.

Diskretisierung der Modelle

Für die numerische Realisierung eines stochastischen Prozesses wird ein diskretes Modell benötigt. Ursprünglich liegt das Modell des betrachteten stochastischen Prozesses allerdings meist in kontinuierlicher Form als stochastische Differentialgleichung vor. Um vom kontinuierlichen Modell zum diskreten Modell zu gelangen, wird auf so genannte *Diskretisierungsschemata* zurückgegriffen. Dabei bezeichnet der Begriff der *Diskretisierung* im Allgemeinen den Übergang von einer (unendlichen) kontinuierlichen zu einer diskreten Betrachtungsweise an nur noch endlich vielen Gitterpunkten. In den meisten Fällen beziehen sich diese Gitterpunkte auf die Zeit- und/oder die Raumachse, so dass das Modell an endlich vielen separaten Zeit- beziehungsweise Ortspunkten betrachtet wird.¹

Wird von einer stochastischen Differentialgleichung ausgegangen, so ist das allgemein bekannte stochastische *Euler-Maruyama-Schema* das einfachste Diskretisierungsschema, das zur approximativen Bestimmung einer Realisierung der stochastischen Differentialgleichung eingesetzt werden kann (vergleiche auch [Gla04] und [KP99]). Auch im Rahmen dieser Diplomarbeit wird dieses Schema zur approximativen Simulation stochastischer Differentialgleichungen verwendet, da es – gerade im Hinblick auf eine einbezogene Korrelation mehrerer stochastischer Prozesse – universell einsetzbar ist. Dabei muss stets im Auge behalten werden, dass beim Euler-Maruyama-Schema, wie bei allen anderen Diskretisierungsschemata, ein gewisser *Diskretisierungsfehler* nicht vermeidbar ist, der mit kleiner werdender Schrittweite der Diskretisierung (Abstand zwischen den Gitterpunkten der Diskretisierung) abnimmt.

Im Einzelfall bestehen exaktere Simulationsmöglichkeiten. So kann, falls die stochastische Differentialgleichung über eine geschlossene Lösung verfügt (wie es beim Aktienkursmodell für endliche Zeitzuwächse der Fall ist (siehe (3.22))), diese Lösung in diskretisierter Form zur Simulation verwendet werden. Außerdem ist es möglich, die Dichtefunktion des Prozesses – falls bekannt – zur Simulation einzusetzen. Beim CIR-Modell können die Zinssätze etwa mit Hilfe der Nichtzentralen Chi-Quadrat-Verteilung simuliert werden (siehe [Gla04]).

Trotz dieser Möglichkeit bietet es sich gerade im Hinblick auf eine Korrelation verschiedener Prozesse an, das Euler-Maruyama-Schema anzuwenden. Dieses Schema verfügt über eine starke Konvergenz der Ordnung 0.5 und eine schwache Konvergenz der Ordnung 1 ([Gla04], [KP99]). Zwar existieren Diskretisierungsverfahren höherer Ordnung wie das Milstein-Verfahren mit der starken Konvergenzordnung 1, jedoch werden diese Verfahren in Bezug auf die Betrachtung mehrerer stochastischer Prozesse mit verschiedenen zugrundeliegenden Wiener-Prozessen deutlich komplizierter. Das Euler-Maruyama-Schema hingegen bleibt für diese Modelle einfach zu handhaben. Daher werden Verfahren höherer

¹Demzufolge wird auch nur endlich viel Zeit- und Speicheraufwand für die numerische Realisierung benötigt.

Ordnung hier außer Acht gelassen und das Euler-Maruyama-Schema wird eingesetzt.

In den folgenden Abschnitten wird in Anlehnung an [Gla04] sowie [Deu04] beschrieben, wie die Diskretisierung eines einzelnen stochastischen Prozesses sowie mehrerer korrelierter stochastischer Prozesse mittels des Euler-Maruyama-Schema aussieht. In Tabelle 5.1 sind die für das Aktienkurs- und das CIR-Modell spezifizierten Diskretisierungsgleichungen abgebildet.

5.1. Das Euler-Maruyama-Schema für einen einzelnen stochastischen Prozess

Das Euler-Maruyama-Schema beruht auf zwei grundlegenden Diskretisierungsschritten. Zum Einen wird in Anlehnung an das klassische Euler-Schema eine Gittereinteilung der Zeitachse vorgenommen, zum Anderen findet eine Übertragung dieser Einteilung auf die Wahrscheinlichkeitsachse statt.

Für die Gitterpunkte der Zeitachse wird eine konstante Zeitschrittweite $\Delta t > 0$ definiert.² Damit handelt es sich um ein äquidistantes Zeitgitter mit den Gitterpunkten

$$(t_i)_{i \in \mathbb{N}_0} = \{i \cdot \Delta t | i \in \mathbb{N}_0\} = \{t_0, t_1, t_2, \dots\} \text{ für feste Zeitschrittweite } \Delta t > 0. \quad (5.1)$$

Für die Einteilung der Wahrscheinlichkeitsachse werden die diskretisierten Zuwächse eines Wiener Prozesses definiert als

$$\Delta W_i = W(t_i) - W(t_{i-1}) = W(i \cdot \Delta t) - W((i-1) \cdot \Delta t) = \sqrt{\Delta t} Z_i \text{ für } i \in \mathbb{N}_0. \quad (5.2)$$

Dabei handelt es sich bei den Z_i um unabhängig standardnormalverteilte³ Zufallszahlen. Eine Approximation des Prozesses, der auf der stochastischen Differentialgleichung

$$dX(t) = a(X, t)dt + b(X, t)dW(t) \quad \text{mit} \quad dW(t) \sim Y(t)\sqrt{dt} \text{ und } Y(t) \sim N(0, 1) \quad (5.3)$$

mit zeitstetigem Wiener-Prozess $W(t)$ und gegebenem Anfangswert $X(0)$ beruht, ergibt sich mit dem eben definierten Gitter über

$$X(t_i) = X(t_{i-1}) + a(X(t_{i-1}), t_{i-1}) \Delta t + b(X(t_{i-1}), t_{i-1}) \Delta W_i \text{ für } i \in \mathbb{N}_0.$$

Dabei ist der Diskretisierungsfehler des Euler-Maruyama-Schemas von der Ordnung $O(\Delta t)$. Das heißt, die Differenz des exakten Wertes $X_{\text{exakt}}(T)$ und des approximierten Wertes $X(T)$ zum Ende des betrachteten Zeitraums ist von der Ordnung $O(\Delta t)$.⁴ Je kleiner die Zeitschrittweite Δt demnach gewählt ist, desto eher nähern sich die diskreten Simulationen an kontinuierliche Realisierungen an. Dabei darf nicht vergessen werden, dass eine Verfeinerung des Zeitgitters einen entsprechend größeren rechnerischen Aufwand zur Folge hat.

5.2. Das Euler-Maruyama-Schema für mehrere korrelierte stochastische Prozesse

Die Korrelation stochastischer Prozesse lässt sich auf die Korrelation der zugrunde liegenden Wiener-Prozesse zurückführen. Um korrelierte Wiener-Prozesse zu erschaffen, fließt

²Die Schrittweite muss nicht zwingend konstant gewählt sein.

³Definition A.0.20.

⁴Driftterm $O(\Delta t)$, Diffusionsterm $O(\sqrt{\Delta t})$.

DISKRETISIERTE SDGL NACH DEM EULER-MARUYAMA-SCHEMA	
$S_i = S(t_i), r_i = r(t_i), i \in \mathbb{N}_0$	
MODELL	DISKRETISIERTE SDGL
Aktienkurs	$S_i = S_{i-1} + S_{i-1} \cdot \underbrace{(\mu\Delta t + \sigma\Delta W_i)}_{\Delta \ln S_i}$ $\Delta \ln S_i \sim N(\mu\Delta t, \sigma^2\Delta t)$
CIR	$r_i = r_{i-1} + \underbrace{b(m - r_{i-1})\Delta t + \sigma\sqrt{r_{i-1}^+}\Delta W_i}_{\Delta r_i}$ $\Delta r_i \sim N(b(m - r_{i-1})\Delta t, \sigma^2 r_{i-1}\Delta t)$
μ, b, m und σ positive Konstanten, $\Delta W_i = \sqrt{\Delta t}Z_i$ diskretisierter Wiener-Prozess mit unabhängig standardnormalverteilten Zufallszahlen Z_i und festem $\Delta t > 0$	

Tabelle 5.1.: Diskretisierte stochastische Differentialgleichung nach dem Euler-Maruyama-Schema.

die Cholesky-Zerlegung der entsprechenden (positiv definiten) Kovarianzmatrix in die Zuwächse der Wiener Prozesse ein (siehe [Gla04] und [Deu04]). Ist eine Kovarianzmatrix Σ gegeben, so wird zunächst die Cholesky-Zerlegung derselben bestimmt, so dass

$$\Sigma = CC^\top. \quad (5.4)$$

Für einen Vektor Z von k unabhängig normalverteilten Zufallsvariablen ist dann durch

$$Z^{korr} = \left(Z_j^{korr} \right)_{j=1,\dots,k} = C \cdot Z \quad (5.5)$$

ein Vektor korrelierter normalverteilter Zufallsvariablen gegeben. Deren Kovarianzen entsprechen denen der Kovarianzmatrix Σ . Mit Hilfe dieser korrelierten Zufallszahlen kann nach (5.2) der Vektor der diskretisierten Zuwächse der korrelierten Wiener Prozesse für festen Zeitzuwachs $\Delta t > 0$ ermittelt werden:

$$\Delta W^{korr} = \sqrt{\Delta t}Z^{korr}. \quad (5.6)$$

Für festes i_0 sind die $\Delta W_{j,i_0}^{korr}$ für $j \in \{1, \dots, k\}$ dann entsprechend der Kovarianzmatrix $\Delta t \cdot \Sigma$ untereinander korreliert. Für festes j handelt es sich um unabhängig normalverteilte Zufallsvariablen mit Mittelwert 0 und Varianz Δt . Damit werden alle zu diskretisierenden Zuwächse angetrieben durch unabhängig standardnormalverteilte Zufallsvariablen $(Z_{j,i})_{j \in \{1,\dots,k\}, i \in \mathbb{N}_0}$, wobei die Zufallszahlen für festes i korreliert sind entsprechend der Kovarianzmatrix Σ . Um die diskretisierten Werte für den gesamten Prozess zu erhalten, werden schließlich lediglich die korrelierten Wiener-Prozesse anstelle der unkorrelierten verwendet:

$$X_j(t_i) = X_j(t_{i-1}) + a_j(X_j(t_{i-1}), t_{i-1})\Delta t + b_j(X_j(t_{i-1}), t_{i-1})\Delta W_{j,i}^{korr} \quad (5.7)$$

mit $\Delta W_{j,i}^{korr} = \sqrt{\Delta t}Z_{j,i}^{korr}$ für $j \in \{1, \dots, k\}$ und $i \in \mathbb{N}_0$.

Satz 5.2.1 [DISKRETISIERUNG ZWEIER KORRELIERTER STOCHASTISCHER PROZESSE]
 Für zwei über eine Kovarianzmatrix

$$\Sigma = \begin{pmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{pmatrix}$$

korrelierte stochastische Prozesse (ρ bezeichnet dabei den Korrelationskoeffizienten der beiden Prozesse, σ_1 und σ_2 die einzelnen Varianzen) wird durch Cholesky-Zerlegung mit $\Sigma = CC^\top$ die Diagonalmatrix

$$C = \begin{pmatrix} \sigma_1 & 0 \\ \rho\sigma_2 & \sqrt{1-\rho^2}\sigma_2 \end{pmatrix} \quad (5.8)$$

berechnet. Damit ist es möglich, die beiden korrelierten Prozesse mit

$$\begin{aligned} \Delta W_{1,i}^{korr} &= \sqrt{\Delta t} Z_{1,i}^{korr} = \sqrt{\Delta t}(C_{1,1} \cdot Z_{1,i} + C_{1,2} \cdot Z_{2,i}) \\ &= \sqrt{\Delta t} \cdot \sigma_1 \cdot Z_{1,i} \\ \text{und } \Delta W_{2,i}^{korr} &= \sqrt{\Delta t} Z_{2,i}^{korr} = \sqrt{\Delta t}(C_{2,1} \cdot Z_{2,i} + C_{2,2} \cdot Z_{2,i}) \\ &= \sqrt{\Delta t} \left(\rho \cdot \sigma_2 \cdot Z_{1,i} + \sqrt{1-\rho^2} \cdot \sigma_2 \cdot Z_{2,i} \right) \end{aligned}$$

einzelnen über (5.7) zu simulieren, denn nach (5.5) gilt für festes i_0

$$Z_{i_0}^{korr} = \left(Z_{j,i_0}^{korr} \right)_{j=1,\dots,k} = C \cdot Z_{i_0} = \begin{pmatrix} \sigma_1 & 0 \\ \rho\sigma_2 & \sqrt{1-\rho^2}\sigma_2 \end{pmatrix} \cdot \begin{pmatrix} Z_{1,i_0} \\ Z_{2,i_0} \end{pmatrix}.$$

Beweis: Im Rahmen der Cholesky-Zerlegung wird nach einer Dreiecksmatrix gesucht, für die

$$CC^\top = \begin{pmatrix} c_{11} & 0 \\ c_{21} & c_{22} \end{pmatrix} \cdot \begin{pmatrix} c_{11} & c_{21} \\ 0 & c_{22} \end{pmatrix} = \Sigma$$

gilt. Durch Lösen der Matrixmultiplikation auf der linken Seite ergibt sich

$$\begin{pmatrix} c_{11}^2 & 0 \\ c_{11}c_{21} & c_{21}^2 + c_{22}^2 \end{pmatrix} = \begin{pmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{pmatrix}.$$

Daraus folgt

$$\begin{aligned} c_{11}^2 = \sigma_1^2 &\Rightarrow c_{11} = \sigma_1 \\ c_{11}c_{21} = \rho\sigma_1\sigma_2 &\Rightarrow c_{21} = \rho\sigma_2 \\ c_{21}^2 + c_{22}^2 = \sigma_2^2 &\Rightarrow c_{22} = \sqrt{\sigma_2^2 - \rho^2\sigma_2^2} = \sqrt{1-\rho^2}\sigma_2. \end{aligned}$$

□

Tabelle 5.2 zeigt die Diskretisierung des korrelierten Systems von stochastischen Differentialgleichungen. Mit diesem System wird auch in Kapitel 9.6 gearbeitet.

DISKRETISIERTE FORM EINES KORRELIERTEN SDGL-SYSTEMS NACH DEM EULER-MARUYAMA-SCHEMA	
$S_i = S(t_i), r_i = r(t_i), i \in \mathbb{N}_0$	
MODELL- KOMPONENTE	DISKRETISIERTE SDGL
Zinszuwächse nach dem CIR- Modell	$\Delta r_i = b(\mathbf{m} - r(t_{i-1})) \Delta t + \sigma_2 \sqrt{r(t_{i-1})^+} \sqrt{\Delta t} Z_{1,i}$
Logarithmische Aktienkurszu- wächse	$\Delta \ln S_i = \mu \Delta t + \sigma_1 \sqrt{\Delta t} \left(\rho Z_{1,i} + \sqrt{1 - \rho^2} Z_{2,i} \right)$
μ, σ_1, b, m und σ_2 positive Konstanten, Korrelationskoeffizient $\rho \in [-1, 1]$, unabhängig standardnormalverteilte Zufallszahlen $Z_{j,i}$ für $j \in \{1, 2\}$ und $i \in \mathbb{N}_0$, festes $\Delta t > 0$	

Tabelle 5.2.: Diskretisierte stochastische Differentialgleichung nach dem Euler-Maruyama-Schema.

Kapitel 6.

Die Simulation der Modelle

6.1. Aktienkurse

In vorigen Kapitel wurde auf eine mögliche Diskretisierung einer stochastischen Differentialgleichung mittels des Euler-Maruyama-Schemas eingegangen. Geht es lediglich um die Simulation eines einzelnen Aktienkurses, so ist es möglich, statt der stochastischen Differentialgleichung die bekannte geschlossene Lösung zu diskretisieren und so eine exaktere Simulation zu erhalten. Da in Kapitel 9.6 allerdings ein Finanzmodell betrachtet wird, das einen Korrelationsterm beinhaltet, muss auf die mittels des Euler-Maruyama-Schemas diskretisierte Form der stochastischen Differentialgleichung zurückgegriffen werden. Für die Simulation einer diskretisierten Lösung beziehungsweise einer diskretisierten stochastischen Differentialgleichung ist es notwendig, standardnormalverteilte Zufallszahlen zu erzeugen, um die Zuwächse des den gesamten Prozess stochastisch antreibenden Wiener-Prozesses zu erstellen. Entsprechend der Ausführungen in Kapitel 5 zur Diskretisierung entsprechen die Zuwächse eines Wiener-Prozesses gerade dem Produkt unabhängig standardnormalverteilter Zufallszahlen mit einem festen Faktor Δt , dem Zeitzuwachs:

$$\Delta W_i = \sqrt{\Delta t} Z_i.$$

Generierung standardnormalverteilter Zufallszahlen

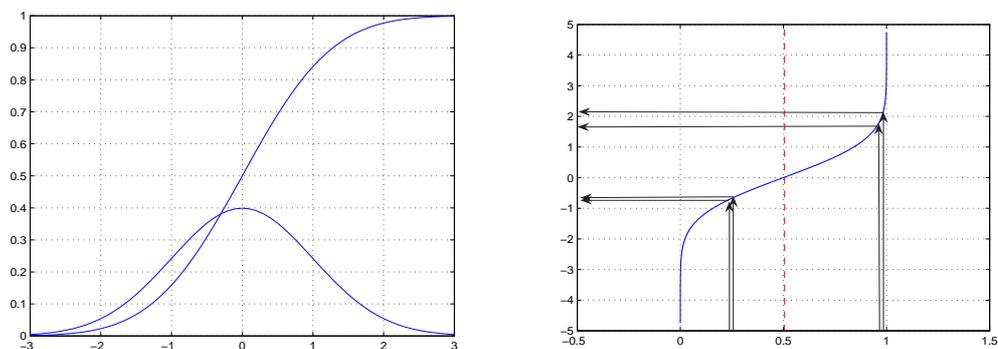
Die folgenden Ausführungen sind an [Gla04] orientiert. Als Grundlage für die Generierung unabhängig standardnormalverteilter Zufallsvariablen werden auf dem Einheitsintervall $(0, 1)$ gleichverteilte Zufallszahlen generiert. Aus diesen werden mittels Transformation standardnormalverteilte Zufallszahlen gewonnen.¹ Für die Transformation wird die Idee der Invertierung der kumulativen Normalverteilung verwendet. In Tabelle 6.1 findet sich eine grobe Skizzierung eines Algorithmus, der eine standardnormalverteilte Zufallszahl zurückgibt. Abbildung 6.1(a) zeigt die Wahrscheinlichkeitsdichte der Standardnormalver-

GENERIERUNG EINER STANDARDNORMALVERTEILTEN ZUFALLSZAHL		
INPUT:	$u \sim U(0, 1)$	Realisierung einer auf dem Einheitsintervall gleichverteilten Zufallsvariable
	$x = \Phi^{-1}(u);$	Transformation mittels inverser (kumulativer) Normalverteilung (Rückgriff auf Näherungsverfahren)
OUTPUT:	$x \sim N(0, 1)$	Standardnormalverteilte Zufallszahl

Tabelle 6.1.: Algorithmus zur Generierung einer standardnormalverteilten Zufallszahl.

¹Vergleiche hierzu zum Beispiel [Gla04] und [Sey00].

teilung und die ihr entsprechende kumulative Normalverteilung. Dabei entspricht die kumulative Normalverteilung $\Phi^{-1}(u)$ gerade dem Integral von $-\infty$ bis u über die Wahrscheinlichkeitsdichte der zugrundeliegenden Normalverteilung $\Phi(x)$. Da für dieses Integral keine geschlossene Lösung existiert, wird auf Näherungsverfahren zurückgegriffen. Dabei sollte im Idealfall $\Phi(\Phi^{-1}(u)) = u$ für alle $u \in [0, 1]$ gelten. Für die Transformation wird die



(a) Wahrscheinlichkeitsdichte der Standardnormalverteilung und die entsprechende kumulative Normalverteilung.

(b) Inverse (kumulative) (Standard-) Normalverteilung und das Prinzip des Transformationsverfahrens.

Abbildung 6.1.: Zur Generierung standardnormalverteilter Zufallszahlen mittels der inversen (kumulativen) Normalverteilung.

inverse (kumulative) Normalverteilung, wie in Abbildung 6.1(b) dargestellt, benötigt. Zur Näherung wird das von Moro [Mor95] erweiterte Approximationsverfahren von Beasley und Springer [BS77] verwendet. Dieses Verfahren basiert auf der ursprünglich von Beasley und Springer verwendeten rationalen Approximation für $0.5 \leq u \leq 0.92$

$$\Phi^{-1}(u) \approx \frac{\sum_{i=0}^3 a_i (u - 0.5)^{2i+1}}{1 + \sum_{i=0}^3 b_i (u - 0.5)^{2i}}$$

und einer von Moro vorgeschlagenen Tschebyscheff-Approximation für $0.92 \leq u < 1$

$$\Phi^{-1}(u) \approx \sum_{i=0}^8 c_i (\log(-\log(1 - u)))^i.$$

Die Konstanten a_i , b_i und c_i dieser Teilapproximationen sowie ein Algorithmus des Verfahrens sind in Tabelle 6.2 abgebildet. Durch Ausnutzung der Punktsymmetrie der inversen (kumulativen) Normalverteilung (siehe Abbildung 6.1(b)) ($\Phi^{-1}(1 - u) = -\Phi^{-1}(u)$ für $u \in (0, 1)$) ist es möglich, analog zum ersten Teil der Funktion im Intervall $[0.5, 1)$ auch den zweiten Teil der Funktion in $(0, 0.5]$ mit den oben angeführten Approximationen zu nähern. Dabei werden auch die Polstellen bei 0 und 1 durch die von Moro eingeführte Tschebyscheff-Approximation gut approximiert. Neben seiner hohen Genauigkeit spricht die relativ kurze Rechenzeit für die Verwendung dieses Approximationsverfahrens.

In Tabelle 6.3 ist ein Algorithmus zur Generierung eines Aktienkurses auf Basis der diskretisierten stochastischen Differentialgleichung für die logarithmischen Zuwächse abgebildet.

DAS APPROXIMATIONSVERFAHREN NACH MORO	
Berechne Approximation von $\Phi^{-1}(u)$ mit $u \in (0, 1)$	
INPUT:	$u \in (0, 1)$
Initialisierung:	$a_0 = 2.50662823884; a_1 = -18.61500062529;$ $a_2 = 41.39119773534; a_3 = -25.44106049637;$ $b_0 = -8.47351093090; b_1 = 23.08336743743;$ $b_2 = -21.06224101826; b_3 = 3.13082909833;$ $c_0 = 0.3374754822726147; c_1 = 0.9761690190917186;$ $c_2 = 0.1607979714918209; c_3 = 0.0276438810333863;$ $c_4 = 0.0038405729373609; c_5 = 0.0003951896511919;$ $c_6 = 0.0000321767881768; c_7 = 0.0000002888167364;$ $c_8 = 0.0000003960315187;$
	$y = u - 0.5;$
Fallunterscheidung:	if $ y < 0.42$ $r = y^2;$ $x = y \cdot \frac{((a_3 \cdot r + a_2) \cdot r + a_1) \cdot r + a_0}{(((b_3 \cdot r + b_2) \cdot r + b_1) \cdot r + b_0) \cdot r + 1};$ else $r = u;$ if $y > 0$ $r = 1 - u;$ $r = \log(-\log(r));$ $x = c_0 + r \cdot (c_1 + r \cdot (c_2 + r \cdot (c_3 + r \cdot (c_4 + r \cdot (c_5 + r \cdot (c_6 + r \cdot (c_7 + r \cdot c_8))))))));$ if $y < 0$ $x = -x;$
OUTPUT:	x Approximation von $\Phi^{-1}(u)$

Tabelle 6.2.: Algorithmus zur Berechnung der inversen (kumulativen) Normalverteilung nach dem Approximationsverfahren von Moro.

SIMULATION EINES AKTIENKURSES	
$S_i = S(t_i)$ für $i \in \{0, \dots, N - 1\}$	
INPUT:	Δt und $S_0 > 0, \mu, \sigma$ und Zeitschrittzahl N
Schleife:	for $i = 1, 2, \dots, N$ Ziehe $Z \sim N(0, 1);$ $\Delta W = \sqrt{\Delta t} Z;$ $\Delta \ln S_i = \mu \Delta t + \sigma \Delta W;$ $S_i = S_{i-1} + S_{i-1} \Delta \ln S_i;$
OUTPUT:	S_1, S_2, \dots, S_N

Tabelle 6.3.: Algorithmus zur Simulation eines Aktienkurses.

6.2. Short Rate

Zur Beschreibung der Zuwächse der Short Rate wird im Rahmen dieser Arbeit das CIR-Modell verwendet. Entsprechend der Diskretisierung mittels des Euler-Maruyama-Schema ist in Tabelle 6.4 ein Algorithmus zur Simulation angegeben. Für die Generierung der benötigten standardnormalverteilten Zufallszahlen können die in den Tabellen 6.1 und 6.2 dargestellten Algorithmen verwendet werden.

SIMULATION EINER SHORT RATE NACH DEM CIR-MODELL	
$r_i = \sum_{k=0}^{i-1} \Delta r_k$ mit $\Delta r_i = r_{i+1} - r_i$ und $r_i = r(t_i)$ für $i \in \{0, \dots, N-1\}$	
INPUT:	Δt und $r_0 > 0$, b , m , σ und Zeitschrittzahl N
Schleife:	for $i = 1, 2, \dots, N$ Ziehe $Z \sim N(0, 1)$; $\Delta W = \sqrt{\Delta t} Z$; $\Delta r_i = b(m - r_{i-1})\Delta t + \sigma\sqrt{r_{i-1}}\Delta W$; $r_i = r_{i-1} + \Delta r_i$;
OUTPUT:	r_1, r_2, \dots, r_N

Tabelle 6.4.: Algorithmus zur Simulation einer Short Rate nach dem CIR-Modell.

6.3. Ein korreliertes Finanzmodell (Aktienkurs und Short Rate)

Soll das in Kapitel 5 Tabelle 5.2 vorgestellte korrelierte Finanzmodell simuliert werden, so kann auf die den letzten Kapiteln entsprechende Simulation der einzelnen Faktoren (Aktienkurs oder Short Rate) zurückgegriffen werden. Lediglich bei der Generierung des Wiener Prozesses für den Aktienkurs muss die diskretisierte Gleichung durch die aus Tabelle 5.2 ersetzt werden. Tabelle 6.5 zeigt einen entsprechenden Algorithmus.

SIMULATION EINES KORRELIERTEN FINANZMODELLS	
(AKTIENKURS & SHORT RATE NACH DEM CIR-MODELL)	
$S_i = S(t_i)$ und $r_i = \sum_{k=0}^{i-1} \Delta r_k$ mit $\Delta r_i = r_{i+1} - r_i$ sowie $r_i = r(t_i)$ für $i \in \{0, \dots, N-1\}$	
INPUT:	Δt und $S_0 > 0$, $r_0 > 0$, μ , σ_1 , b , m , σ_2 , ρ und N
Schleife:	for $i = 1, 2, \dots, N$ Ziehe $Z_1 \sim N(0, 1)$; Ziehe $Z_2 \sim N(0, 1)$; $\Delta W_1 = \sqrt{\Delta t} Z_1$; $\Delta W_2 = \sqrt{\Delta t} (\rho Z_1 + \sqrt{1 - \rho^2} Z_2)$; $\Delta r_i = b(m - r_{i-1})\Delta t + \sigma\sqrt{r_{i-1}}\Delta W_1$; $\Delta \ln S_i = \mu\Delta t + \sigma\Delta W_2$; $r_i = r_{i-1} + \Delta r_i$; $S_i = S_{i-1} + S_{i-1}\Delta \ln S_i$;
OUTPUT:	Korrelierte Raten r_1, r_2, \dots, r_N und S_1, S_2, \dots, S_N

Tabelle 6.5.: Algorithmus zur Simulation eines korrelierten Finanzmodells bestehend aus einem Aktienkurs und einer Short Rate nach dem CIR-Modell.

Kapitel 7.

Dichte-Approximationen

Das später in Kapitel 8.4 vorgestellte Histogramm-Matching-Verfahren zur Parameterschätzung basiert auf dem Vergleich von Dichte-Approximationen in Form von Histogrammen. In diesem Kapitel sollen daher Dichte-Approximationen (verschiedener Dimensionen) auf Grundlage unterschiedlicher Basisfunktionen vorgestellt werden.

In 7.1 wird die Dichte-Approximation mittels des Tensorprodukt-Ansatzes erklärt, bevor in 7.2 der Dünngitter-Ansatz für Dichte-Approximationen vorgestellt wird (vergleiche [BG04], [GSZ92], [Gar04], [Mer05]).

Im Folgenden seien N Datenpunkte $\underline{x}_j \in \mathbb{R}^d$ ($1 \leq j \leq N$) der endlichen Dimension d gegeben, auf deren Grundlage mittels der vorgestellten Dichte-Approximationen eine möglichst gute Näherung $\rho(\underline{x})$ mit $\underline{x} = (x_1, \dots, x_d)^T \in \mathbb{R}^d$ für die zugrunde liegende (unbekannte) wahre Dichtefunktion gesucht wird.

7.1. Der Tensorprodukt-Ansatz

Gemäß dem Tensorprodukt-Ansatz wird die zu approximierende Dichtefunktion mittels einer endlichen Anzahl von Basisfunktionen $\varphi_{\underline{l}, \underline{i}}(\underline{x})$ und Gewichtskoeffizienten $w_{\underline{l}, \underline{i}}$ auf $\Omega \subset \mathbb{R}^d$ diskret dargestellt durch:

$$\rho_{\underline{l}}^{PA}(\underline{x}) = \sum_{i_1=1}^{n_{l_1}} \cdots \sum_{i_d=1}^{n_{l_d}} w_{\underline{l}, \underline{i}} \cdot \varphi_{\underline{l}, \underline{i}}(\underline{x}) \quad \text{für } \underline{x} \in \Omega. \quad (7.1)$$

Der (Multi-)Index $\underline{l} = (l_1, \dots, l_d)^T \in \mathbb{N}^d$ fasst die Verfeinerungslevel bezüglich der einzelnen Richtungen $r \in \{1, 2, \dots, d\}$ zusammen. Diese Verfeinerungslevel bestimmen insbesondere über den (Multi-)Index $\underline{i} = (i_1, \dots, i_d)^T$ die Anzahl der Gewichte und Basisfunktionen. Die Tensorprodukt-Basis

$$\varphi_{\underline{l}, \underline{i}}(\underline{x}) = \prod_{r=1}^d \varphi_{l_r, i_r}(x_r) = \varphi_{l_1, i_1}(x_1) \cdot \varphi_{l_2, i_2}(x_2) \cdot \dots \cdot \varphi_{l_d, i_d}(x_d) \quad (7.2)$$

setzt sich aus n_{l_r} eindimensionalen Funktionen $\varphi_{l_r, i_r}(x_r)$ pro Richtung $r \in \{1, 2, \dots, d\}$ zusammen ($1 \leq i_r \leq n_{l_r}$).

Anmerkung 7.1.1 [DIE FORM DER TENSORPRODUKT-BASIS]

Dienen die äußeren Grenzen des Trägers Ω der Tensorprodukt-Basis auch als Stützstellen für die eindimensionalen Basisfunktionen $\varphi_{l_r, i_r}(x_r)$, so wird von einer geschlossenen Form der Tensorprodukt-Basis gesprochen. Andernfalls handelt es sich um eine offene oder um eine halboffene Form der Tensorprodukt-Basis – je nachdem ob zumindest eine der beiden

gegenüberliegenden Grenzen Stützstelle ist.

In Abbildung 7.2(a) ist eine halboffene Form, in Abbildung 7.4(a) eine offene und in Abbildung 7.4(b) eine geschlossene Form der Tensorproduktbasis dargestellt. Um welche Basisfunktionen es sich dabei genau handelt, wird in Kapitel 7.1.1 erläutert.

Für die weiteren Ausführungen werden folgende Annahmen getroffen:

Annahme 7.1.2 [ANZAHL DER EINDIMENSIONALEN BASISFUNKTIONEN UND GEWICHTSKOEFFIZIENTEN PRO RICHTUNG $r \in \{1, 2, \dots, d\}$]

Es sei $n_{l_r} = O(2^{l_r})$ mit

$$n_{l_r} = \begin{cases} 2^{l_r} + 1 & \text{mit } l_r \geq 0 & \text{bei geschlossener Form der Basis,} \\ 2^{l_r} & \text{mit } l_r \geq 0 & \text{bei halboffener Form der Basis,} \\ 2^{l_r} - 1 & \text{mit } l_r \geq 1 & \text{bei offener Form der Basis.} \end{cases} \quad (7.3)$$

Damit kommt es zu einer ungefähren Verdopplung der Anzahl der eindimensionalen Basisfunktionen $\varphi_{l_r, i_r}(x_r)$ von Verfeinerungslevel zu Verfeinerungslevel. Eine Erhöhung des Verfeinerungslevels bezüglich einer Richtung $r \in \{1, 2, \dots, d\}$ führt dann außerdem zu einer (ungefähren) Verdopplung der Anzahl ($\prod_{r=1}^d n_{l_r}$) der mehrdimensionalen Basisfunktionen $\varphi_{\underline{l}, \underline{i}}(\underline{x}) = \prod_{r=1}^d \varphi_{l_r, i_r}(x_r)$ und der entsprechenden Gewichtskoeffizienten $w_{\underline{l}, \underline{i}}$.

Annahme 7.1.3 [SCHRITTWEITE IN RICHTUNG $r \in \{1, 2, \dots, d\}$]

Es bezeichne b_r die Breite des Träger-Intervalls der Tensorprodukt-Basis $\varphi_{\underline{l}, \underline{i}}(\underline{x})$ bezüglich der Richtung $1 \leq r \leq d$. Dann ist die Schrittweite in Richtung r gegeben durch

$$h_{l_r} = \begin{cases} \frac{b_r}{n_{l_r}-1} & \text{bei geschlossener Form der Basis,} \\ \frac{b_r}{n_{l_r}} & \text{bei halboffener Form der Basis,} \\ \frac{b_r}{n_{l_r}+1} & \text{bei offener Form der Basis.} \end{cases}$$

Also ist mit (7.3) stets

$$h_{l_r} = \frac{b_r}{2^{l_r}}. \quad (7.4)$$

7.1.1. Die Wahl der Basisfunktionen

Zur Vereinfachung wird im Folgenden davon ausgegangen, dass sich das Gebiet, in dem sich die N Datenpunkte befinden und auf dem die Basisfunktionen definiert sein sollen, auf den d -dimensionalen Einheitswürfel (Abbildung 7.1 ¹) skaliert werden kann. Das heißt, von jetzt an wird

$$\Omega = [0, 1]^d$$

als Trägergebiet für die Tensorprodukt-Basis angenommen. Dies impliziert, dass $b_r = 1$ und somit nach (7.4)

$$h_{l_r} = 2^{-l_r}$$

für alle Richtungen $1 \leq r \leq d$ gilt.

¹Dieses Bild basiert auf dem Bild „Parallelprojektionen der 0- bis 5-dimensionalen Würfelanaloge“ aus der freien Enzyklopädie Wikipedia und steht unter der GNU-Lizenz für freie Dokumentation. Der Urheber des Bildes ist Karl Bednarik (22.11.2002).

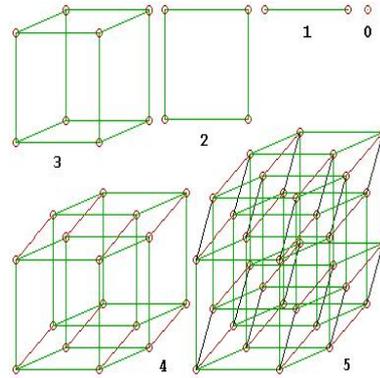


Abbildung 7.1.: Parallelprojektionen der 0- bis 5-dimensionalen Würfelanaloge.

7.1.2. Die Basis der charakteristischen Funktionen

Das Trägergebiet Ω wird in jede der Richtungen $r \in \{1, 2, \dots, d\}$ in n_{l_r} disjunkte Einzelintervalle I_{l_r, i_r} der festen Breite h_{l_r} unterteilt:

$$I_{l_r, i_r} = \begin{cases} [i_r \cdot h_{l_r}, (i_r + 1) \cdot h_{l_r}), & 0 \leq i_r < n_{l_r} - 1 \\ [(n_{l_r} - 1) \cdot h_{l_r}, n_{l_r} \cdot h_{l_r}] = [(n_{l_r} - 1) \cdot h_{l_r}, 1], & i_r = n_{l_r} - 1. \end{cases} \quad (7.5)$$

Da hier lediglich die „linken“ äußeren Grenzen als zusätzliche Stützstellen für die Tensorprodukt-Basis dienen und es sich damit um eine halboffene Form der Basis handelt, gilt entsprechend (7.3)

$$n_{l_r} = 2^{l_r}.$$

Die der Tensorprodukt-Basis über (7.2) zugrunde liegenden eindimensionalen Basisfunktionen $\varphi_{l_r, i_r}(x_r)$ werden nun als die charakteristischen Funktionen $\chi_{I_{l_r, i_r}}(x_r)$ der Teilintervalle aus (7.5) gewählt:

$$\varphi_{l_r, i_r}(x_r) : [0, 1] \rightarrow \mathbb{R} \quad \text{mit} \quad \varphi_{l_r, i_r}(x_r) = \chi_{I_{l_r, i_r}}(x_r) = \begin{cases} 1, & x_r \in I_{l_r, i_r} \\ 0, & \text{sonst} \end{cases}$$

für $l_r \in \mathbb{N}$ und $0 \leq i_r < n_{l_r}$. So ergibt sich bezüglich der Richtung r ein eindimensionales halboffenes Gitter G_{l_r} mit äquidistanter Schrittweite h_{l_r} und den n_{l_r} Stützstellen

$$x_{l_r, i_r} = i_r \cdot h_{l_r} \quad \text{für} \quad 0 \leq i_r \leq n_{l_r} - 1 = 2^{l_r} - 1.$$

Das Gesamtgitter ist ein Rechteckgitter mit Schrittweite $h_{l_r} = 2^{-l_r}$ in Richtung r , das sich aus den Punkten

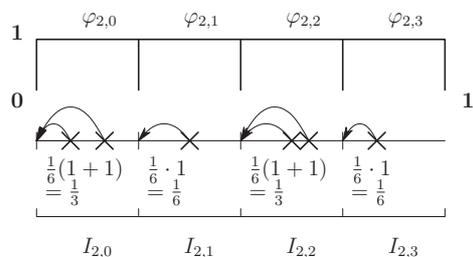
$$x_{l, \underline{i}} = (x_{l_1, i_1}, x_{l_2, i_2}, \dots, x_{l_d, i_d})^T$$

zusammensetzt. Dabei wird deutlich, dass der (Multi-)Index \underline{i} gerade den Ort des Gitterpunktes bestimmt. Dieses Gitter

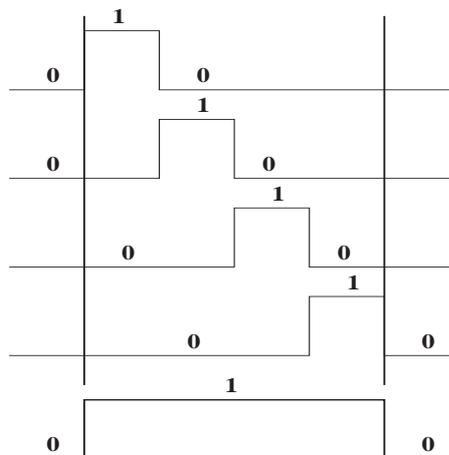
$$G_l = G_{l_1} \times \dots \times G_{l_d} = \{x_{l, \underline{i}} \mid i_r = 0, 1, \dots, 2^{l_r} - 1, \quad r \in \{0, 1, \dots, d\}\}$$

ist zwar in Bezug auf die einzelnen Richtungen äquidistant, kann aber abhängig von den Verfeinerungsleveln l_1, l_2, \dots, l_d unterschiedliche Schrittweiten für die verschiedenen Richtungen haben.

Abbildung 7.2 veranschaulicht die eindimensionale Tensorprodukt-Basis der charakteristischen Funktionen zum Verfeinerungslevel $l_1 = 2$.



(a) Veranschaulichung des Einzugsbereichs der Stützstellen und des Aufbaus der Gewichtskoeffizienten mittels exemplarischer Datenpunkte.



(b) Die eindimensionalen charakteristischen Basisfunktionen $\varphi_{2,i}(x_{2,i})$ für $i = 0, \dots, 3$ und ihre Summe (siehe (7.12)).

Abbildung 7.2.: Die eindimensionale Tensorprodukt-Basis der charakteristischen Funktionen zum Verfeinerungslevel $l_1 = 2$.

7.1.3. Die Basis der stückweise linearen Hutfunktionen

Das Trägergebiet wird in jeder Richtung r in sich überlappende Einzelintervalle unterteilt:

$$I_{l_r, i_r} = [(i_r - 1) \cdot h_{l_r}, (i_r + 1) \cdot h_{l_r}] \cap [0, 1].$$

Auf diesen Intervallen werden die eindimensionalen Basisfunktionen definiert als

$$\begin{aligned} \varphi_{l_r, i_r}(x_r) &= \begin{cases} 1 - |\frac{1}{h_{l_r}} \cdot x_r - i_r|, & x_r \in I_{l_r, i_r} \\ 0, & \text{sonst} \end{cases} \\ &= \begin{cases} \frac{1}{h_{l_r}} \cdot x_r, & x_r \in [(i_r - 1) \cdot h_{l_r}, i_r \cdot h_{l_r}] \cap [0, 1] \\ 1 - (\frac{1}{h_{l_r}} \cdot x_r - i_r), & x_r \in [i_r \cdot h_{l_r}, (i_r + 1) \cdot h_{l_r}] \cap [0, 1] \\ 0, & \text{sonst} \end{cases} \end{aligned}$$

mit $0 \leq i_r \leq 2^{l_r}$ für die geschlossene Form der Tensorprodukt-Basis und $1 \leq i_r < 2^{l_r}$ für die offene Form (vergleiche Abbildung 7.3).

Im geschlossenen Fall sind es demnach genau $n_{l_r} = 2^{l_r} + 1$ Funktionen, während es im offenen lediglich $n_{l_r} = 2^{l_r} - 1$ sind. Ferner gilt für die Schrittweiten entsprechend (7.4)

$$h_{l_r} = \left. \begin{cases} \frac{1}{n_{l_r} - 1} & \text{im geschlossenen Fall} \\ \frac{1}{n_{l_r} + 1} & \text{im offenen Fall} \end{cases} \right\} = 2^{-l_r}.$$

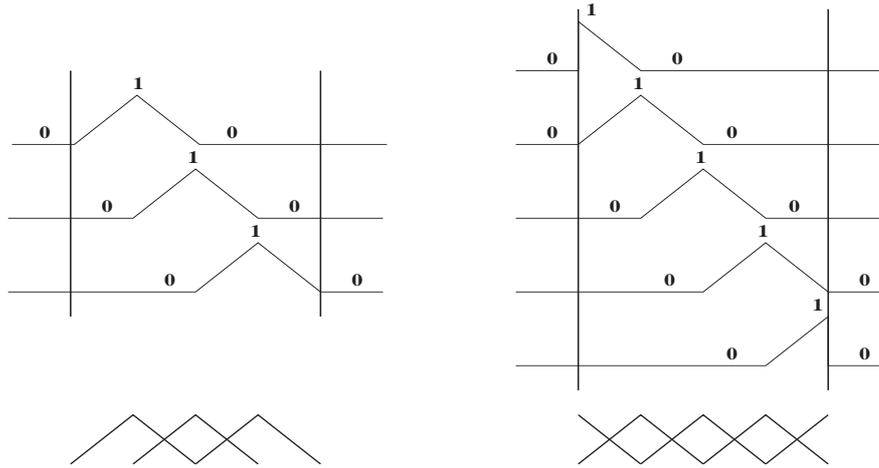


Abbildung 7.3.: Dargestellt sind links die stückweise linearen Hutfunktionen, die zusammen die offene, rechts diejenigen, die die geschlossene eindimensionale Tensorprodukt-Basis zum Verfeinerungslevel $l_1 = 2$ bilden.

Auch hier ergibt sich ein (je nach Wahl der Basis geschlossenes oder offenes) Gitter G_{l_r} bestehend aus den $\prod_{r=1}^d n_{l_r}$ Stützstellen $x_{\underline{l}, \underline{i}} = (x_{l_1, i_1}, x_{l_2, i_2}, \dots, x_{l_d, i_d})^T$ mit

$$x_{l_r, i_r} = i_r \cdot h_{l_r} \quad \begin{cases} \text{für } 0 \leq i_r \leq n_{l_r} - 1 = 2^{l_r} & \text{im geschlossenen Fall} \\ \text{und} \\ \text{für } 1 \leq i_r < n_{l_r} + 1 = 2^{l_r} & \text{im offenen Fall.} \end{cases}$$

Bei dem resultierenden Gesamtgitter

$$G_{\underline{l}} = G_{l_1} \times \dots \times G_{l_d} = \begin{cases} \{x_{\underline{l}, \underline{i}} \mid i_r = 0, 1, \dots, 2^{l_r}, r \in \{1, \dots, d\}\} & \text{im geschlossenen Fall} \\ \{x_{\underline{l}, \underline{i}} \mid i_r = 1, \dots, 2^{l_r} - 1, r \in \{1, \dots, d\}\} & \text{im offenen Fall} \end{cases}$$

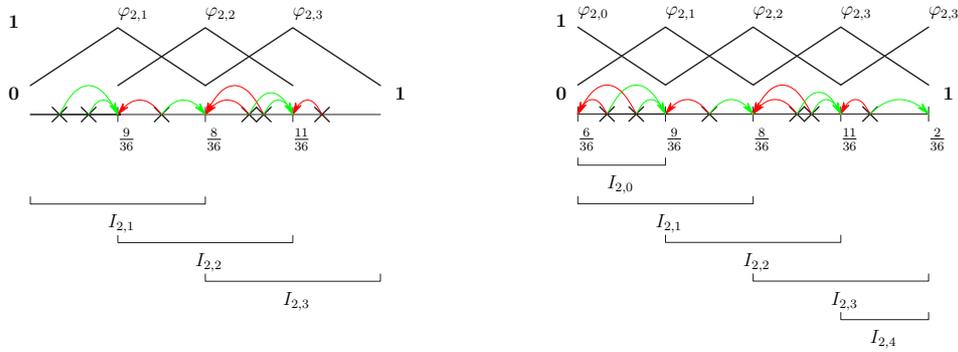
handelt es sich wieder um ein in Bezug auf die einzelnen Richtungen äquidistantes Rechteckgitter, dessen Schrittweiten h_{l_r} in die einzelnen Richtungen sich jedoch in Abhängigkeit von den Verfeinerungsleveln unterscheiden können.

Abbildung 7.4 veranschaulicht den Einzugsbereich der Stützstellen und die resultierenden Gewichte der stückweise linearen Hutfunktionen im Eindimensionalen zum Verfeinerungslevel 2.

Die $\varphi_{\underline{l}, \underline{i}}(\underline{x})$ bilden eine nodale Basis (Knotenbasis) für den Raum $V_{\underline{l}}$ der über $G_{\underline{l}}$ stückweise d -linearen Hutfunktionen

$$V_{\underline{l}} = \begin{cases} \text{span}\{\varphi_{\underline{l}, \underline{i}}(\underline{x}) \mid 0 \leq i_r \leq 2^{l_r} \text{ für } r \in \{1, 2, \dots, d\}\} & \text{im geschlossenen Fall,} \\ \text{span}\{\varphi_{\underline{l}, \underline{i}}(\underline{x}) \mid 1 \leq i_r \leq 2^{l_r} - 1 \text{ für } r \in \{1, 2, \dots, d\}\} & \text{im offenen Fall.} \end{cases}$$

Abbildung 7.5 zeigt die eindimensionalen (offenen) nodalen Basen von V_1 , V_2 und V_3 . Beispiele für zweidimensionale nodale Basen finden sich in Abbildung 7.6. Dabei sind hier, obwohl nicht zwingend vorgeschrieben, die Verfeinerungslevel in beide Richtungen identisch.



(a) Der offene Fall.

(b) Der geschlossene Fall.

Abbildung 7.4.: Veranschaulichung des Einzugsbereichs der Stützstellen und des Aufbaus der Gewichtskoeffizienten der stückweise linearen Hutfunktionen im Ein-dimensionalen mittels exemplarischer Datenpunkte.

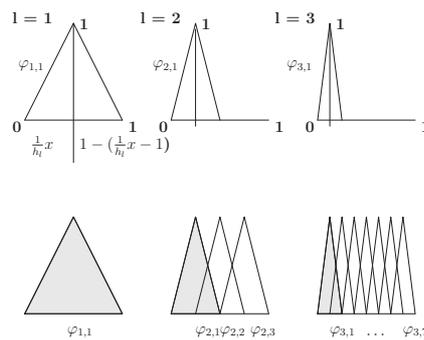
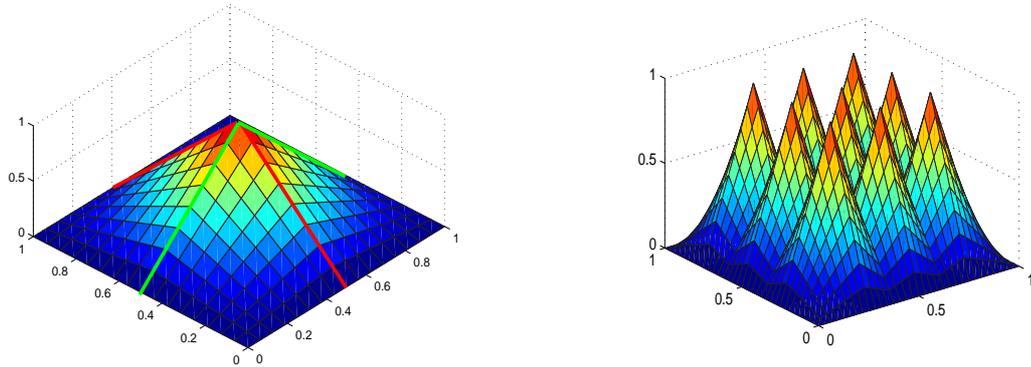


Abbildung 7.5.: Stückweise lineare Hutfunktionen als Basisfunktionen in einer Dimension (φ_{l_r, i_r} für feste Richtung r) zum Verfeinerungslevel 1, 2 und 3 (skaliert auf $[0, 1]$).



(a) Zweidimensionale Basisfunktion $\varphi_{\underline{l},1}$ zum Verfeinerungslevel $\underline{l} = (1, 1)^T$. Die rote Funktion entspricht der eindimensionalen Basisfunktion $\varphi_{l_1,1}$ der ersten Richtung, die grüne Funktion die der zweiten Richtung ($\varphi_{l_2,1}$).

(b) Zweidimensionale Basisfunktionen $\varphi_{\underline{l},i}$ zum Verfeinerungslevel $\underline{l} = (2, 2)^T$ für $\underline{i} = (i_1, i_2)^T$ mit $1 \leq i_r \leq 3$ ($1 \leq r \leq 2$).

Abbildung 7.6.: Stückweise lineare Hutfunktionen als Basisfunktionen in zwei Dimensionen (skaliert auf $[0, 1] \times [0, 1]$).

7.1.4. Die Bestimmung der Gewichtskoeffizienten

Von folgenden Eigenschaften der Tensorprodukt-Basis, die insbesondere für die Bestimmung der Gewichtskoeffizienten maßgeblich sind, wird im Weiteren ausgegangen:

- Jeder Punkt wird durch die Tensorprodukt-Basis (höchstens) einfach gewertet:

$$\sum_{i_1=1}^{n_{l_1}} \cdots \sum_{i_d=1}^{n_{l_d}} \varphi_{\underline{l},\underline{i}}(\underline{x}) = \sum_{i_1=1}^{n_{l_1}} \cdots \sum_{i_d=1}^{n_{l_d}} \varphi_{l_1,i_1}(x_1) \cdot \varphi_{l_2,i_2}(x_2) \cdot \cdots \cdot \varphi_{l_d,i_d}(x_d) \leq 1 \quad (7.6)$$

Gleichheit gilt nur für die geschlossene Form der Tensorprodukt-Basis. Ansonsten kann es abhängig von der Wahl der Funktionen $\varphi_{l_r,i_r}(x_r)$ dazu kommen, dass Datenpunkte, die im äußersten Einzugsgebiet der Basisfunktionen liegen, nur anteilig gewertet werden (siehe dazu auch Abbildung 7.4). Unter der Voraussetzung, dass alle N Datenpunkte im Einzugsbereich der Basisfunktionen liegen, gilt mit (7.6)

$$\sum_{j=1}^N \sum_{i_1=1}^{n_{l_1}} \cdots \sum_{i_d=1}^{n_{l_d}} \varphi_{\underline{l},\underline{i}}(x_j) = \hat{N} \leq N. \quad (7.7)$$

Dabei bezeichnet \hat{N} den tatsächlich gewerteten Anteil der N Datenpunkte. Für den geschlossenen Fall gilt gerade $\hat{N} = N$.

- Die Integration einer einzelnen Basisfunktion über ein Gebiet Ω , das ihr Träger-

Intervall $I_{\underline{l},i} \subset \Omega$ vollständig enthält², führt zu

$$\begin{aligned}
 \int_{\Omega} \varphi_{\underline{l},i}(\underline{x}) \, d\underline{x} &= \int_{\Omega_1} \cdots \int_{\Omega_d} \varphi_{l_1,i_1}(x_1) \cdots \varphi_{l_d,i_d}(x_d) \, dx_1 \cdots dx_d \\
 &= b_1 2^{-l_1} \cdot b_2 2^{-l_2} \cdots b_d 2^{-l_d} \\
 &= (b_1 \cdot b_2 \cdots b_d) \cdot 2^{-l_1} \cdot 2^{-l_2} \cdots 2^{-l_d} \\
 &= (b_1 \cdot b_2 \cdots b_d) \cdot 2^{-(l_1+l_2+\dots+l_d)} \\
 &= b \cdot 2^{-|\underline{l}|}
 \end{aligned} \tag{7.8}$$

mit $b = b_1 \cdot b_2 \cdots b_d$.

Anmerkung 7.1.4

Diese Eigenschaften sind, wie leicht nachzuvollziehen ist, sowohl von der Tensorprodukt-Basis der charakteristischen Funktionen als auch der stückweise d -linearen Hutfunktionen erfüllt.

Um die Gewichtskoeffizienten für die Tensorproduktbasis geeignet zu wählen, spielt zusätzlich folgende Dichte-Eigenschaft eine entscheidende Rolle:

Satz 7.1.5

Für eine beliebige Dichtefunktion $\rho(\underline{x})$ gilt

$$\rho(\underline{x}) \geq 0 \text{ sowie } \int_{-\infty}^{\infty} \rho(\underline{x}) \, d\underline{x} = 1 \text{ (vergleiche A.3)}. \tag{7.9}$$

Da es sich bei der zu approximierenden Funktion um eine Dichtefunktion handelt, muss diese Eigenschaft erfüllt sein. Mit den am Anfang des Kapitels aufgeführten Eigenschaften der Tensorproduktbasis wird folgendes Resultat erhalten:

Satz 7.1.6

Die Dichteigenschaft (7.9) ist im Produkt-Ansatz genau dann erfüllt, wenn die Gewichtskoeffizienten $w_{\underline{l},i}$ gewählt sind als

$$w_{\underline{l},i} = \frac{2^{|\underline{l}|}}{b \cdot \hat{N}} \sum_{j=1}^N \varphi_{\underline{l},i}(\underline{x}_j). \tag{7.10}$$

²Das heißt, für jede Richtung $1 \leq r \leq d$ gilt, dass das jeweilige Träger-Intervall I_{l_r,i_r} in Ω_r enthalten ist: $I_{l_r,i_r} \subset \Omega_r$ für $1 \leq r \leq d$.

Beweis: Mit (7.10) gilt

$$\begin{aligned}
 \int \rho_{\underline{l}}^{PA}(\underline{x}) \, d\underline{x} &\stackrel{(7.1)}{=} \int \sum_{i_1=1}^{n_{l_1}} \cdots \sum_{i_d=1}^{n_{l_d}} w_{\underline{l},i} \cdot \varphi_{\underline{l},i}(\underline{x}) \, d\underline{x} \\
 &= \sum_{i_1=1}^{n_{l_1}} \cdots \sum_{i_d=1}^{n_{l_d}} w_{\underline{l},i} \int \varphi_{\underline{l},i}(\underline{x}) \, d\underline{x} \\
 &\stackrel{(7.8)}{=} \sum_{i_1=1}^{n_{l_1}} \cdots \sum_{i_d=1}^{n_{l_d}} w_{\underline{l},i} \cdot b \cdot 2^{-|\underline{l}|} \\
 &\stackrel{(7.10)}{=} \sum_{i_1=1}^{n_{l_1}} \cdots \sum_{i_d=1}^{n_{l_d}} \frac{2^{|\underline{l}|}}{b \cdot \hat{N}} \sum_{j=1}^N \varphi_{\underline{l},i}(\underline{x}_j) \cdot b \cdot 2^{-|\underline{l}|} \\
 &= \sum_{i_1=1}^{n_{l_1}} \cdots \sum_{i_d=1}^{n_{l_d}} \frac{1}{\hat{N}} \sum_{j=1}^N \varphi_{\underline{l},i}(\underline{x}_j) \\
 &= \frac{1}{\hat{N}} \sum_{i_1=1}^{n_{l_1}} \cdots \sum_{i_d=1}^{n_{l_d}} \sum_{j=1}^N \varphi_{\underline{l},i}(\underline{x}_j) \\
 &\stackrel{(7.7)}{=} \frac{1}{\hat{N}} \cdot \hat{N} \\
 &= 1.
 \end{aligned} \tag{7.11}$$

□

Anmerkung 7.1.7 [AUSSAGEN DIE CHARAKTERISTISCHE BASIS BETREFFEND]

In die Wertung durch die charakteristische Basis gehen nur diejenigen Datenpunkte ein, die sich im gesamt betrachteten Quader $\Omega = [0, 1]^d$ befinden (siehe dazu Abbildung 7.2(b) unten). Es gilt

$$\sum_{i_1=1}^{n_{l_1}} \cdots \sum_{i_d=1}^{n_{l_d}} \varphi_{\underline{l},i}(\underline{x}) = \begin{cases} 1, & \underline{x} \in I, \\ 0, & \text{sonst.} \end{cases} \tag{7.12}$$

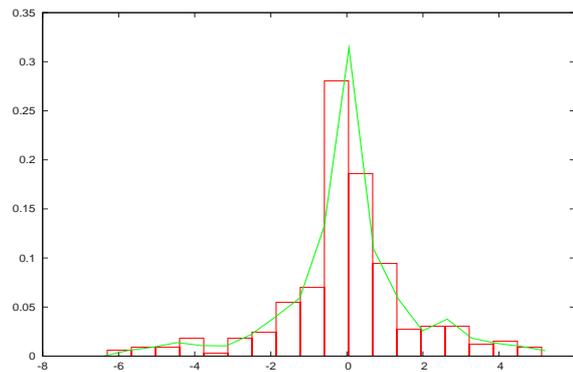
Die betroffenen Datenpunkte werden quasi „gezählt“ und die $2^{|\underline{l}|}$ Gewichte $w_{\underline{l},i}$ der charakteristischen Basis stellen dementsprechend die mit $2^{|\underline{l}|}$ skalierten relativen Häufigkeiten der sich im jeweiligen Teil-Quader $Q_{\underline{l},i} = I_{l_1,i_1} \times I_{l_2,i_2} \times \dots \times I_{l_r,i_r}$ befindenden Datenpunkte dar (siehe dazu Abbildung 7.2(a)).

Anmerkung 7.1.8 [AUSSAGEN DIE STÜCKWEISE d -LINEAREN HUT-BASIS BETREFFEND]

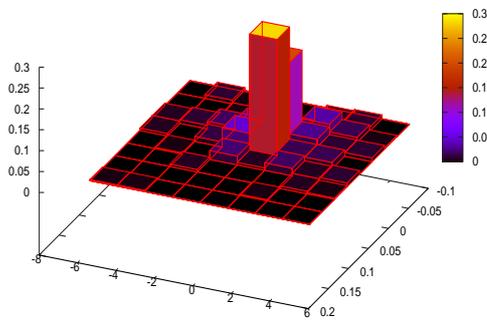
Auch hier gehen nur die im betrachteten Gebiet liegenden Datenpunkte in die Wertung ein. Allerdings wird der einzelne Datenpunkt nicht mehr nur einfach von einer der Basisfunktionen „gezählt“ (wie es bei den charakteristischen Funktionen der Fall ist), sondern er wird anteilig entsprechend seiner Lage von allen umliegenden Basisfunktionen ausgewertet (vergleiche Abbildung 7.4).

Mit Hilfe dieser ermittelten Gewichte und der vorher festgelegten Basisfunktionen lässt sich nun mit (7.1) eine Approximation der wahren Dichte angeben.

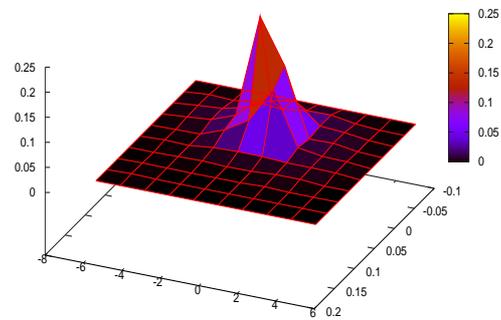
In Abbildung 7.7 sind Beispiele ein- und zweidimensionaler Dichte-Approximationen mittels des Tensorprodukt-Ansatzes auf Grundlage charakteristischer und stückweise linearer Funktionen abgebildet.



(a) Eindimensionale Dichte-Approximationen (rotes Histogramm: charakteristische Basisfunktionen, grüne Kurve: stückweise linearer Basisfunktionen).



(b) Zweidimensionale Dichte-Approximation auf Grundlage charakteristischer Basisfunktionen



(c) Zweidimensionale Dichte-Approximation auf Grundlage stückweise linearer Basisfunktionen

Abbildung 7.7.: Dichte-Approximationen mittels des Tensorprodukt-Ansatzes auf Grundlage charakteristischer und stückweise linearer Funktionen. Zugrunde liegen dieselben Datenpunkte.

7.2. Der Dünngitter-Ansatz

Beim Tensorprodukt-Ansatz führt eine Erhöhung der Dimension zu einem gleichzeitigen exponentiellen Anstieg der Anzahl von Basisfunktionen. Diese mit der Zahl der Dimensionen exponentiell wachsende Komplexität wird auch als der „Fluch der Dimension“ bezeichnet. Um diesen zu „brechen“, können dünne Gitter eingesetzt werden. Beispielsweise ist entsprechend Anmerkung 7.1.2 beim Tensorprodukt-Ansatz die Anzahl der Basisfunktionen pro Richtung $r \in \{1, \dots, d\}$ in Abhängigkeit vom Verfeinerungslevel l_r durch $n_{l_r} = O(2^{l_r})$ gegeben. Damit ergibt sich mit festem Level $l = l_r$ für alle Richtungen r die Gesamtanzahl $n = O\left((2^l)^d\right)$ der Basisfunktionen. Für ein festes Level kommt es demzufolge bei einer Erhöhung der Dimension zum exponentiellen Anstieg der Anzahl von Basisfunktionen.

Beim nun vorgestellten Dünngitter-Ansatz liegen zwar die selben Basisfunktionen wie beim Tensorprodukt-Ansatz zugrunde; aber bei der Differenzentechnik wird lediglich eine Auswahl dieser Basisfunktionen zum Aufbau des dünnen Gitters verwendet. Es werden insgesamt nur $n = O\left(2^l (\log 2^l)^{d-1}\right)$ Basisfunktionen verwendet. Damit bietet sich der Einsatz des Dünngitter-Ansatzes gerade in Bezug auf höhere Dimensionen an, denn einerseits verspricht dieser Ansatz ungefähr dieselbe Genauigkeit der Dichte-Approximation bezüglich desselben Levels bei kleinerem Aufwand. Andererseits ermöglicht die Verwendung des Dünngitter-Ansatzes bei gleichem Aufwand den Einsatz eines feineren Gitters für die Dichteapproximation.

Im Rahmen dieser Arbeit wird die Kombinationstechnik zur Generierung von dünnen Gittern angewendet. Die Kombinationstechnik basiert auf vollen Produktgittern mit äquidistanter Schrittweite bezüglich aller Richtungen.

7.2.1. Herleitung der Dünnen Gitter über die Differenzen der eindimensionalen nodalen Basisfunktionen

In 7.1.3 wurde mittels der stückweisen d -linearen Hutfunktionen eine nodale Basis für den Raum $V_{\underline{l}}$ besprochen. Jetzt soll eine weitere auf den stückweisen d -linearen Hutfunktionen basierende Basis vorgestellt werden, die $V_{\underline{l}}$ aufspannt. Es handelt sich um eine so genannte hierarchische Basis. Diese Basis wird für den offenen Fall eingeführt. Der geschlossene Fall wird weitgehend analog entwickelt.

$$\mathcal{I}_{\underline{l}} := \{\underline{i} \in \mathbb{N}^d \mid 1 \leq i_r \leq 2^{l_r} - 1, \ i_r \text{ ungerade für } r \in \{1, 2, \dots, d\}\} \quad (7.13)$$

sei die Menge der ungeraden Indizes, die im Folgenden verwendet wird. Die hierarchische Basis von $V_{\underline{l}}$ ist definiert als

$$B_{\underline{l}} = \{\varphi_{k, \underline{i}}(\underline{x}) \mid \underline{i} \in \mathcal{I}_{\underline{l}}, \ 0 \leq k_r \leq l_r, \ r \in \{1, 2, \dots, d\}\}.$$

Im Unterschied zu den nodalen Basen bilden die hierarchischen Basen eine aufsteigende Folge von Mengen. Bei einer Erhöhung des Verfeinerungslevels bezüglich einer der Richtungen $r \in \{1, 2, \dots, d\}$ wird die Basis nicht komplett ausgetauscht, sondern es kommen neue Funktionen hinzu. In Abbildung 7.8 findet sich eine Gegenüberstellung von einer eindimensionalen hierarchischen Basis zum Verfeinerungslevel $l = 3$ zu einer entsprechenden nodalen Basis.

Mit Hilfe der in (7.13) definierten Indexmengen können die Differenzräume

$$W_{\underline{l}} = V_{\underline{l}} \setminus \bigcup_{r=1}^d V_{\underline{l}-\underline{e}_r}$$

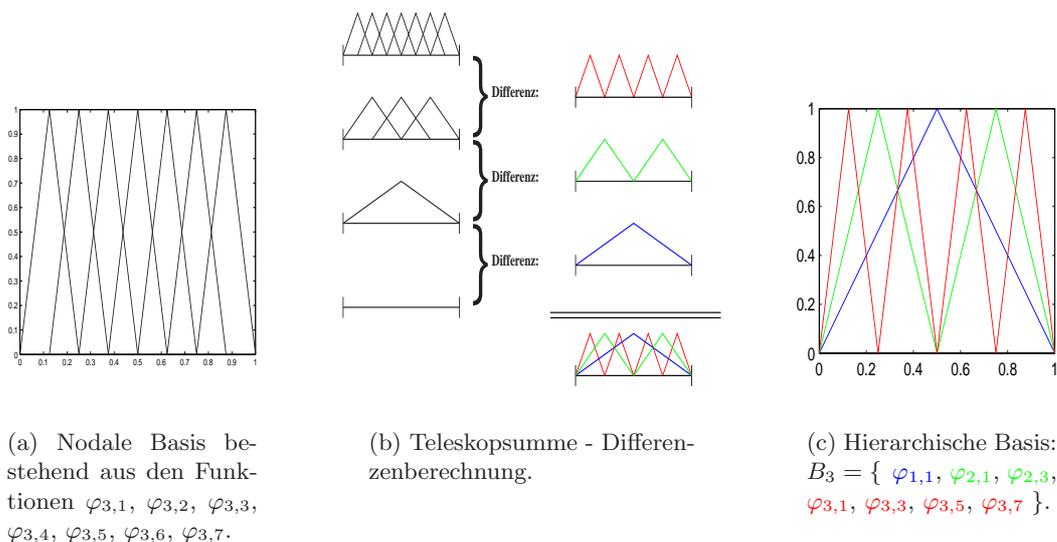


Abbildung 7.8.: Eindimensionale Basen zum Verfeinerungslevel $l = 3$.

aufgespannt werden. Für diese Definition muss zusätzlich angenommen werden, dass $V_{\underline{l}} = \emptyset$, falls $l_r = -1$ für mindestens ein $r \in \{1, 2, \dots, d\}$ gilt. Dabei bezeichnet $\underline{0}$ den d -dimensionalen Nullvektor und \underline{e}_r den r -ten d -dimensionalen Einheitsvektor. Es gilt

$$W_{\underline{l}} = \text{span}\{\varphi_{\underline{l}, \underline{i}}(\underline{x}) \mid \underline{i} \in \mathcal{I}_{\underline{l}}\}.$$

Mit diesen Differenzräumen lässt sich nun eine Definition der Räume $V_{\underline{l}}$ als direkte Summe ihrer Teilräume („Multilevel-Teilraumzerlegung“) angeben:

$$V_{\underline{l}} = \bigoplus_{k_1=1}^l \cdots \bigoplus_{k_d=1}^{l_d} W_{(k_1, \dots, k_d)} = \bigoplus_{\underline{k} \leq \underline{l}} W_{\underline{k}},$$

dabei gilt für zwei d -dimensionale (Multi-)Indizes \underline{a} und \underline{b} , dass

$$\underline{a} \leq \underline{b} \Leftrightarrow \forall_{1 \leq j \leq d} a_j \leq b_j$$

Mittels Abbildung 7.8(b) ist zu sehen, dass diese Dünngitter-Herleitung der Zerlegung einer Folge von vollen Basen in ihre Teleskopsummenglieder entspricht. Dabei werden für die Differenzen die jeweils (von der Höhe ihrer Verfeinerungslevel her) aufeinanderfolgenden Basen voneinander abgezogen. Die Summe dieser einzelnen Glieder führt gerade wieder auf das entsprechende volle Gitter.

In Abbildung 7.9 finden sich zwei Tableaus. Das linke zeigt die zweidimensionalen $V_{\underline{l}}$, das rechte die Differenzräume $W_{\underline{l}}$. Durch Addition der entsprechenden Differenzräume des rechten Schemas ergeben sich die Räume $V_{\underline{k}}$ für festes \underline{k} im linken Schema. Zum Beispiel ergibt sich $V_{(2,3)}$ durch $W_{(1,1)} \oplus W_{(1,2)} \oplus W_{(1,3)} \oplus W_{(2,1)} \oplus W_{(2,2)} \oplus W_{(2,3)}$. Das volle zweidimensionale Gitter zum Verfeinerungslevel $\underline{l} = (3, 3)$ kann entweder mittels $V_{\underline{l}}$ direkt in 7.9(a) abgelesen oder mittels Addition aller in 7.9(b) über der entsprechenden Diagonale liegenden $W_{\underline{l}}$ erhalten werden.

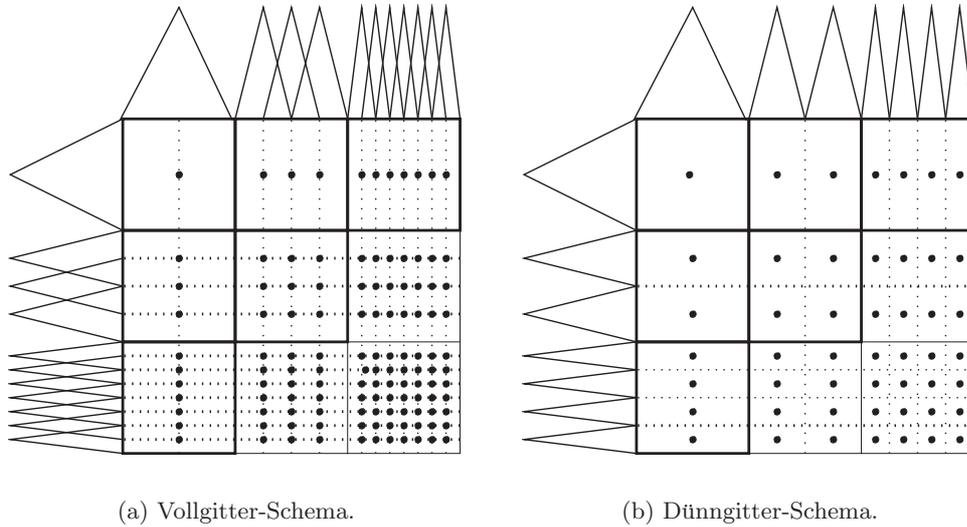


Abbildung 7.9.: Vergleich der zweidimensionalen Schemata des vollen Gitters (Tensorprodukt-Ansatz) und des dünnen Gitters bis zum Verfeinerungslevel $\underline{l} = (l_1, l_2)^T = (3, 3)^T$. Eingezeichnet sind die Stützstellen und ihr jeweiliger Einzugsbereich.

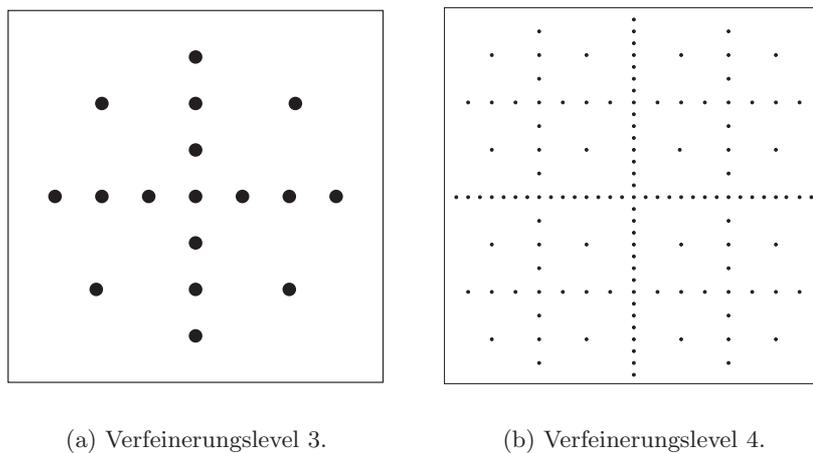
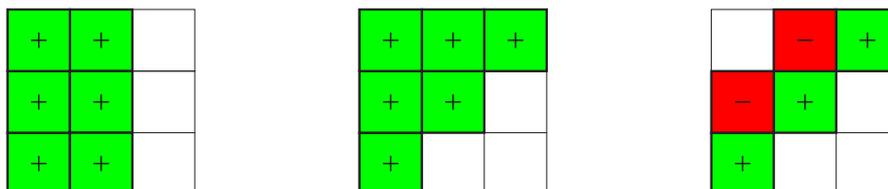


Abbildung 7.10.: (Offene) Dünne Gitter verschiedener Verfeinerungslevel.

7.2.2. Die Kombinationstechnik zur Erstellung von Dünnen Gittern

Für zwei Dimensionen lässt sich anhand der Darstellungen in den Abbildungen 7.9(a) und 7.10(a) leicht nachvollziehen, dass die dünnen Gitter zum Verfeinerungslevel \underline{l} sich durch Subtraktion der Gitter $G_{\underline{l}}$ auf der Superdiagonalen (der oberen Nebendiagonale) von der Summe derer auf der Hauptdiagonalen ergeben (siehe Abbildung 7.11(c)).



(a) Bildung des Vollen Gitters zum Verfeinerungslevel $\underline{l} = (2, 3)$ mittels des Differenzen-Schemas in 7.9(b).

(b) Bildung des Dünnes Gitters zum Verfeinerungslevel $l = 3$ mittels des Differenzen-Schemas in 7.9(b).

(c) Bildung des Dünnes Gitters zum Verfeinerungslevel $l = 3$ auf Grundlage der Kombinationstechnik mittels des Vollgitters-Schemas in 7.9(a).

Abbildung 7.11.: Bildung zweidimensionaler Gitter mittels der in 7.9 abgebildeten Schemata.

Allgemein wird die Dünngitter-Kombinationslösung (zum Verfeinerungslevel l) f_l auf dem Dünngitter-Raum $V_l = \bigoplus_{|\underline{k}| \leq l} W_{\underline{k}}$ durch folgende Überlagerungen der Lösungen von vollen Gittern erhalten:

$$\begin{aligned}
 f_l(\underline{x}) &= \sum_{l \leq |\underline{k}| \leq l+d-1} (-1)^{l+d-|\underline{k}|-1} \binom{d-1}{|\underline{k}|-l} f_{\underline{k}}(\underline{x}) \\
 &= \sum_{q=0}^{d-1} (-1)^q \binom{d-1}{q} \sum_{|\underline{k}|=l-q} f_l(\underline{x}).
 \end{aligned} \tag{7.14}$$

Dabei stellen die Funktionen $f_l(\underline{x})$ diskrete lokale Lösungen dar, die gerade in den Räumen V_l der stückweise d -linearen Funktionen auf dem Gitter G_l existieren. Insbesondere gilt sind die $f_l(\underline{x})$ (mittels eines Finite Elemente-Ansatzes) gegeben als

$$f_l(\underline{x}) = \sum_{i_1=1}^{2^{l_1}} \cdots \sum_{i_d=1}^{2^{l_d}} c_{l,i} \cdot \phi_{l,i}(\underline{x}).$$

7.2.3. Die Gewichtskoeffizienten im Dünngitter-Ansatz

Satz 7.2.1 [DÜNNGITTER-DICHTE-APPROXIMATION MITTELS DIFFERENZENBILDUNG]

Mit den durch

$$\Delta_l(\underline{x}) = \sum_{k_1=0}^1 \cdots \sum_{k_d=0}^1 (-1)^{|\underline{k}|} \underbrace{(2^{|\underline{l}|-|\underline{k}|} - 1)}_{\text{Skalierung}} \rho_{l-\underline{k}}^{PA}(\underline{x}). \tag{7.15}$$

definierten Differenzen ergibt sich die Dünngitter-Dichte-Approximation über

$$\rho_l^{DG}(\underline{x}) = \sum_{|\underline{k}| \leq l+d-1} \underbrace{(2^l - 1)^{-1}}_{\text{Skalierung}} \Delta_{\underline{k}}(\underline{x}). \quad (7.16)$$

Satz 7.2.2 [DIE KOMBINATIONSTECHNIK IN DER DICHTE-APPROXIMATION]
 Wird (7.14) für $f_l(\underline{x}) = \rho_l^{DG}(\underline{x})$ und $f_l(\underline{x}) = \rho_l^{PA}(\underline{x})$ angewendet, so ergibt sich

$$\begin{aligned} \rho_l^{DG}(\underline{x}) &= \sum_{l \leq |\underline{k}| \leq l+d-1} (-1)^{l+d-|\underline{k}|-1} \binom{d-1}{|\underline{k}|-l} \rho_{|\underline{k}|-1}^{PA}(\underline{x}) \\ &= \sum_{q=0}^{d-1} (-1)^q \binom{d-1}{q} \sum_{|\underline{k}|=l-q} \rho_{|\underline{k}|-1}^{PA}(\underline{x}). \end{aligned} \quad (7.17)$$

Mittels der Kombinationstechnik (7.17) lässt sich feststellen, dass die in (7.10) für den Tensorprodukt-Ansatz definierten Gewichte auch für den Dünngitter-Ansatz Gültigkeit haben:

Satz 7.2.3

Mit den Gewichten

$$w_{\underline{l}, \underline{i}} = \frac{2^{|\underline{l}|}}{b \cdot \hat{N}} \sum_{j=1}^N \varphi_{\underline{l}, \underline{i}}(\underline{x}_j) \quad (7.18)$$

ist die Dichteigenschaft (7.9) für den Dünngitter-Ansatz erfüllt. Es gilt

$$\int \rho_l^{DG}(\underline{x}) \, d\underline{x} = 1.$$

Beweis: Mit (7.17) und (7.18) ist:

$$\begin{aligned} \int \rho_l^{DG}(\underline{x}) \, d\underline{x} &= \int \sum_{q=0}^{d-1} (-1)^q \binom{d-1}{q} \sum_{|\underline{k}|=l-q} \rho_{|\underline{k}|-1}^{PA}(\underline{x}) \, d\underline{x} \\ &= \sum_{q=0}^{d-1} (-1)^q \binom{d-1}{q} \sum_{|\underline{k}|=l-q} \int \rho_{|\underline{k}|-1}^{PA}(\underline{x}) \, d\underline{x} \\ &\stackrel{(7.11)}{=} \sum_{q=0}^{d-1} (-1)^q \binom{d-1}{q} \sum_{|\underline{k}|=l-q} 1 \\ &= 1. \end{aligned}$$

□

Satz 7.2.4

Auch die Differenzenbildung (7.16) erfüllt die Dichteigenschaft. Es gilt

$$\int \rho_l^{DG}(\underline{x}) \, d\underline{x} = \sum_{|\underline{k}| \leq l+d-1} \underbrace{(2^l - 1)^{-1}}_{\text{Skalierung}} \Delta_{\underline{k}}(\underline{x}) = 1$$

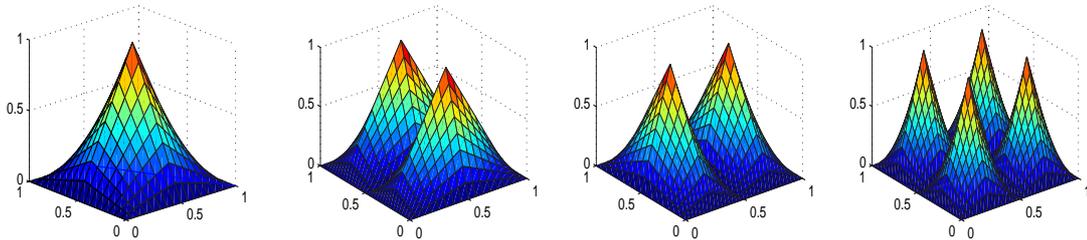
Beweis:

$$\begin{aligned}
 \int \rho_t^{DG}(\underline{x}) \, d\underline{x} &= \int \sum_{|\underline{k}| \leq l+d-1} \underbrace{(2^l - 1)^{-1}}_{\text{Skalierung}} \Delta_{\underline{k}}(\underline{x}) \, d\underline{x} \\
 &= \sum_{|\underline{k}| \leq l+d-1} \underbrace{(2^l - 1)^{-1}}_{\text{Skalierung}} \int \Delta_{\underline{k}}(\underline{x}) \, d\underline{x} \\
 &= \sum_{|\underline{k}| \leq l+d-1} \underbrace{(2^l - 1)^{-1}}_{\text{Skalierung}} \sum_{p_1=0}^1 \cdots \sum_{p_d=0}^1 (-1)^{|\underline{p}|} \cdot \underbrace{(2^{|\underline{k}|-|\underline{p}|} - 1)}_{\text{Skalierung}} \\
 &= 1
 \end{aligned}$$

mit

$$\begin{aligned}
 \int \Delta_l(\underline{x}) \, d\underline{x} &= \int \sum_{p_1=0}^1 \cdots \sum_{p_d=0}^1 (-1)^{|\underline{p}|} \underbrace{(2^{|\underline{l}|-|\underline{p}|} - 1)}_{\text{Skalierung}} \rho_{\underline{l}-\underline{p}}^{PA}(\underline{x}) \, d\underline{x} \\
 &= \sum_{p_1=0}^1 \cdots \sum_{p_d=0}^1 (-1)^{|\underline{p}|} \underbrace{(2^{|\underline{l}|-|\underline{p}|} - 1)}_{\text{Skalierung}} \int \rho_{\underline{l}-\underline{p}}^{PA}(\underline{x}) \, d\underline{x} \\
 &\stackrel{(7.11)}{=} \sum_{p_1=0}^1 \cdots \sum_{p_d=0}^1 (-1)^{|\underline{p}|} \cdot \underbrace{(2^{|\underline{l}|-|\underline{p}|} - 1)}_{\text{Skalierung}} \cdot 1.
 \end{aligned}$$

□

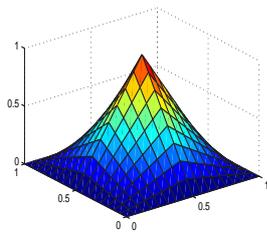


(a) Hierarchische Basisfunktion für $\underline{l} = (1, 1) : (\varphi_{1,1} \cdot \varphi_{1,1})$.

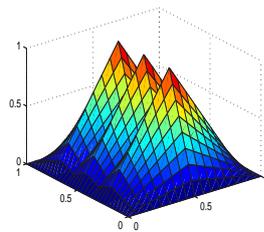
(b) Hierarchische Basisfunktion für $\underline{l} = (1, 2) : (\varphi_{1,1} \cdot \varphi_{2,1} + \varphi_{1,2} \cdot \varphi_{2,2})$.

(c) Hierarchische Basisfunktion für $\underline{l} = (2, 1) : (\varphi_{2,1} \cdot \varphi_{1,1} + \varphi_{2,2} \cdot \varphi_{1,2})$.

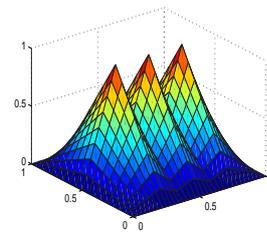
(d) Hierarchische Basisfunktion für $\underline{l} = (2, 2) : (\varphi_{2,1} \cdot \varphi_{2,1} + \varphi_{2,2} \cdot \varphi_{2,2} + \varphi_{2,3} \cdot \varphi_{2,3} + \varphi_{2,4} \cdot \varphi_{2,4})$.



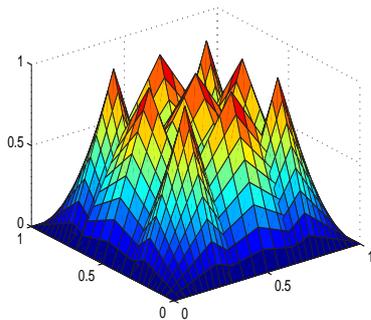
(e) Nodale Basisfunktion für $l_1 = 1, l_2 = 1 : (\varphi_{1,1} \cdot \varphi_{1,1})$.



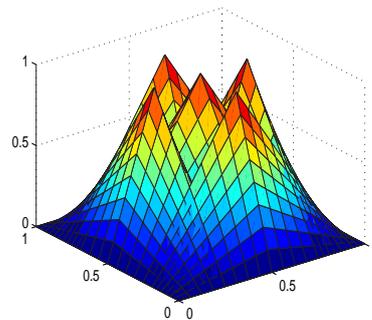
(f) Nodale Basisfunktion für $l_1 = 1, l_2 = 2 : (\varphi_{1,1} \cdot \varphi_{2,1} + \varphi_{1,2} \cdot \varphi_{2,2})$.



(g) Nodale Basisfunktion für $l_1 = 2, l_2 = 1 : (\varphi_{2,1} \cdot \varphi_{1,1} + \varphi_{2,2} \cdot \varphi_{1,2})$.



(h) Zusammenstellung der nodalen Basisfunktionen die für die Bildung eines Vollen Gitters nach Abbildung 7.11(a) für $\underline{l} = (2, 2)$ gebraucht werden.



(i) Zusammenstellung der nodalen Basisfunktionen die für die Bildung eines Dünnes Gitters nach Abbildung 7.11(b) für $\underline{l} = (2, 2)$ gebraucht werden.

Abbildung 7.12.: Die zweidimensionalen nodalen und hierarchischen Basisfunktionen für die Bildung eines Vollen sowie eines Dünnes Gitters nach den in Abbildung 7.11 veranschaulichten Methoden für $\underline{l} = (2, 2)$. Die Kombinationstechnik verwendet die in 7.12(e) bis 7.12(g) dargestellten nodalen Basisfunktionen.

Kapitel 8.

Die Parameterschätzverfahren

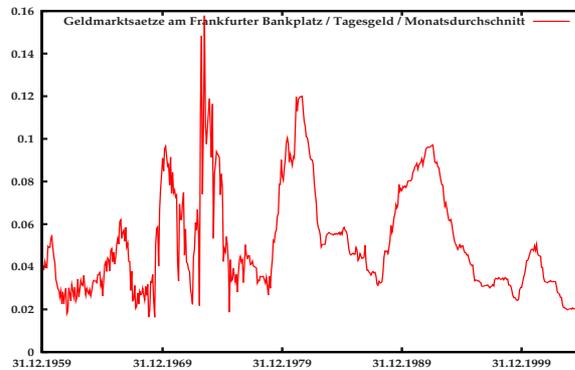
In Kapitel 4 wurden stochastische Modelle zur Beschreibung von Aktienkursen und Zinssätzen vorgestellt und diskutiert. Für eine adäquate Beschreibung derselben ist im Anschluss an die Auswahl eines Modells die Festlegung der Parameter ausschlaggebend. Ziel ist es, mittels der in Kapitel 6 beschriebenen Verfahren simulierte Datensätze generieren zu können, die die am Markt beobachteten Daten möglichst gut approximieren.

In einem ersten Ansatz wird versucht, über theoretische Betrachtungen zu „optimalen“ Parameterwerten für ein Modell zu gelangen. Dementsprechend ist auch die für das Vasicek- sowie das CIR-Modell geltende Parameterbeschränkung $2bm \geq \sigma^2$ hergeleitet (vergleiche die Modellbeschreibungen im Kapitel 4.3), da Zinssätze in der Realität lediglich nichtnegative Werte annehmen. Weil in Bezug auf die vorliegenden ökonomischen Modelle keine konkreten Parameterwerte aus der Realität ableitbar sind, muss auf eine Stichproben-Analyse (in Form historischer Daten) zurückgegriffen werden. Bei einer Parameterschätzung auf Basis einer Stichprobe muss beachtet werden, dass aufgrund des endlichen Umfangs der Stichprobe nur ein „best fit“ und keine genaue Bestimmung der gesuchten Parameterwerte möglich ist. Die Güte einer solchen „best fit“-Schätzung wird maßgeblich vom Grad der Repräsentativität der Stichprobe bestimmt. Neben der Quantität ist dabei auch die Qualität der Stichprobe ausschlaggebend.

Auf die Stichproben-Problematik werden wir in Kapitel 8.1 eingehen. In Kapitel 8.2 werden die innerhalb dieser Arbeit zum Einsatz kommenden parameterspezifischen Parameterschätzverfahren beschrieben. Sie nutzen die speziellen Eigenschaften einzelner Parameter zu ihrer Schätzung auf Grundlage historischer Daten. In Kapitel 8.3 wird ein Überblick über die allgemeineren Parameterschätzverfahren gegeben. Hier muss zwischen den Prozess-spezifischen und den gänzlich allgemein gehaltenen Verfahren unterschieden werden. Das im Rahmen dieser Arbeit entwickelte Histogramm-Matching-Verfahren, das auf den Dichte-Approximationen aus Kapitel 7 basiert, gehört zu der allgemein gehaltenen Verfahrensgruppe und ist damit prinzipiell universell einsetzbar. Es wird in Kapitel 8.4 genau erklärt. Das zum Vergleich herangezogene Maximum-Likelihood-Verfahren (vergleiche [Lo86], [Sø97], [FMW03]) hingegen gehört zu der Gruppe der Prozess-spezifischen Parameterschätzverfahren. Es wird in Kapitel 8.5 im Hinblick auf das CIR-Modell beschrieben.

8.1. Die Repräsentativität der historischen Stichprobe

Zeitpunkt	Werte von x(t)		relative Differenz zur Vorperiode in %
	Wert	Wertestab	
2006-05-17	2,60		-0,4
2006-05-16	2,61		0,0
2006-05-15	2,61		0,0
2006-05-14		kein Nachweis	
2006-05-13		kein Nachweis	
2006-05-12	2,61		0,4
2006-05-11	2,60		0,4
2006-05-10	2,59		2,8
2006-05-09	2,52		5,0
2006-05-08	2,40		0,0
2006-05-07		kein Nachweis	
2006-05-06		kein Nachweis	
2006-05-05	2,40		-3,2
2006-05-04	2,48		-5,0
2006-05-03	2,61		-0,4
2006-05-02	2,62		-1,1
2006-05-01		kein Nachweis	
2006-04-30		kein Nachweis	
2006-04-29		kein Nachweis	
2006-04-28	2,65		1,1
2006-04-27	2,62		-0,4
2006-04-26	2,63		0,4
2006-04-25	2,62		0,0
2006-04-24	2,62		0,0



(a) Wertetabelle der Tagesdurchschnitte für den Zeitraum vom 24.04.2006 bis zum 17.05.2006.

(b) Die Monatsdurchschnitte der letzten 60 Jahre.

Abbildung 8.1.: Ausschnitte der Zeitreihen des Tagesgeldsatzes am Frankfurter Bankplatz (Quelle: [Zei]).

Die für die Parameterschätzung verwendete Stichprobe sollte die Verhältnisse der Grundgesamtheit zutreffend widerspiegeln. In Bezug auf die Erfüllung dieses Anspruch der *Repräsentativität* sind sowohl Quantität als auch Qualität der Stichprobe entscheidend:

- *Quantität der Stichprobe:* Das schwache Gesetz der großen Zahlen sowie der Zentrale Grenzwertsatz (siehe Satz A.0.8 und Satz A.0.23) zeigen, dass ein sicherer Rückschluss auf statistische Gesetzmäßigkeiten der Grundgesamtheit nur mittels eines angemessen großen Umfangs der Stichprobe sicher möglich ist. Allerdings sind dem Auflösungslevel der Daten durch die Bedingungen in der Realität gewisse Grenzen gesetzt. Wird etwa die in Abbildung 8.1(a) dargestellte Wertetabelle der Zeitreihe der Tagesdurchschnitte des Tagesgeldsatzes am Frankfurter Bankplatz (als Approximation der Short Rate) betrachtet, fallen direkt die nicht vorhandenen Datenwerte (zum Beispiel für den 14.05.2006) ins Auge. Diese Lücken in der Zeitreihe entstehen durch Wochenenden und Feiertage, an denen auch Börsen und Banken geschlossen bleiben. Verursacht durch diese Umstände können Daten aus der Finanzwelt höchstens auf monatlichem Auflösungslevel ohne Lücken in den Zeitreihen betrachtet werden (siehe Abbildung 8.1(b)).
- *Qualität der Stichprobe:* Neben dem quantitativen Aspekt darf die Qualität einer Stichprobe nicht außer Acht gelassen werden. Die Stichprobe sollte von einem ähnlichen ökonomischen Umfeld geprägt sein, wie es für die mit Hilfe des Modells zu generierenden Daten anzunehmen ist. Wie anhand von Abbildung 8.1(b) zu erkennen ist, muss daher je nach Zeitreihe bei der Festlegung des historischen Datenfensters darauf geachtet werden, nicht zu weit in die Vergangenheit zurückzublicken. Dargestellt ist die Zeitreihe der Monatsdurchschnitte der Tagesgeldsätze am Frankfurter Bankplatz seit Ende 1959.¹ Deutlich zu erkennen ist, dass die Marktsituation vor

¹Informationen der Deutschen Bundesbank ([Zei]): „Geldmarktsätze werden im Allgemeinen nicht offiziell

1979 eine deutlich höhere Volatilität der Zinssätze hervorgerufen hatte, als es danach der Fall war. Dies liegt vor allem am Europäischen Währungssystem (EWS), das 1979 als Nachfolger des Europäischen Wechselkursverbundes eingeführt wurde und bis einschließlich 1998 die Grundlage der Währungspolitik zwischen den Ländern der Europäischen Gemeinschaft gebildet hat (für nähere Informationen siehe zum Beispiel [EuL]).^{2 3} Aufgrund dieses Einschnitts der historischen Zeitreihe ist es nicht sinnvoll, für die Parameterschätzung von Short-Rate-Modellen Tagesgeldsätze vor 1979 zu verwenden. Der Zeitrahmen der Stichprobe ist somit in etwa auf die letzten 30 Jahre beschränkt. Damit ist wegen der Beschränkung des Auflösungslevels der mögliche Umfang der Stichprobe (die Quantität) begrenzt.

Es bleibt zu prüfen, inwieweit Quantität sowie Qualität der historischen Stichprobe im konkret vorliegenden Fall Probleme bei der Parameterschätzung verursachen (siehe Kapitel 9.6 und 10).

festgesetzt oder notiert. Die hier angegebenen Sätze sind ungewichtete Monatsdurchschnitte, die auf Angaben vom Frankfurter Bankplatz basieren.“ „Beginnend mit März 1970 wurden die Reihen auf eine neue, verbreiterte Erhebungsgrundlage gestellt. Bis Juni 1990 wurden sie nach der Deutschen Zinsmethode 360/360 Tage berechnet, ab Juli 1990 nach der Zinsmethode act/360 ermittelt.“

²Das EWS zielte darauf ab, in Europa durch die Einführung fester, jedoch anpassungsfähiger Wechselkurse eine währungsstabile Zone zwischen den beteiligten Ländern zu schaffen.

1999 wurde das EWS innerhalb der Mitgliedsländer der Währungsunion durch die Einführung des Euro abgelöst. In den übrigen EU-Ländern trat das EWS II mit dazugehörigem Wechselkursmechanismus II an seine Stelle.

³Mit der Absicht, die Preise stabil zu halten, setzen Notenbanken die Leitzinsen als geldpolitische Mittel ein. Diese haben direkte Auswirkung auf die Tagesgeldsätze.

8.2. Parameterspezifische Parameterschätzverfahren

In diesem Kapitel werden Parameterschätzverfahren vorgestellt, die die analytischen Eigenschaften einzelner Parameter für eine Schätzung derselben ausnutzen. Solche Verfahren sind parameterspezifisch und lassen sich daher nicht beliebig auf andere Parameter übertragen. Im Rahmen dieser Arbeit kommen folgende auf historischen Stichproben basierende parameterspezifische Schätzverfahren zur Anwendung:

1. Die Schätzung des Driftparameters μ des Aktienkurs-Modells über die historische mittlere Rendite (skalierter Mittelwert der logarithmischen Kursänderungen),
2. Die Schätzung des Diffusionsparameters σ des Aktienkurs-Modells über die historische Volatilität,
3. Die Schätzung des Mean-Reversion-Niveaus m der Short-Rate-Modelle, denen ein Mean-Reversion-Prozess zugrunde liegt, über den historischen Mittelwert.

Zu 1.: Die diskretisierten logarithmischen Kursänderungen $\Delta \ln S_i = \mu \Delta t + \sigma \Delta W_i$ sind normalverteilt mit $N(\mu \Delta t, \sigma^2 \Delta t)$ (vergleiche Kapitel 4.1 sowie Tabelle 5.1). Der Erwartungswert ist demnach nur von dem zu schätzenden Parameter μ sowie dem entsprechend der Diskretisierungsannahmen (siehe Kapitel 5) konstanten Zeitzuwachs Δt abhängig. Die Schätzung des Parameters μ kann daher über eine Approximation des Erwartungswertes erfolgen. Da es sich bei den logarithmischen Kursänderungen $\Delta \ln S_i$ um eine normalverteilte Zufallsvariable handelt, ist es möglich, den Erwartungswert über den Mittelwert der vorliegenden historischen Stichprobe zu approximieren (vergleiche das schwache Gesetz der großen Zahlen (Satz A.0.8)).

Für eine Stichprobe von $N + 1$ historischen Aktienkurswerten S_0, \dots, S_N mit $S_i = S(i \cdot \Delta t)$ ergibt sich ein unverzerrter Schätzwert $\hat{\mu}$ des Driftparameters mit $x_i = \ln \left(\frac{S(i \cdot \Delta t)}{S((i-1) \cdot \Delta t)} \right)$ für $i = 1, \dots, N$ über

$$\hat{\mu} = \frac{1}{\Delta t} \mathbb{E} \left[\ln \left(\frac{S(t + \Delta t)}{S(t)} \right) \right] \approx \frac{1}{\Delta t} \cdot \bar{x} = \frac{1}{\Delta t} \cdot \frac{1}{N} \sum_{i=1}^N x_i.$$

Zu 2.: Ein Schätzwert des Diffusionsparameters σ kann analog über eine Approximation der Varianz erfolgen, denn diese ist ebenfalls nur vom zu schätzenden Parameter σ sowie dem konstanten Zeitzuwachs Δt abhängig. Die Varianz kann als historische Varianz mit Hilfe des Mittelwertes der historischen Aktienkurswerte über

$$\hat{\sigma}^2 = \frac{1}{\Delta t} \text{var} \left[\ln \left(\frac{S(t + \Delta t)}{S(t)} \right) \right] \approx \frac{1}{\Delta t} \cdot \frac{1}{N-1} \sum_{i=1}^N (x_i - \bar{x})^2$$

unverzerrt geschätzt werden.

In Tabelle 8.1 ist ein Schätzalgorithmus für den Drift- sowie den Diffusionsparameter des Aktienkurs-Modells entsprechend 1. und 2. angegeben.

Zu 3.: Zur Schätzung des langfristigen Mean-Reversion-Niveaus m von Short-Rate-Modellen, denen ein Mean-Reversion-Prozess zugrunde liegt, kann der Mittelwert der historischen Datenwerte verwendet werden. Für eine Stichprobe von $N + 1$ historischen Daten r_0, \dots, r_N mit $r_i = r(i \cdot \Delta t)$ ergibt sich

$$\hat{m} = \frac{1}{N+1} \sum_{i=0}^N r_i.$$

In Tabelle 8.2 ist ein entsprechender Schätzalgorithmus angegeben.

Da sowohl für das Vasicek- als auch für das CIR-Modell $E[r(t)] \xrightarrow{t \rightarrow \infty} m$ gilt (vergleiche (4.29) und (4.38)), liefert diese Schätzung für einen ausreichend großen Umfang der historischen Stichprobe entsprechend des schwachen Gesetzes der großen Zahlen (Satz A.0.8) gute Ergebnisse. Eine numerische Analyse dieser Schätzung folgt in Kapitel 9.1.

HISTORISCHE SCHÄTZER FÜR DIE PARAMETER DES AKTIENKURS-MODELLS	
INPUT:	Historische Aktienkurswerte S_0, \dots, S_N mit $S_i = S(i \cdot \Delta t)$, Δt
Initialisierung:	Relative historische Aktienkurswerte: $x_i = \frac{S_i - S_{i-1}}{S_{i-1}}$ ($i = 1, \dots, N$); $a_1 = x_1$; $b_1 = 0$;
Schleife:	für $i = 2, \dots, N$ $a_i = a_{i-1} + (x_i - a_{i-1})/i$; $b_i = b_{i-1} + (i-1)(x_i - a_{i-1})^2/i$;
	$b_N = b_N/(N-1)$;
OUTPUT:	Driftparameter $\hat{\mu} = \frac{1}{\Delta t} \bar{x} = \frac{1}{\Delta t} a_N$, Diffusionsparameter $\hat{\sigma} = \sqrt{s^2/\Delta t} = \sqrt{b_N/\Delta t}$

Tabelle 8.1.: Schätzalgorithmus für den Drift- und den Diffusionsparameter des Aktienkurs-Modells mittels historischer Stichprobe.

MITTELWERT-SCHÄTZER FÜR DAS MEAN-REVERSION-NIVEAU	
INPUT:	Historische Short-Rate-Daten r_0, \dots, r_N mit $r_i = r(i \cdot \Delta t)$, Δt
Initialisierung:	$m_0 = r_0$;
Schleife:	für $i = 1, \dots, N$ $m_i = m_{i-1} + r_i$;
	$m_N = m_N/(N+1)$;
OUTPUT:	Mean-Reversion-Niveau $\hat{m} = \bar{r} = m_N$

Tabelle 8.2.: Schätzalgorithmus für das Mean-Reversion-Niveau von Short-Rate-Modellen mittels historischer Stichprobe.

8.3. Allgemeinere Parameterschätzverfahren

Parameterspezifische Parameterschätzverfahren (Kapitel 8.2) existieren nur für eine kleine Anzahl an Modellparametern. Für die Schätzung aller übrigen Parameter muss daher auf allgemeinere Parameterschätzverfahren zurückgegriffen werden. In dieser Arbeit werden zwei dieser allgemeineren auf historischen Stichproben basierenden Parameterschätzverfahren analysiert:

- *Das Histogramm-Matching-Verfahren.* Es basiert auf den in Kapitel 7 beschriebenen Dichte-Approximationen in Form diskreter Histogramme. Die Schätzung der Parameter erfolgt über den Vergleich eines – auf Grundlage der historischen Stichprobe erstellten – „historischen“ Histogramms mit „simulierten“ Histogrammen, die für die verschiedenen Parametersätze des Parametersuchraums generiert werden. Der geschätzte Parametersatz ist derjenige, für den der Histogramm-Vergleich die kleinste Abweichung liefert.
- *Das Maximum-Likelihood-Verfahren* (vergleiche [Lo86], [Sø97], [FMW03]). Dabei wird die Übergangsdichte des diskretisierten Modells beziehungsweise eine Approximation derselben als Likelihood-Funktion für die Schätzung der Parameter mittels der historischen Stichprobe verwendet.

Während beim Maximum-Likelihood-Verfahren die Kenntnis der Übergangsdichte, einer Approximation der Übergangsdichte oder einer gleichwertigen Prozess-Information für die Festlegung der Likelihood-Funktion unverzichtbar ist, funktioniert das Histogramm-Matching-Verfahren auch ohne nähere Prozess- beziehungsweise Parameter-Informationen. Damit ist das Histogramm-Matching-Verfahren grundsätzlich das allgemeinere der beiden Verfahren. Zusätzlich kann durch eine parameterspezifische Skalierung eine erhebliche Verbesserung der Schätzgüte erreicht werden (vergleiche 8.4.1 und 9.2). Da die Verfahren in Kapitel 9 im Hinblick auf eine Schätzung der Parameter b und σ des CIR-Modells analysiert werden sollen, werden in den folgenden Kapiteln jegliche Verfahrensspezifizierungen in Bezug auf das CIR-Modell vorgestellt.

Sollen diese allgemeinen Parameterschätzverfahren auch für den in Kapitel 4.4 vorgestellten Mehr-Faktor-Fall des CIR-Modells eingesetzt werden, so muss zunächst die Hürde der nicht am Markt beobachtbaren Teilraten der Short Rate überwunden werden. Diese Problematik sowie ihre Auswirkungen hinsichtlich einer möglichen Schätzung mittels der Histogramm-Matching-Verfahren werden in Kapitel 8.6 näher erörtert.

Im Rahmen der Parameterschätzung mittels der Maximum-Likelihood-Methode existieren, abgesehen von dem hier betrachteten Maximum-Likelihood-Verfahren auf Grundlage der approximierten Übergangsdichte des diskretisierten Modells, zahlreiche andere Ansätze, die auf weiteren Approximationsmöglichkeiten der wahren Likelihoodfunktion basieren (siehe unter anderem [Ped95], [SC97], [KPS04], [Pou99] und [ASK02]). Außerdem gibt es Verfahren, die auf der Generalisierten beziehungsweise der effizienten Momentenmethode basieren (siehe etwa [GT96]). Diese haben zwar dieselbe asymptotische Effizienz wie die Maximum-Likelihood-Methode, allerdings versagen sie bei Parameterschätzung anhand kleiner Datenmengen (vergleiche [DS01]). Neben diesen parametrischen Schätzmethoden gibt es auch einige nichtparametrische Ansätze (siehe etwa [AS02], [Sta97]).

8.4. Das allgemeine Histogramm-Matching-Verfahren

In Kapitel 7 wurden Dichte-Approximationen auf Grundlage verschiedener Basisfunktionen vorgestellt. Dabei wurde für den mehrdimensionalen Fall sowohl der Tensorprodukt- als auch der Dünngitter-Ansatz beschrieben.

Wird die Dichte mittels charakteristischer Basisfunktionen approximiert, ergibt sich die allgemein bekannte Form eines Histogramms. In diesem Zusammenhang wird ab jetzt von *naiven Histogrammen* gesprochen. Die Resultate der anderen Dichte-Approximationen werden in dieser Arbeit ebenfalls als *Histogramme* bezeichnet – auch wenn sie nicht die typische Histogramm-Form haben.

In diesem Kapitel wird das allgemeine Histogramm-Matching-Verfahren beschrieben. Es nutzt das Prinzip der Dichte-Approximation zur Schätzung der Parameter eines diskretisierten Modells mittels einer historischen Stichprobe. In Kapitel 8.4.1 werden die möglichen, für die Parameter des CIR-Modells spezifischen Skalierungen erklärt und diskutiert. Bei den numerischen Analysen wird sich herausstellen, dass diese Skalierungen zu einer tatsächlichen Verbesserung der Schätzungen führen (siehe Kapitel 9.2). In Kapitel 8.4.2 wird ein iteratives Histogramm-Matching-Verfahren auf Grundlage der Vergleiche eindimensionaler Histogramme, in Kapitel 9.4 ein paralleles Histogramm-Matching-Verfahren basierend auf den Vergleichen mehrdimensionaler Histogramme beschrieben. In diesem Zusammenhang ist auch das (Dünngitter-)Histogramm-Matching-Verfahren einzuordnen, das auf der Approximation mehrdimensionaler Dichtefunktionen mittels des Dünngitter-Ansatzes beruht. Die verschiedenen Histogramm-Matching-Verfahren basieren alle auf dem in Tabelle 8.3 angegebenen und im Anschluss näher erläuterten allgemeinen Histogramm-Matching-Verfahren. Eine genaue Analyse der Verfahrensparameter erfolgt in Kapitel 9. Das Histogramm-Matching-Verfahren basiert auf dem Vergleich des so genannten *historischen Histogramms*, das aus dem historischen Datensatz (der Stichprobe) generiert wird, mit den *simulierten Histogrammen*, die aus simulierten Datensätzen aufgebaut werden. Für jeden Parametersatz (eines gegebenen Suchraums) wird ein simuliertes Histogramm erstellt und mit dem historischen Histogramm verglichen. Der geschätzte Parametersatz ist derjenige, dessen simuliertes Histogramm die kleinste Abweichung zum historischen Histogramm aufweist. Dabei liegt die Annahme zugrunde, dass die Histogramme die wahren Dichtefunktionen gut approximieren und infolgedessen aus einer relativ kleinen Abweichung der Histogramme auch auf eine entsprechende gute Übereinstimmung der Parameter geschlossen werden kann. Eine genaue Übereinstimmung der Histogramme ist allerdings erst für eine sehr große Datenanzahl ($N \rightarrow \infty$) möglich. Sie ist aufgrund des unzureichenden Umfangs historischer Stichproben auszuschließen. Es wird daher nach einem „best fit“ bezüglich der Histogramm-Abweichungen gesucht.

ALLGEMEINES HISTOGRAMM-MATCHING-VERFAHREN	
Bezeichnungen:	Dat - Datensatz; H - Histogramm; hist - historische Datenquelle; sim - simulierte Datenquelle.
INPUT:	(1) Stichprobe Dat_{hist} .
Initialisierung:	(2) Wähle die Intervallgrenzen und die Anzahl der Teilintervalle k_r pro Richtung $r \in \{1, \dots, d\}$ (Teilintervallbreite $h_r = \frac{b_r}{k_r}$). (3) Wähle den Parametersuchraum Θ (und den Startparametersatz θ_{start}).
	(4) Ermittle H_{hist} aus Dat_{hist} .
Schleife:	(5) Starte mit θ_{start} und durchlaufe den Parametersuchraum Θ : (6) Simuliere Dat_{sim} (θ) entsprechend aktuellem θ . (7) Ermittle H_{sim} (θ) aus Dat_{sim} (θ). (8) Ermittle die Abweichung der Histogramme: $\ \mathbf{H}_{sim}(\theta) - \mathbf{H}_{hist}\ _2$.
OUTPUT:	(9) Parametersatz $\hat{\theta}$, bei dem die kleinste Histogramm-Abweichung auftrat („best fit“): $\hat{\theta} = \arg \min_{\theta} \ \mathbf{H}_{sim}(\theta) - \mathbf{H}_{hist}\ _2$

Tabelle 8.3.: Algorithmus des allgemeinen Histogramm-Matching-Verfahrens.

Bevor das in Tabelle 8.3 beschriebene Histogramm-Matching-Verfahren spezifiziert wird (siehe Kapitel 8.4.1 bis 8.4.3), werden die einzelnen Unterpunkte umrissen:

Zu (1): Die Problematik der Auswahl einer repräsentativen Stichprobe in Form historischer Daten **Dat_{hist}** wurde bereits in Kapitel 8.1 erläutert. Unter Umständen kann eine parameterspezifische Skalierung der historischen sowie der entsprechenden simulierten Datenwerte zu besseren Resultaten der Parameterschätzung führen. Die Skalierungen des historischen Datensatzes sind vor der Festlegung der Intervallgrenzen und der Teilintervallbreite durchzuführen. In Kapitel 8.4.1 wird die Skalierung bezüglich der Parameterwerte des CIR-Modells beschrieben. Eine numerische Analyse findet sich in Kapitel 9.2.

Zu (2): Die Intervallgrenzen beschränken das Grundintervall der zu erstellenden Histogramme. Sie werden anhand der historischen Stichprobe festgelegt. In Kapitel 9.3.4 wird untersucht, inwieweit es sinnvoll ist, für diese Festlegung der Intervallgrenzen nur einen gewissen Prozentsatz der historischen Datenwerte einzubeziehen, um so „Ausreißer“ (Datenwerte, die weit von den übrigen abweichen) auszuschließen und das Verfahren effizienter zu gestalten. Je schmaler das betrachtete Grundintervall der Histogramme bezüglich der Richtung r ist (b_r), desto geringer fällt auch die für eine kleinere Teilintervallbreite $h_r = \frac{b_r}{k_r}$ benötigte Anzahl k_r der Teilintervallen für die Richtung r aus. In Kapitel 9 werden Analysen über die sinnvolle Festlegung dieser Teilintervallbreiten beziehungsweise die Festlegung der Anzahl der Teilintervalle durchgeführt.

Zu (3): Bei der Definition des Parametersuchraums Θ ist darauf zu achten, mit welchen Parameterwerten Daten generiert werden können, die den real zu beobachtenden Werten nicht widersprechen und den Voraussetzungen der Modelle genügen. Zum Beispiel werden die entsprechend des CIR-Modells generierten Zinsraten für zu große Parameterwerte der

Mean-Reversion-Stärke b ($b \cdot \Delta t \geq 1$) oft direkt über das Mean-Reversion-Niveau gezogen. Da dieses Verhalten verglichen mit realen Beobachtungen nicht sinnvoll erscheint, wird der Parametersuchraum auf Werte mit $b \cdot \Delta t < 1$ eingeschränkt. Außerdem ist darauf zu achten, dass die Beschränkung $2bm \geq \sigma^2$ (vergleiche Kapitel 4.3.2) einzuhalten ist, um negative Werte der generierten Short Rate zu verhindern.

Zu (4): Der Aufbau der Histogramme mittels Dichte-Approximationen wurde in Kapitel 7 erklärt. Je nach Spezifizierung des Verfahrens handelt es sich um ein- oder mehrdimensionale Histogramme. Darüber hinaus steht die Wahl der Basisfunktionen offen und bezüglich der mehrdimensionalen Dichte-Approximationen kann einerseits auf den Tensorprodukt-, andererseits auf den Dünngitter-Ansatz zurückgegriffen werden.

Zu (5): Das Durchlaufen des Parametersuchraums kann auf verschiedene Arten erfolgen. Die unkomplizierteste Methode ist das simple „Durchforsten“ des gesamten Suchraums, dabei entspricht der Startparametersatz der unteren Grenze des Suchraums. Diese Methode ist zwar einfach und robust, jedoch steigt dabei der Aufwand exponentiell mit der Anzahl an Parametern. Um genauere Ergebnisse zu erlangen, kann adaptive Verfeinerung eingesetzt werden. Das Verfahren ist dann weiterhin relativ einfach, allerdings muss ein Verfeinerungs- oder Stoppkriterium (Änderung der in (8) ermittelten Abweichung sinkt unter einen gewissen Schwellwert ε) vorgegeben werden. Ferner besteht die Möglichkeit, das Newton-Verfahren für die Suche nach den optimalen Parameterwerten durch Minimumsuche einzusetzen. Das Newton-Verfahren verfügt gegenüber den anderen beiden zwar über eine höhere Genauigkeit und Schnelligkeit. Allerdings ist es weniger robust und benötigt darüber hinaus eine Ableitung.

Zu (6): Die Simulationen werden auf Grundlage des diskretisierten Modells generiert. Dabei wird im Rahmen dieser Arbeit das Euler-Maruyama-Schema verwendet (vergleiche Kapitel 5). Der Simulationszeitraum T entspricht dabei dem historischen Zeitraum, die Anzahl der zu simulierenden Zeitschritte $N - 1$ stimmt mit der der historischen Stichprobe überein. Um ein möglichst glattes Vergleichshistogramm \mathbf{H}_{sim} und damit auch eine entsprechend glatte Fehlerkurve bezüglich der Histogrammabweichungen zu erhalten, wird pro Parametersatz nicht bloß ein Datensatz für die Ermittlung des zugehörigen Histogramms generiert. Vielmehr wird eine größere Anzahl von Datensätzen simuliert und mit Hilfe einer Mittelung für die Erstellung der simulierten Histogramme verwendet. Welche Anzahl dabei sinnvoll ist und inwieweit das mit dem Datensatzumfang N zusammenhängt, wird ebenfalls in Kapitel 9 erläutert.

Zu (7): Entsprechend (4).

Zu (8) und (9): Abschließend soll die für den Vergleich der Histogramme notwendige Ermittlung der Histogramm-Abweichungen erläutert werden. Auf Grundlage der L_2 -Norm ergibt sich der geschätzte Parametersatz allgemein über

$$\hat{\theta} = \arg \min_{\theta} \|\mathbf{H}_{\text{sim}}(\theta) - \mathbf{H}_{\text{hist}}\|_2.$$

Wie auch Tabelle 8.3 bezeichnet \mathbf{H}_{hist} dabei das historische Histogramm, während $\mathbf{H}_{\text{sim}}(\theta)$ die simulierten Histogramme in Abhängigkeit vom aktuellen Parametersatz θ darstellt. Die exakte Ermittlung der Histogramm-Abweichungen hängt von den zugrunde liegenden Basisfunktionen der Dichte-Approximationen sowie der Anzahl an Dimensionen ab:

- Für Histogramme auf Grundlage charakteristischer Basisfunktionen, der naiven Form von Histogrammen, kann der Fehler über die L_2 -Norm der relativen Häufigkeiten pro Teilintervall bestimmt werden. Für eindimensionale Histogramme wird die Abweichung über

$$\|\mathbf{H}_{\text{sim}}(\theta) - \mathbf{H}_{\text{hist}}\|_2 = \sqrt{\sum_{i=1}^k (h_i^{\text{sim}} - h_i^{\text{hist}})^2}$$

ermittelt, für zweidimensionale Histogramme mittels

$$\|\mathbf{H}_{\text{sim}}(\theta) - \mathbf{H}_{\text{hist}}\|_2 = \sqrt{\sum_{i=1}^{k_1} \sum_{j=1}^{k_2} (h_{ij}^{\text{sim}} - h_{ij}^{\text{hist}})^2}.$$

Dabei bezeichnet k die Anzahl der Teilintervalle im eindimensionalen Histogramm, k_1 und k_2 die Anzahlen der Teilintervalle bezüglich der verschiedenen Richtungen der zweidimensionalen Histogramme. Die h_i beziehungsweise h_{ij} geben die relativen Häufigkeiten der sich im Intervall I_i beziehungsweise I_{ij} befindenden Datenwerte an.

- Bei eindimensionalen Histogrammen, die auf stückweise linearen Basisfunktionen gründen, kann auf den Vergleich der Geraden in den einzelnen Teilintervallen zurückgegriffen werden. Um eine Fallunterscheidung bezüglich sich schneidender Geraden zu umgehen, wird eine zweifache L_2 -Norm verwendet:

$$\|\mathbf{H}_{\text{sim}}(\theta) - \mathbf{H}_{\text{hist}}\|_2 = \sqrt{\sum_{i=1}^k \left(\sqrt{\int_{I_i} (g_i^{\text{sim}}(x) - g_i^{\text{hist}}(x))^2 dx} \right)^2}.$$

Dabei bezeichnet k die Anzahl der Teilintervalle und g_i die für das i -te Intervall I_i ermittelbare Gerade. Durch die innere L_2 -Norm entfällt die Fallunterscheidung, die entsprechend der Lage der Geraden zueinander zu treffen wäre.

Analog kann in zwei Dimensionen die Abweichung der Histogramme über den Vergleich der für die einzelnen Teilquader I_{ij} ermittelbaren Ebenen erfolgen. Auch hier unterbindet eine zusätzliche L_2 -Norm die Notwendigkeit einer Fallunterscheidung bezüglich der Lage der Ebenen zueinander:

$$\|\mathbf{H}_{\text{sim}}(\theta) - \mathbf{H}_{\text{hist}}\|_2 = \sqrt{\sum_{i=1}^{k_1} \sum_{j=1}^{k_2} \left(\sqrt{\int_{I_{ij}} (e_{ij}^{\text{sim}}(x, y) - e_{ij}^{\text{hist}}(x, y))^2 dx dy} \right)^2}.$$

k_1 bezeichnet dabei die Anzahl der Teilintervalle in Bezug auf die erste, k_2 diejenige bezüglich der zweiten Richtung. Desweiteren beschreiben die e_{ij} die im Teilquader I_{ij} aufgespannten Ebenen. Die Funktionsgleichungen der zum Vergleich eingesetzten Geraden beziehungsweise Ebenen sind für das jeweilige Teilintervall (-quader) mit Hilfe der durch die Dichte-Approximation ermittelten Werte zu bestimmen.

- Für das mehrdimensionale Dünngitter-Histogramm-Matching (auf Grundlage stück-

weise linearer Basisfunktionen) kann die Abweichung nach (7.7) über

$$\begin{aligned}
 \|\mathbf{H}_{\text{sim}} - \mathbf{H}_{\text{hist}}\|_2 &= \|\mathbf{f}_{\text{sim}} - \mathbf{f}_{\text{hist}}\|_2 = \left\| \sum_{i=1}^{n_g} \varphi_i g_i^{\text{sim}} - \sum_{i=1}^{n_g} \varphi_i g_i^{\text{hist}} \right\|_2 \\
 &= \left\| \sum_{i=1}^{n_g} \varphi_i (g_i^{\text{sim}} - g_i^{\text{hist}}) \right\|_2 \leq \left\| \sum_{i=1}^{n_g} \varphi_i \right\|_2 \cdot \left\| \sum_{i=1}^{n_g} g_i^{\text{sim}} - g_i^{\text{hist}} \right\|_2 \\
 &\leq \left\| \sum_{i=1}^{n_g} \varphi_i \right\|_2 \cdot \sum_{i=1}^{n_g} |g_i^{\text{sim}} - g_i^{\text{hist}}| \\
 &\leq C \cdot \sum_{i=1}^{n_g} |g_i^{\text{sim}} - g_i^{\text{hist}}|
 \end{aligned}$$

durch einen Vergleich der Gewichtskoeffizienten nach oben mit einer Konstanten C abgeschätzt werden. Dabei ist n_g die Anzahl der Gewichtskoeffizienten. Hier wird auch der Vorteil des Dünngitter-Verfahrens deutlich. Wird eine für alle Richtungen konstante Anzahl k der Teilintervalle festgelegt, so sind es in Bezug auf den Dünngitter-Ansatz lediglich $O(k \log k^2)$ statt $\prod_{r=1}^d k_r$ (Tensorprodukt-Ansatz) Auswertungen pro Histogramm-Abweichung.

8.4.1. Skalierungen für das Histogramm-Matching-Verfahren

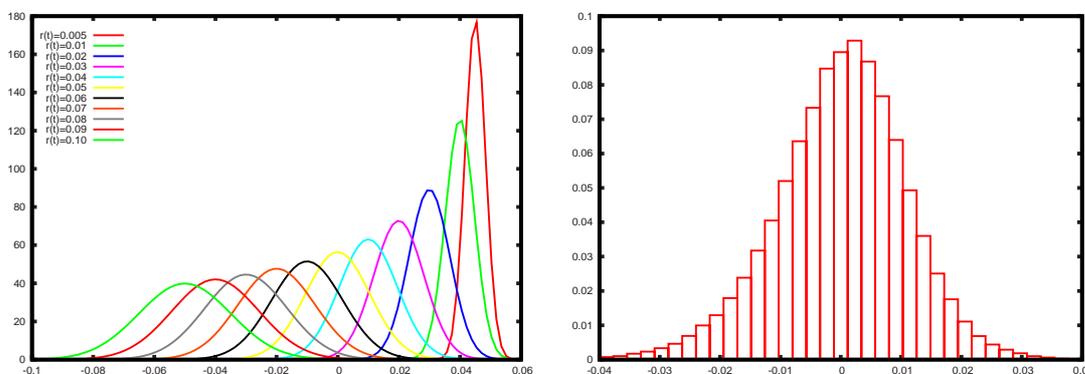
In Kapitel 5 wurde festgestellt, dass die mittels des Euler-Maruyama-Schemas diskretisierten Zuwächse des CIR-Modells normalverteilt sind mit Erwartungswert

$$E[\Delta r_i] = b(m - r_{i-1})\Delta t \tag{8.1}$$

und Varianz

$$\text{var}[\Delta r_i] = \sigma^2 r_{i-1} \Delta t. \tag{8.2}$$

Damit hängen sowohl der Erwartungswert als auch die Varianz der Übergangsdichten des diskretisierten CIR-Modells vom Wert $r_{i-1} = r(t_{i-1})$ der Short Rate zum vorausgegangenen Diskretisierungszeitpunkt t_{i-1} ab.



(a) Die Dichtefunktionen der unskalierten Zuwächse (Übergangsdichten des diskretisierten CIR-Modells) für feste Ausgangswerte der Short Rate.

(b) Approximation der „Überlagerung der Übergangsdichten“ des diskretisierten CIR-Modells (siehe 8.2(a)) in Form eines naiven Histogramms basierend auf entsprechend generierten Datensätzen (Datensatzumfang $N = 300$, Simulationsanzahl $M = 500$).

Abbildung 8.2.: Die Dichtefunktionen bezüglich der unskalierten Zuwächse des CIR-Modells für $b = 1.0$, $\sigma = 0.15$, $m = 0.05$ und $\Delta t = 0.1$.

Die Übergangsdichten für festen Ausgangswert $r_{i-1} = r(t_{i-1})$

In Abbildung 8.2(a) sind die Dichtefunktionen der diskretisierten Zuwächse Δr_i des CIR-Modells für die festen Ausgangswerte $r_{i-1} \in \{0.005, 0.01, 0.02, \dots, 0.10\}$ mit übereinstimmendem Mean-Reversion-Niveau $m = 0.05$, Parameterwerten $b = 1.0$ und $\sigma = 0.15$ sowie dem konstanten Zeitzuwachs $\Delta t = 0.1$ abgebildet. Sie entsprechen den Übergangsdichten des diskretisierten CIR-Modells. Es ist zu erkennen, dass sowohl die Erwartungswerte als auch die Varianzen dieser normalverteilten Dichtefunktionen in Abhängigkeit von den Ausgangswerten r_{i-1} variieren. Stimmt der Ausgangswert mit dem Mean-Reversion-Niveau m des Modells überein ($r_{i-1} = 0.05$ (gelbe Kurve)), beträgt der Erwartungswert genau $E[\Delta r_i] = 0$. Ist r_{i-1} größer als das Mean-Reversion-Niveau ($r_{i-1} > 0.05$ (zum Beispiel $r_{i-1} = 0.06$ (schwarze Kurve))), rutscht der Erwartungswert der Dichtefunktion nach links ab. Hier wird die *Mean Reversion* sichtbar: Die Wahrscheinlichkeit für einen negativen Zuwachs wird aufgrund des negativen Mean-Reversion-Zugs größer (siehe die Verlagerung

des Mittel- beziehungsweise Erwartungswertes). Basierend auf einem hohen Diffusionsanteil kann es dennoch zu einem positiven Zuwachs kommen (siehe die Breite der einzelnen Dichtefunktionen). Analog sind die Verschiebungen für $r_{i-1} < m$ (zum Beispiel $r_{i-1} = 0.04$ (türkisfarbene Kurve)) nach rechts zu deuten.

Auch der *Level-Effekt* (siehe Kapitel 4.3) ist wiederzuerkennen: die Varianz der einzelnen Dichtefunktionen (betrachte die Breite und Höhe der einzelnen Dichtefunktionen) verhält sich proportional zum Level r_{i-1} , dem Ausgangswert der Short Rate.

An dieser Stelle lässt sich festhalten, dass auf Grundlage der Zuwächse für einen konstanten Ausgangswert r_{i-1} die Histogramme die Approximation einer einzelnen Dichtefunktion aus Abbildung 8.2(a) darstellen würden. Da die Erwartungswerte der einzelnen Dichtefunktionen aus Abbildung 8.2(a) für festes r_{i-1} und bekanntes m mit (8.1) nur vom Parameter b , die Varianz nach 8.2 vom Parameter σ abhängen, könnten diese Parameter sehr gut mittels des Histogramm-Matching-Verfahren geschätzt werden. Für eine solche Schätzung anhand der einzelnen Dichtefunktionen liegen allerdings zu wenig Daten mit dem selben Ausgangswert vor.

Die Überlagerungsdichte für variierende Ausgangswerte $r_{i-1} = r(t_{i-1})$

Für den Aufbau der approximativen Histogramme (siehe etwa Abbildung 8.2(b)) werden alle beobachteten beziehungsweise generierten Zuwächse des CIR-Modells verwendet. Folglich variiert der Ausgangswert für nahezu jeden Zuwachs. Die Histogramme approximieren also keine einzelne Übergangsdichte (siehe Abbildung 8.2(a)) sondern eine Überlagerung verschiedener Übergangsdichten, denn die Werte entspringen je nach Ausgangswert den einzelnen Dichtefunktionen. Aufgrund der Mean-Reversion-Eigenschaft kann dabei davon ausgegangen werden, dass (abhängig von der Höhe der Mean-Reversion-Stärke b) Zuwächse um 0 gehäuft auftreten, da die Short Rate immer wieder zurück zum Mean-Reversion-Niveau m und damit zum Zuwachs 0 des Mean-Reversion-Terms gezogen wird.⁴ Die Werte für die Überlagerung stammen also vor allem von den nahe um 0 liegenden einzelnen Dichtefunktionen aus Abbildung 8.2(a). Nur in seltenen Fällen treten Werte entsprechend der weiter entfernt liegenden Dichtefunktionen auf.

So ist auch das approximative Histogramm aus Abbildung 8.2(b) zu erklären. Dabei liegen das selbe Mean-Reversion-Niveau $m = 0.05$, die selben Parameterwerte $b = 1.0$ und $\sigma = 0.15$ sowie der selbe konstante Zeitzuwachs $\Delta t = 0.1$ wie bei den einzelnen Dichtefunktionen aus Abbildung 8.2(a) zugrunde (Anfangsrate $r_0 = 0.0499$, Datensatzumfang $N = 300$, Simulationsanzahl $M = 500$). Aufgrund der Konzentrierung der Werte der Short Rate um das Mean-Reversion-Niveau ist das Histogramm nach links und rechts beschränkt. Durch die verschiedenen Formen der einzelnen Dichtefunktionen (bestimmt durch die Erwartungswerte und die Varianzen) kommt es zu einer leicht schiefen Kurve. Mit (8.1) und (8.2) ist klar, dass sowohl der Erwartungswert als auch die Varianz dieser so genannten *Überlagerungsdichte* nicht nur von den Parametern b und σ sondern auch von der Verteilung der r_{i-1} abhängen. Durch die Überlagerung sind die Auswirkungen der einzelnen Parameter an den approximativen Histogrammen nur schwach zu erkennen. Dies gilt vor allem für den Parameter b . Die Parameterschätzung mittels Histogramm-Matching-Verfahren, die auf diesen Überlagerungsdichten beruht, kann demnach unpräzise sein. An dieser Stelle können Skalierungen eingesetzt werden: mit dem Ziel, die Zuwächse derart zu skalieren, dass die Auswirkungen der einzelnen Parameter wieder gut an den Histogrammen zu erkennen sind und sich daher auch die Parameter besser schätzen lassen.

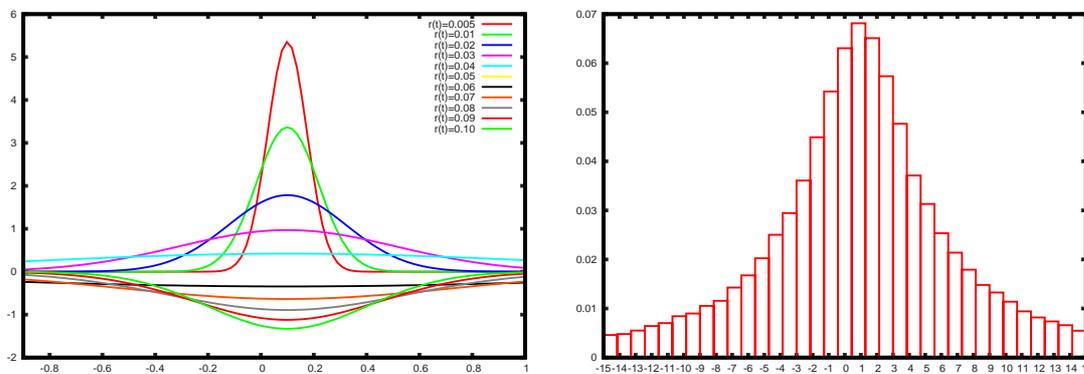
⁴Vergleiche auch die nichtzentrale Chi-Quadrat-Verteilung der Short Rate nach dem CIR-Modell.

Die Skalierung der Zuwächse bezüglich des Parameters b

Da für die Lage der Erwartungswerte $E[\Delta r_i]$ der einzelnen Übergangsdichten unter anderem auch der Parameter b verantwortlich ist, bietet es sich an, die Zuwächse Δr_i so zu skalieren, dass diese Erwartungswerte nur noch von b und nicht mehr vom Ausgangswert r_{i-1} abhängig sind. Für den Parameter b werden daher fortan die folgenden skalierten Zuwächse betrachtet

$$\frac{\Delta r_i}{(m - r_{i-1}) \Delta t} \sim N \left(b, \frac{\sigma^2 \cdot r_{i-1}}{(m - r_{i-1})^2 \cdot \Delta t} \right). \quad (8.3)$$

Damit hat auch der Parameter σ keinerlei Einfluss mehr auf die Lage des Erwartungswertes der Überlagerungsdichte. Dieser stimmt nun genau mit dem Erwartungswert $E[\Delta r_i]$ der einzelnen Übergangsdichten überein und ist damit lediglich noch von b abhängig.



(a) Skalierte Übergangsdichten der Zuwächse des diskretisierten CIR-Modells für verschiedene vorausgehende Werte der Short Rate.

(b) Überlagerung der Übergangsdichten in approximativer Histogrammdarstellung (Datensatzumfang $N = 300$, Simulationsanzahl $M = 500$).

Abbildung 8.3.: Die Dichtefunktionen der bezüglich des Parameters b nach (8.3) skalierten Zuwächse für $b = 1.0$, $\sigma = 0.15$, $m = 0.05$ und $\Delta t = 0.1$.

In Abbildung 8.3(a) sind Übergangsdichten der skalierten Zuwächse für feste Ausgangswerte r_{i-1} abgebildet. Abbildung 8.3(b) zeigt ein approximatives Histogramm der skalierten Zuwächse, das eine Überlagerung einzelner Übergangsdichten darstellt. Durch die Skalierung liegt der Erwartungswert der Überlagerungsdichte ebenso wie die Erwartungswert $E[\Delta r_i]$ der einzelnen Übergangsdichten über dem wahren Wert von b . Gleichzeitig hat der Parameter b bei dieser Skalierung (über die Ausgangswerte r_{i-1}) kaum noch Einfluss auf die Breite der Überlagerungsdichte (vergleiche die Varianz in (8.3)). Beim Histogramm-Matching-Verfahren ist daher für die Schätzung des Parameters b vor allem die Lage des Erwartungswertes maßgeblich.

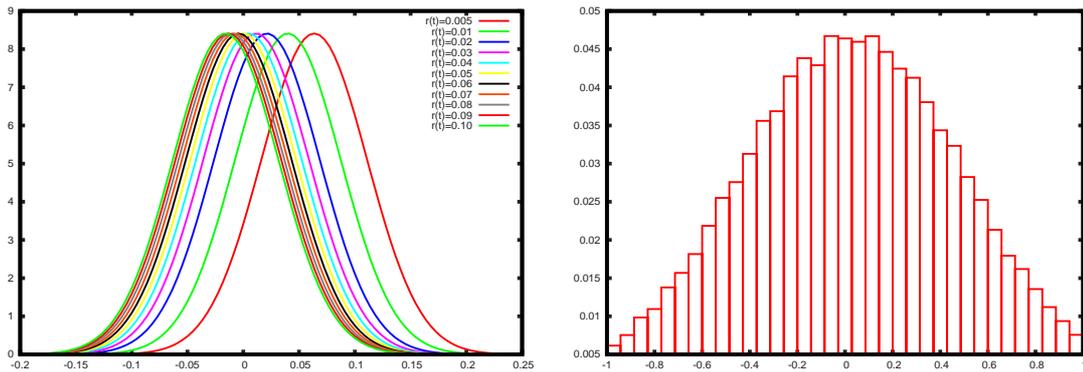
Die Skalierung der Zuwächse bezüglich des Parameters σ

Für den Diffusionsparameter σ bringt eine Skalierung nur leichte Besserung. Entsprechend der Überlegungen zum Parameter b soll erreicht werden, dass die Varianzen $\text{var}[\Delta r_i]$ der einzelnen Übergangsdichten schließlich unabhängig von r_{i-1} sind, so dass σ allein für diese

Varianzen ausschlaggebend ist. Dies wird durch eine Skalierung entsprechend

$$\frac{\Delta r_i}{\sqrt{r_i \Delta t}} \sim N \left(\frac{b \cdot (m - r_{i-1}) \cdot \sqrt{\Delta t}}{\sqrt{r_i}}, \sigma^2 \right). \quad (8.4)$$

erreicht. Die Breite der Überlagerungsdichte wird durch die Skalierung zu einem großen Teil vom Parameter σ beziehungsweise der Varianz $\text{var}[\Delta r_i]$ der einzelnen Übergangsdichten bestimmt.⁵



(a) Skalierte Übergangsdichten der Zuwächse des diskretisierten CIR-Modells für verschiedene vorausgehende Werte der Short Rate.

(b) Überlagerung der Übergangsdichten in approximativer Histogrammdarstellung (Datensatzumfang $N = 300$, Simulationsanzahl $M = 500$).

Abbildung 8.4.: Die Dichtefunktionen der bezüglich des Parameters σ nach (8.4) skalierten Zuwächse für $b = 1.0$, $\sigma = 0.15$, $m = 0.05$ und $\Delta t = 0.1$.

⁵Zusätzlich wird die Varianz natürlich durch die Lage der einzelnen Mittelwerte der Überlagerungsdichten beeinflusst. Allerdings befinden sich diese r_{i-1} und damit auch die Erwartungswerte aufgrund des Mean-Reversion Prinzips in einem gewissen Bereich.

8.4.2. Das iterative Histogramm-Matching-Verfahren

Um alle Parameter des CIR-Modells auf Grundlage eindimensionaler Histogramm-Matching-Verfahren schätzen zu können, muss auf ein iteratives Verfahren zurückgegriffen werden. In Tabelle 8.4 ist ein iterativer Schätzalgorithmus für das CIR-Modell mittels eindimensionaler Histogramm-Matching-Schätzungen angegeben.

Als Startwert kann einerseits eine Schätzung von σ als Varianz der mit \sqrt{r} skalierten historischen Zuwächse über

$$\hat{\sigma} = \sqrt{\text{var} \left[\frac{dr}{\sqrt{r}} \right] / dt}$$

herangezogen werden. Andererseits kann ein Startwert für den Parameter b über eine Schätzung des Erwartungswertes der mit $(m-r)$ skalierten historischen Zuwächse bestimmt werden:

$$\hat{b} = E \left[\frac{dr}{(m-r)} \right] / dt.$$

Das Abbruchkriterium orientiert sich an den Veränderungen der geschätzten Werte. An dieser Stelle besteht zusätzlich die Möglichkeit einer adaptiven Verfeinerung der eindimensionalen Parametergitter.

ITERATIVER SCHÄTZALGORITHMUS FÜR DAS CIR-MODELL MITTELS EINDIMENSIONALER HISTOGRAMM-MATCHING-VERFAHREN	
INPUT:	Historische Short-Rate-Daten r_0, \dots, r_N , Zeitzuwachs Δt
Initialisierung:	Schätze das historische Mean-Reversion-Niveau m (Algorithmus in Tabelle 8.2); Schätze Startwert b_{neu} (Mittelwert der bezüglich b skalierten historischen Datenzuwächse) $\xrightarrow{\text{gehe zu}} (1)$; oder σ_{neu} (Standardabweichung der bezüglich σ skalierten historischen Datenzuwächse) $\xrightarrow{\text{gehe zu}} (2)$;
	Eindimensionales Histogramm-Matching-Verfahren (Tabelle 8.3): (1) σ_{neu} auf Grundlage des Startwertes b_{neu} ; (2) b_{neu} auf Grundlage des Startwertes σ_{neu} ;
Schleife:	DO $\sigma_{alt} = \sigma_{neu}$ und $b_{alt} = b_{neu}$; Eindimensionales Histogramm-Matching-Verfahren: (1) σ_{neu} auf Grundlage des Startwertes b_{neu} ; (2) b_{neu} auf Grundlage des Startwertes σ_{neu} ; WHILE ($ b_{neu} - b_{alt} > \varepsilon_b$ && $ \sigma_{neu} - \sigma_{alt} > \varepsilon_\sigma$);
OUTPUT:	Geschätzte Parameter $\hat{b} = b_{neu}$ und $\hat{\sigma} = \sigma_{neu}$

Tabelle 8.4.: Iterativer Schätzalgorithmus für das CIR-Modell mittels eindimensionaler Histogramm-Matching-Verfahren (Reihenfolge von (1) und (2) hängt von der Wahl des Startwertes ab).

8.4.3. Paralleles (Dünngitter-)Histogramm-Matching-Verfahren

Die in 8.4.1 eingeführten Skalierungen ermöglichen den Aufbau eines zweidimensionalen Histogramms, für das pro Achse die in Bezug auf einen zu schätzenden Parameter skalierten Zuwächse einzuordnen sind. Demnach existiert bei einer Parallelschätzung für jeden der beiden Parameter des CIR-Modells eine Achse. Tabelle 8.5 zeigt einen parallelen Schätzalgorithmus für das CIR-Modell mittels (paralleler) zweidimensionaler Histogramm-Matching-Verfahren. Werden die zweidimensionalen Histogramme auf Grundlage des Dünngitter-Ansatzes erstellt (vergleiche Kapitel 7), so wird vom Dünngitter-Histogramm-Matching-Verfahren gesprochen. Hier besteht, wie beim iterativen Verfahren, zusätzlich die Möglichkeit einer adaptiven Verfeinerung des zweidimensionalen Parametergitters.

PARALLELER SCHÄTZALGORITHMUS FÜR DAS CIR-MODELL MITTELS ZWEIDIMENSIONALER HISTOGRAMM-MATCHING-VERFAHREN	
INPUT:	Historische Short-Rate-Daten r_0, \dots, r_N , Zeitzuwachs Δt
Initialisierung:	Schätze das historische Mean-Reversion-Niveau m (Algorithmus in Tabelle 8.2);
	Zweidimensionales Histogramm-Matching-Verfahren für σ_{neu} und b_{neu} (Tabelle 8.3, Kapitel 8.4);
OUTPUT:	Geschätzte Parameter $\hat{b} = b_{neu}$ und $\hat{\sigma} = \sigma_{neu}$;

Tabelle 8.5.: Paralleler Schätzalgorithmus für das CIR-Modell mittels zweidimensionaler Histogramm-Matching-Verfahren.

8.5. Das (diskretisierte) Maximum-Likelihood-Verfahren

Als Vergleichsverfahren für die vorgestellten Histogramm-Matching-Verfahren wird die Maximum-Likelihood-Schätzung des diskretisierten CIR-Modells eingesetzt. Diese Schätzmethode beruht auf der bereits 1986 von Andrew W. Lo [Lo86] vorgestellten Maximum-Likelihood-Schätzung generalisierter Itô-Prozesse mittels diskret generierter Daten. Aufgegriffen worden ist diese Idee unter anderem von Michael Sørensen, der sich mit der Spezifizierung verschiedener Schätzfunktionen bezüglich der diskret beobachteten Daten beschäftigt hat. Die Arbeitsgruppe Tom Fischer, Angelika May und Brigitte Walther referenzieren Sørensen [Sø97] in ihrer Arbeit zur „Anpassung eines CIR-1-Modells zur Simulation der Zinsstrukturkurve“ [FMW03]. Nach Lo ist die Likelihood-Funktion als Lösung einer besonderen funktionalen partiellen Differentialgleichung charakterisiert. Insbesondere wird bei dieser Form der Schätzung davon ausgegangen, dass die Zustandsvariable, im vorliegenden Fall die Short Rate, direkt am Markt zu beobachten ist. Ansonsten, so merkt Lo an, werde die zugrunde liegende Analysis stark verkompliziert (siehe dazu auch die in Kapitel 8.6 erläuterte Problematik im Zwei-Faktor-Fall des CIR-Modells). Dabei wird eine Approximation der wahren Likelihood-Funktion verwendet. Statt der tatsächlichen Dichtefunktion, die bezüglich des CIR-Modells auf einer Nichtzentralen Chi-Quadrat-Verteilung beruht, werden die Übergangsdichten der (mittels des Euler-Maruyama-Schemas) diskretisierten stochastischen Differentialgleichung eingesetzt. Dadurch ergibt sich die für die weitere Verwendung vorteilhafte relativ einfache (asymptotische) Normalverteilung. ([Sø97],[FMW03])

Das Verfahren

Bei einer Maximum-Likelihood-Schätzung wird zunächst auf Basis der beobachteten diskreten Datenwerte $x = (x_1, \dots, x_N)$ mit $x_j = x(t_j)$ ein statistisches Modell gesucht, das den Zufallsmechanismus, der den beobachteten Daten zugrunde liegt, bis auf einen unbekanntem Parametersatz $\theta \in \Theta \subset \mathbb{R}^k$ möglichst gut wiedergibt (siehe [See05]). Auf Grundlage dieses statistischen Modells wird eine Likelihood-Funktion in Form einer Dichterepräsentation gesucht, die das Verhalten und sämtliche Informationen des Prozesses beschreibt. Mit Hilfe von Maximierung werden dann die für den vorgegebenen Datensatz optimalen Parameterwerte in Bezug auf das Modell bestimmt.

Die Likelihood-Funktion

Das statistische Modell ist im vorliegenden Fall durch die stochastische Differentialgleichung des CIR-Modells gegeben. Im Idealfall kann die Likelihood-Funktion durch die exakte Dichtefunktion des beschriebenen Prozesses definiert werden, wenn das Modell durch eine stochastische Differentialgleichung gegeben ist. Dann wird auch von der *wahren Likelihood-Funktion* (engl. *true maximum likelihood estimator*) gesprochen. Ansonsten bietet sich der Einsatz der Übergangsdichten für die Likelihood-Funktion an; denn wenn es sich, wie im vorliegenden Fall, bei dem durch die stochastische Differentialgleichung beschriebenen Prozess $X(t)$ um einen Markov-Prozess handelt, kann die Dichtefunktion als das Produkt der einzelnen (bedingten) Übergangsdichten betrachtet werden ([Lo86]):

$$\rho(X) = \rho_1(X_1) \prod_{i=2}^N \rho(X_i, t_i | X_{i-1}, t_{i-1}). \quad (8.5)$$

Die Übergangsdichten enthalten sämtliche den Prozess betreffende Informationen (siehe Kapitel 3.5) und können als Basis einer Maximum-Likelihood-Schätzung dienen. Die entscheidende Aufgabe besteht in ihrer Bestimmung. Die Übergangsdichten sind durch die Lösung einer entsprechenden funktionalen partiellen Differentialgleichung charakterisiert. Doch die Lösung dieser partiellen Differentialgleichung ist oft mühsam, denn die Dichtefunktionen sind nicht immer gut analytisch handhabbar. Oft wird dann auf eine Lösung entsprechend „educated guesses“ (fundierter Vermutungen) zurückgegriffen ([Lo86]). Eine möglichst gute Approximation der Übergangsdichten wird gesucht. Im Fall des CIR-Modells kann, so auch [FMW03], eine Approximation der Übergangsdichten mittels Normalverteilung mit Erwartungswert $r_{i-1} + b(m - r_{i-1})\Delta t = r_{i-1} + (a - b \cdot r_{i-1})\Delta t$ und Varianz $\sigma^2 r_{i-1}\Delta t$ verwendet werden. Aufgrund dieser Approximation sind derartige Schätzer allerdings verzerrt (demnach nicht erwartungstreu). Die approximierten Likelihood-Funktionen konvergieren jedoch für eine große Anzahl simulierter Pfade und eine kleine Schrittweite $h = \Delta t$ (große Beobachtungsdichte) gegen die wahre Likelihood-Funktion. Ihr Einsatz ist somit gerechtfertigt. ([FMW03])

Die Normalverteilung als Approximation der wahren Übergangsdichten des CIR-Modells

Wie bereits erläutert, kann beim CIR-Modell auf die Möglichkeit einer approximierten Likelihood-Funktion mittels normalverteilter Übergangsdichten mit Mittelwert $\tilde{\mu} = r_{i-1} + (a - b \cdot r_{i-1})\Delta t$ und Varianz $\tilde{\sigma}^2 = \sigma^2 r_{i-1}\Delta t$ zurück gegriffen werden. Die Likelihood-Funktion $L_N : \Theta \rightarrow [0, \infty)$ ist dann entsprechend (8.5) gegeben durch

$$\begin{aligned} L_N(\theta) &= \prod_{i=1}^N \frac{1}{\sqrt{2\pi\tilde{\sigma}^2}} \exp\left[-\frac{(r_i - \tilde{\mu})^2}{2\tilde{\sigma}^2}\right] \\ &= \left[\frac{1}{\sqrt{2\pi\sigma\sqrt{\Delta t}}}\right]^N \prod_{i=1}^N \frac{1}{\sqrt{r_{i-1}}} \exp\left[-\frac{1}{2\sigma^2\Delta t} \sum_{i=1}^N \frac{(r_i - (r_{i-1} + (a - b \cdot r_{i-1})\Delta t))^2}{r_{i-1}}\right] \end{aligned} \quad (8.6)$$

für den Parametersatz $\theta = (a, b, \sigma^2) \in \Theta$. Statt dieser Likelihood-Funktion wird im Folgenden jedoch die Log-Likelihood-Funktion $l_N : \Theta \rightarrow \mathbb{R}^k \cup \{-\infty\}$ mit

$$\begin{aligned} l_N(\theta) &= -N \ln(\sqrt{2\pi}) - \frac{N}{2} \ln(\sigma^2 \Delta t) + \sum_{i=1}^N \ln\left(\frac{1}{\sqrt{r_{i-1}}}\right) \\ &\quad - \frac{1}{2\sigma^2\Delta t} \sum_{i=1}^N \frac{(r_i - (r_{i-1} + (a - b \cdot r_{i-1})\Delta t))^2}{r_{i-1}} \end{aligned} \quad (8.7)$$

für $L_N(\theta) > 0$ und $l_N(\theta) = -\infty$ für $L_N(\theta) = 0$ verwendet. Sie übernimmt – als streng monotone Transformation der Likelihood-Funktion – deren Eigenschaft, für den jeweils fest gewählten Parametersatz θ ein Stützbarkeitsmaß in Bezug auf die beobachteten Daten darzustellen. Gesucht wird der Parametersatz $\hat{\theta} = \{a, b, \sigma^2\} \in \Theta \subset \mathbb{R}^k$, mit dem die Likelihood-Funktion beziehungsweise die entsprechende Log-Likelihood-Funktion maximiert wird:

$$L_N(\hat{\theta}) = \max_{\theta \in \Theta} L_N(\theta) \quad \text{beziehungsweise} \quad l_N(\hat{\theta}) = \max_{\theta \in \Theta} l_N(\theta).$$

Um die Log-Likelihood-Funktion bezüglich ihrer Parameterwerte zu maximieren, werden die entsprechenden partiellen Ableitungen gebildet:

$$\begin{aligned}\frac{\partial l_N(\theta)}{\partial \sigma^2} &= -\frac{n}{2} \cdot \frac{1}{\sigma^2 \Delta t} \cdot \Delta t - (-1) \cdot \frac{1}{2\sigma^2 \Delta t} \sum_{i=1}^n \frac{(r_i - (r_{i-1} + (a - b \cdot r_{i-1})\Delta t))^2}{r_{i-1}} \\ &= -\frac{n}{2\sigma^2} + \frac{1}{2\sigma^4 \Delta t} \sum_{i=1}^n \frac{(r_i - (r_{i-1} + (a - b \cdot r_{i-1})\Delta t))^2}{r_{i-1}}\end{aligned}\quad (8.8)$$

$$\begin{aligned}\frac{\partial l_N(\theta)}{\partial a} &= \sum_{i=1}^n -2 \cdot \frac{1}{2\sigma^2 \Delta t r_{i-1}} (r_i - (r_{i-1} + (a - b \cdot r_{i-1})\Delta t)) \cdot (-\Delta t) \\ &= \sum_{i=1}^n \frac{1}{\sigma^2 r_{i-1}} (r_i - (r_{i-1} + (a - b \cdot r_{i-1})\Delta t))\end{aligned}\quad (8.9)$$

$$\begin{aligned}\frac{\partial l_N(\theta)}{\partial b} &= \sum_{i=1}^n -2 \cdot \frac{1}{2\sigma^2 \Delta t r_{i-1}} (r_i - (r_{i-1} + (a - b \cdot r_{i-1})\Delta t)) \cdot (r_{i-1} \Delta t) \\ &= \sum_{i=1}^n -\frac{1}{\sigma^2} (r_i - (r_{i-1} + (a - b \cdot r_{i-1})\Delta t)).\end{aligned}\quad (8.10)$$

Aus $\frac{\partial l_N(\theta)}{\partial \sigma^2} = 0$ ergibt sich für den Parameter $\sigma^2 > 0$:

$$\begin{aligned}0 &= -\frac{n}{2\sigma^2} + \frac{1}{2\sigma^4 \Delta t} \sum_{i=1}^n \frac{(r_i - (r_{i-1} + (a - b \cdot r_{i-1})\Delta t))^2}{r_{i-1}} \\ \Leftrightarrow \sigma^2 &= \frac{1}{n\Delta t} \sum_{i=1}^n \frac{1}{r_{i-1}} (r_i - (r_{i-1} + (a - b \cdot r_{i-1})\Delta t))^2.\end{aligned}\quad (8.11)$$

Das Auflösen von $\frac{\partial l_N(\theta)}{\partial b} = 0$ nach a ergibt mit $\sigma^2 > 0$:

$$\begin{aligned}0 &= \sum_{i=1}^n -\frac{1}{\sigma^2} (r_i - (r_{i-1} + (a - b \cdot r_{i-1})\Delta t)) \\ \Leftrightarrow 0 &= \sum_{i=1}^n (r_i - r_{i-1}) + \sum_{i=1}^n (-a\Delta t) + \sum_{i=1}^n (br_{i-1}\Delta t) \\ \Leftrightarrow a &= \frac{1}{n\Delta t} \left[(r_n - r_0) + b\Delta t \sum_{i=1}^n r_{i-1} \right].\end{aligned}\quad (8.12)$$

Das Einsetzen von (8.12) in $\frac{\partial l_N(\theta)}{\partial a} = 0$ führt zu

$$\begin{aligned}
 0 &= \sum_{i=1}^n -\frac{1}{\sigma^2 r_{i-1}} \left(r_i - r_{i-1} - \left(\frac{1}{n\Delta t} \left[(r_n - r_0) + b\Delta t \sum_{j=1}^n r_{j-1} \right] - br_{i-1} \right) \Delta t \right) \\
 \Leftrightarrow 0 &= \sum_{i=1}^n \frac{r_i - r_{i-1}}{r_{i-1}} - \sum_{i=1}^n \frac{1}{r_{i-1}} \left(\frac{1}{n\Delta t} (r_n - r_0) + \frac{b}{n} \sum_{j=1}^n r_{j-1} \right) \Delta t + \sum_{i=1}^n b \frac{r_{i-1}}{r_{i-1}} \Delta t \\
 \Leftrightarrow 0 &= \sum_{i=1}^n \left(\frac{r_i}{r_{i-1}} - 1 \right) - \frac{1}{n} (r_n - r_0) \sum_{i=1}^n \frac{1}{r_{i-1}} - \frac{b\Delta t}{n} \sum_{i=1}^n \left(\frac{1}{r_{i-1}} \sum_{j=1}^n r_{j-1} \right) + b\Delta t n \\
 \Leftrightarrow 0 &= \sum_{i=1}^n \frac{r_i}{r_{i-1}} - n - \frac{1}{n} (r_n - r_0) \sum_{i=1}^n \frac{1}{r_{i-1}} - b\Delta t \left(-n + \frac{1}{n} \left(\sum_{i=1}^n r_{i-1} \right) \left(\sum_{i=1}^n \frac{1}{r_{i-1}} \right) \right) \\
 \Leftrightarrow b &= \frac{\frac{1}{n} (r_n - r_0) \sum_{i=1}^n \frac{1}{r_{i-1}} - \sum_{i=1}^n \frac{r_i}{r_{i-1}} + n}{\Delta t \left(n - \frac{1}{n} \left(\sum_{i=1}^n r_{i-1} \right) \left(\sum_{i=1}^n \frac{1}{r_{i-1}} \right) \right)}. \tag{8.13}
 \end{aligned}$$

Für den Parameter a lässt sich aus 8.12 und 8.13 schließen

$$\begin{aligned}
 a &= \frac{1}{n\Delta t} \left[(r_n - r_0) + \frac{\frac{1}{n} (r_n - r_0) \sum_{i=1}^n \frac{1}{r_{i-1}} - \sum_{i=1}^n \frac{r_i}{r_{i-1}} + n}{\Delta t \left(n - \frac{1}{n} \left(\sum_{i=1}^n r_{i-1} \right) \left(\sum_{i=1}^n \frac{1}{r_{i-1}} \right) \right)} \Delta t \sum_{i=1}^n r_{i-1} \right] \\
 \Leftrightarrow a &= \frac{(r_n - r_0) \left[n \left(\sum_{i=1}^n r_{i-1} \right)^{-1} - \frac{1}{n} \sum_{i=1}^n \frac{1}{r_{i-1}} \right]}{\Delta t \left[n^2 \left(\sum_{i=1}^n r_{i-1} \right)^{-1} - \sum_{i=1}^n \frac{1}{r_{i-1}} \right]} \tag{8.14}
 \end{aligned}$$

$$\begin{aligned}
 &+ \frac{\frac{1}{n} (r_n - r_0) \sum_{i=1}^n \frac{1}{r_{i-1}} - \sum_{i=1}^n \frac{r_i}{r_{i-1}} + n}{\Delta t \left[n^2 \left(\sum_{i=1}^n r_{i-1} \right)^{-1} - \sum_{i=1}^n \frac{1}{r_{i-1}} \right]} \\
 \Leftrightarrow a &= \frac{n(r_n - r_0) \left(\sum_{i=1}^n r_{i-1} \right)^{-1} - \sum_{i=1}^n \frac{r_i}{r_{i-1}} + n}{\Delta t \left[n^2 \left(\sum_{i=1}^n r_{i-1} \right)^{-1} - \sum_{i=1}^n \frac{1}{r_{i-1}} \right]}. \tag{8.15}
 \end{aligned}$$

Preis Anpassung im Fall realer Daten

Werden die Parameter des Short-Rate-Modells anhand realer (am Markt beobachtbarer) Daten geschätzt, so steht noch aus, den bereits in Kapitel 4.2.1 erläuterten Maßwechsel vom realen zum äquivalenten risikoneutralen Wahrscheinlichkeitsmaß zu spezifizieren. In diesem Zusammenhang muss der Marktpreis des Risikos (Definition 4.2.24) direkt ermittelt

werden (vergleiche Folgerung 4.2.15). Andernfalls sind die von diesem Marktpreis beeinflussten, unter dem realen Wahrscheinlichkeitsmaß geschätzten Parameter so anzupassen, dass die reale und die auf Grundlage des Short-Rate-Modells generierte Zinsstruktur für den aktuellen Zeitpunkt $t = 0$ möglichst gut übereinstimmen. Diese Anpassung kann über die Preise der entsprechenden Nullkupon-Anleihen erfolgen.⁶ Sind die theoretischen Preise, die basierend auf den im realen Wahrscheinlichkeitsraum geschätzten Parametern und einer Näherung für $r(0)$ (etwa durch den aktuellen Ein-Tages-Zinssatz [FMW03]) ermittelt werden, durch $P_j = P(0, \tau_j, r(0))$ gegeben (vergleiche (4.24)) und sind die entsprechenden realen Preise mit $\hat{P}_j = \hat{P}(0, \tau_j, r(0))$ bezeichnet, so kann eine Anpassung (engl. Fitting) der theoretischen an die realen Preise mittels der Methode der kleinsten Quadrate erfolgen. ([FMW03])

Im CIR-Modell ist lediglich der Parameter $a (= b \cdot \mathbf{m})$ vom Maßwechsel betroffen; es gilt: $\tilde{a} = a - \lambda_0 \sigma$ mit $\lambda_0 = \frac{\lambda(t,r)}{\sqrt{r}}$ ([CIR85], [FMW03]). Der entsprechende Parameter \tilde{a} der theoretischen Welt kann durch die Methode der kleinsten Quadrate ermittelt werden:

$$\tilde{a} = \arg \min_a \sum_{j=1}^k (P_j - \hat{P}_j)^2 .$$

Dabei werden lediglich die aktuellen Preise berücksichtigt. Um auch historische Daten einzubeziehen, kann folgende Schätzung verwendet werden:

$$\tilde{a} = \arg \min_a \sum_{i=1}^N \sum_{j=1}^k (P(t_i, \tau_j, r(t_i)) - \hat{P}(t_i, \tau_j, r(t_i)))^2 .$$

Die theoretischen Preise bezüglich des CIR-Modells sind durch (4.39) gegeben.

⁶Für das Verhältnis der stetigen Kassazinsen, die entsprechend Definition 4.2.5 die Zinsstruktur bilden, und der Preise der Nullkupon-Anleihen vergleiche (4.4).

8.6. Die Schätzproblematik des Zwei-Faktor-Falls des CIR-Modells

Entsprechend der in Kapitel 4.4 vorgestellten Mehr-Faktor-Modelle setzt sich die Short Rate im Zwei-Faktor-Fall als Summe zweier unkorrelierter Teilraten $r(t) = \sum_{i=1}^2 r_i(t)$ (4.43) zusammen. Die diskretisierten stochastischen Differentialgleichungen dieser Teilraten sind durch

$$\begin{aligned}\Delta r_{1,i} &= b_1 (m_1 - r_{1,i-1}) \Delta t + \sigma_1 \sqrt{r_{1,i-1}} \Delta W_{1,i} \\ \Delta r_{2,i} &= b_2 (m_2 - r_{2,i-1}) \Delta t + \sigma_2 \sqrt{r_{2,i-1}} \Delta W_{2,i}\end{aligned}$$

basierend auf zwei unabhängigen (unkorrelierten) diskretisierten Wiener-Prozessen $\Delta W_{1,i}$ und $\Delta W_{2,i}$ für $i \in \mathbb{N}_0$ gegeben (vergleiche Kapitel 4.4.2 (4.47) sowie Kapitel 5, Tabelle 5.1). Dabei sind b_1, b_2, m_1, m_2 sowie σ_1 und σ_2 positive Konstanten. Wie im eindimensionalen Fall des CIR-Modells (vergleiche Kapitel 5, Tabelle 5.1) sind diese diskretisierten Zuwächse der beiden Teilraten normalverteilt entsprechend

$$\begin{aligned}\Delta r_{1,i} &\sim N(b_1 (m_1 - r_{1,i-1}) \Delta t, \sigma_1^2 r_{1,i-1} \Delta t) \\ \Delta r_{2,i} &\sim N(b_2 (m_2 - r_{2,i-1}) \Delta t, \sigma_2^2 r_{2,i-1} \Delta t).\end{aligned}$$

Während für den Ein-Faktor-Fall der Short-Rate-Modelle der Tagesgeldsatz als Approximation der Short Rate zur Schätzung herangezogen werden kann, ist dies für den Mehr-Faktor-Fall nicht möglich. In Kapitel 4.2.1 (4.4) wurde dargelegt, dass sich die Zinsstruktur über

$$R(t, T) = -\frac{1}{T-t} \ln P(r, t; T)$$

aus der Short Rate bestimmen lässt. Für den Mehr-Faktor-Fall des CIR-Modells mit k Teilraten bedeutet dies nach (4.48)

$$R(t, T) = \sum_{j=1}^k \left(-\frac{A_j(\tau)}{\tau} + \frac{B_j(\tau)}{\tau} r_j(t) \right)$$

$$\text{mit } A_j(\tau; b_j, m_j, \sigma_j) = \frac{2b_j m_j}{\sigma_j^2} \ln \left[\frac{2h_j e^{(b_j+h_j)\tau/2}}{(h_j + b_j)(e^{h_j\tau} - 1) + 2h_j} \right]$$

$$\text{und } B_j(\tau; b_j, \sigma_j) = \frac{2(e^{h_j\tau} - 1)}{(h_j + b_j)(e^{h_j\tau} - 1) + 2h_j}$$

$$\text{mit } h_j(b_j, \sigma_j) = \sqrt{b_j^2 + 2\sigma_j^2}.$$

Dementsprechend ist für die Schätzung der Short Rate im Falle zweier Teilraten $r = r_1 + r_2$ ($k = 2$) die Zinsstruktur für zwei verschiedene Fälligkeitsdaten zu betrachten:

$$\implies R(t, T_1), R(t, T_2) \quad \text{für } k = 2.$$

Die Zuwächse dieser Zinssätze mit den Fälligkeitsdaten T_1 und T_2 lassen sich über

$$\Delta R(i, T_1) = \sum_{j=1}^2 \left(-\frac{A_j(\tau_1)}{\tau_1} + \frac{B_j(\tau_1)}{\tau_1} \Delta r_{j,i} \right) \quad (8.16)$$

$$\Delta R(i, T_2) = \sum_{j=1}^2 \left(-\frac{A_j(\tau_2)}{\tau_2} + \frac{B_j(\tau_2)}{\tau_2} \Delta r_{j,i} \right) \quad (8.17)$$

bestimmen. Da die $A_j(\tau; b_j, m_j, \sigma_j)$ und $B_j(\tau; b_j, \sigma_j)$ allerdings auch von den unbekanntem (zu schätzenden) Parametern m_j , b_j und σ_j für $j = 1, 2$ abhängen, können die am Markt beobachteten Zinssätze $R(t, T_1)$ und $R(t, T_2)$ nicht für die Teilraten der Short Rate transformiert werden. Eine Schätzung der Parameter kann somit nicht direkt über die beiden Teilraten erfolgen.

Eine weitere Ansatzmöglichkeit besteht darin, die Parameter über die $R(t, T_1)$ und $R(t, T_2)$ zu schätzen. Das heißt, die Histogramme werden nicht für die diskreten Zuwächse Δr_i der Short Rate generiert, sondern für diejenigen der beiden Kassazinssätze: $\Delta R(i, T_1)$ und $\Delta R(i, T_2)$. Außerdem wird nicht mehr die Überlagerung der Übergangsdichten der diskretisierten Zuwächse der Short Rate approximiert, sondern die der Übergangsdichten der diskretisierten Zuwächse der beiden Zinssätze $R(t, T_1)$ und $R(t, T_2)$. Diese diskretisierten Zuwächse sind normalverteilt entsprechend

$$\begin{aligned} \Delta R(i, T_1) &\sim N \left(\sum_{j=1}^2 \frac{(-A_j(\tau_1) + B_j(\tau_1)b_j(m_j - r_{j,i-1})\Delta t)}{\tau_1}, \sum_{j=1}^2 \frac{(B_j(\tau_1)\sigma_j^2 r_{j,i-1}\Delta t)}{\tau_1} \right) \\ \Delta R(i, T_2) &\sim N \left(\sum_{j=1}^2 \frac{(-A_j(\tau_2) + B_j(\tau_2)b_j(m_j - r_{j,i-1})\Delta t)}{\tau_2}, \sum_{j=1}^2 \frac{(B_j(\tau_2)\sigma_j^2 r_{j,i-1}\Delta t)}{\tau_2} \right). \end{aligned}$$

Damit sind, wie auch bei den diskretisierten Zuwächsen Δr_i der Short Rate, sowohl die Erwartungswerte $E[\Delta R(i, T_1)]$ und $E[\Delta R(i, T_2)]$ als auch die Varianzen $\text{var}[\Delta R(i, T_1)]$ und $\text{var}[\Delta R(i, T_2)]$ der Verteilungen von den jeweils vorausgehenden Werten $r_{j,i-1}$ der Teilraten für $j = 1, 2$ abhängig. Zudem kommt erschwerend hinzu, dass die zu schätzenden Parameter b_1, b_2, σ_1 und σ_2 sowohl in die Erwartungswerte als auch in die Varianzen einfließen. Auch die Mean-Reversion-Niveaus m_1 und m_2 der beiden Teilraten sind in den Erwartungswerten enthalten. Es gelten insgesamt folgende Abhängigkeiten:

$$E(m_1, m_2, b_1, b_2, \sigma_1, \sigma_2, r_{1,i-1}, r_{2,i-1}) \quad \text{sowie} \quad \text{var}(b_1, b_2, \sigma_1, \sigma_2, r_{1,i-1}, r_{2,i-1}).$$

Die Schätzungen des Ein-Faktor-Falls des CIR-Modells mittels der Histogramm-Matching-Verfahren basierten bisher auf einer Mittelwertschätzung für das Mean-Reversion-Niveau m . Diese Schätzung ist für die beiden Mean-Reversion-Niveaus m_1 und m_2 auf Grund nicht am Markt beobachtbarer Teilraten nicht möglich. Daher müssen auch diese Parameter mittels der Histogramm-Matching-Verfahren geschätzt werden. Die Schätzung der Parameter über die Histogramm-Matching-Verfahren gestaltet sich allerdings schwierig. Für die Schätzung der Parameter des Ein-Faktor-Falls des CIR-Modells wurden in Kapitel 8.4.1 Skalierungen eingeführt, aufgrund derer es möglich ist, die Parameter einzeln über den Erwartungswert beziehungsweise die Varianz zu schätzen. Zudem erlauben diese Skalierungen den Aufbau zweidimensionaler Histogramme, die für eine parallele Schätzung der Parameter genutzt werden können. Wie in Kapitel 9.2 dargelegt, führt die Einführung der Skalierungen zu einer deutlichen Besserung der Schätzergebnisse. Werden jetzt die Zuwächse der Kassazinsen ((8.16) und (8.17)) betrachtet, so ist eine solche Skalierung nicht mehr möglich. Es gelingt nicht, den Erwartungswert beziehungsweise die Varianz so zu skalieren, dass sie unabhängig von den zuletzt erreichten Werten der Teilraten ist. Das hat einerseits zur Folge, dass keine Verbesserung der Schätzung erreicht werden kann; andererseits entfällt die Möglichkeit, mehrdimensionale Histogramme für die Schätzung nutzen zu können.

Demnach müssten alle Parameter aus dem Vergleich der beiden eindimensionalen Histogramme für die Zuwächse $\Delta R(i, T_1)$ und $\Delta R(i, T_2)$ geschätzt werden.

Die einzelnen Parameter können allerdings in einer parallelen Schätzung mittels eindimensionaler Histogramm-Matching-Verfahren kaum gefasst werden. Dies liegt daran, dass die Erwartungswerte $E[\Delta R(i, T_1)]$ und $E[\Delta R(i, T_2)]$ der Verteilungen der diskretisierten Zuwächse sowohl von allen sechs Parameterwerten $m_1, m_2, b_1, b_2, \sigma_1$ und σ_2 als auch den vorhergehenden Werten $r_{1,i-1}$ und $r_{2,i-1}$ der Teilraten abhängen – und auch bei den Varianzen $\text{var}[\Delta R(i, T_1)]$ und $\text{var}[\Delta R(i, T_2)]$ eine Abhängigkeit von fast allen Parameterwerten b_1, b_2, σ_1 und σ_2 (bis auf die Mean-Reversion-Niveaus) und den vorhergehenden Werten $r_{1,i-1}$ und $r_{2,i-1}$ der Teilraten besteht.

8.7. Schätzung eines korrelierten Finanzmodells

Die Parameterschätzung für die einzelnen stochastischen Prozesse erfolgt zuerst getrennt. Im Anschluss können die Korrelationskoeffizienten mittels der diskretisierten stochastischen Differentialgleichungen, den geschätzten Parameterwerten und den historischen Datenwerten ermittelt werden. Ein Algorithmus für den Korrelationskoeffizienten zweier stochastischer Prozesse ist in Tabelle 8.6 angegeben.

SCHÄTZER FÜR DEN KORRELATIONSKOEFFIZIENTEN ZWEIER STOCHASTISCHER PROZESSE	
INPUT:	Historische Daten der Prozesse x_0, \dots, x_N und y_0, \dots, y_N , Zeitschrittweite Δt , die stochastischen Differentialgleichungen dx und dy , Parameterschätzwerte $\hat{\Theta}$;
Initialisierung:	Ermittlung der historischen Zuwächse: $z_{1,i} = x_i - x_{i-1}$ und $z_{2,i} = y_i - y_{i-1}$ für $i = 1, \dots, N$;
Schleife:	für $i = 1, \dots, N$ Herausfiltern der historischen (korrelierten) Zufallszahlen $w_{1,i}$ und $w_{2,i}$ durch Einsetzen der Parameterschätzwerte;
Initialisierung:	$p_1 = w_{1,1} \cdot w_{2,1}$;
Schleife:	für $i = 2, \dots, N$ $p_i = p_{i-1} + w_{1,i} \cdot w_{2,i}$;
	$p_N = p_N / N$;
OUTPUT:	Korrelationskoeffizient $\hat{\rho} = p_N$;

Tabelle 8.6.: Schätzalgorithmus für den Korrelationskoeffizienten zweier stochastischer Prozesse.

Kapitel 9.

Numerische Ergebnisse

9.1. Schätzung des Mean-Reversion-Niveaus

Die in Kapitel 8.2 erläuterte Schätzung des Mean-Reversion-Niveaus über den Mittelwert der historischen Daten (vergleiche Tabelle 8.2) liefert, wie bereits für großes N zu erwarten ist (da $E[r(t)] \xrightarrow{t \rightarrow \infty} m$ nach (4.38)), für das CIR-Modell gute Ergebnisse. Im Vergleich zu den mittels des Maximum-Likelihood-Verfahrens erlangten Schätzwerten (vergleiche Kapitel 8.5) liefert die Mittelwertschätzung die genaueren Schätzergebnisse. In Abbildung 9.1 sind die relativen L_2 -Fehler $\frac{\|m - \hat{m}\|_2}{|m|}$ der beiden Verfahren in Abhängigkeit vom Datensatzumfang N abgebildet. Zur Erstellung dieser Kurven wurde je Datensatzumfang N über 100 Schätzungen gemittelt. Beide Verfahren sind von der Ordnung $\frac{1}{2}$. Die Kurve der Mittelwertschätzung liegt leicht unter der der Maximum-Likelihood-Schätzung.

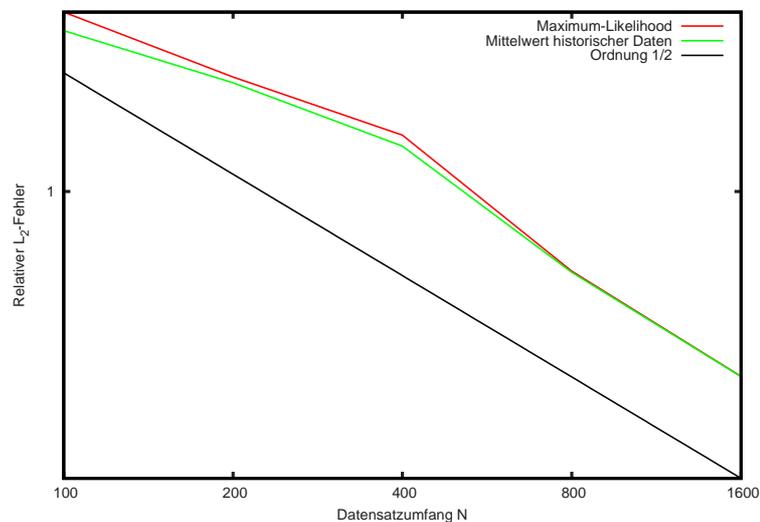


Abbildung 9.1.: Relativer L_2 -Fehler bezüglich der verschiedenen Schätzmethoden des Mean-Reversion-Niveaus des CIR-Modells in Abhängigkeit vom Datensatzumfang N (Mittelung über 100 Schätzungen, wahre Parameterwerte $m = 0.05$ ($b = 1.0$, $\sigma = 0.15$), $\Delta t = 0.1$).

9.2. Skalierungen

Die in Kapitel 8.4.1 für das CIR-Modell erörterten parameterspezifischen Skalierungen führen zu einer deutlichen Verbesserung der Schätzergebnisse. In Abbildung 9.2 sind die relativen L_2 -Fehlerkurven der eindimensionalen Histogramm-Matching-Verfahren auf Grundlage einerseits der unskalierten andererseits der skalierten Datenwerte abgebildet. Die Verfahren sind entsprechend der Diskretisierung mittels des Euler-Maruyama-Schemas von der Ordnung $\frac{1}{2}$. Die Fehlerkurven der skalierten Schätzungen sind um eine Konstante nach unten verschoben.

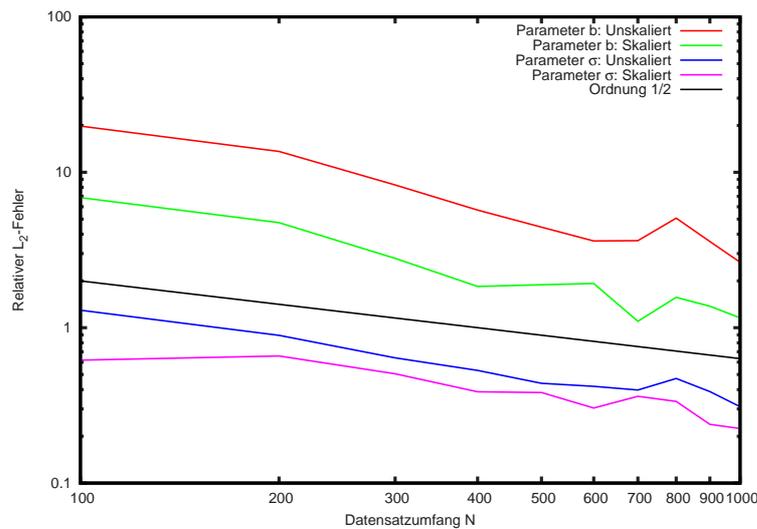


Abbildung 9.2.: Relativer L_2 -Fehler bezüglich der eindimensionalen Histogramm-Matching-Verfahren auf Grundlage der unskalierten sowie der parameterspezifisch skalierten Datenwerte in Abhängigkeit vom Datensatzumfang N (Mittelung über 100 Schätzungen, wahre Parameterwerte $b = 1.0$, $\sigma = 0.15$, $m = 0.05$, Simulationsanzahl $M = 500$, Teilintervallanzahl $k = \sqrt{N}$, Anfangsrate 0.0499, konstanter Zeitzuwachs $\Delta t = 0.1$).

9.3. Verfahrensparameter des eindimensionalen Histogramm-Matching-Verfahrens

Zunächst werden die Verfahrensparameter des eindimensionalen Histogramm-Matching-Verfahrens analysiert, die auch von den in Kapitel 8.4.2 vorgestellten iterativen Histogramm-Matching-Verfahren verwendet werden. Im Anschluss werden die Verfahrensparameter für die in Kapitel 9.4 erörterten parallelen Histogramm-Matching-Verfahren auf Grundlage zweidimensionaler Histogramme untersucht.

Die Analysen bezüglich des eindimensionalen Histogramm-Matching-Verfahrens basieren auf der Annahme, dass nur der zu schätzende Parameter unbekannt ist. Erst bei der Konvergenzuntersuchung zu den vorgestellten iterativen Histogramm-Matching-Verfahren wird von der grundsätzlichen Unkenntnis aller Parameter (CIR-Modell: m , b und σ) ausgegangen.

Die folgenden Betrachtungen beziehen sich auf den Einsatz von Histogrammen auf Grundlage charakteristischer Basisfunktionen. In 9.3.6 werden die Auswirkungen eines Basiswechsels hin zu den stückweise linearen Basisfunktionen dargelegt. Die zuvor basierend auf den charakteristischen Basisfunktionen durchgeführten Analysen der Verfahrensparameter führen allerdings für die stückweise linearen Basisfunktionen zu weitgehend übereinstimmenden Resultaten und werden daher nicht noch einmal getrennt aufgeführt.

Wie bereits in Kapitel 8 festgestellt wurde, haben folgende Verfahrensparameter einen entscheidenden Einfluss auf den Erfolg der Histogramm-Matching-Verfahren:

- *Der Datensatzumfang N .* Er entspricht der Anzahl der historischen Datenwerte und bestimmt gleichzeitig den Umfang der – für die Erstellung der simulierten Histogramme – generierten Datensätze.
- *Die Simulationsanzahl M .* Es handelt sich dabei um die Anzahl der pro festgehaltenem Parametersatz für die Erstellung des simulierten Histogramms generierten Datensätze.
- *Die Intervallgrenzen.* Sie bestimmen, wieviel Prozent der (historischen) Datenwerte für die Erstellung der Histogramme einbezogen werden. Durch die Wahl der Intervallgrenzen ist die Breite b des Trägerintervalls der Histogramme festgelegt.
- *Die Anzahl der Teilintervalle k .* Nach Festlegung der Intervallgrenzen ist sie für die Breite h der Teilintervalle entscheidend. Es gilt $h = \frac{b}{k}$. Diese Breite ist derart auszubalancieren, dass weder ein „Under-“ noch ein Overfitting-Effekt auftritt, der eine gute Schätzung unmöglich machen würde.

Im Folgenden werden Analysen hinsichtlich dieser Verfahrensparameter durchgeführt – mit dem Ziel, geeignete Werte beziehungsweise Regeln für eine Wahl dieser Verfahrensparameter zu finden, die auch – jeweils im Anschluss – verwendet beziehungsweise angewendet werden.

Dabei liegen in der Regel folgende Werte zugrunde:

Wahrer Parameterwerte:	m	0.05
	b	1.0
	σ	0.15
Anfangswert der Short Rate	r_0	0.0499
Verfahrensparameter:		
(Konstanter) Zeitzuwachs (9.3.1)	Δt	0.1
Datensatzumfang (9.3.2)	N	variiert
Simulationsanzahl (9.3.3)	M	$\begin{cases} 25.600/N & N < 512 \\ 100 & N \geq 512 \end{cases}$
Prozentsatz der hinsichtlich der Wahl der Intervallgrenzen nicht beachteten historischen Daten (9.3.4)	p	$\begin{cases} 5\% & \text{bezüglich des Parameters } b \\ 1\% & \text{bezüglich des Parameters } \sigma \end{cases}$
Anzahl der Teilintervalle (9.3.5)	k	$\begin{cases} \sqrt{N} & \text{bezüglich des Parameters } b \\ 4 & \text{bezüglich des Parameters } \sigma \\ & \text{beim Einsatz charakteristischer Basisfunktionen} \\ 5 & \text{bezüglich des Parameters } \sigma \\ & \text{beim Einsatz stückweise linearer Basisfunktionen} \end{cases}$
Anzahl der Schätzungen, über die gemittelt wird:		100

Abweichungen von diesen Werten werden an der entsprechenden Stelle kenntlich gemacht.

9.3.1. Der diskretisierte Zeitzuwachs Δt

Bevor der Stichprobenumfang N näher betrachtet wird, soll kurz die Rolle des diskretisierten Zeitzuwachses Δt erläutert werden. Im Rahmen dieser Arbeit wird mit einem für alle Datensatzumfänge N konstanten Zeitzuwachs $\Delta t = \frac{\tau}{N}$ gearbeitet. τ entspricht der Anzahl von Jahren, die der betrachtete (beziehungsweise der entsprechend zu generierende) Zeitraum umfasst. Für festes N und festes τ ergibt sich somit ein Zeitgitter, bei dem zeitlich äquidistant entfernte Datenwerte (etwa auf monatlicher Basis für $\Delta t = \frac{1}{12}$) beobachtet und generiert werden. Eine Steigerung von N zieht für festes τ einen kleiner werdenden Zeitzuwachs $\Delta t = \frac{\tau}{N}$ nach sich. Diese Verfeinerung der Datenaufösung (monatlich, täglich, minütlich,...) beziehungsweise des Zeitgitters führt zu einer Verbesserung der Approximationsgüte der Diskretisierung (hier: Euler-Maruyama-Schema, vergleiche Kapitel 5). Um die weiteren Betrachtungen losgelöst von der Diskretisierungsgüte für Δt zu machen, wird der Zeitzuwachs konstant für alle N gewählt. Dies wird erreicht, indem τ für konstantes $\Delta t (= \frac{\tau}{N})$ an N angepasst wird. Entsprechend der am Markt beobachtbaren Daten wird im Folgenden von $\Delta t \approx \frac{1}{12}$ ausgegangen. Für die Analysen der Verfahren ist

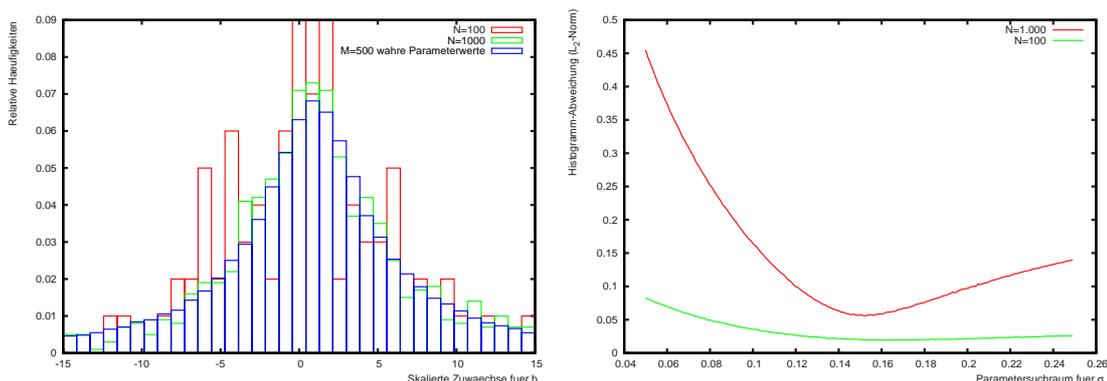
$$\Delta t = 0.1$$

gewählt.

9.3.2. Der Datensatzumfang N

Wie bereits in Kapitel 8 festgestellt, hat sowohl die Qualität als auch die Quantität der betrachteten Stichprobe in Form historischer Daten entscheidenden Einfluss auf die Güte der Schätzwerte. Der Umfang N des historischen Datensatzes ist ausschlaggebend für die Güte der historischen Dichteapproximation in Form des historischen Histogramms. Damit trägt der Verfahrensparameter N maßgeblich zum Erfolg oder Misserfolg der Parameterschätzung mittels der Histogramm-Matching-Verfahren bei.

In Abbildung 9.3(a) sind zum Vergleich zwei (naive) historische Histogramme abgebildet, denen eine unterschiedliche Anzahl historischer Daten zugrunde liegt. Das rote Histogramm baut auf einem Datensatz des Umfangs $N = 100$ auf, das grüne basiert auf $N = 1.000$ Datenwerten. Dabei gilt für beide Datensätze $\Delta t = 0.1$. Im Vergleich der beiden Histogramme mit dem blauen Histogramm, das eine hinreichend gute Approximation der wahren Dichtefunktion darstellt, schneidet das grüne Histogramm deutlich besser ab. Je höher der Datensatzumfang N ist, desto besser kann die wahre Dichtefunktion approximiert werden. Abbildung 9.3(b) zeigt die Histogramm-Abweichungen für die Schätzung des Parame-



(a) Historische Histogramme der für b skalierten Zuwächse des CIR-Modells für $N = 100$ (rot) und $N = 1.000$ (grün) sowie ein simuliertes Histogramm mit $N = 300$ und $M = 500$ (blau).

(b) Histogramm-Abweichungskurven (L_2 -Norm) für $N = 100$ (grün) und $N = 1.000$ (rot). Der wahre Parameterwert entspricht $\sigma = 0.15$, die Anzahl der für die simulierten Histogramme generierten Datensätze ist $M = 500$.

Abbildung 9.3.: Analyse des Datensatzumfangs N , Abbildungen auf Grundlage des naiven eindimensionalen Histogramm-Matching-Verfahrens.

ters σ aus dem CIR-Modell mittels des naiven eindimensionalen Histogramm-Matching-Verfahrens für verschiedene Datensatzumfänge N . Die grüne Kurve basiert auf einem historischen Datensatz des Umfangs $N = 100$. Die rote ist auf Grundlage von $N = 1.000$ historischen Datenwerten entstanden. Die rote Kurve ist spitzer und trifft den wahren Parameter ($\sigma = 0.15$) besser. Dies liegt darin begründet, dass ein – aufgrund des höheren Datensatzumfangs N (vergleiche die Erläuterungen zu Abbildung 9.3(a)) – glatteres historisches Histogramm sich hinsichtlich der simulierten Histogramme besser einordnen lässt.

9.3.3. Die Simulationsanzahl M

Für die Erstellung der simulierten Histogramme werden M einzelne Datensätze des Umfangs N generiert, über die anschließend gemittelt wird. Damit hängen die simulierten Histogramme ebenso wie die historischen Histogramme bezüglich ihrer (Dichte-)Approximationsgüte von N , aber auch von der Simulationsanzahl M ab. Vielmehr können simulierte Histogramme, die auf einem unbefriedigenden Datenumfang N basieren, mittels ausreichend großer Simulationsanzahl M ausgeglichen (geglättet) werden. Je mehr Datensätze für den Aufbau der simulierten Histogramme verwendet werden (je höher also M ist), desto besser approximieren die simulierten Histogramme die wahren Dichtefunktionen. Damit wird die vom Parametersatz abhängende Kurve der Histogramm-Abweichungen glatter und spitzer.

In Abbildung 9.4 sind zwei Histogramm-Abweichungs-Kurven dargestellt, denen das selbe historische Histogramm zugrunde liegt. Die Erstellung ihrer simulierten Histogramme basiert allerdings auf unterschiedlichen Mengen generierter Datensätze. Für die simulierten Histogramme der roten Kurve wurde lediglich ein Datensatz pro Parametersatz generiert. Für die simulierten Histogramme der grünen Kurve wurde je Parametersatz über $M = 1.000$ Datensätze gemittelt. Die Anzahl der Zeitschritte ist mit $N = 500$ genügend hoch gewählt, um den Effekt nicht durch ein unzureichend glattes historisches Histogramm zu verwischen.

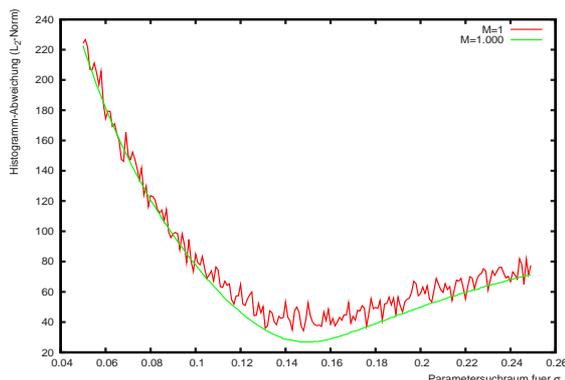


Abbildung 9.4.: Analyse der Simulationsanzahl M , Histogramm-Abweichungskurven (L_2 -Norm) auf Grundlage eines historischen Datensatzes des Umfangs $N = 500$ und simulierter Histogramme, die auf einer unterschiedlichen Anzahl von generierten Datensätzen basieren: $M = 1$ (rot) und $M = 1.000$ (grün). Der wahre Parameterwert entspricht $\sigma = 0.15$.

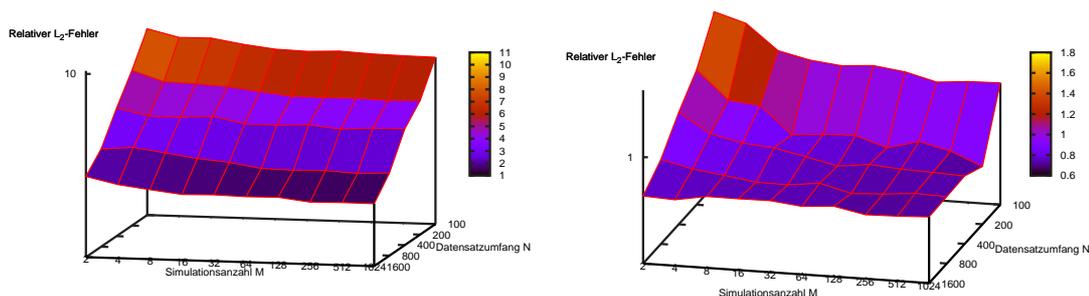
Für die Erstellung glatter simulierter Histogramme ist also entsprechend dieser Feststellungen sowie denen des vorausgehenden Abschnitts die Größe $N \cdot M$ entscheidend.

In Abbildung 9.5 ist der relative L_2 -Fehler des naiven eindimensionalen Histogramm-Matchings für den Parameter b und den Parameter σ des CIR-Modells in Abhängigkeit vom Datensatzumfang N und der Simulationsanzahl M abgebildet.

Für glatte Fehlerkurven ist über 100 Schätzungen gemittelt worden. Außerdem wurden, um die Kurve entlang der Achse der Simulationsanzahl M noch mehr zu glätten, in Bezug auf festes N die für die simulierten Histogramme generierten Datensätze bei jeder Erhöhung von M erhalten. Das heißt, die für niedrigeres M simulierten Datensätze werden jeweils weiterverwendet und lediglich die noch fehlende Anzahl wird neu generiert. Für eine höhere Glattheit der Kurven entlang der Achse bezüglich des Datensatzumfangs werden die historischen Datensätze aufeinander aufgebaut. Bei einer Erhöhung von N werden die

bereits generierten historischen Zeitschritte jeweils weiterverwendet – nur die ausstehende Zeitschrittanzahl wird neu dazu generiert.

Für festes N ist eine Sättigung der Fehlerkurven in den Abbildungen 9.5(a) und 9.5(b) zu beobachten. Die simulierten Histogramme erreichen einen Glattheitsgrad, der nicht weiter zu verbessern ist. Je größer N ist, desto schneller tritt diese Sättigung ein, denn um ausreichend glatte simulierte Histogramme für den Vergleich zu erstellen, werden (entsprechend der Größe $N \cdot M$) für einen höheren Datensatzumfang N weniger Datensätze benötigt.



(a) Für den Parameter b .

(b) Für den Parameter σ .

Abbildung 9.5.: Analyse der Simulationsanzahl M , relativer L_2 -Fehler (eindimensionales naives Histogramm-Matching) in Abhängigkeit von M und N .

Als „Faustregel“ für die anschließenden Berechnungen lässt sich die Simulationsanzahl M in Abhängigkeit von N über

$$M = \begin{cases} 25.600/N & N < 512 \\ 100 & N \geq 512 \end{cases} \quad (9.1)$$

wählen. Alternativ könnte M für beliebiges N konstant sehr hoch gewählt werden. Dies würde aber einen größeren Rechenaufwand bedeuten.

9.3.4. Die Intervallgrenzen

Um mögliche „Ausreißer“ (einzelne Datenwerte an den Rändern, die stark von den übrigen abweichen) von vornherein von den Betrachtungen auszuschließen und eine Verschlechterung der Güte des Histogramms beziehungsweise unnötigen Aufwand zur Garantierung dieser Güte zu verhindern, wird nur ein gewisser Anteil der Datenwerte, bestimmt durch die Wahl der Grenzen des Grundintervalls der Histogramme, einbezogen. Diese Grenzen werden in Abhängigkeit des (skalierten) historischen Datensatzes so gewählt, dass ein fester Prozentsatz der kleinsten und größten Datenwerte ausgeschlossen wird. Alternativ könnte die Auswahl des einbezogenen Prozentsatzes sukzessive vom höchsten Teilintervall aus erfolgen. Dabei würden parallel das Gebiet der sich links beziehungsweise rechts anschließenden Teilintervalle nach und nach einbezogen, bis der erwünschte Prozentsatz erreicht ist. Diese Hinzunahme von Teilintervallen endet auf der jeweiligen Seite, sobald sich dort kein Teilintervall mehr anschließt.

In Abbildung 9.6 sind die relativen L_2 -Fehler des eindimensionalen naiven Histogramm-Matching-Verfahrens für die Parameter b und σ des CIR-Modells in Abhängigkeit von N

und dem nicht in die Beobachtungen eingeschlossenen Prozentsatz der historischen Daten je Seite dargestellt. Für den Parameter b (Abbildung 9.6(a)) ist ein starker Abfall des Fehlers bis zur 2%-Quote festzustellen. Für den Parameter σ hingegen ist in Abbildung 9.6(b) kein solch starker Abfall zu erkennen. Dies liegt darin begründet, dass durch die Skalierungen die Datenwerte für den Parameter b vor allem an den Seiten stärker auseinander gezogen werden und ein „Ausreißer“ deswegen sehr weit von den übrigen Datenwerten abweicht. Für den Parameter b werden im Folgenden 90% des historischen Datensatzes in die

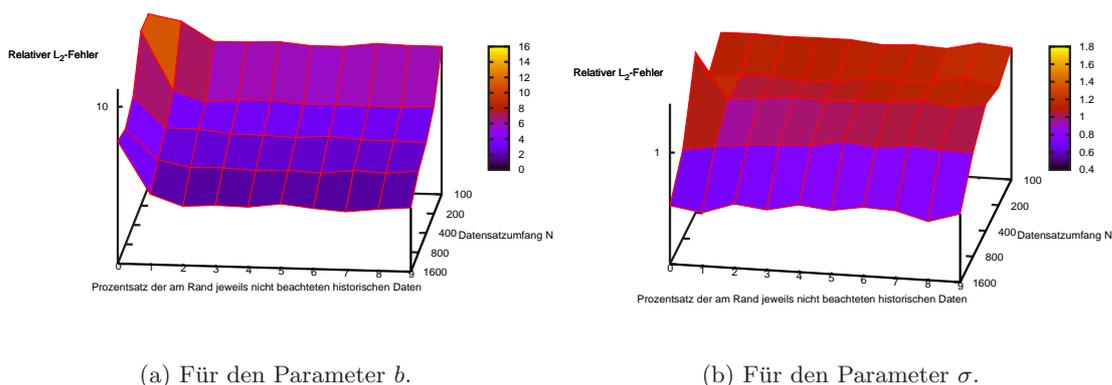


Abbildung 9.6.: Analyse der Wahl der Intervallgrenzen, relativer L_2 -Fehler (eindimensionales naives Histogramm-Matching) in Abhängigkeit vom Prozentsatz der am Rand jeweils nicht beachteten historischen Daten und N .

Betrachtung und für die Wahl der Intervallgrenzen einbezogen. Dies erscheint sinnvoll, da auch nach der 2%-Marke meist noch ein leichter Abfall des Fehlers bis zu 5% beobachtbar bleibt. Im Gegensatz zu den Auswirkungen der Skalierung bezüglich des Parameters b liegen die Datenwerte nach der Skalierung für σ dicht beieinander. „Ausreißer“ entfernen sich nur in Ausnahmefällen weiter. Die starke unregelmäßige Erhöhung des relativen L_2 -Fehlers in Abbildung 9.6(b) bei $N = 400$ und einer Beachtung aller Daten (Prozentsatz der jeweils nicht beachteten historischen Daten beträgt 0) ist auf ebensolche extremen „Ausreißer“ zurückzuführen. Um auch diesen vereinzelt „Ausreißern“ gerecht zu werden, bietet es sich an, in Bezug auf den Parameter σ ungefähr 1% der historischen Daten pro Seite von den Betrachtungen und der Wahl der Intervallgrenzen auszuschließen. Demnach werden für σ im Folgenden 98% der skalierten historischen Daten für den Aufbau des historischen Histogramms und die Wahl der Intervallgrenzen verwendet. Die Intervallgrenzen für die auf den skalierten Datenwerten basierenden Histogramme werden ab jetzt anhand des historischen Datensatzes entsprechend eines an den Seiten nicht beachteten Prozentsatzes von

$$p = \begin{cases} 5\% & \text{bezüglich des Parameters } b \\ 1\% & \text{bezüglich des Parameters } \sigma \end{cases} \quad (9.2)$$

festgesetzt.

9.3.5. Die Anzahl k der Teilintervalle

Nachdem die Intervallgrenzen mittels der historischen Daten festgelegt sind und somit die Breite b des Grundintervalls gegeben ist, hängt die Breite h der Teilintervalle lediglich

noch von der Anzahl k der Teilintervalle ab ($h = \frac{b}{k}$). Die Breite der Teilintervalle muss so „ausbalanciert“ werden, dass sich keine zu starke Verdichtung der Datenwerte innerhalb eines einzelnen Teilintervalls (h zu groß) ergibt. Außerdem ist auszuschließen, dass infolge einer zu feinen Intervalleinteilung nur noch einzelne oder gar keine Datenwerte mehr in den einzelnen Teilintervallen liegen. Dann wird von Overfitting bezüglich der einzelnen Datenwerte gesprochen.

In den Abbildungen 9.7 und 9.8 sind die relativen L_2 -Fehler des eindimensionalen Histogramm-Matching-Verfahrens für die Parameter b und σ des CIR-Modells in Abhängigkeit von der Anzahl k von Teilintervallen dargestellt. Dabei sind die Intervallgrenzen festgehalten, so dass die Anzahl der Teilintervalle Rückschlüsse auf die Teilintervallbreite zulässt.

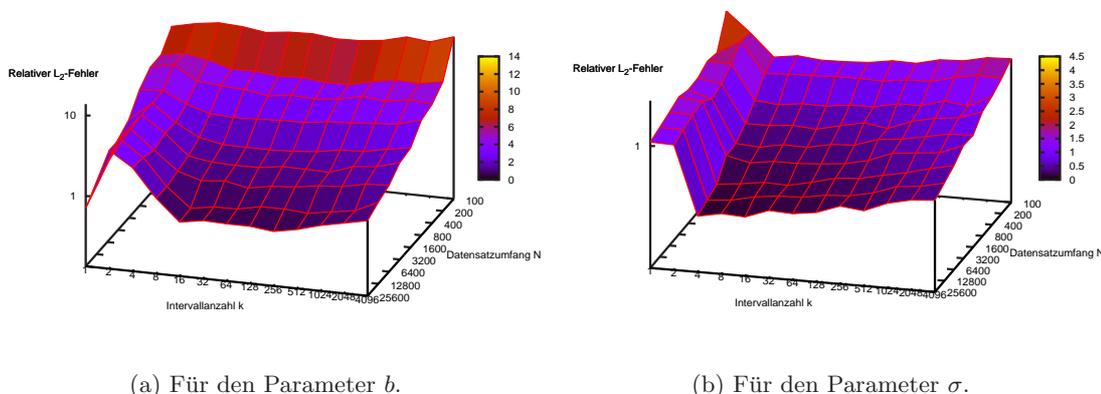
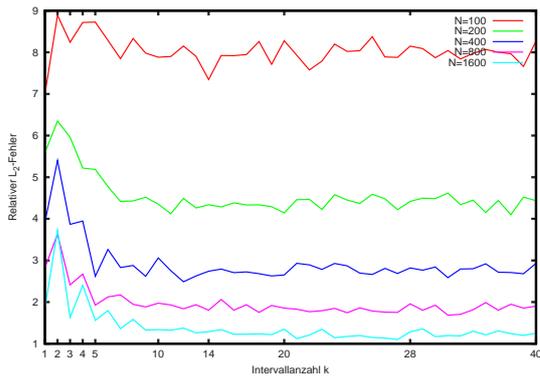


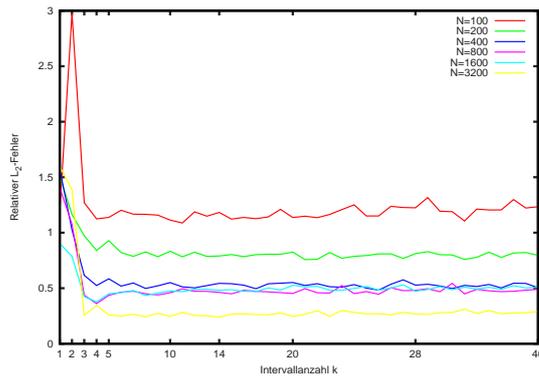
Abbildung 9.7.: Analyse der Anzahl k der Teilintervalle, relativer L_2 -Fehler (eindimensionales naives Histogramm-Matching) gegen die Anzahl der Teilintervalle k (legt die Teilintervallbreite fest) und N .

Für den Parameter b ist eine Abhängigkeit der Teilintervallbreite h und des Datensatzumfanges N in Bezug auf das Verhalten der Fehlerkurve zu erkennen (siehe Abbildung 9.7(a) und 9.8(a) für die Fehlerkurven auf Grundlage der charakteristischen Basisfunktionen sowie 9.8(c) für Fehlerkurven auf Grundlage der stückweise linearen Basisfunktionen). Für optimale Fehlerwerte gilt $k = \frac{h}{b} \approx \sqrt{N}$. Bei einem höheren Datensatzumfang N kann die Breite der Teilintervalle kleiner gewählt werden, denn dann existieren mehr Datenwerte, die dem Overfitting-Effekt entgegenwirken.

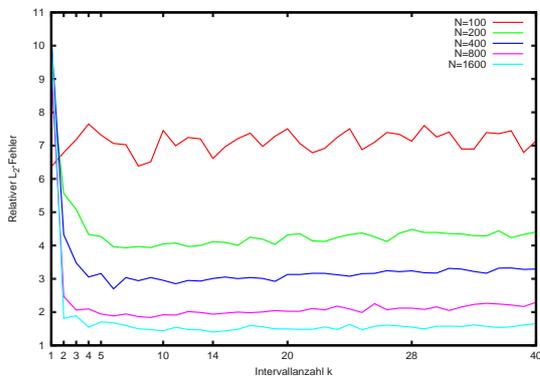
Für den Parameter σ ist ein solcher Zusammenhang nicht zu erkennen (siehe Abbildung 9.7(b) und 9.8(b) für die Fehlerkurven auf Grundlage der charakteristischen Basisfunktionen sowie 9.8(d) für Fehlerkurven auf Grundlage der stückweise linearen Basisfunktionen). Der Fehler ist sogar am geringsten für Histogramme mit sehr kleiner Anzahl von Teilintervallen. Das lässt sich dadurch erklären, dass der Parameter σ im Wesentlichen über die Varianz der approximierten Dichtefunktion entscheidet. Ist der Parameter σ größer als der wahre Wert gewählt, so fließt ein entsprechend kleinerer Prozentsatz von Daten in das Histogramm ein. Die Schätzung erfolgt – insbesondere für die Teilintervallanzahl Eins – lediglich über diesen Prozentsatz der sich im Grundintervall befindenden Daten und nicht über eine wirkliche Approximation der Dichtefunktion. Erst ab einer Teilintervallanzahl größer Eins erfolgt die Schätzung tatsächlich über eine Approximation der Dichtefunktion. Bei einer Anzahl von zwei Teilintervallen fällt die „Spitze“ der Dichteapproximation teilweise in das rechte und teilweise in das linke Teilintervall. Dadurch ist der Fehler des Histogramm-Matching-Verfahrens sehr groß. Erst bei drei Teilintervallen stabilisiert sich



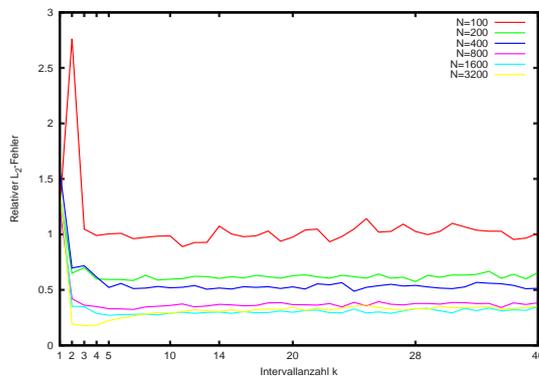
(a) Für den Parameter b , auf Grundlage charakteristischer Basisfunktionen.



(b) Für den Parameter σ , auf Grundlage charakteristischer Basisfunktionen.



(c) Für den Parameter b , auf Grundlage stückweise linearer Basisfunktionen.



(d) Für den Parameter σ , auf Grundlage stückweise linearer Basisfunktionen.

Abbildung 9.8.: Analyse der Anzahl k der Teilintervalle, relativer L_2 -Fehler (eindimensionales Histogramm-Matching, verschiedene Basisfunktionen) in Abhängigkeit der Anzahl der Teilintervalle k (legt die Teilintervallbreite fest) für verschiedene Datensatzumfänge N .

die Lage der „Spitze“. Sie liegt dann meist im mittleren Teilintervall. Die Schätzung mittels dreier Teilintervalle erfolgt sowohl über den eingeschlossenen Prozentsatz als auch über die Verteilung auf die drei Teilintervalle. Die Varianz scheint dadurch sehr gut zu „fassen“ zu sein. So ist σ über diese Aufteilung besser zu schätzen als über eine feinere Intervallaufteilung. Bei näherer Betrachtung der Fehlerkurven scheint es für den Parameter σ sinnvoll, die Anzahl an Teilintervallen konstant für alle N zu wählen ($k = 4$ für die naiven Histogramme und $k = 5$ für die Histogramme auf Grundlage stückweise linearer Funktionen). Basierend auf diesen Erkenntnissen wird im Folgenden die Anzahl k von Teilintervallen festgelegt auf

$$k = \begin{cases} \sqrt{N} & \text{bezüglich des Parameters } b \\ 4 & \text{bezüglich des Parameters } \sigma \text{ beim Einsatz charakteristischer Funktionen} \\ 5 & \text{bezüglich des Parameters } \sigma \text{ beim Einsatz stückweise linearer Funktionen.} \end{cases} \quad (9.3)$$

9.3.6. Die Basisfunktionen

Wie in Kapitel 7 erläutert, kann die wahre Dichtefunktion sowohl mit naiven Histogrammen auf Grundlage charakteristischen Basisfunktionen als auch mit Histogrammen basierend auf stückweise linearen Funktionen approximiert werden. Bei den bisherigen Analysen sind meist die charakteristischen Funktionen als Basisfunktionen zum Einsatz gekommen. In Abbildung 9.8 sind neben den Fehlerkurven der eindimensionalen Histogramm-Matching-Verfahren auf Grundlage der charakteristischen Basisfunktionen auch diejenigen basierend auf den stückweise linearen Funktionen abgebildet.

Abbildung 9.9 zeigt die relativen L_2 -Fehlerkurven bezüglich des eindimensionalen Histogramm-Matching-Verfahrens auf Grundlage der charakteristischen sowie der stückweise linearen Funktionen in Abhängigkeit vom Datensatzumfang N . Alle Fehlerkurven haben die selbe Ordnung. Dabei liegen die Fehlerkurven der stückweise linearen Basisfunktionen meist leicht unter denen der charakteristischen Basisfunktionen. Hier ist keine wesentliche Verbesserung des eindimensionalen Histogramm-Matching-Verfahrens durch den Einsatz der stückweise linearen Basisfunktionen festzustellen. Interessant bleibt die Frage, ob dies auch für den Einsatz beim iterativen Histogramm-Matching-Verfahren und dem parallelen (zweidimensionalen) Histogramm-Matching-Verfahren so bleibt (siehe Kapitel 9.3.7 Tabelle 9.1 und Kapitel 9.4 Abbildung 9.14). Wie sich herausstellen wird, zeigen die stückweise linearen Basisfunktionen bei den iterativen Verfahren eine größere Wirkung. Der Verbesserungseffekt „vervielfacht“ sich dabei dadurch, dass die jeweiligen Schätzwerte für die Schätzung des nächsten Parameterwertes „mitgenommen“ werden. Da anzunehmen ist, dass für eine Dichteapproximation mittels stückweise linearer Basisfunktionen die Verfahrensparameter nur leicht von denen der charakteristischen Funktionen abweichen, werden diese hier nicht gesondert untersucht.

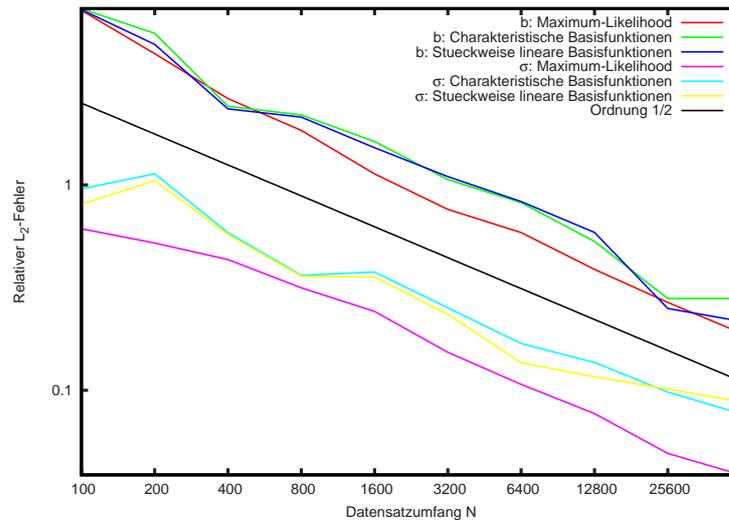


Abbildung 9.9.: Analyse der Basisfunktionen, relativer L_2 -Fehler (eindimensionales Histogramm-Matching, verschiedene Basisfunktionen) in Abhängigkeit vom Datensatzumfang N . (Mittelung über 100 Schätzungen, wahre Parameter $b = 1.0$, $\sigma = 0.15$, $m = 0.05$, Anfangswert $r_0 = 0.0499$ und konstanter Zeitzuwachs $\Delta t = 0.1$, Simulationsanzahl nach (9.1), Wahl der Intervallgrenzen nach (9.2) und Anzahl der Teilintervalle nach (9.3)).

9.3.7. Die Iterationsanzahl und die Startwertwahl des iterativen Histogramm-Matching-Verfahrens basierend auf eindimensionalen Histogrammen

In Tabelle 9.1 sind die L_2 -Fehler des in Kapitel 8.4.2 vorgestellten iterativen Histogramm-Matching-Verfahrens in Abhängigkeit von der Wahl der Basisfunktionen, der Wahl des Startwertes (b oder σ (vergleiche Kapitel 8.4.2), Tabelle 8.4), der Anzahl von Iterationen sowie dem Datensatzumfang N abgebildet.

Für eine steigende Anzahl von Iterationen ist keine systematische Besserung des Fehlers zu erkennen. Unter der Voraussetzung, dass die simulierten Histogramme ausreichend glatt sind, würde ein erneuter Vergleich des historischen Histogramms mit den simulierten Histogrammen keine wesentliche Veränderung bringen. Daher genügt beim iterativen Verfahren für eine ausreichend hohe Simulationsanzahl M (vergleiche die Untersuchungen in 9.3.3) eine einzige Iteration. Das heißt, für jeden zu schätzenden Parameter wird nur einmal ein eindimensionales Histogramm-Matching-Verfahren gestartet.

In Bezug auf die Wahl der Startwerte scheint, unabhängig von der Wahl der Basisfunktionen, für die Schätzung des Parameters σ der Startwert b bessere Ergebnisse zu liefern. Für die Schätzung des Parameters b hingegen bringt σ als Startwert die genaueren Schätzergebnisse hervor.

Bezüglich der Wahl der Basisfunktionen lässt sich feststellen, dass die stückweise linearen Basisfunktionen meist zu einem kleineren Fehler führen als die charakteristischen Basisfunktionen. Diese Beobachtung unterstützt die Resultate aus Kapitel 9.3.6.

RELATIVE L_2 -FEHLER DES ITERATIVEN HISTOGRAMM-MATCHING-VERFAHRENS (siehe Kapitel 8.4.2)							
	b ALS STARTWERT					σ ALS STARTWERT	
	Anzahl der Iterationen						
N	1	2	3	4	5	1	2
CHARAKTERISTISCHE BASISFUNKTIONEN							
PARAMETER b							
100	7.59523	7.92445	7.81425	7.91458	7.71059	7.65261	7.58081
200	3.85681	3.74374	4.04081	3.99805	3.90512	3.66998	3.80173
400	2.57391	2.66927	2.53414	2.6487	2.82013	2.60408	2.57694
800	1.9685	1.97642	1.95256	1.98037	2.05015	1.94454	1.94856
1600	1.31696	1.38067	1.34629	1.34048	1.39194	1.25	1.31696
PARAMETER σ							
100	0.651121	0.66233	0.681604	0.682622	0.650587	0.668383	0.657225
200	0.702031	0.693772	0.719085	0.702476	0.702031	0.694372	0.720002
400	0.736122	0.752727	0.746566	0.750555	0.749398	0.756913	0.75783
800	0.573004	0.66065	0.637541	0.657489	0.627384	0.653197	0.640041
1600	0.463756	0.542627	0.532486	0.535672	0.524272	0.562423	0.547405
STÜCKWEISE LINEARE BASISFUNKTIONEN							
PARAMETER b							
100	7.42883	6.981	7.07438	7.29405	7.3133	7.27259	7.10963
200	3.52225	3.52225	3.62931	3.30955	3.40266	3.61421	3.375
400	2.52488	2.49687	2.50624	2.54951	2.55869	2.45904	2.58904
800	2.01556	2	1.87916	1.97642	1.95256	1.92841	1.91213
1600	1.34629	1.36931	1.29301	1.29301	1.32877	1.19242	1.31101
PARAMETER σ							
100	0.640204	0.64318	0.626055	0.639336	0.625222	0.651121	0.66233
200	0.620987	0.654578	0.659861	0.656167	0.652931	0.702031	0.693772
400	0.784441	0.803422	0.80247	0.80273	0.817559	0.736122	0.752727
800	0.63765	0.690712	0.676747	0.677362	0.683587	0.573004	0.66065
1600	0.454071	0.528034	0.524735	0.527902	0.525264	0.463756	0.542627

Tabelle 9.1.: Analyse des iterativen Histogramm-Matching-Verfahrens, relative L_2 -Fehler (siehe Kapitel 8.4.2 Tabelle 8.4) bei unterschiedlicher Iterationsanzahl und variierenden Startwerten.

9.4. Verfahrensparameter des zweidimensionalen Histogramm-Matching-Verfahrens

Da beim parallelen Histogramm-Matching-Verfahren für das CIR-Modell entsprechend der Erläuterungen in Kapitel 8.4.3 zweidimensionale Histogramme verwendet werden, ist es sinnvoll, die Wahl der Verfahrensparameter zumindest teilweise erneut zu überprüfen.

- *Der Datensatzumfang N .* Für die Erstellung eines glatten zweidimensionalen Histogramms wird ein größerer Datensatzumfang N benötigt, als es bei den eindimensionalen Histogrammen der Fall war. Dies liegt daran, dass die Informationskraft der einzelnen Datenwerte durch eine feinere Aufteilung des Grundgebiets immer mehr verlorengelht, da die Datenwerte weiter auseinandergezogen werden (vergleiche die Erläuterungen aus Kapitel 9.3.5 zur Anzahl der Teilintervalle).
- *Die Simulationsanzahl M .* Da, wie gerade erläutert, die Hinzunahme einer Dimension für das Histogramm zu einem „Auseinanderziehen“ der vorhandenen Daten führt, wird für die Erstellung ausreichend glatter simulierter Histogramme bei den zweidimensionalen Histogrammen, im Vergleich zu den eindimensionalen Histogrammen, bei gleichem Datensatzumfang N eine größere Anzahl M an simulierten Datensätzen benötigt.
- *Die Intervallgrenzen.* Die Intervallgrenzen können weiter entsprechend der in Kapitel 9.3.4 (9.2) festgehaltenen Regel für jede Achse, das heißt bezüglich jedes Parameters, aus den skalierten Datenwerten bestimmt werden.
- *Die Anzahl k_q der Teilquader.* Sie wird über die Anzahl der Teilintervalle je Richtung bestimmt: $k_q = k_1 \cdot k_2$. Die Anzahlen k_1 und k_2 entscheiden nach der Festlegung der Intervallgrenzen über die Breite der Teilquader in die jeweilige Richtung $r \in \{1, 2\}$ ($h_r = \frac{b_r}{k_r}$). Wird die Anzahl k_1 von Teilintervallen bezüglich der ersten Richtung anhand der optimalen Werte für die eindimensionalen Histogramme festgelegt (vergleiche (9.3)), so werden die Datenwerte aufgrund der zweiten Dimension trotzdem weiter auseinandergezogen. Ein Overfitting-Effekt kann schneller entstehen. Die Anzahl der Teilintervalle pro Richtung muss daher für die zweidimensionalen Histogramme erneut untersucht werden.

Bei den folgenden Analysen liegen in der Regel die Werte der Tabelle am Anfang von Kapitel 9.3 zugrunde. Abweichungen von diesen Werten werden an den jeweiligen Stellen kenntlich gemacht.

9.4.1. Die Simulationsanzahl M

Ähnlich wie beim eindimensionalen Histogramm-Matching-Verfahren wird für eine Erhöhung der Simulationsanzahl M eine Senkung des relativen L_2 -Fehlers bezüglich der geschätzten Parameter erwartet. Für den Parameter b kann dies in Abbildung 9.10(a) beobachtet werden.

Allerdings ist für die Schätzung von σ kein solches Verhalten zu erkennen (siehe Abbildung 9.10(b)). Dies liegt daran, dass der Parameter σ auch nach der Skalierung nur schlecht aus den zweidimensionalen Histogrammen zu schätzen ist. Die Änderungen der simulierten Histogramme für variierendes σ sind im Verhältnis zu den im Bezug auf das historische

Histogramm ermittelten Abweichungen sehr gering. Die Anzahl N der historischen Daten muss für eine Schätzung von σ über das zweidimensionale Histogramm erheblich höher sein, um gute Ergebnisse zu erzielen. In diesem Zusammenhang sind in Abbildung 9.11 die

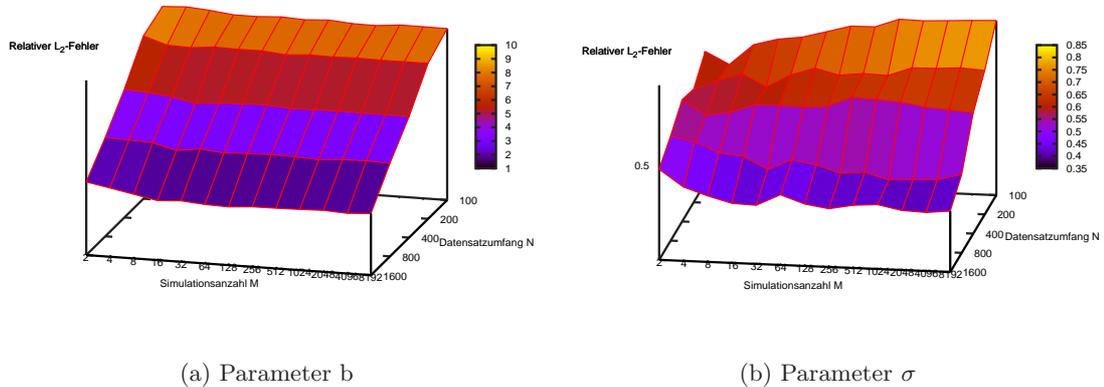


Abbildung 9.10.: Analyse der Simulationsanzahl M , relativer L_2 -Fehler (paralleles (zweidimensionales) naives Histogramm-Matching) in Abhängigkeit von M und N .

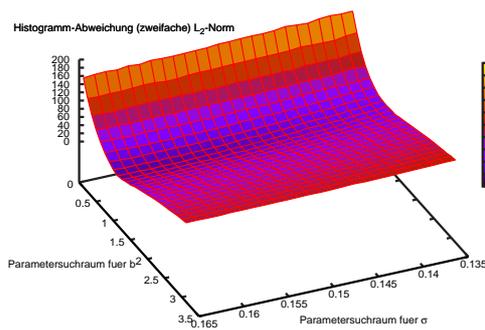
Histogramm-Abweichungskurven entsprechend (zweifacher) L_2 -Norm und ihre Isolinien für eine Parallelschätzung der Parameter b und σ des CIR-Modells mittels des Tensorprodukt-Ansatzes auf Grundlage charakteristischer Basisfunktionen für verschiedene Datensatzumfänge N dargestellt. Die wahren Parameter sind $b = 1.0$ und $\sigma = 0.15$. Für $N = 1.000$ ist die Histogramm-Abweichungskurve parallel zur Achse des Parametersuchraums von σ sehr flach. An den Isolinien ist erkennbar, dass der Parameter b im Gegensatz zu σ gut geschätzt wird. Ein wirkliches Tal der Kurve bezüglich σ (und eine gute Schätzung) ergibt sich erst für sehr großes N (etwa für $N = 10.000$ – siehe Abbildung 9.11(c) und 9.11(d)).

9.4.2. Die Anzahl k_q der Teilquader

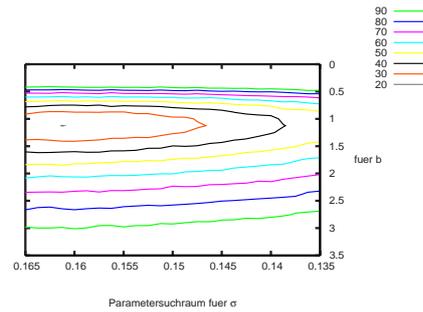
Durch die Erhöhung der Dimension der Histogramme ist eine erneute Untersuchung in Bezug auf die Teilintervallbreite sinnvoll, denn es handelt sich nun für die selben Teilintervallbreiten um wesentlich mehr Teilintervalle, auf die die Datensätze verteilt werden. So kann es schnell zum Overfitting-Effekt kommen.

In Abbildung 9.12 sind die relativen L_2 -Fehler des parallelen naiven Histogramm-Matching-Verfahrens für die Parameter b und σ in Abhängigkeit von der für beide Richtungen identischen Anzahl der Teilintervalle $k = k_1 = k_2$ und N dargestellt. Für den Parameter b ist entsprechend der eindimensionalen Beobachtungen (vergleiche Kapitel 9.3.5, Abbildung 9.7(a)) wieder eine Abhängigkeit von k und N in Bezug auf das Verhalten der Fehlerkurve zu erkennen. Für optimale Fehlerwerte ist die Intervallanzahl auf $k (= k_1 = k_2) \approx \frac{1}{2}\sqrt{N}$ zu setzen. Das entspricht einer Anzahl von $k_q = \frac{1}{4}\sqrt{N}$ Teilquadern.

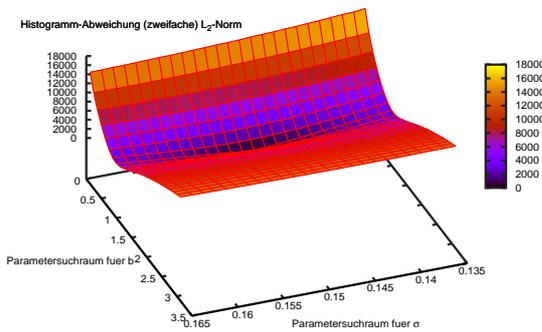
Für den Parameter σ ist hier (wie bereits im Eindimensionalen) keine solche Abhängigkeit zu beobachten. In Bezug auf die Abbildung 9.12(b) scheint es, als seien die optimalen Fehlerwerte unabhängig von N für $k = 8$ zu erreichen. Dass die Fehlerkurve für kleines N ein anderes Verhalten aufweist, kann auch darauf zurückzuführen sein, dass mit diesem Datensatzumfang nicht ausreichend Werte für eine gute Schätzung existieren. Die Fehler-



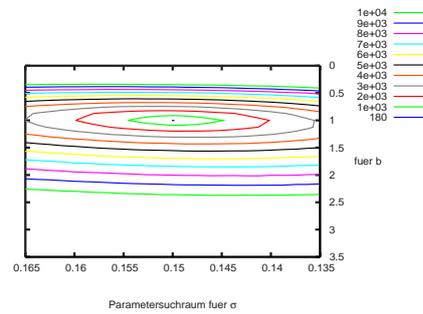
(a) Datenumfang $N = 1.000$.



(b) Datenumfang $N = 1.000$.



(c) Der Datenumfang ist mit $N = 10.000$ sehr groß gewählt.



(d) Der Datenumfang ist mit $N = 10.000$ sehr groß gewählt.

Abbildung 9.11.: Analyse des Datensatzumfangs N , Histogramm-Abweichungen entsprechend (zweifacher) L_2 -Norm für eine Parallelschätzung der Parameter b und σ des CIR-Modells mittels des Tensorprodukt-Ansatzes auf Grundlage charakteristischer Basisfunktionen. Die wahren Parameter sind $b = 1.0$ und $\sigma = 0.15$.

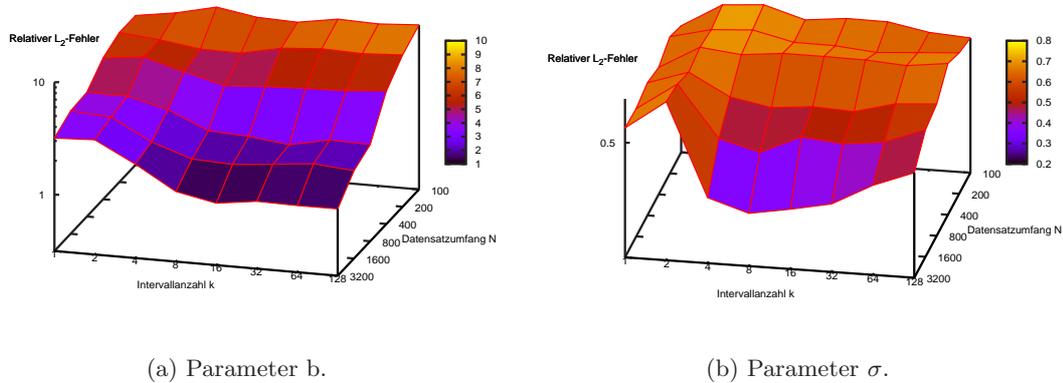


Abbildung 9.12.: Analyse der Anzahl k_q der Teilquader, relativer L_2 -Fehler (paralleles (zweidimensionales) naives Histogramm-Matching) gegen die Anzahl der Teilintervalle k pro Richtung (konstant für beide Richtungen $k = k_1 = k_2$) (legt die Teilintervallbreite für eine Richtung fest) und N .

kurve fällt dem entsprechend willkürlich aus.

Als Anzahl der Teilintervalle pro Richtung wird im Folgenden

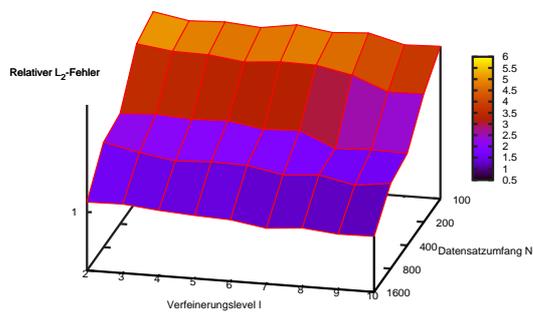
$$k(= k_1 = k_2) = \begin{cases} \frac{1}{2}\sqrt{N} & \text{bezüglich des Parameters } b \\ 8 & \text{bezüglich des Parameters } \sigma \end{cases} \quad (9.4)$$

gesetzt.

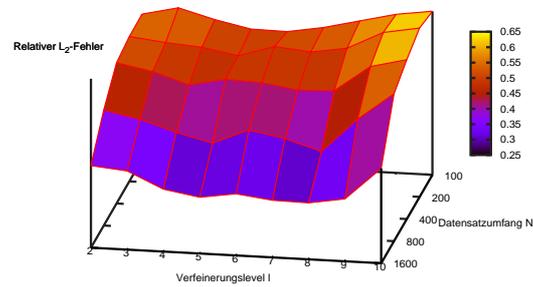
9.4.3. Die Levelanzahl beim Einsatz des Dünngitter-Ansatzes entsprechend der Kombinationstechnik

Beim parallelen Histogramm-Matching-Verfahren auf Grundlage der Kombinationstechnik des Dünngitter-Ansatzes ist durch eine Verfeinerung des Intervallgitters (eine Erhöhung der Anzahl von Teilintervallen) kein so schneller Eintritt des Overfitting-Effekts zu erwarten. Die Anwendung der Kombinationstechnik, bei der die Dichte-Approximationen verschieden feiner Intervallgitter zusammengesetzt werden, wirkt diesem Effekt entgegen. Der Verlust der Aussagekraft der Daten, der bei den „normalen“ zweidimensionalen Dichte-Approximationen proportional mit der Teilintervallanzahl zusammenhängt, wird hier durch die Kombinationen aufgefangen. Die durch die Datenpunkte gegebenen Informationen werden anhand verschiedener Intervallgitter ausgewertet und die Resultate wieder miteinander verknüpft. Dadurch gehen weniger Informationen verloren.

In Abbildung 9.13 finden sich die relativen L_2 -Fehler des zweidimensionalen Dünngitter-Histogramm-Matching-Verfahrens auf Grundlage der Kombinationstechnik in Abhängigkeit von der Levelanzahl l und dem Datensatzumfang N . In Bezug auf den Parameter σ tritt der Overfitting-Effekt für $N = 1.600$ nicht ab ungefähr $k = 32$ wie beim „normalen“ parallelen Verfahren auf, sondern erst für ein Level $l \geq 8$ auf. Für die einzelnen Richtungen bedeutet das eine Anzahl von $k \geq 256$ Teilintervallen. Das bedeutet für das feinste Kombinationsgitter eine Teilquaderanzahl $k_q \geq 512$ (zum Vergleich: beim entsprechenden vollen Gitter wären es $k_q = 262.144$ Teilquader bedeuten).



(a) Parameter b .



(b) Parameter σ .

Abbildung 9.13.: Analyse der Levelanzahl, relativer L_2 -Fehler des zweidimensionalen Dünngitter-Histogramm-Matching-Verfahrens auf Grundlage der Kombinationstechnik in Abhängigkeit von der Levelanzahl l und dem Datensatzumfang N .

Für den Parameter b wird der Overfitting-Effekt bis zu dem in Abbildung 9.13(a) dargestellten Level $l = 10$ gar nicht beziehungsweise kaum erfasst. Er tritt erst mit dem Level $l = 11$ auf.

9.5. Vergleich der Verfahren

In diesem Kapitel sollen die Güte und der Aufwand der Histogramm-Matching-Verfahren untereinander sowie mit dem Maximum-Likelihood-Verfahren verglichen werden. Dafür werden die Parameter des in Kapitel 4.3.2 vorgestellten CIR-Modells geschätzt.

Die Abbildungen 9.14(a) und 9.14(b) zeigen die relativen L_2 -Fehler folgender Verfahren:

- *Maximum-Likelihood* [MAX-LIK]: Maximum-Likelihood-Verfahren ([Lo86], [Sø97])
- *Iterativ, Char., Startwert b* [IT(CH)]: iteratives Histogramm-Matching-Verfahren mit dem Startwert b basierend auf eindimensionalen Dichte-Approximationen mit charakteristischen Basisfunktionen
- *Iterativ, Stckw., Startwert b* [IT(ST)]: iteratives Histogramm-Matching-Verfahren mit dem Startwert b basierend auf eindimensionalen Dichte-Approximationen mit stückweise linearen Basisfunktionen
- *Iterativ, Char., Startwert s* [IT(CH)]: iteratives Histogramm-Matching-Verfahren mit dem Startwert σ basierend auf eindimensionalen Dichte-Approximationen mit charakteristischen Basisfunktionen
- *Iterativ, Stckw., Startwert s* [IT(ST)]: iteratives Histogramm-Matching-Verfahren mit dem Startwert σ basierend auf eindimensionalen Dichte-Approximationen mit stückweise linearen Basisfunktionen
- *Parallel, Char.* [PAR(CH)]: paralleles Histogramm-Matching-Verfahren basierend auf zweidimensionalen Dichte-Approximationen mittels des Tensorprodukt-Ansatzes mit charakteristischen Basisfunktionen
- *Parallel, Stckw.* [PAR(ST)]: paralleles Histogramm-Matching-Verfahren basierend auf zweidimensionalen Dichte-Approximationen mittels des Tensorprodukt-Ansatzes mit stückweise linearen Basisfunktionen
- *Dünngitter-Ansatz, Level 8* [DG(8)]: paralleles Histogramm-Matching-Verfahren basierend auf zweidimensionalen Dichte-Approximationen mittels des Dünngitter-Ansatzes (Kombinationstechnik) mit Level 8
- *Dünngitter-Ansatz, Level 10* [DG(10)]: paralleles Histogramm-Matching-Verfahren basierend auf zweidimensionalen Dichte-Approximationen mittels des Dünngitter-Ansatzes (Kombinationstechnik) mit Level 10

Die Verfahrensparameter der Histogramm-Matching-Verfahren sind entsprechend der Analysen in den Kapiteln 9.3 und 9.4 gewählt.

Die Verfahren sind alle von der Ordnung $\frac{1}{2}$. In Bezug auf die Schätzung des Parameters b (Abbildung 9.14(a)) ergibt sich für Datensätze, deren Umfang N kleiner 400 ist, der kleinste relative L_2 -Fehler $\left(\frac{\|b-\hat{b}\|_2}{|b|}\right)$ beim parallelen Histogramm-Matching-Verfahren, das auf dem Dünngitter-Ansatz (Kombinationstechnik) des Levels 10 basiert. Für Datensätze, die zwischen 400 und 10.000 Datenwerten umfassen, liefert das Maximum-Likelihood-Verfahren im Mittel die besten Schätzergebnisse. Für Datensatzumfänge größer 10.000 deutet es sich an, dass die parallelen Histogramm-Matching-Verfahren auf Grundlage des Dünngitter-Ansatzes im Vergleich zum Maximum-Likelihood-Verfahren bessere Ergebnisse liefern.

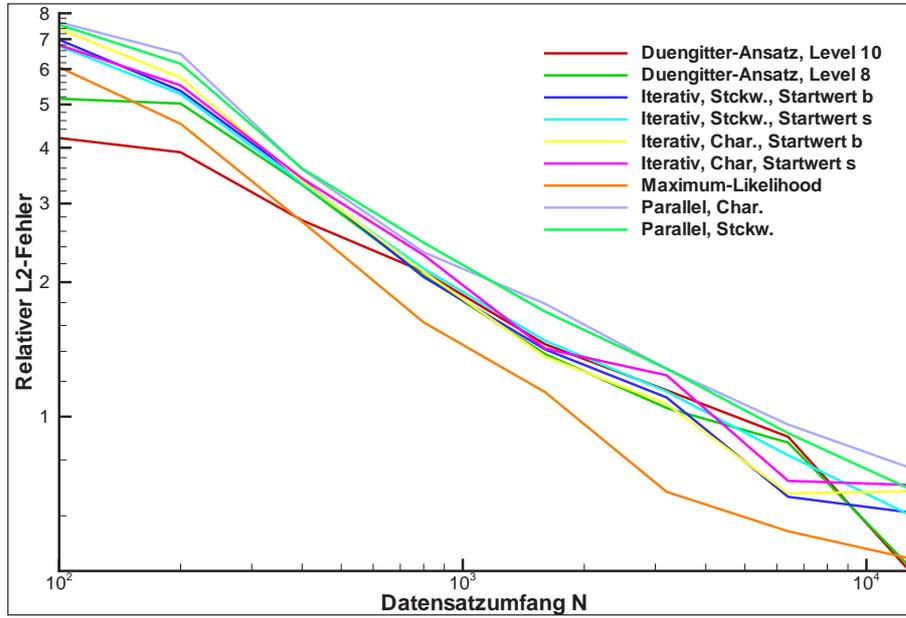
Bei einem Vergleich der Histogramm-Matching-Verfahren untereinander bringen die iterativen sowie die auf der Kombinationstechnik basierenden parallelen Dünigitter-Histogramm-Matching-Verfahren gegenüber den parallelen Verfahren auf Grundlage des klassischen Tensorprodukt-Ansatzes im Mittel die besseren Ergebnisse hervor.

Dass die iterative Schätzung gut funktioniert, lässt sich anhand der in den Abbildungen 9.11(b), 9.11(d), 9.16(a) und 9.16(b) dargestellten Isolinien der Histogramm-Abweichungskurven der parallelen Histogramm-Matching-Verfahren auf Grundlage des klassischen Tensorprodukt-Ansatzes nachvollziehen. Die Achsen der Isolinien sind parallel zu den Achsen der Parametersuchräume. Zumindest einer der beiden Parameter lässt sich demnach gut unabhängig vom vorgegebenen Wert des anderen schätzen. Für die in den Abbildungen 9.11(b) und 9.11(d) betrachteten Isolinien handelt es sich dabei um den Parameter b . Für die bei der historischen Schätzung entstandenen Isolinien in den Abbildungen 9.16(a) und 9.16(b) ist es hingegen der Parameter σ . Abbildung 9.17(b) zeigt eine Fehlerkurve für festes b . Die Schätzung der beiden Parameter kann also sehr gut separiert erfolgen. Daher liefern die iterativen Verfahren gute Schätzergebnisse.

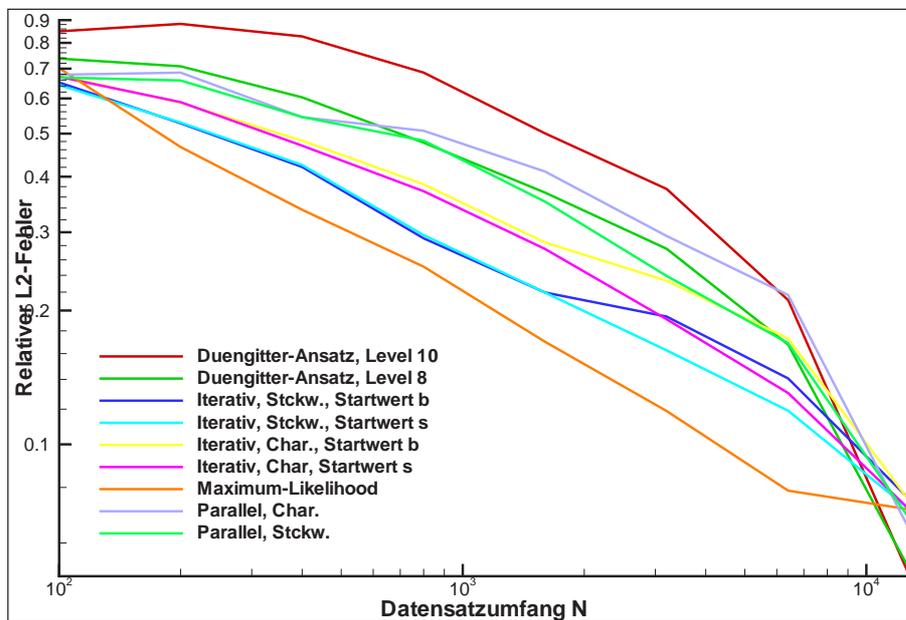
Bei den parallelen Verfahren liegt die Schwierigkeit im Auseinanderziehen der Daten durch die zweite Achse. Die Informationen werden mehr gestreut und verlieren an Aussagekraft – die Schätzergebnisse werden ungenauer. Um dieselbe Genauigkeit wie bei den iterativen Verfahren zu erhalten, sind mehr historische Datenwerte erforderlich. Ausgeglichen werden kann dieser Informationsverlust der parallelen Verfahren mittels der Kombinationstechnik. Hierbei kommt es zu einer Auswertung der Daten bezüglich verschiedener Intervallaufteilungen. Diese verschiedenen Informationsauswertungen werden zur letztendlichen Dichte-Approximation zusammengesetzt. Die auf dieser Kombinationstechnik beruhenden Verfahren schneiden für kleine Datensatzumfänge in Bezug auf den Parameter b am besten ab.

Bezüglich des Diffusionsparameters σ ergeben sich für kleine Datensatzumfänge $N < 150$ die kleinsten relativen L_2 -Fehler $\left(\frac{\|\sigma - \hat{\sigma}\|_2}{|\sigma|}\right)$ bei den iterativen Histogramm-Matching-Verfahren auf Grundlage der stückweise linearen Basisfunktionen. Bei Datensätzen, die mehr Datenwerte umfassen, dominiert lange das Maximum-Likelihood-Verfahren. Ab $N = 10.000$ scheint es, dass die Histogramm-Matching-Verfahren im Mittel wieder die besseren Schätzergebnisse liefern.

Wie auch in Bezug auf die Schätzung des Parameters b verfügen die iterativen Histogramm-Matching-Verfahren verglichen mit den parallelen Verfahren auf Basis des klassischen Tensorprodukt-Ansatzes über eine höhere mittlere Genauigkeit. Die Histogramm-Matching-Verfahren auf Basis der Kombinationstechnik schneiden verhältnismäßig schlecht ab. Das liegt daran, dass bei der Kombinationstechnik die Intervallanzahl in beide Richtungen übereinstimmt und daher ein Kompromiss hinsichtlich der beiden Parameter b und σ gefunden werden muss (vergleiche Kapitel 9.4.3). Die Fehlerkurve bezüglich des Levels 8 liegt verhältnismäßig weiter unter der für das Level 10. Noch bessere Ergebnisse sind entsprechend der Analysen in Kapitel 9.4.3 hinsichtlich des Parameters σ für eine weiter sinkende Levelanzahl bis zu Level 4 zu erwarten. Im Gegensatz dazu können die Fehlerkurven der beiden Level 8 und 10 für den Parameter b betrachtet werden. Hier verhält es sich genau umgekehrt. Die Fehlerkurve des Levels 10 liegt unterhalb der Fehlerkurve des Levels 8. Dies entspricht auch den in Kapitel 9.4.3 gemachten Beobachtungen. An dieser Stelle wäre es somit sinnvoll die Level der beiden Richtungen unabhängig voneinander zu wählen. Tabelle 9.2 gibt die Zeit für eine Schätzung eines Datensatzes vom Umfang $N = 100$ mittels des entsprechenden Verfahrens an. Außerdem sind für die Histogramm-Matching-Verfahren die jeweilige Gesamtanzahl der Teilintervalle beziehungsweise der Gewichtskoeffizienten sowie die Anzahl der Auswertungen pro Datenpunkt angegeben. Die Abkürzungen sind



(a) Parameter b .



(b) Parameter σ .

Abbildung 9.14.: Vergleich der Verfahren hinsichtlich der relativen L_2 -Fehler.

in der Einleitung zu diesem Kapitel aufgelistet. Das Maximum-Likelihood-Verfahren hat mit $1.9e^{-02}$ Sekunden die kürzeste Rechenzeit. Darauf folgen die iterativen Verfahren mit $3.5e^{-01}$ Sekunden. Die parallelen Histogramm-Matching-Verfahren auf Grundlage des klassischen Tensorprodukt-Ansatzes benötigen fast das sechsfache an Zeit. Mit Abstand am längsten brauchen die Histogramm-Matching-Verfahren auf Basis der Kombinationstechnik, bei denen eine Auswertung der Daten für verschiedene Gitter erfolgen muss. Bei Level 8 handelt es sich um 13, bei Level 10 um 17 einzelne Gitterauswertungen pro Datenpunkt im Gegensatz zu der einfachen Auswertung der Datenpunkte bezüglich der anderen Verfahren. Zusätzlich zu diesem erhöhten Aufwand muss beachtet werden, dass bei der Kombinationstechnik weitaus höhere Level für die selbe Genauigkeit nötig sind und demzufolge auch eine sehr viel größere Anzahl an Vergleichsoperationen für die Ermittlung der Abweichung der Dichte-Approximationen ausgewertet werden muss. Für das Level 8 handelt es sich um 10.265 Vergleiche von Gewichtskoeffizienten im Gegensatz zu lediglich 40 Vergleichen des optimalen parallelen Verfahrens auf Grundlage des klassischen Tensorprodukt-Ansatzes.

	MAX-LIK	IT(CH)	IT(ST)	PAR(CH)	PAR(ST)	DG(8)	DG(10)
Rechenzeit	$1.9e^{-02}$	$3.5e^{-01}$	$3.5e^{-01}$	2.09	2.56	$3.35e^{+01}$	$1.3749e^{+02}$
Teilintervallanzahl / Gitterpunkte	–	$\sqrt{100} + 4 = 14$	$\sqrt{100} + 5 = 15$	$\frac{1}{2}\sqrt{100} \cdot 8 = 40$	$\frac{1}{2}\sqrt{100} \cdot 8 = 40$	10265	23579
Auswertungsanzahl je Datenpunkt	1	1	1	1	1	13	17

Tabelle 9.2.: Vergleich der Rechenzeit in Sekunden für eine durchgeführte Schätzung der Parameter des CIR-Modells mittels der verschiedenen Verfahren auf Grundlage optimierter Verfahrensparameter. Die Erklärung der Kürzel der Verfahren ist in der Einleitung des Kapitels gegeben.

Level	2	3	4	5	6	7	8	9	10
Rechenzeit	3.27	4.63	6.34	8.5	$1.232e^{+01}$	$1.942e^{+01}$	$3.35e^{+01}$	$6.63e^{+01}$	$1.3749e^{+02}$
Gitterpunkte	7	29	95	273	732	1813	4375	10265	23579
Vergleich Gitter des TPA	$3^2 = 9$	$7^2 = 49$	$15^2 = 225$	$31^2 = 961$	$63^2 = 3969$	$127^2 = 16129$	$255^2 = 65025$	$511^2 = 261121$	$1023^2 = 1046529$

Tabelle 9.3.: Vergleich der Rechenzeit in Sekunden für eine durchgeführte Schätzung der Parameter des CIR-Modells mit Dünngitter-Histogramm-Matching-Verfahren (Kombinationstechnik) auf Grundlage verschiedener Level (TPA = Tensorprodukt-Ansatz).

Im Vergleich liefern die Histogramm-Matching-Verfahren auf Basis der Kombinationstechnik häufig die besten Ergebnisse, dafür ist ihr Speicher- und Rechenaufwand bei optimaler Verfahrensparameterwahl allerdings auch am höchsten. Hier könnte mit einer Kombinationstechnik angesetzt werden, die verschiedene Levelzahlen für die einzelnen Richtungen zulässt; denn wie in Kapitel 9.4.3 festgestellt wurde, ist die optimale Levelanzahl von den Eigenschaften der für die einzelnen Parameter skalierten Zuwächse abhängig.

9.6. Schätzung eines korrelierten Finanzmodells anhand realer Daten

Zum Abschluss soll ein korreliertes Finanzmodell bestehend aus Aktienkurs und CIR-Modell basierend auf einer historischen Stichprobe geschätzt werden. Dafür wird der Deutschen Aktienindex (DAX-Index) und der Tagesgeldsatz (als Approximation der Short Rate) betrachtet. Die realen Daten entstammen der Zeitreihendatenbank der Deutschen Bundesbank ([Zei]). Es handelt sich um die Monatsendstände des DAX-Index und die monatliche Durchschnittswerte des Tagesgeldsatzes am Frankfurter Bankplatz. Der Zeitraum ist entsprechend der Beobachtungen in Kapitel 8.1 von Januar 1979 bis April 2006 gewählt. Verglichen werden die Schätzergebnisse der folgenden Verfahren:

- das Maximum-Likelihood-Verfahren ([Lo86], [Sø97])
- ein iteratives Histogramm-Matching-Verfahren mit dem Startwert b basierend auf eindimensionalen Dichte-Approximationen mit stückweise linearen Basisfunktionen
- ein iteratives Histogramm-Matching-Verfahren mit dem Startwert σ basierend auf eindimensionalen Dichte-Approximationen mit stückweise linearen Basisfunktionen
- ein paralleles Histogramm-Matching-Verfahren basierend auf zweidimensionalen Dichte-Approximationen mittels des Tensorprodukt-Ansatzes mit stückweise linearen Basisfunktionen
- ein paralleles Histogramm-Matching-Verfahren basierend auf zweidimensionalen Dichte-Approximationen mittels des Dünngitter-Ansatzes (Kombinationstechnik) mit Level 8
- ein paralleles Histogramm-Matching-Verfahren basierend auf zweidimensionalen Dichte-Approximationen mittels des Dünngitter-Ansatzes (Kombinationstechnik) mit Level 10

Die Verfahrensparameter der Histogramm-Matching-Verfahren sind den Analysen der Kapitel 9.3 und 9.4 entsprechend gewählt. Die Schätzung erfolgt in mehreren Schritten (vergleiche Kapitel 8.7):

1. Schätzung des Mean-Reversion-Niveaus (Algorithmus in Tabelle 8.2) (entfällt beim Maximum-Likelihood-Verfahren)
2. Schätzung der übrigen Parameter des CIR-Modells (Histogramm-Matching-Verfahren (Kapitel 8.4) oder Maximum-Likelihood-Verfahren (Kapitel 8.5))
3. Schätzung der Parameter des Aktienkursmodells (Algorithmus in Tabelle 8.1)
4. Schätzung des Korrelationsparameters (mittels der diskretisierten stochastischen Differentialgleichungen, den geschätzten Parameterwerten und den historischen Datenwerten) (Algorithmus in Tabelle 8.6)

In Tabelle 9.4 sind die Schätzergebnisse der Verfahren abgebildet. Entgegen unseren Erwartungen fallen die geschätzten Parameterwerte sehr klein aus. Dies trifft sowohl auf die Ergebnisse des Maximum-Likelihood-Verfahrens als auch der Histogramm-Matching-Verfahren zu. Aufgrund verschiedener Angaben in der Literatur (siehe etwa [FMW03])

SCHÄTZUNG DES MEAN-REVERSION-NIVEAUS						
Parameter	MAX-LIK	MW-HIST				
m	0.0507219	0.0529729				
SCHÄTZUNG DES CIR-MODELLS						
Parameter	MAX-LIK	IT(ST, $b = 0$)	IT(ST, $\sigma = 0.0380313$)	PAR(ST)	DG(8)	DG(10)
b	0.0547233	0.005	0.001	0.001	0.003	0.002
σ_1	0.0407374	0.031	0.03	0.019	0.026	0.022
SCHÄTZUNG DES AKTIENKURS-MODELLS						
Parameter	HIST					
μ	0.109162					
σ_2	0.209315					
SCHÄTZUNG DES KORRELATIONSPARAMETERS						
Parameter	MAX-LIK	IT(ST, b)	IT(ST, σ)	PAR(ST)	DG(8)	DG(10)
ρ	-0.039207	-0.0482993	-0.0496341	-0.0783696	-0.0574288	-0.0677766

Tabelle 9.4.: Schätzungen des korrelierten Finanzmodells bestehend aus dem DAX-Index (Monatsendaende) und dem Tagesgeldsatz von Januar 1979 bis April 2006.

sowie durchgeführter Untersuchungen bezüglich der Auswirkungen der Parameter des CIR-Modells wurde bei den Analysen der Verfahrensparameter von den wahren Parameterwerten $b = 1.0$ und $\sigma = 0.15$ ausgegangen. Solche Werte hätten wir auch bei einer Schätzung anhand realer Daten erwartet. Dies ist jedoch nicht der Fall. Zur Überprüfung der Funktionalität des Histogramm-Matching-Verfahrens ist in Abbildung 9.15 ein naives historisches Histogramm für den Parameter b auf Grundlage des historischen Tagesgeldsatzes abgebildet. Entsprechend der Skalierung der Datenwerte (vergleiche Kapitel 8.4.1) sollte

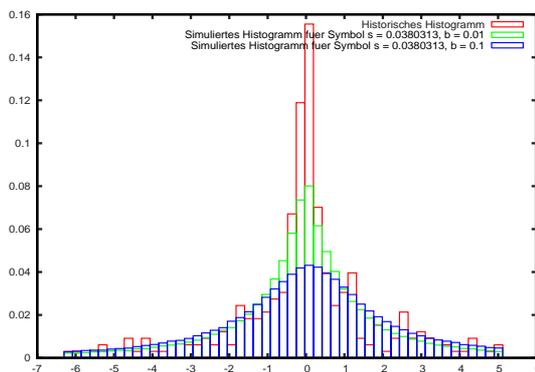
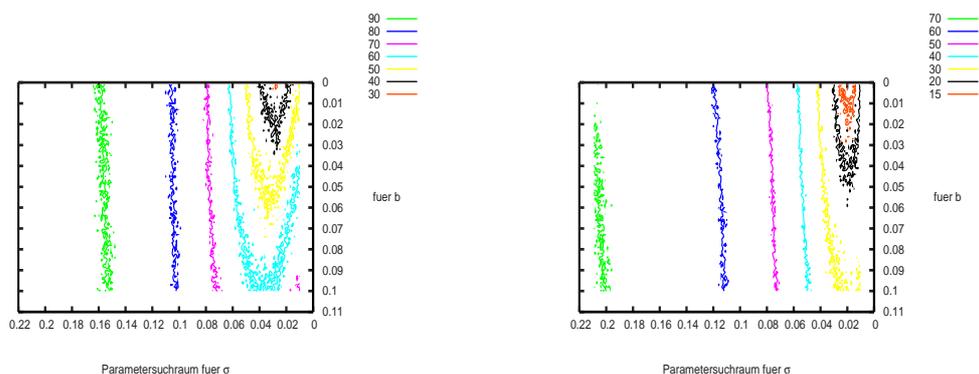


Abbildung 9.15.: Eindimensionale Histogramme zur Parameterschätzung anhand der realen Tagesgeldsätze von Januar 1979 bis April 2006.

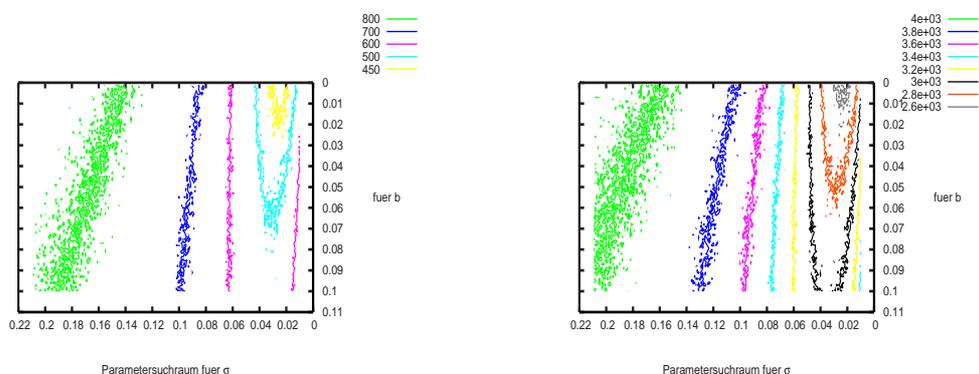
der Erwartungswert der durch das Histogramm approximierten Dichtefunktion mit dem wahren Parameterwert b übereinstimmen. Im abgebildeten Histogramm hat das Intervall $[0; 0.25)$ den höchsten Wert. Das passt zu den durch die Histogramm-Matching-Verfahren ermittelten Werten. Weiterhin können die Histogramm-Abweichungs-Kurven auf besonde-

re Auffälligkeiten untersucht werden. Abbildung 9.16 zeigt die Isolinien der Histogramm-Abweichungs-Kurven der zweidimensionalen Histogramm-Matching-Verfahren. In Abbildung 9.17 sind die Histogramm-Abweichungs-Kurven des iterativen Verfahrens mit Startwert b zu sehen. Grundsätzlich scheint sich die geringste Histogramm-Abweichung bezüglich des Parameters b allerdings in der Nähe des Wertes Null zu befinden.



(a) Histogramm-Matching-Verfahren basierend auf charakteristischen Basisfunktionen.

(b) Histogramm-Matching-Verfahren basierend auf stückweise lineare Basisfunktionen.



(c) Dünngitter-Histogramm-Matching-Verfahren Level 9.

(d) Dünngitter-Histogramm-Matching-Verfahren Level 11.

Abbildung 9.16.: Isolinien der zweidimensionalen (Dünngitter-)Histogramm-Abweichungs-Kurve (L_2 -Norm) der parallelen Histogramm-Matching-Verfahren auf Grundlage der realen Tagesgeldsätze von Januar 1979 bis April 2006.

Es stellt sich die Frage, ob das CIR-Modell mit solch kleinen Parameterwerten sinnvoll ist und ob es überhaupt für die angemessene Modellierung der aktuellen Short Rate geeignet ist. Die in Abbildung 9.18 dargestellten Zeitreihen basieren bis zum aktuellen Zeitpunkt auf historischen und im Anschluss auf den mit den geschätzten Parameterwerten simulierten Datenwerten. Bei der Short Rate fallen die Schwankungen des Prozesses infolge des kleinen Diffusionsparameters sehr gering aus. So entfernt sich die Short Rate trotz der äußerst kleinen Mean-Reversion-Stärke b kaum vom Mean-Reversion-Niveau $\hat{m} = 0.0507219$.

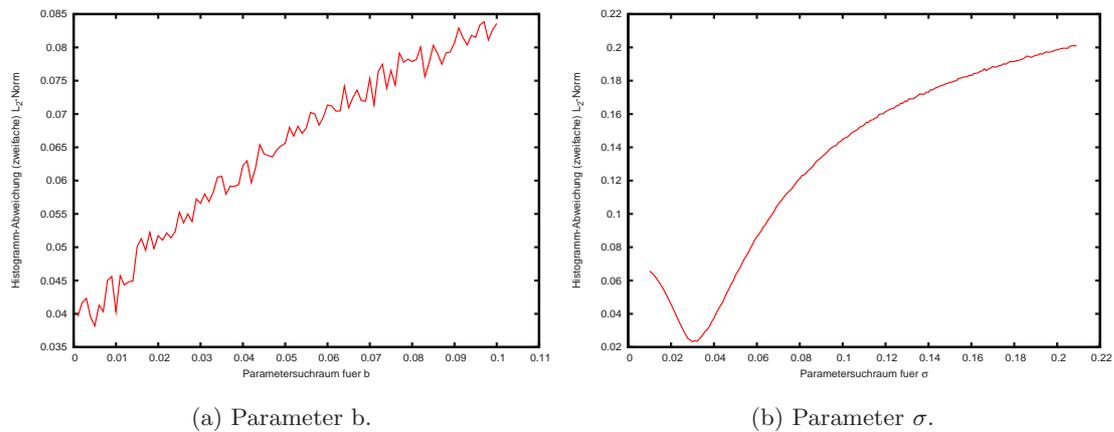


Abbildung 9.17.: Eindimensionale Histogramm-Abweichungs-Kurve der iterativen Histogramm-Matching-Verfahren (L_2 -Norm) auf Grundlage der realen Tagesgeldsätze von Januar 1979 bis April 2006.

Die eigentliche Mean Reversion ist dabei unbedeutend. Daher scheint eine Modellierung mittels des CIR-Modells ungeeignet.

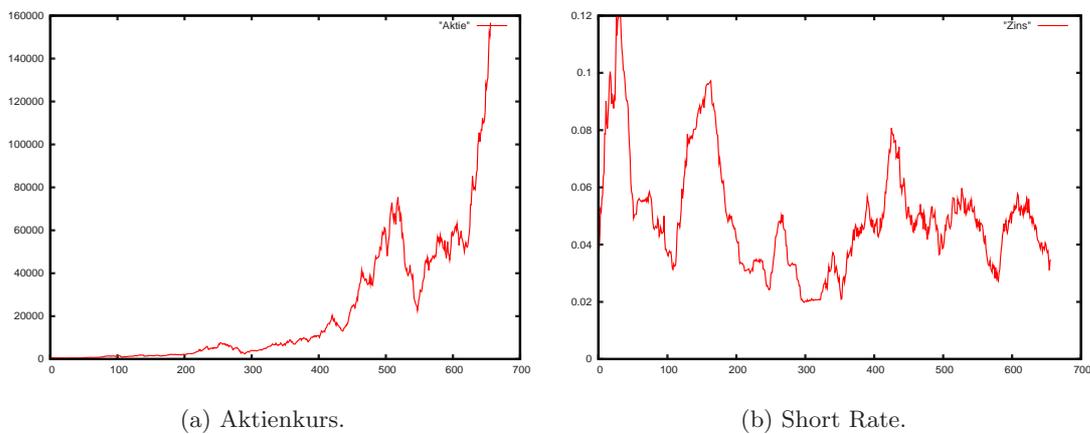


Abbildung 9.18.: Zeitreihen basieren bis zum 328. Wert auf historischen und im Anschluss auf den mit den (mittels des Maximum-Likelihood-Verfahrens) geschätzten Parameterwerten simulierten Datenwerten.

Kapitel 10.

Ausblick

In dieser Arbeit wurden Histogramm-Matching-Verfahren zur Parameterschätzung vorgestellt, analysiert und mit dem Maximum-Likelihood-Verfahren verglichen. Zur Schätzung mehrerer Parameter wurden ein iteratives Histogramm-Matching-Verfahren basierend auf eindimensionalen Dichte-Approximationen, ein paralleles Histogramm-Matching-Verfahren basierend auf zweidimensionalen Dichte-Approximationen mittels des Tensorprodukt-Ansatzes sowie ein paralleles Histogramm-Matching-Verfahren basierend auf zweidimensionalen Dichte-Approximationen mittels des Dünngitter-Ansatzes entwickelt. Dabei kommt beim Dünngitter-Ansatz die Kombinationstechnik zum Einsatz.

Bezogen auf das CIR-Modell hat sich herausgestellt, dass eine Schätzung der Parameter sehr gut separiert erfolgen kann. Daher können mittels der iterativen Histogramm-Matching-Verfahren gute Schätzergebnisse erhalten werden.

Bei den parallelen Verfahren liegt die Schwierigkeit im Auseinanderziehen der Daten durch die zweite Achse. Die Informationen werden mehr gestreut und verlieren an Aussagekraft – die Schätzergebnisse werden ungenauer. Es kann schneller zum so genannten Overfitting-Effekt kommen. Im Vergleich mit den iterativen Histogramm-Matching-Verfahren schneiden die parallelen Histogramm-Matching-Verfahren auf Grundlage des Tensorprodukt-Ansatzes daher hinsichtlich ihrer Schätzgüte schlechter ab; denn um dieselbe Genauigkeit zu erhalten, sind bei den parallelen Verfahren mehr historische Datenwerte erforderlich.

Ausgeglichen werden kann dieser Informationsverlust der parallelen Histogramm-Matching-Verfahren durch den Einsatz der Kombinationstechnik. Hierbei kommt es zu einer Auswertung der Daten bezüglich verschiedener Intervallaufteilungen. Diese verschiedenen Informationsauswertungen werden zur letztendlichen Dichte-Approximation zusammengesetzt. Die auf der Kombinationstechnik beruhenden Dünngitter-Histogramm-Matching-Verfahren weisen – für eine in Bezug auf die betrachteten Datenpunkte optimal ausgewählte Levelanzahl – im Vergleich zu den anderen Histogramm-Matching-Verfahren die besten Schätzeigenschaften vor.

Allerdings benötigen sie dafür einen erheblich höheren Rechen- und Speicheraufwand als die anderen Verfahren. Dies ist einerseits darauf zurückzuführen, dass jeder Datenpunkt bezüglich der einzelnen zu kombinierenden Gitter mehrfach ausgewertet werden muss. Andererseits muss beim Einsatz der Kombinationstechnik eine vergleichsweise hohe Auflösung des Gitters verwendet werden, um dieselbe Genauigkeit wie beim Tensorprodukt-Ansatz zu erreichen.

Bei einem Vergleich mit dem Maximum-Likelihood-Verfahren (sowie den anderen Histogramm-Matching-Verfahren) stellte sich heraus, dass die auf der Kombinationstechnik beruhenden Dünngitter-Histogramm-Matching-Verfahren für kleine sowie sehr große Datensatzumfänge im Mittel die besten Schätzergebnisse liefern. Dabei interessieren vor allem die kleinen Datensätze, denn ihr Umfang ist mit dem der am Markt zu beobachtbaren Daten zu vergleichen.

Die Vorteile der Histogramm-Matching-Verfahren gegenüber dem Maximum-Likelihood-Verfahren liegen darin begründet, dass

- sie für beliebige stochastische Prozesse universell einsetzbar sind,
- für ihren Einsatz keine genaue Kenntnis der Verteilung des Prozesses notwendig ist,
- sie prinzipiell gut in der Zahl der zu schätzenden Parameter skalieren.

Ihr Nachteil besteht in einem wesentlich höheren Speicher- und Rechenaufwand.

Innerhalb dieser Arbeit wurde bei der Verwendung der Kombinationstechnik von einer für alle Richtungen konstanten Levelanzahl ausgegangen. Durch die Skalierungen der Datenwerte bedingt, würde sich allerdings eine Anpassung der Level bezüglich der einzelnen Richtungen anbieten. Durch eine solche Levelspezifizierung sind hinsichtlich der gesamt betrachteten Güte der Schätzungen der einzelnen Parameter bessere Ergebnisse zu erwarten.

Im numerischen Teil dieser Arbeit wurden die Parameterschätzverfahren speziell für das CIR-Modell untersucht. Wie sich allerdings anhand realer Daten in abschliessenden Analysen herausgestellt hat, eignet sich das CIR-Modell nicht unbedingt für eine angemessene Modellierung der aktuellen Short Rate. Aufgrund dieses Ergebnisses liegt es nahe, für die Modellierung der Short Rate andere Modelle zu untersuchen.

Anhang A.

Wichtiges aus der Wahrscheinlichkeitstheorie und Statistik

Für ein tieferes Verständnis der wahrscheinlichkeitstheoretischen Begriffe wird auf [Bau02] verwiesen. Hier soll nur ein knapper Überblick der in dieser Diplomarbeit verwendeten Begriffe und Resultate aus der Wahrscheinlichkeitstheorie und Statistik gegeben werden.

(Ω, \mathcal{A}, P) sei der im Folgenden zugrundegelegte Wahrscheinlichkeitsraum mit der Grundmenge (*Ergebnismenge*) Ω , der σ -Algebra (*Ereignismenge*) \mathcal{A} und dem *Wahrscheinlichkeitsmaß* P . Dabei gilt $P(\Omega) = 1$ sowie $0 \leq P(A) \leq 1$ für ein Ereignis $A \in \mathcal{A}$.

Definition A.0.1 [(P-)FAST SICHER]

Gilt für eine Menge $A \in \mathcal{A}$, dass $P(A) = 1$, so ist diese Menge (*P*-)fast sicher.

Eine Funktion, die jedem Ergebnis eine Wahrscheinlichkeit zuordnet, wird (*Wahrscheinlichkeits-*)*Dichte* genannt. Die Betrachtung dieser Wahrscheinlichkeiten für alle Elemente des Ergebnisraumes definiert die (*Wahrscheinlichkeits-*)*Verteilung*.

Definition A.0.2 [VERTEILUNG UND DICHTEN IM DISKRETEN FALL]

Im diskreten Fall liegen in der Ergebnismenge Ω nur endlich viele oder abzählbar unendlich viele Werte ω_i . Die (*Wahrscheinlichkeits-*)*Verteilung* ist dann gegeben durch die endlich vielen (beziehungsweise abzählbar unendlich vielen) *Wahrscheinlichkeiten*

$$P(X = x_i) = p(x_i) \quad \text{mit} \quad x_i = X(\omega_i) \quad \text{für alle } i. \quad (\text{A.1})$$

Die Funktion $x \mapsto p(x)$ wird als (*Wahrscheinlichkeits-*)*Dichte* bezeichnet.

Definition A.0.3 [VERTEILUNG UND DICHTEN IM STETIGEN FALL; ZUFALLSVARIABLE]

Im stetigen Fall kann die Verteilung einer reellen Zufallsvariablen X nicht durch die Wahrscheinlichkeiten $P(X = x_i)$ in (A.1) gegeben sein. Vielmehr ist entscheidend, mit welcher Wahrscheinlichkeit die Zufallsvariable X in ein bestimmtes Werte-Intervall fällt. Hier wird mit den Wahrscheinlichkeiten $P(a \leq X \leq b)$ gearbeitet. Diese Wahrscheinlichkeiten können mittels der (*Wahrscheinlichkeits-*)*Dichte* p über

$$P(a \leq X \leq b) = \int_a^b p(x) dx \quad (\text{A.2})$$

ermittelt werden. Gilt diese Identität für eine nichtnegative, integrierbare Funktion p , die zusätzlich die Eigenschaft der Normierung auf eins

$$\int_{-\infty}^{+\infty} p(x) dx = 1 \quad (\text{A.3})$$

erfüllt, so wird durch diese eine Zufallsvariable X definiert.

Definition A.0.4 [(KUMULATIVE) VERTEILUNGSFUNKTION IM STETIGEN FALL]

Mit einem nach unten unendlichen Intervall $(-\infty, x]$ ergibt sich aus (A.2) die sogenannte (kumulative) Verteilungsfunktion $x \mapsto F(x)$ mit

$$F(x) = P(X \leq x) = \int_{-\infty}^x p(y)dy.$$

Definition A.0.5 [ERWARTUNGSWERT]

Der Erwartungswert einer reellen Zufallsvariable $X : \Omega \rightarrow \mathbb{R}$ ist als das Integral bezüglich des Wahrscheinlichkeitsmaßes P definiert:

$$E[X] = \int_{\Omega} X dP$$

Mit Hilfe der (Wahrscheinlichkeits-)Dichte kann der Erwartungswert auch als das Integral

$$E[X] = \int_{-\infty}^{+\infty} xp(x)dx \tag{A.4}$$

ausgedrückt werden.

Bemerkung A.0.6 [ERWARTUNGSWERT IM DISKRETEN FALL]

Im diskreten Fall entspricht der Erwartungswert der Summe aller möglichen, mit der jeweiligen (Wahrscheinlichkeits-)Dichte gewichteten Werte:

$$E[X] = \sum_i x_i p(x_i)dx$$

Satz A.0.7 [LINEARITÄT DER ERWARTUNGSWERTE]

Die Erwartungswerte sind linear. Für m Zufallsvariablen X_k ($k = 1, \dots, m$) gilt

$$E \left[\sum_k X_k \right] = \sum_k E[X_k].$$

Satz A.0.8 [DAS SCHWACHE GESETZ DER GROSSEN ZAHLEN]

Der Mittelwert einer Stichprobe \bar{X}_n konvergiert im stochastischen Sinne gegen den Erwartungswert $E[X]$:

$$\lim_{n \rightarrow \infty} (|\bar{X}_n - E[X]| > \varepsilon) = 0.$$

Definition A.0.9 [VARIANZ]

Die Varianz entspricht dem Erwartungswert der quadratischen Abweichung der Zufallsvariable X von ihrem Erwartungswert:

$$\text{var}[X] = E[(X - E[X])^2] \tag{A.5}$$

Definition A.0.10 [STANDARDABWEICHUNG]

Die Wurzel aus der Varianz heißt Standardabweichung:

$$\text{std}[X] = \sqrt{\text{var}[X]} \stackrel{(A.5)}{=} \sqrt{E[(X - E[X])^2]}$$

Definition A.0.11 [KOVARIANZ, KORRELATIONSKOEFFIZIENT]

Die Kovarianz zweier Zufallsvariablen X und Y ist definiert durch

$$\text{cov}[X, Y] = E[(x - E[X])(y - E[Y])]. \tag{A.6}$$

Der Korrelationskoeffizient $\rho(X, Y)$ (kurz: $\rho_{X, Y}$) zweier Zufallsvariablen X und Y entspricht dem Quotienten von Kovarianz und Standardabweichungen.

$$\rho(X, Y) = \frac{\text{cov}[X, Y]}{\sqrt{\text{var}[X]}\sqrt{\text{var}[Y]}} \quad (\text{A.7})$$

Bemerkung A.0.12

Die Kovarianz ist sowohl ein Maß für die Stärke der Schwankungen der einzelnen Zufallsvariablen, als auch für die Stärke der gegenseitigen Abhängigkeit dieser Zufallsvariablen. Damit ist sie ein Maß für die Varianzen und die Korrelation, die die lineare Abhängigkeit zweier Zufallsvariablen beschreibt. Daher stammt auch der Name Kovarianz.

Satz A.0.13 [WERTEBEREICH DES KORRELATIONSKOEFFIZIENTEN]

Für zwei Zufallsvariablen X und Y gilt aufgrund der Cauchy-Schwarz-Ungleichung $(E[XY])^2 \leq E[X]^2 E[Y]^2$ für den Korrelationskoeffizienten

Insbesondere bedeutet der Wert 1 (−1) einen vollständig positiven (negativen) linearen Zusammenhang zwischen den betrachteten Zufallsvariablen. X und Y sind unkorreliert, wenn $\rho(X, Y) = 0$ ist.

Satz A.0.14 [VERSCHIEBUNGSSATZ FÜR DIE KOVARIANZ]

Die Kovarianz zweier Zufallsvariablen X und Y entspricht der Differenz des Erwartungswertes des Produktes und des Produktes der Erwartungswerte der einzelnen Zufallsvariablen:

$$\text{cov}[X, Y] = E[XY] - E[X]E[Y]$$

Satz A.0.15 [SYMMETRIE DER KOVARIANZ UND DES KORRELATIONSKOEFFIZIENTEN]

Die Kovarianz und damit per Definition auch der Korrelationskoeffizient zweier Zufallsvariablen X und Y ist symmetrisch.

$$\text{cov}[X, Y] = \text{cov}[Y, X] \quad \text{sowie} \quad \rho(X, Y) = \rho(Y, X)$$

Die Kovarianz einer Zufallsvariable zu sich selbst entspricht ihrer Varianz. (Dies folgt aus A.0.14.) Der Korrelationskoeffizient einer Zufallsvariable zu sich selbst ist aufgrund eines vollständigen linearen Zusammenhangs 1. (Siehe A.0.13.)

$$\text{cov}[X, X] = \text{var}[X] \quad \text{sowie} \quad \rho(X, X) = 1$$

Satz A.0.16 [KOVARIANZ- UND KORRELATIONSMATRIX]

Gegeben seien m Zufallsvariablen X_k ($k = 1, \dots, m$). Die Kovarianzen sowie die Korrelationen zwischen diesen m Zufallsvariablen lassen sich in der Kovarianzmatrix (C) beziehungsweise der Korrelationsmatrix (K) darstellen. Dies sind $m \times m$ -Matrizen, Auf der Hauptdiagonalen der Kovarianzmatrix stehen aufgrund der Symmetrie die Varianzen der einzelnen Zufallsvariablen, während sich bei der Korrelationsmatrix aufgrund der vollständigen positiven Korreliertheit einer Zufallsvariable mit sich selbst auf der Hauptdiagonalen lauter Einsen ($\rho(X, X) = 1$) befinden. Weiterhin gilt wegen den Symmetrieeigenschaften, dass oberhalb der Hauptdiagonalen jeweils die gleichen Informationen in Form der Kovarianzen beziehungsweise Korrelationskoeffizienten stehen wie unterhalb derselben.

$$C = \begin{pmatrix} \text{cov}[X_1, X_1] & \cdots & \text{cov}[X_1, X_m] \\ \vdots & \ddots & \vdots \\ \text{cov}[X_m, X_1] & \cdots & \text{cov}[X_m, X_m] \end{pmatrix} = \begin{pmatrix} \text{var}[X_1] & \cdots & \text{cov}[X_1, X_m] \\ \vdots & \ddots & \vdots \\ \text{cov}[X_m, X_1] & \cdots & \text{var}[X_m] \end{pmatrix}$$

$$K = \begin{pmatrix} \rho_{11} & \cdots & \rho_{1m} \\ \vdots & \ddots & \vdots \\ \rho_{m1} & \cdots & \rho_{mm} \end{pmatrix} = \begin{pmatrix} 1 & \cdots & \rho_{1m} \\ \vdots & \ddots & \vdots \\ \rho_{m1} & \cdots & 1 \end{pmatrix}$$

mit $\text{cov}[X_i, X_j] = \text{cov}[X_j, X_i]$ und $\rho_{ij} = \rho_{ji}$ (dabei ist $\rho_{ij} = \rho_{X_i, X_j}$) für $i, j \in \{1, 2, \dots, m\}$.

Definition A.0.17 [GLEICHVERTEILUNG]

Eine Zufallsvariable X heißt gleichverteilt, wenn ihre Dichte p überall gleich, also konstant ist. Für den Fall einer stetigen Verteilung auf einem Intervall $[a, b]$ ergibt sich direkt aus der Eigenschaft der Normierung auf eins die konstante Dichte $p = \frac{1}{b-a}$.

$$1 = \int_{-\infty}^{+\infty} \underbrace{p(x)}_{=p \text{ konstant}} dx = p \cdot \int_a^b 1 dx = p \cdot (b-a) \implies p = \frac{1}{b-a}$$

Bei einer stetigen Gleichverteilung auf einem Intervall $[a, b]$ ergibt sich daher:

$$\begin{aligned} \mathbb{E}[X] &= \int_{-\infty}^{+\infty} xp(x)dx = \frac{1}{b-a} \int_a^b x dx = \frac{1}{b-a} \frac{b^2 - a^2}{2} = \frac{a+b}{2} \\ \text{var}[X] &= \int_{-\infty}^{+\infty} (x - \mathbb{E}[x])^2 p(x)dx = \frac{1}{b-a} \int_a^b \left(x - \frac{a+b}{2}\right)^2 dx = \frac{(b-a)^2}{12} \end{aligned}$$

Für das Intervall $[0, 1]$ beträgt der Erwartungswert also $\frac{1}{2}$ und die Varianz $\frac{1}{12}$.

Definition A.0.18 [NORMALVERTEILUNG]

Die Dichte der Normalverteilung (auch: Gaußverteilung) ist von der Form

$$p(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}.$$

Sie wird von den beiden Parametern σ und μ vollständig bestimmt. Vielmehr entsprechen diese beiden Parameter gerade dem Erwartungswert und der Standardabweichung (der Wurzel aus der Varianz) dieser Verteilung.

$$\begin{aligned} \mathbb{E}[X] &= \int_{-\infty}^{+\infty} xp(x)dx = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{+\infty} xe^{-\frac{(x-\mu)^2}{2\sigma^2}} dx = \mu \\ \text{var}[X] &= \int_{-\infty}^{+\infty} (x - \mathbb{E}[x])^2 p(x)dx = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{+\infty} (x - \mu)^2 e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx = \sigma^2 \end{aligned}$$

Ist eine Zufallsvariable X gemäß der Normalverteilung mit Erwartungswert μ und Varianz σ^2 verteilt, so kann die Kurzschreibweise

$$X \sim N(\mu, \sigma^2)$$

genutzt werden.

Definition A.0.19 [(KUMULATIVE) NORMALVERTEILUNG]

Die (kumulative) Normalverteilung $x \mapsto F(x)$ ist gegeben durch

$$F(x) = P(X \leq x) = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^x e^{-\frac{(y-\mu)^2}{2\sigma^2}} dy$$

Definition A.0.20 [DICHTE DER STANDARDNORMALVERTEILUNG]

Die Standardnormalverteilung entspricht einer Normalverteilung mit Erwartungswert 0 und Varianz 1. Die Dichte ist demnach

$$p(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}.$$

Ist eine Zufallsvariable X standardnormalverteilt, so wird (entsprechend Definition A.0.18) die Kurzschreibweise

$$X \sim N(0, 1)$$

genutzt.

Definition A.0.21 [(KUMULATIVE) STANDARDNORMALVERTEILUNG]

Die (kumulative) Standardnormalverteilung $x \mapsto F(x)$ ist gegeben durch

$$F(x) = P(X \leq x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x e^{-\frac{y^2}{2}} dy.$$

Satz A.0.22 [SYMMETRIE DER STANDARDNORMALVERTEILUNG]

Die Dichte $p(x)$ der Standardnormalverteilung ist symmetrisch um den Erwartungswert 0.

$$p(x) = -p(x) \tag{A.8}$$

Die (kumulative) Standardnormalverteilung $x \mapsto F(x)$ ist folglich ebenfalls symmetrisch. Es gilt

$$F(-x) = 1 - F(x). \tag{A.9}$$

Satz A.0.23 [AUSSAGE DES ZENTRALER GRENZWERTSATZ]

Die Summe einer großen Anzahl unkorrelierter Zufallszahlen ist - unabhängig von der Verteilung der zugrundeliegenden Zufallszahlen - annähernd normalverteilt.

Bemerkung A.0.24 [NORMALVERTEILUNG VON MITTELWERTEN]

Mit dem zentralen Grenzwertsatz lässt sich schliessen, dass die über die Summe von Zufallszahlen definierten Mittelwerte annähernd normalverteilt sind, vorausgesetzt es handelt sich um unkorrelierte Zufallszahlen, deren Anzahl ausreichend groß ist.

Definition A.0.25 [LOGNORMALVERTEILUNG]

Ist eine Zufallsvariable X lognormalverteilt mit den Parametern μ und σ , dann entspricht die (kumulierte) Lognormalverteilung $x \mapsto F_{\ln}(x)$, die gegeben ist durch $F_{\ln}(x) = P(X \leq x)$, gerade der (kumulierten) Normalverteilung mit $F(\ln(x)) = P(X \leq \ln(x))$.

$$F_{\ln}(x) = F(\ln(x)) = \frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^{\ln(x)} e^{-\frac{(y-\mu)^2}{2\sigma^2}} dy$$

Eine Zufallsvariable ist also genau dann lognormalverteilt, wenn ihr Logarithmus normalverteilt ist. Die Dichte entspricht

$$p(x) = \frac{1}{\sqrt{2\pi}\sigma} \frac{1}{x} e^{-\frac{(\ln(x)-\mu)^2}{2\sigma^2}}. \tag{A.10}$$

Da der Faktor $1/x$ auftaucht, ergibt sich die Dichte allerdings nicht durch einfaches Ersetzen von x durch $\ln(x)$. Mit dieser Dichte ergibt sich für die (kumulierte) Lognormalverteilung auch

$$F_{\ln}(x) = \frac{1}{\sqrt{2\pi}\sigma} \int_0^x \frac{1}{y} e^{-\frac{(\ln(y)-\mu)^2}{2\sigma^2}} dy, \tag{A.11}$$

denn eine lognormalverteilte Zufallsvariable ist nur für nichtnegative Werte definiert. Diese Form der (kumulierte) Lognormalverteilung stimmt mit der obigen überein, da

$$\begin{aligned} F_{\ln}(x) &= \frac{1}{\sqrt{2\pi\sigma}} \int_0^x \frac{1}{y} e^{-\frac{(\ln(y)-\mu)^2}{2\sigma^2}} dy \\ &= \frac{1}{\sqrt{2\pi\sigma}} \int_{\ln(0)=-\infty}^{\ln(x)} e^{-\frac{(u-\mu)^2}{2\sigma^2}} du \end{aligned}$$

mit $u = \ln(y) \implies du = \frac{1}{y} dy, y = e^u$ gilt.

Satz A.0.26 [ERWARTUNGSWERT UND VARIANZ DER LOGNORMALVERTEILUNG]

$$\begin{aligned} \mathbb{E}[X] &= \int_{-\infty}^{+\infty} xp(x)dx \stackrel{(A.11)}{=} \int_0^{+\infty} xp(x)dx \\ &\stackrel{(A.10)}{=} \frac{1}{\sqrt{2\pi\sigma}} \int_0^{+\infty} e^{-\frac{(\ln(x)-\mu)^2}{2\sigma^2}} dx = e^{\mu+\frac{1}{2}\sigma^2} \quad (A.12) \\ \text{var}[X] &= \int_{-\infty}^{+\infty} (x - \mathbb{E}[X])^2 p(x)dx \stackrel{(A.11)}{=} \int_0^{+\infty} (x - \mathbb{E}[X])^2 p(x)dx \\ &\stackrel{(A.10),(A.12)}{=} \frac{1}{\sqrt{2\pi\sigma}} \int_0^{+\infty} \left(x - e^{\mu+\frac{1}{2}\sigma^2}\right)^2 e^{-\frac{(\ln(x)-\mu)^2}{2\sigma^2}} dx = e^{2\mu}(e^{2\sigma^2} - e^{\sigma^2}) \end{aligned}$$

Definition A.0.27 [WIENER-PROZESS (AUCH: STANDARD BROWNSCHE BEWEGUNG)]

Ein Wiener-Prozess (auch: Standard Brownsche Bewegung) ist ein stetiger stochastischer Prozess, dem folgende Eigenschaften zugrundeliegen:

- $W_0 = 0$ P -fast sicher,
- $W_t \sim N(0, t)$ für alle $t \geq 0$,
- Alle Zuwächse ΔW sind unabhängig voneinander.

Insbesondere gilt für die Zuwächse des Wiener Prozesses

$$W_t - W_s \sim N(0, t - s) \text{ für alle } s < t.$$

Literaturverzeichnis

- [AS02] AÏT-SAHALIA, YACINE: *Maximum-Likelihood Estimation of Discretely-Sampled Diffusions: A Closed-Form Approximation Approach*. *Econometrica*, 70:223–262, 2002.
- [ASK02] AÏT-SAHALIA, YACINE und ROBERT KIMMEL: *Estimating Affine Multifactor Term Structure Models Using Closed-Form Likelihood Expansions*. NBER Working Paper No. T0286, December 2002.
- [Bau02] BAUER, HEINZ: *Wahrscheinlichkeitstheorie*. de Gruyter, Berlin/New York, 5. Auflage, 2002.
- [BDT90] BLACK, FISCHER, EMANUEL DERMAN und WILLIAM TOY: *A One-Factor Model of Interest Rates and its Application to Treasury Bond Options*. *Financial Analysts Journal*, 46:33–39, Januar-Februar 1990.
- [BG04] BUNGARTZ, HANS-JOACHIM und MICHAEL GRIEBEL: *Sparse Grids*. *Acta Numerica*, Seiten 1–121, 2004.
- [BK91] BLACK, FISCHER und PIOTR KARASINSKI: *Bond and Option Pricing when Short Rates are Lognormal*. *Financial Analysts Journal*, 47:52–59, Juli-August 1991.
- [BM01] BRIGO, DAMIANO und FABIO MERCURIO: *Interest Rate Models*. Springer-Verlag, Berlin/Heidelberg/New York, 2001.
- [BS77] BEASLEY, J.D. und S.G. SPRINGER: *The percentage points of the normal distribution*. *Applied Statistics*, 26:115–121, 1977.
- [CGY04] CHAN, TONY F., MICHAEL GRIEBEL und ANDY M. YIP: *Data Clustering using Active Contours on Sparse Grids*. Working paper, Oktober 2004.
- [CIR85] COX, JOHN C., JONATHAN E. JR. INGERSOLL und STEPHEN A. ROSS: *A Theory of the Term Structure of Interest Rates*. *Econometrica*, 53:385–407, März 1985.
- [CS92] CHEN, REN-RAW und LOUIS SCOTT: *Pricing Interest Rate Options in a Two Factor Cox-Ingersoll-Ross Model of the Term Structure*. *The Review of Financial Studies*, 5:613–636, 1992.
- [CS95] CHEN, REN-RAW und LOUIS SCOTT: *Interest Rate Options in Multi-Factor Cox-Ingersoll-Ross Models of the Term Structure*. *The Journal of Derivatives*, 3:53–72, 1995.
- [Deu04] DEUTSCH, H.-P.: *Derivative und Interne Modelle*. Schäffer-Poeschel Verlag, Stuttgart, 3. Auflage, 2004.

- [DK96] DUFFIE, DARRELL und RUI KAN: *A Yield-Factor Model of Interest Rates*. *Mathematical Finance*, 64:379–406, Oktober 1996.
- [Dot78] DOTHAN, L. URI: *On the Term Structure of Interest Rates*. *Journal of Financial Economics*, 6:59–69, März 1978.
- [DS01] DUFFEE, GREGORY R. und RICHARD H. STANTON: *Estimation of Dynamic Term Structure Models*. 2001.
- [EuL] Europa-Lexikon der Bundesregierung: [http://www.bundesregierung.de\(/E/e/447680/Lexikon.htm\)](http://www.bundesregierung.de(/E/e/447680/Lexikon.htm).
- [FMW03] FISCHER, TOM, ANGELIKA MAY und BRIGITTE WALTHER: *Anpassung eines CIR-1-modells zur Simulation der Zinsstrukturkurve*. *Blätter der DGVM XXVI*, 2:193–206, Oktober 2003.
- [FMW04] FISCHER, TOM, ANGELIKA MAY und BRIGITTE WALTHER: *Anpassung eines CIR-k-modells zur Simulation der Zinsstrukturkurve*. *Blätter der DGVM XXVI*, 3:369–387, Mai 2004.
- [Gar04] GARCKE, JOCHEN: *Maschinelles Lernen durch Funktionsrekonstruktion mit verallgemeinerten dünnen Gittern*. Doktorarbeit, Rheinische Friedrich-Wilhelms-Universität Bonn, 2004.
- [Gla04] GLASSERMAN, PAUL: *Monte Carlo Methods in Financial Engineering*. Springer-Verlag, New York, 4. Auflage, 2004.
- [GP99] GRILL, WOLFGANG und HANS PERCZYNSKI: *Wirtschaftslehre des Kreditwesens*. Verlag Gehlen, Bad Homburg vor der Höhe, 33. Auflage, 1999.
- [GSZ92] GRIEBEL, MICHAEL, MICHAEL SCHNEIDER und CHRISTOPH ZENGER: *A combination technique for the solution of sparse grid problems*, Seiten 263–281. *Iterative Methods in Linear Algebra*. IMACS, Elsevier, North Holland, 1992.
- [GT96] GALLANT, A. RONALD und GEORGE E. TAUCHEN: *Which moments to match?* *Econometric Theory*, 12:657–681, 1996.
- [Hau05] HAUGH, MARTIN: *Continuous-Time Short Rate Models*. <http://www.columbia.edu/mh2078/TS05.html>, 2005.
- [HL86] HO, THOMAS S. Y. und SANG-BIN LEE: *Term Structure Movements and the Pricing of Interest Rate Contingent Claims*. *The Journal of Finance*, 41:1011–1029, Dezember 1986.
- [Hul01] HULL, JOHN C.: *Optionen, Futures und andere Derivative*. Internationale Standardlehrbücher der Wirtschafts- und Sozialwissenschaften, Oldenbourg, 4. Auflage, 2001.
- [HW90] HULL, JOHN C. und ALAN WHITE: *Pricing Interest Rate Derivative Securities*. *The Review of Financial Studies*, 3:573–592, 1990.
- [IWFK96] IKEDA, NOBUYUKI, SINZO WATANABE, MASATOSHI FUKUSHIMA und HIROSHI KUNITA: *Ito's Stochastic Calculus and Probability Theory*. Springer-Verlag, New York, 4. Auflage, 1996.

- [KP99] KLOEDEN, PETER E. und ECKHARD PLATEN: *Numerical Solution of Stochastic Differential Equations*. Springer-Verlag, Berlin/Heidelberg/New York, 3. Auflage, 1999.
- [KPS04] KELLY, LEAH, ECKHARD PLATEN und MICHAEL SØRENSEN: *Estimation for discretely observed diffusions using transform functions*. Journal of Applied Probability, 41A:99–118, 2004.
- [Kwo98] KWOK, YUE-KUEN: *Mathematical Models of Financial Derivatives*. Springer-Verlag, Singapore, 1. Auflage, 1998.
- [LL96] LAMBERTON, D. und B. LAPEYRE: *Introduction to Stochastic Calculus Applied to Finance*. Chapman & Hall/CRC Press LLC, Boca Raton, Florida, 1996.
- [Lo86] LO, ANDREW W.: *Maximum Likelihood Estimation of generalized Ito Processes with discretely sampled data*. NBER Technical Paper Series, 59, 1986.
- [LVA06] Landesvermessungsamt NRW: <http://www.lverma.nrw.de>, April 2006.
- [Mer05] MERTENS, THOMAS: *Optionspreisbewertung mit dünnen Gittern*. Diplomarbeit, Rheinische Friedrich-Wilhelms-Universität Bonn, März 2005.
- [Mis04] MISCIA, ORAZIO DI: *Term structure of interest models: concept and estimation problem in a continuous-time setting*. Working paper, School of Economics and Management at the University of Aarhus, November 2004.
- [Mor95] MORO, B.: *The full monte*. Risk, 8:57–58, Februar 1995.
- [MR97] MUSIELA, MAREK und MAREK RUTKOWSKI: *Martingale Methods in Financial Modelling*. Springer-Verlag, Berlin, 1997.
- [Nef96] NEFTCI, SALIH N.: *An Introduction to the Mathematics of Financial Derivatives*. Academic Press, San Diego/London, 1996.
- [Ped95] PEDERSEN, A.R.: *A new approach to maximum likelihood estimation for stochastic differential equations based on discrete observations*. Scandinavian Journal of Statistics, 22:55–71, 1995.
- [Pou99] POULSEN, ROLF: *Approximate maximum likelihood estimation of discretely observed diffusion processes*. Tech. Rep. 29, Centre for Analytical Finance, University of Aarhus., 1999.
- [Reb96] REBONATO, RICARDO: *Interest Rate Option Models*. John Wiley & Sons, New York, 1996.
- [SC97] SANTA-CLARA, PEDRO: *Simulated Likelihood Estimation of Diffusions with an Application to Short Term Interest Rate*. Working Paper, Anderson School of Management, Januar 1997.
- [Sch05] SCHMIDT, THORSTEN: *Zinsstrukturmodelle*. <http://www.math.uni-leipzig.de/~tschmidt>, Mathematisches Institut der Universität Leipzig, Januar 2005.
- [See05] SEEBER, GILG: *Algorithmische Mathematik 6 – Statistik*. <http://stat.uibk.ac.at/gs/LV/uibk/AlgMath6>, Institut für Statistik der Universität Innsbruck, Januar 2005.

- [Sey00] SEYDEL: *Einführung in die numerische Berechnung von Finanz-Derivaten*. Springer-Verlag, Berlin/Heidelberg/New York, 2000.
- [Sil86] SILVERMAN, BERNARD W.: *Density Estimation for Statistics and Data Analysis*. Chapman and Hall, 1986.
- [Sø97] SØRENSEN, M.: *Estimating Functions for Discretely Observed Diffusions; A review*. IMS Lecture Notes - Monograph Series, 32:305–325, 1997.
- [Sta97] STANTON, RICHARD H.: *A Nonparametric Model of Term Structure Dynamics and the Market Price of Interest Rate Risk*. Journal of Finance, 52:1973–2002, 1997.
- [Vas77] VASICEK, OLDRICH: *An Equilibrium Characterization of the Term Structure*. Journal of Financial Economics, 5:177–188, November 1977.
- [Zei] Zeitreihendatenbank der Deutschen Bundesbank: <http://www.bundesbank.de>.

Index

- (*P*-)fast sicher, 141
- (Kumulative) Normalverteilung, 144
- (Kumulative) Standardnormalverteilung, 145
- (Kumulative) Verteilungsfunktion, 142
- (Rest-)Laufzeit, 32

- Abzinsungspapiere, 9
- Aktie, 9
 - Nominalwert~, 9
 - Stück~, 9
- Aktionär, 9
- Anleger, 8
- Anleihe
 - Kupon~, 11
 - Null-Kupon~, 11
- Arbitragefreie Modelle, 46

- Börsenfähiges Wertpapier, 7
- Bond, 11

- Cox-Ingersoll-Ross-Modell, 52
 - CIR-Modell, 52

- Diskont-Anleihe, 38
- Diskretisierung, 59
- Diskretisierungsfehler, 59
- Diskretisierungsschema, 59
- Drift, 20

- Effekte, 7
- Emission, 8
- Emittent, 8
- Equilibrium models, 44
- Erwartungswert, 142
- Euler-Maruyama-Schema, 59

- Fälligkeitszeitpunkt (Fälligkeit), 13
- Faktor-Modell
 - Ein~, 33
 - Mehr~, 33

- Festverzinsliche Anleihe, 11
- Finanzmärkte, 8
 - Devisenmarkt, 9
 - Geldmarkt (Bankengeldmarkt), 8
 - Kapitalmarkt, 8
 - Kassamarkt, 9
- Floating Rate Notes, 11
- Forward Rate, 32
 - diskret, 36
 - instantan, 37

- Gaußverteilung, 144
- Gedächtnislosigkeit eines Prozesses, 14
- Gesamtlaufzeit, 32
- Gläubiger, 10
- Gleichgewichtsmodelle, 44
- Gleichverteilung, 144

- Histogramm, 93
 - Historisches ~, 93
 - Naives ~, 93
 - Simuliertes ~, 93

- Irrfahrt, 14
- Itô
 - ~-Prozess, 19
 - Lemma von ~, 20
 - Stochastische DGL von ~, 19

- Kassazins
 - diskret, 36
 - instantan, 37
 - stetig, 36
- Korrelationskoeffizient, 143
- Korrelationsmatrix, 143
- Kovarianz, 142
- Kovarianzmatrix, 143

- Level-Effekt, 45
- Lognormale Modelle, 46

- Markov-Prozess, 13
- Marktpreis des Risikos, 40
- Martingaleigenschaft, 19
- Mean Reversion, 44
 - ~-Effekt, 25
 - ~-Niveau, 25
 - ~-Stärke, 25
- Mitgliedschaftspapiere, 7
- No-arbitrage models, 46
- Nominal
 - ~wert, 9
 - ~wertaktie, 9
 - ~zins, 12
- Normalverteilung, 144
- Normierung auf eins, 141
- Pfad, 13
- Prämien-Anleihe, 38
- Prinzip der Arbitragefreiheit, 32
- Prozess
 - Stetig stochastischer \sim , 13
 - Stochastischer \sim , 13
- Pull-to-par-Phänomen, 38
- Random Walk, 14
 - ~-Vektor (RW-Vektor), 14
- repräsentative Stichprobe, 88
- Risiko-Transfer, 8
- Schrittvektoren, 14
- Schuldner, 10
- Schuldverschreibung, 10
- Schwache Markteffizienz, 13
- Selbstähnlichkeit, 16
- Short Rate, 37
- Spot Rate, 32
 - diskret, 36
 - instantan, 37
 - stetig, 36
- Standard Brownsche Bewegung, 146
- Standardabweichung, 142
- Straight Bonds, 11
- Term structure, 32
- Terminzins
 - diskret, 36
 - instantan, 37
 - stetig, 36
- Unkorreliertheit, 143
- Urkunde, 7
- variabel verzinsliche Anleihe, 11
- Varianz, 142
- Vasicek-Modell, 48
- Volatilität, 20
- Wertpapier, 7
 - festverzinsliches, 12
 - Geld~, 7
 - Kapital~, 7
 - Waren~, 7
- Wiener-Prozess, 18, 146
- Zero-Bond, 11
- Zins
 - Geld~, 8
 - Kapital~, 9
 - Kassazins, 32
 - Terminzins, 32
- Zinsstruktur, 32
- Zinsstrukturkurve
 - buckelig, 33
 - fallend, 33
 - flach, 33
 - invers, 33
 - normal, 33
 - steigend, 33
- Zinsstrukturkurven, 33
- Zinsstrukturmodelle, 33
- Zinstragende Papiere, 9
- Zufallsvariable, 141