

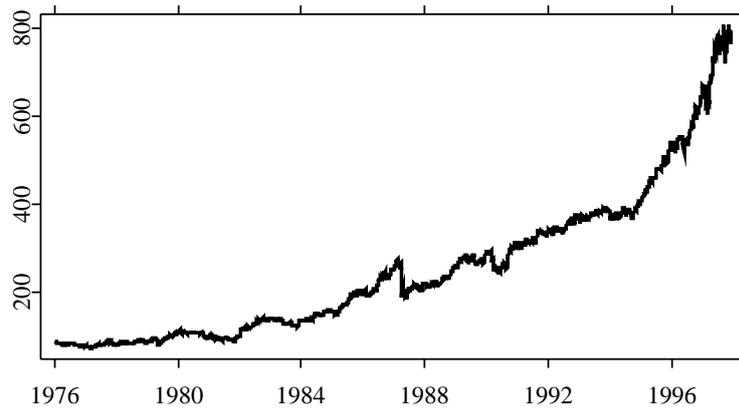
J. Franke
W. Härdle
C. Hafner

Einführung in die Statistik der Finanzmärkte



Springer

Einführung in die
Statistik der Finanzmärkte



Jürgen Franke
Wolfgang Härdle
Christian Hafner

Jürgen Franke
Universität Kaiserslautern
Fachbereich Mathematik
Postfach 3049
67653 Kaiserslautern
franke@mathematik.uni-kl.de

Wolfgang Härdle
Humboldt-Universität zu Berlin
Wirtschaftswissenschaftliche Fakultät
Institut für Statistik und Ökonometrie
Spandauer Str. 1
10178 Berlin
haerdle@wiwi.hu-berlin.de

Christian Hafner
R&D Energy Markets
Electrabel
Traverse d'Esopé 6
B-1348 Louvain-la-Neuve

Vorwort

Bis vor etwa zwanzig Jahren war die sogenannte Finanzmathematik im Vergleich zu anderen Anwendungsfeldern der Mathematik ein wenig anspruchsvolles Gebiet. Dies änderte sich schlagartig im Zuge der bahnbrechenden Arbeiten von Black, Scholes und Merton, für die sie 1997 den Nobelpreis für Wirtschaftswissenschaften bekommen haben, und der Bedeutung, die derivative Finanzinstrumente seither in den Finanzmärkten gewonnen haben. Die Bewertung solcher komplexer Anlageformen und die Einschätzung der damit verbundenen Risiken erfordern in einem durch die Globalisierung und die damit einhergehenden Verflechtungen nationaler Märkte immer komplizierter gewordenen Umfeld anspruchsvolle mathematische und statistische Modelle und Methoden.

Betont wurde die Notwendigkeit zur Weiterentwicklung und Verbesserung der mathematischen Grundlagen eines objektiven Risikomanagements durch Turbulenzen der Finanzmärkte in den neunziger Jahren - von den großen Pleiten im Derivatehandel wie Metallgesellschaft oder Orange County über die Asienkrise bis zum Fall des Long-Term Capital Management Funds. Dadurch sahen sich die Gesetzgeber zu Gegenmaßnahmen veranlasst. In Kontinentaleuropa wird diese noch lange nicht abgeschlossene Entwicklung im wesentlichen durch die Basler Bankenaufsicht (Basel Committee on Banking Supervision) beeinflusst, deren Empfehlungen EU-weit als Grundlage für Gesetzgebungsinitiativen dienen, mit denen Finanzinstitutionen zu einem dezidierten, alle Geschäftsbereiche umfassenden Risikomanagement verpflichtet werden. Daraus resultiert ein explodierender Bedarf an Fachleuten im Financial Engineering, die intern Risiken kontrollieren und profitable Investitionsmöglichkeiten entdecken können und die außerdem Sorge tragen, dass den gesetzlichen Vorschriften genüge getan wird. Interne Modelle, die vom Bundesaufsichtsamt für Kreditwesen in Kooperation mit den Landesämtern begutachtet werden, tragen dazu bei, ein auf das jeweilige Institut zugeschnittenes kostenoptimales Risikomanagement zu ermöglichen.

In Zukunft werden die hierfür entwickelten Ansätze auch für andere Firmen als Banken und ähnliche Finanzinstitutionen relevant, da das Gesetz zur Kontrolle und Transparenz im Unternehmensbereich (KonTraG) unter anderem alle Aktiengesellschaften zu einem dezidierten Risikomanagement verpflichtet, was auf jeden Fall auch ein Finanzrisikomanagement beinhaltet.

Der vorliegende Text soll die nötigen mathematischen und statistischen Grundlagen für eine Tätigkeit im Financial Engineering vermitteln. Unser Ziel ist es dabei, eine überschaubare Einführung in wichtige Ideen aus den verschiedensten Bereichen der Finanzmathematik und Finanzstatistik zu geben. Wir streben keine umfassende Abdeckung aller in der Praxis relevanten Details an, und wir umgehen auch bewusst die technischen Feinheiten der modernen, auf dem Methodenapparat der stochastischen Analysis aufbauenden Finanzmathematik. Für beide Zwecke gibt es bereits eine Vielzahl von Büchern. Wir wollen stattdessen Studie-

renden der Mathematik, der Statistik und der Wirtschaftswissenschaften bzw. entsprechend vorgebildeten, interessierten Studierenden anderer Fachrichtungen einen ersten Einstieg in die Modellierung und mathematische Analyse von Finanzdaten geben. Gleichzeitig richtet sich das Buch an Praktiker, die ihr im Beruf erworbenes Wissen vertiefen oder verbreitern wollen. Neben einer Einführung in die klassische Theorie der Bewertung von Derivaten legen wir dabei besonderen Wert auf die statistischen Aspekte beim Einsatz finanzmathematischer Verfahren, d.h. die Auswahl geeigneter Modelle, sowie ihre Anpassung und Validierung anhand von Daten.

Das vorliegende Buch besteht aus drei Teilen. Die ersten beiden sind so organisiert, dass sie unabhängig voneinander gelesen werden und als Stoff für je eine Lehrveranstaltung im Gesamtumfang von ca. 30 Stunden (2 Semesterwochenstunden) dienen können. Dabei nehmen wir bewusst ein gewisses Maß an Redundanz in Kauf, wenn dasselbe Thema, wenn auch unter unterschiedlichen Gesichtspunkten, in beiden Teilen angesprochen wird. Der dritte Teil stellt ausgewählte Anwendungen auf aktuelle Probleme aus der Praxis vor. Sowohl die *Bewertung von Optionen* wie auch die *Statistische Modellierung von Finanzzeitreihen* sind mehrfach Thema von Seminaren und Vorlesungen im Schwerpunkt Financial Mathematics des internationalen Studienprogramms Mathematics International der Universität Kaiserslautern (www.mathematik.uni-kl.de) sowie im VWL, BWL und Statistik Studiengang an der Humboldt-Universität zu Berlin (ise.wiwi.hu-berlin.de) gewesen. Gleichzeitig waren sie Grundlage von Weiterbildungskursen, die die Autoren im In- und Ausland für Praktiker aus Banken gehalten haben.

Der erste Teil *Bewertung von Optionen* behandelt die klassische Preistheorie für derivative Finanzinstrumente. Neben der Black-Scholes-Gleichung für gewöhnliche europäische und amerikanische Optionen und ihre numerische Lösung über die Approximation durch Binomialprozesse nach Cox, Ross und Rubinstein werden auch einige gängige exotische Optionen und ihre Bewertung angesprochen. Stochastische Modelle für die Zinsentwicklung und die Preisbestimmung von Zinsderivaten schließen diesen Abschnitt ab. Das nötige Handwerkszeug aus der stochastischen Analysis, insbesondere der Wiener-Prozess, stochastische Differentialgleichungen und die Itô-Formel, werden heuristisch motiviert und nicht rigoros hergeleitet. Auch auf fortgeschrittene Methoden der modernen Finanzmathematik wie die Martingaltheorie und die daraus resultierende elegante Charakterisierung von Arbitragefreiheit in vollständigen Märkten verzichten wir, um den Text auch für Nichtmathematiker bzw. Mathematiker ohne entsprechende vertiefte Kenntnisse der Theorie stochastischer Prozesse verständlich zu machen.

Der zweite Teil *Statistische Modellierung von Finanzzeitreihen* stellt die nunmehr klassische auf Engle, Bollerslev u.a. zurückgehende Finanzzeitreihenanalyse vor. Ausgehend von den traditionellen linearen Prozessen, insbesondere der ARMA- und ARIMA-Prozesse, die seit langem ein populäres Handwerkszeug der Ökonometrie bei der Modellierung von Wirtschaftsdaten sind, wird motiviert, warum Finanzzeitreihen in der Regel nicht auf diese Weise beschrieben werden können. Als alternative, mit den ARMA-Prozessen konzeptionell verwandte Modellklasse werden Zeitreihenmodelle mit stochastischer Volatilität betrachtet: neben den gebräuchlichen ARCH- und GARCH-Prozessen auch aktuelle Verallgemeinerungen, die eine asymmetrische Verteilung der Renditen zulassen. Als Unterstützung bei der explorativen Datenanalyse und beim Auffinden und Validieren einfacher parametrischer Modelle

gilt unser besonderes Augenmerk nichtparametrischen Modellen für Finanzzeitreihen und ihre Anpassung an Daten mit Hilfe von Kernschätzern und anderen Glättungsverfahren.

Im dritten Teil *Spezifische Finanzanwendungen* werden aktuelle praktische Fragen wie flexible Optionsbewertung, Risikoquantifizierung, Generierung von Handelssignalen sowie Kredit scoring und Rating angesprochen. Dabei werden zum einen die in den ersten beiden Teilen vorgestellten Konzepte eingesetzt, um Lösungsvorschläge zu entwickeln, zum anderen aber auch spezifische Erweiterungen und Alternativen zu nichtparametrischen Zeitreihenmodelle wie semiparametrische Ansätze und neuronale Netze betrachtet.

Der vorliegende Text ist als e-book konzipiert. Das Buch befindet sich als HTML- und PDF-file auch auf beiliegender CD. Alle Beispiele, Tabellen und Graphen können interaktiv reproduziert und eventuell verändert werden. Eine Netzversion befindet sich auf dem e-book Server www.quantlet.com. Auch auf diesem Server können die Beispiele interaktiv ausgeführt werden.

Das vorliegende Buch wäre nicht möglich ohne die kooperativen Vorschläge von P. Cizek, M. Fengler, Z. Hlavka, E. Kreuzberger, S. Klinke, D. Mercurio und D. Peithmann. Der erste Teil des Buches ist aus einem Weiterbildungskurs hervorgegangen, der gemeinsam mit G. Maercker, K. Schindler und N. Siedow entwickelt worden ist. Wir danken all diesen Kolleginnen und Kollegen. Ganz besonderen Dank möchten wir aussprechen an T. Kleinow, der den Text in allen Phasen begleitet hat, die e-book Plattform geschaffen und betreut hat und der mit zahlreichen Kommentaren die Präsentation des Buches mitbeeinflusst hat.

Für die Redaktion des Textes danken wir Beate Siegler und Anja Ossetrova.

Kaiserslautern, Berlin und Louvain-la-Neuve, Mai 2001

Inhaltsverzeichnis

I. Bewertung von Optionen	1
1. Finanzderivate	3
2. Grundlagen des Optionsmanagements	11
2.1. Arbitragebeziehungen	11
2.2. Portefeuille-Versicherung	22
3. Grundlegende Begriffe der Wahrscheinlichkeitstheorie	29
3.1. Reellwertige Zufallsgrößen	29
3.2. Erwartungswert und Varianz	31
3.3. Schiefe und Kurtosis	33
3.4. Zufallsvektoren, Abhängigkeit, Korrelation	33
3.5. Bedingte Wahrscheinlichkeiten und Erwartungswerte	35
3.6. Literaturhinweise	36
4. Stochastische Prozesse in diskreter Zeit	37
4.1. Binomialprozesse	37
4.2. Trinomialprozesse	40
4.3. Allgemeine Irrfahrten	41
4.4. Geometrische Irrfahrten	43
4.5. Binomialmodelle mit zustandsabhängigen Zuwächsen	44
5. Stochastische Integrale und Differentialgleichungen	47
5.1. Der Wiener-Prozess	47
5.2. Stochastische Integration	50
5.3. Stochastische Differentialgleichungen	52
5.4. Der Aktienkurs als stochastischer Prozess	54
5.5. Itô's Lemma	55
5.6. Literaturhinweise	58
6. Black-Scholes-Optionsmodell	59
6.1. Die Black-Scholes-Differentialgleichung	59
6.2. Die Black-Scholes-Formel für europäische Optionen	65
6.3. Risikomanagement mit Hedge-Strategien	69

6.3.1.	Delta-Hedgen	72
6.3.2.	Gamma und Theta	75
6.3.3.	Rho und Vega	77
6.3.4.	Historische und implizierte Volatilität	78
6.4.	Literaturhinweise	81
7.	Das Binomialmodell für europäische Optionen	83
7.1.	Der Cox-Ross-Rubinstein-Ansatz zur Optionsbewertung	84
7.2.	Diskrete Dividendenerträge	86
7.2.1.	Dividenden als Prozentsatz des Aktienkurses	89
7.2.2.	Dividenden als feste Beträge	89
7.3.	Literaturhinweise	92
8.	Amerikanische Optionen	95
8.1.	Arbitragebeziehungen für amerikanische Optionen	95
8.2.	Das Trinomialmodell für amerikanische Optionen	102
8.3.	Literaturhinweise	106
9.	Exotische Optionen und Zinsderivate	107
9.1.	Beispiele für exotische Optionen	107
9.1.1.	Zusammengesetzte Optionen, Optionen auf Optionen	107
9.1.2.	Chooser Optionen oder “Wie es euch gefällt”-Optionen	108
9.1.3.	Barrier-Optionen	108
9.1.4.	Asiatische Optionen	109
9.1.5.	Lookback-Optionen	111
9.2.	Modelle für den Zinssatz und Zinsderivate	112
9.2.1.	Wert eines Bonds bei bekannten zeitabhängigem Zinssatz	112
9.2.2.	Stochastische Zinsmodelle	113
9.2.3.	Die Bondbewertungs-Gleichung	113
9.2.4.	Lösung der Zerobond-Bewertungsgleichung	115
9.2.5.	Bewertung von Bond-Optionen	116
9.3.	Literaturhinweise	116
II.	Statistische Modellierung von Finanzzeitreihen	117
10.	Einführung: Definitionen und Konzepte	119
10.1.	Einige Definitionen	120
10.2.	Statistische Analyse deutscher Aktienrenditen	126
10.3.	Erwartungsbildung und Markteffizienz	127
10.4.	Ökonomische Modelle: Ein kurzer Überblick	133
10.4.1.	Aktienkurse: Das CAPM	133
10.4.2.	Wechselkurse: Die Zinsparitätentheorie	134
10.4.3.	Zinsstruktur: Das Cox-Ingersoll-Ross Modell	136
10.4.4.	Derivate: Das Black-Scholes Modell	137

10.4.5. Der Marktpreis des Risikos	139
10.5. Die Irrfahrt-Hypothesen	142
10.6. Einheitswurzeltests	144
10.6.1. Dickey-Fuller Tests	144
10.6.2. Der KPSS Test auf Stationarität	146
10.6.3. Varianzquotiententests	148
10.7. Literaturhinweise	150
11. ARIMA Zeitreihenmodelle	151
11.1. Moving Average Prozesse	152
11.2. Autoregressive Prozesse	153
11.3. ARMA Modelle	156
11.4. Partielle Autokorrelationen	157
11.5. Schätzung der Momentfunktionen	160
11.5.1. Schätzung der Mittelwertfunktion	160
11.5.2. Schätzung der Kovarianzfunktion	162
11.5.3. Schätzung der ACF	162
11.6. Portmanteau Statistiken	163
11.7. Schätzung von $AR(p)$ Modellen	164
11.8. Schätzung von $MA(q)$ und $ARMA(p, q)$ Modellen	165
11.9. Literaturhinweise	168
12. Zeitreihen mit stochastischer Volatilität	171
12.1. ARCH- und GARCH Modelle	171
12.1.1. ARCH(1): Definition und Eigenschaften	172
12.1.2. Schätzung von ARCH(1) Modellen	178
12.1.3. ARCH(q): Definition und Eigenschaften	182
12.1.4. Schätzung von ARCH(q) Modellen	183
12.1.5. Generalisiertes ARCH (GARCH)	184
12.1.6. Schätzung von GARCH(p, q) Modellen	187
12.2. Erweiterungen der GARCH-Modelle	189
12.2.1. Exponential GARCH	190
12.2.2. Threshold ARCH Modelle	191
12.2.3. Risiko und Rendite	192
12.2.4. Schätzergebnisse für DAX-Renditen	193
12.3. Literaturhinweise	194
13. Nichtparametrische Konzepte für Finanzzeitreihen	197
13.1. Nichtparametrische Regression	199
13.2. Konstruktion des Schätzers	201
13.3. Asymptotische Normalität	203
13.4. Literaturhinweise	217
III. Spezifische Finanzanwendungen	219

14. Optionsbewertung mit flexiblen Volatilitätsschätzern	221
14.1. Optionsbewertung mit ARCH-Modellen	221
14.2. Eine Monte-Carlo-Studie	228
14.3. Anwendung zur Bewertung von DAX-Calls	232
14.4. Literaturhinweise	234
15. Value at Risk und Backtesting	237
15.1. Vorhersage und VaR-Modelle	237
15.2. Backtesting mit erwartetem Shortfall	239
15.3. Backtesting in Aktion	240
15.4. Literaturhinweise	246
16. Neuronale Netze	247
16.1. Vom Perzeptron zum nichtlinearen Neuron	248
16.2. Backpropagation	255
16.3. Neuronale Netze in der nichtparametrischen Regressionsanalyse	257
16.4. Vorhersage von Finanzzeitreihen mit neuronalen Netzen	262
16.5. Risikoquantifizierung mit neuronalen Netzen	265
17. Volatilitätsrisiko von Optionsportfolios	271
17.1. Einführung	271
17.2. Datenbeschreibung	273
17.3. Hauptkomponentenanalyse der VDAX-Dynamik	275
17.4. Stabilitätsanalyse der VDAX Dynamik	277
17.5. Messung des implizierten Volatilitätsrisikos	279
18. Nichtparametrische Schätzer für Kreditausfallwahrscheinlichkeiten	283
18.1. Logistische Regression	283
18.2. Semiparametrische Modelle für das Kredit scoring	285
18.3. Kredit scoring mit neuronalen Netzen	287
Literaturverzeichnis	289

Teil I.

Bewertung von Optionen

1. Finanzderivate

Die klassische Finanzmathematik beschäftigt sich in erster Linie mit grundlegenden Finanzinstrumenten oder Anlageformen wie Aktien, Devisen, festverzinslichen Anleihen, Rentenwerten und anderen mehr. Unter *Finanzderivaten* oder *derivativen Finanzinstrumenten* (*derivatives*, *derivative securities* oder *contingent claims*) versteht man Anlageformen, die von einfacheren Finanzinstrumenten abgeleitet werden. Der Wert des Derivats hängt vom Wert des zugrundeliegenden Instruments (*underlying*) ab. In diesem Abschnitt betrachten wir mit Terminkontrakten und Optionen zwei einfache Formen von Finanzderivaten sowie einige Kombinationen von ihnen. Da die moderne Finanzmathematik auch in der deutschen Praxis von englischen Begriffen dominiert wird, erwähnen wir auch jeweils den englischen Fachausdruck, wenn der nicht offensichtlich mit dem deutschen übereinstimmt.

Einfache Finanzderivate waren bereits um die Jahrhundertwende an den europäischen Börsen populär, um in der Zeit zwischen den Weltkriegen an Bedeutung zu verlieren. In den siebziger Jahren begann zuerst in den USA die Renaissance der abgeleiteten Finanzinstrumente gleichzeitig mit den bahnbrechenden Arbeiten von Black, Scholes und Merton, die die Berechnung eines angemessenen Preises für solche Anlageformen ermöglichten. Ihr allgemeiner finanzwissenschaftlicher Ansatz, der nicht nur die Bewertung von Finanzderivaten erlaubt, sondern überall dort Anwendungen findet, wo in der Finanzwelt das mit komplexen Investitionsformen verbundene Risiko gemessen und kontrolliert werden soll, wurde 1997 mit dem Nobelpreis für Wirtschaftswissenschaften gewürdigt. Gleichzeitig gab er den Anstoß zur Entwicklung der modernen analytischen Finanzmathematik, deren Grundlagen im ersten Teil dieses Buches beschrieben wird. Wir konzentrieren uns hierbei auf die mathematische Modellierung und Analyse. Dieser Abschnitt beschränkt sich daher darauf, in knapper Form die dafür nötige Fachterminologie aus der Finanzwelt einzuführen. Zahlreiche für die Praxis, nicht aber für die mathematische Behandlung von Finanzderivaten wichtige Details übergehen wir dabei und verweisen auf die einschlägige Fachliteratur, z.B. [Hull \(2000\)](#), ?.

Eine besonders einfache Form von Finanzderivaten sind Terminkontrakte (*forward contract* oder *future contract*). Dabei handelt es sich um einen Vertrag zwischen zwei Parteien, zu einem bestimmten Zeitpunkt in der Zukunft eine Transaktion zu einem bereits bei Vertragsabschluss festgelegten Preis durchzuführen. Gegenstand des Terminkontrakts können Aktien, Devisen und Anleihen, aber auch landwirtschaftliche Produkte (Getreide, Fleisch) oder Rohstoffe (Öl, Kupfer, elektrische Energie) sein.

DEFINITION 1.1 (Terminkontrakt)

Ein Terminkontrakt ist ein Vertrag zwischen zwei Parteien, bei der sich der Käufer bzw. der Verkäufer des Kontraktes zu einem Zeitpunkt t verpflichtet, zu einem bestimmten zukünftigen

Zeitpunkt $T > t$, dem Ende der Laufzeit (*expiration date* oder *maturity*), ein Objekt zu einem heute vereinbarten Preis K , dem Terminkurs (*delivery price*), zu kaufen (Terminkauf) bzw. zu verkaufen (Terminverkauf).

Der Wert $V_{K,T}(S_t, \tau)$ eines Terminkontrakts zur Zeit t hängt von dem aktuellen Wert S_t des zugrundeliegenden Objekts, das Gegenstand der vereinbarten Transaktion ist, von der Restlaufzeit $\tau = T - t$ und von den Parametern K, T des Kontrakts ab.

In der Praxis unterscheidet man bei Terminkontrakten zwischen *Forwards* und *Futures*. Forwards werden zwischen den beiden Vertragspartnern ausgehandelt, und zwischen dem Abschluss des Terminkontrakts und dem Ende der Laufzeit kommt es zu keinen weiteren Zahlungen. Futures werden an Terminbörsen gehandelt. Als Absicherung gegen hohe Verluste aufgrund der Zahlungsunfähigkeit eines der Vertragspartner am Ende der Laufzeit dient hier ein täglicher Gewinn-Verlust-Ausgleich. Unter bestimmten Bedingungen, die über kürzere Zeiträume hinweg approximativ erfüllt sind, unterscheiden sich Forwards und Futures nicht in ihrem Wert.

BEISPIEL 1.1

Ein Investor erwirbt am 1. September einen Forward-Kontrakt mit dem Inhalt, in 90 Tagen 1000000 EUR zum Umtauschkurs von 1.2 USD zu kaufen. Falls der Kurs am Ende der 90 Tage auf 1.3 gestiegen ist, gewinnt der Investor, da 1000000 EUR dann am Markt für USD 1300000 verkauft werden können.

In der obigen Notation ist also $t = 1.$ September, $\tau = 90$ Tage, $T = 30.$ November und $K = \text{USD } 1200000.$

DEFINITION 1.2 (Spot-Preis, Forward-Preis, Future-Preis)

Der aktuelle Kurs S_t eines Objekts (Aktie, Währung, Rohstoff, ...) wird oft auch als Spot-Preis bezeichnet. Der Forward-Preis F_t des Objektes ist der Terminkurs, für den der Wert eines Forward-Kontraktes zur Zeit $t = 0$ ist, d.h. er löst die Gleichung $V_{F_t,T}(S_t, \tau) = 0$. Entsprechend ist der Future-Preis auf der Grundlage der Wertentwicklung eines Future-Kontrakts definiert.

Den Wert eines Forwards, aus dem sich sofort der Forward-Preis ergibt, werden wir später berechnen. Aus der Wertgleichheit von Forwards und Futures ergibt sich sofort, dass unter den entsprechenden Bedingungen auch Forward-Preis und Future-Preis zusammenfallen. Beim Abschluss eines Terminkontraktes zur Zeit $t = 0$ wird meist als Terminkurs $K = F_0$ gewählt. Der Kontrakt hat dann anfangs für Käufer und Verkäufer den Wert 0, und es finden keine Zahlungen statt. Im Laufe der Zeit werden dann Terminkurs K und aktueller Forward-Preis F_t voneinander abweichen.

Bei einem Terminkontrakt wechselt das zugrundeliegende Objekt am Ende der Laufzeit auf jeden Fall den Besitzer. Im Gegensatz dazu gewährt eine Option einem der beiden Vertragspartner das Recht zu entscheiden, ob es überhaupt zu einer Transaktion kommt. Offensichtlich macht es dabei einen Unterschied, ob die Entscheidung beim Käufer oder beim Verkäufer liegt. Daher wird zwischen zwei grundlegenden Typen von Optionen unterschieden: zwischen Kaufoptionen und Verkaufsoptionen. Abgesehen davon gleichen europäische

Optionen einem Forward-Kontrakt. Daneben gibt es amerikanische Optionen, bei denen der Inhaber des Ausübungsrechts sich auch schon vor Ende der Laufzeit zum Kauf oder Verkauf entschließen kann. Die Bezeichnungswiese hat historische Gründe und hängt nicht davon ab, wo die betreffenden Derivate gehandelt werden.

DEFINITION 1.3 (Kaufoption - Call, Verkaufsoption - Put)

Eine europäische Kaufoption oder europäischer Call ist ein Vertrag zwischen zwei Parteien, der dem Käufer das Recht gibt, zu einem bestimmten zukünftigen Zeitpunkt $T > t$, dem Ende der Laufzeit oder Verfallszeitpunkt (expiration date oder maturity), ein Objekt zu einem bei Vertragsschluss vereinbarten Preis K , dem Ausübungskurs (strike price oder exercise price), zu kaufen. Übt er das Recht nicht aus, verfällt die Option zur Zeit T ohne weitere Konsequenzen.

Eine europäische Verkaufsoption oder europäischer Put gibt dem Käufer das Recht, das Objekt zur Zeit T zum vereinbarten Preis K zu verkaufen.

Der Inhaber eines amerikanischen Calls bzw. eines amerikanischen Puts kann sein Recht, das Objekt zu kaufen bzw. zu verkaufen, jederzeit vom Erwerb der Option bis zum Verfallszeitpunkt T ausüben.

Die in der Definition eingeführten Optionsformen werden auch *Standardoptionen (plain vanilla options)* genannt. Daneben gibt es zahlreiche komplexere Finanzderivate, und es werden laufend neue entwickelt. Diese Finanzinstrumente werden meist nicht an den Terminbörsen gehandelt, sondern von Banken für einzelne Kunden maßgeschneidert entworfen. Man spricht hier von OTC-Derivaten (OTC = "over the counter"). Hierzu gehören z.B. Optionen auf Optionen, die dem Käufer zu einem Zeitpunkt T das Recht geben, eine Option mit Verfallszeitpunkt $T' > T$ zu kaufen oder zu verkaufen. Die mathematische Behandlung dieser *exotischen Optionen* ist besonders aufwendig, wenn der aktuelle Wert des Derivats nicht nur vom aktuellen Kurs S_t des zugrundeliegenden Objekts abhängt wie bei den Standardoptionen, sondern wenn der gesamte Kursverlauf $S_{t'}, 0 \leq t' \leq t$, eine Rolle spielt.

Beispiele für solche pfadabhängigen Optionen sind asiatische, Knockout- oder Lookback-Optionen: Bei asiatischen Optionen hängt der Ausübungskurs vom Mittelwert der Kurse des zugrundeliegenden Objekts über einen bestimmten Zeitraum ab. Knockout-Optionen verfallen, sobald der Kurs des zugrundeliegenden Objekts eine vorgegebene Schranke über- oder unterschreitet. Bei Lookback-Optionen hängt der Ausübungskurs vom Maximum oder Minimum der Kurse in einem bestimmten Zeitraum ab.

Die Bedeutung von Optionen reicht weit über den direkten Handel an den Terminbörsen und ihren Stellenwert als OTC-Derivate hinaus. Viele komplexe Finanzderivate, die auf den ersten Blick nichts mit Optionen zu tun haben, lassen sich durch geeignete Kombinationen mehrerer Optionen mit unterschiedlichen Parametern duplizieren. Auch viele klassische Finanzinstrumente wie z.B. Wandelanleihen oder Bezugsrechte bei Aktien besitzen einen Optionsanteil, und selbst in der Versicherungswirtschaft spielt die Optionsbewertungstheorie bei der Festlegung angemessener Prämien eine immer größere Rolle.

Um sich mit Terminkontrakten, Standardoptionen und einfachen Kombination von Optionen vertraut zu machen, ist es am einfachsten, zunächst nur den *Pay-off*, d.h. den Wert des Derivats am Verfalltag T , zu betrachten. Bei einem Terminkauf ist der Pay-off einfach

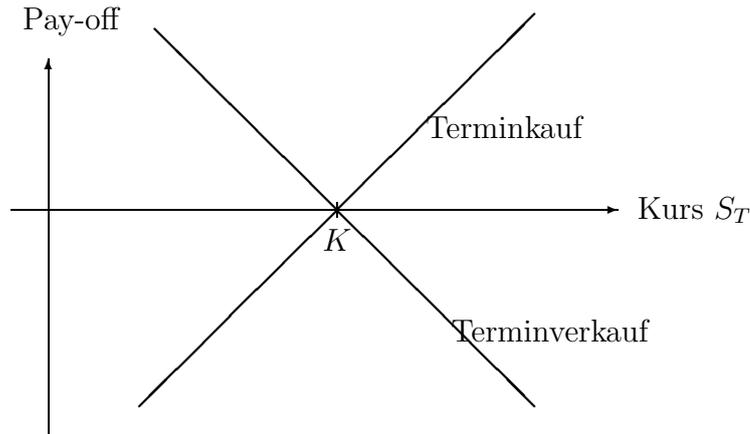


Abbildung 1.1.: Wert des Derivates am Verfallstag

$S_T - K$, wobei S_T der Spotpreis des zugrundeliegenden Objekts am Ende der Laufzeit ist. Der Käufer des Objekts erzielt einen Gewinn, wenn dieser aktuelle Kurs oberhalb des Terminkurses, den er zu zahlen hat, liegt. Er kann das für K erworbene Objekt sofort auf dem freien Markt wieder veräußern und dafür S_T einnehmen. Entsprechend ist der Pay-off eines Terminverkaufs $K - S_T$. Diese Pay-off Funktionen sind in [Abbildung 1.1](#) dargestellt.

Bei einem Call ist der Pay-off:

$$\max\{S_T - K, 0\} = (S_T - K)^+.$$

Der Besitzer der Option wird sein Kaufrecht nur ausüben, wenn der Ausübungskurs K unterhalb des aktuellen Marktpreises S_T für das Objekt liegt, und in diesem Fall denselben Ertrag wie bei einem Terminkauf erzielen. Andernfalls lässt er sein Recht verfallen. Bei einem Put ist der Pay-off entsprechend

$$\max\{K - S_T, 0\} = (K - S_T)^+.$$

Für den Erwerb einer Kaufoption, die ja ein Recht darstellt, ist zur Zeit $t = 0$ (im Gegensatz zu einem Terminkontrakt) stets ein positiver Betrag $C(S_0, T)$ zu zahlen. Zur Veranschaulichung einer Option wird oft auch dieser ursprünglich gezahlte Preis vom Pay-off abgezogen. Die so bestimmte Ertrags- oder Gewinnfunktion, für einen Call $(S_T - K)^+ - C(S_0, T)$, ist aber finanzmathematisch nicht korrekt, da hier Einnahmen zur Zeit T mit Ausgaben zur Zeit 0 verrechnet werden. Den korrekten Gesamtertrag erhält man durch Aufzinsen der zur Zeit 0 gezahlten Optionsprämie auf die Zeit T , denn der Investor hätte die Prämie statt für den Call auch für eine risikolose festverzinsliche Anlage verwenden können. Bei kontinuierlicher Verzinsung mit konstantem Zinssatz r wäre der Ertrag eines Calls $(S_T - K)^+ - C(S_0, T)e^{rT}$.

BEISPIEL 1.2

Wir betrachten zur Zeit 0 den Verkauf einer Kaufoption (Call short) mit Ausübungskurs K und Call-Prämie C_0 . Der Pay-off und die Gewinnfunktion sind in [Abbildung 1.2](#) bzw. in [Abbildung 1.3](#) dargestellt.

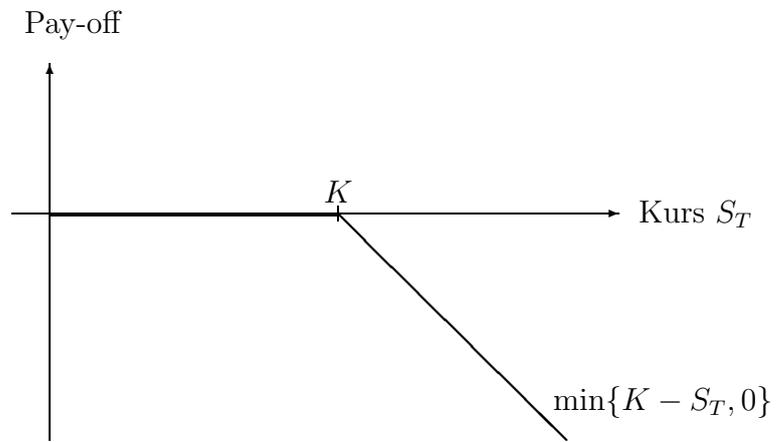


Abbildung 1.2.: Pay-off beim Verkauf einer Kaufoption

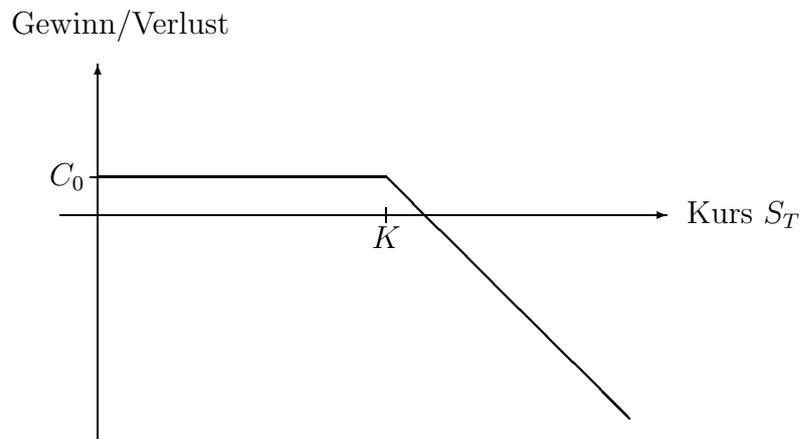


Abbildung 1.3.: Gewinn/Verlust beim Verkauf einer Kaufoption

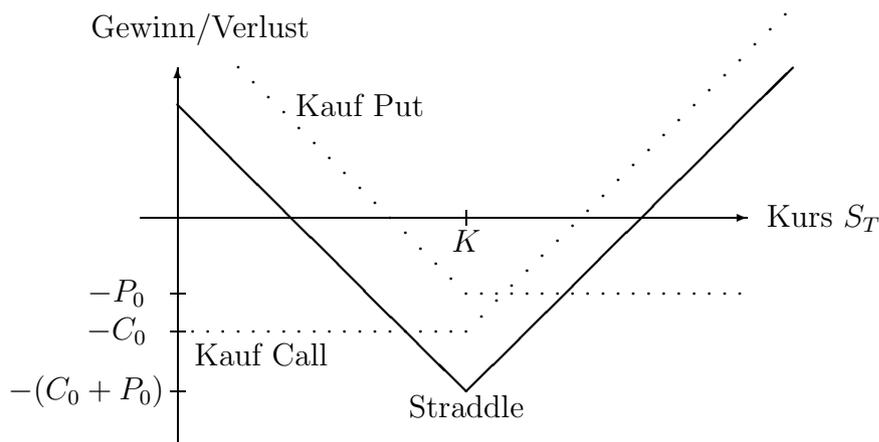


Abbildung 1.4.: Gewinn/Verlust beim Kauf eines Straddles

BEISPIEL 1.3

Für den Kauf eines Straddles, das heißt eines Portfolios aus einem Call und einem Put mit demselben Ausübungspreis K , ist der Gewinn/Verlust in Abbildung 1.4 dargestellt. C_0 bzw. P_0 bezeichnen wieder die Call- bzw. Putprämie.

Ein weiteres grundlegendes Finanzinstrument, das bei der Optionsbewertung eingesetzt wird, sind *Anleihen* oder *Bonds*. Neben den reinen Zinserträgen erhält der Käufer eventuell auch Couponzahlungen zu festen Zeitpunkten. Wir betrachten vor allem Zerobonds, d.h. Anleihen ohne Couponerträge.

DEFINITION 1.4 (Zerobond)

Ein Zerobond ist eine festverzinsliche Anleihe ohne Couponzahlungen. Der Käufer legt zur Zeit 0 einen Betrag B_0 an und erhält am Ende T der Laufzeit den Betrag B_T zurück, der sich aus B_0 zuzüglich des Zinsertrags ergibt. Der Wert des Zerobonds am Ende der Laufzeit wird auch als sein nomineller Wert bezeichnet.

Der Kauf eines Zerobonds entspricht dem Anlegen eines Geldbetrags zu einem festen Zinssatz für eine feste Zeitspanne. Der Verkauf eines Zerobonds entspricht analog dem Leihen eines Geldbetrags. Bonds werden gehandelt und können zum aktuellen Wert B_t , d.h. B_0 zuzüglich der bis zur Zeit t aufgelaufenen Zinsen, vorzeitig veräußert werden.

In der Praxis werden Zinsen zu diskreten Zeitpunkten (jährlich, halbjährlich, ...) gutgeschrieben und dann weiter verzinst. Bei jährlicher Verzinsung mit Zinssatz r p.a. (per annum) wächst das Anfangskapital B_0 in n Jahren einschließlich Zins und Zinseszins auf $B_n^{(1)} = B_0(1+r)^n$. Werden die Zinsen k -mal jährlich gutgeschrieben mit dann entsprechend geringerem Zinssatz $\frac{r}{k}$ pro $\frac{1}{k}$ Jahre, so ist das Kapital nach $n = \frac{nk}{k}$ Jahren $B_n^{(k)} = B_0(1 + \frac{r}{k})^{nk}$. Für $k \rightarrow \infty$ erhält man den Grenzfall der *stetigen Verzinsung*. Dabei wächst das Anfangskapital nach n Jahren auf $B_n = B_0 \cdot e^{nr}$. Im Falle der stetigen Verzinsung wird r auch als *short*

rate bezeichnet. Bei der Bewertung von Finanzderivaten wird meist von stetiger Verzinsung ausgegangen. Der Unterschied zur diskreten Verzinsung ist vor allem für häufige Gutschriften gering. Aus einem Kapital von $B_0 = 1000$ EUR wird bei einem Jahreszinssatz von $r = 10\%$ zum Beispiel innerhalb eines Jahres ($n = 1$) 1100 EUR bei jährlicher und 1105.17 EUR bei kontinuierlicher Verzinsung.

Um selbst diese kleinen Abweichungen zu umgehen, modifiziert man den stetigen Zinssatz geringfügig. Geht man von einer jährlichen Verzinsung r_1 , die am Jahresende gutgeschrieben wird, aus, so erhält man bei kontinuierlicher Verzinsung dasselbe Kapital nach n Jahren, d.h. $B_n = B_n^{(1)}$, wenn man als stetigen Zinssatz $r = \log(1 + r_1)$ wählt.

Im folgenden wird stets von einer stetigen Verzinsung ausgegangen. Um Geldbeträge, die zu verschiedenen Zeitpunkten gezahlt werden, vergleichen zu können, müssen sie auf denselben Zeitpunkt aufgezinst bzw. abgezinst werden, d.h. der Zinsertrag über die betreffende Zeitspanne wird addiert oder subtrahiert. Bei kontinuierlicher Verzinsung wird aus einem Kapital B zur Zeit t (mit $\Delta t > 0$)

Aufzinsung : $B e^{r\Delta t}$ zur Zeit $t + \Delta t$

Abzinsung : $B e^{-r\Delta t}$ zur Zeit $t - \Delta t$

Zum Schluss dieses Abschnitts werden noch einige Sprechweisen aus der Finanzwelt eingeführt. Ein *Portfolio* oder *Portefeuille* ist eine Kombination mehrerer Finanzinstrumente, deren Wertentwicklung als Ganzes gesehen wird. Ein einzelnes Finanzinstrument in einem Portfolio wird auch *Position* genannt. Ein Investor nimmt eine *long position* ein, wenn er ein Finanzinstrument kauft, und eine *short position*, wenn er es verkauft. Ein *Call long* ist z.B. der Kauf einer Kaufoption, ein *Put long* der Kauf einer Verkaufsoption, ein *Forward short* der Verkauf eines Forward-Kontraktes.

Ein Investor *schließt eine Position* in seinem Portfolio, wenn er die zukünftige Wertentwicklung des Portfolios unabhängig von dem betreffenden Finanzinstrument macht. Wird das Instrument frei gehandelt, kann er es einfach verkaufen (z.B. eine Aktie oder einen Bond) oder zurückzahlen (z.B. einen verkauften Bond, d.h. geliehenes Geld) und damit ganz aus dem Portfolio entfernen. Manche Finanzinstrumente lassen sich aber nicht auf diese Weise schließen. Der Investor kann dann seinem Portfolio ein entgegengesetztes Finanzinstrument hinzufügen, so dass die Werte der beiden Positionen sich gegenseitig aufheben und für das Gesamtportfolio keine Rolle mehr spielen.

BEISPIEL 1.4

Hat der Investor beispielsweise am 1. Februar einen über ein Jahr laufenden Forward-Kontrakt über 1000000 USD mit Terminkurs 1200000 EUR gekauft und möchte er diese Position vorzeitig am 1. Juni schließen, so kann er an diesem Termin einen entsprechenden Forward-Kontrakt über denselben USD-Betrag, mit demselben Terminkurs und demselben Fälligkeitstag 31. Januar verkaufen. Die Werte des gekauften und des verkauften Forwards addieren sich stets zu 0.

Short selling ist eine Handelsstrategie, bei der der Investor Objekte, z.B. Aktien, die ihm nicht gehören, verkauft und sie später zurückkauft. In der Praxis wird hierzu ein Broker eingeschaltet, der einen Partner vermittelt, in dessen Besitz sich die betreffenden Objekte befinden und der sie für eine gewisse Zeitspanne ausleiht. Der Investor, der sich für das

short selling entscheidet, verpflichtet sich, dem Besitzer der ausgeliehenen Objekte sämtliche Erträge wie z.B. Dividenden zu erstatten, die in der Zwischenzeit bis zur Rückgabe anfallen.

BEISPIEL 1.5

Entschließt sich ein Investor zum Beispiel für das Short selling von 1000 Aktien, so leiht er sie sich vom Besitzer und verkauft sie sofort für $1000 S_0$, wobei S_t den jeweilige Kurs zur Zeit t bezeichnet. Später, zur Zeit $t > 0$, schließt er die Position, indem er die Aktien für $1000 S_t$ zurückkauft und sie dem ursprünglichen Besitzer zurückgibt. Wird zur Zeit t_0 , $0 < t_0 < t$, eine Dividende D je Aktie ausgeschüttet, so muss der Short seller dem Aktienbesitzer $1000 \cdot D$ zahlen. Short selling wirft nur dann Gewinn ab, wenn S_t deutlich unterhalb von S_0 liegt. Außerdem ist das Short selling in der Praxis zahlreichen Restriktionen unterworfen. Für das folgende ist aber nur wichtig, dass im Prinzip die Möglichkeit besteht, Aktien und andere Finanzobjekte zu verkaufen, die man gar nicht besitzt.

2. Grundlagen des Optionsmanagements

2.1. Arbitragebeziehungen

In diesem Abschnitt betrachten wir den wichtigen Begriff der Arbitragefreiheit. Unter *Arbitrage* verstehen wir einen risikolosen Gewinn. In einem idealen Finanzmarkt, in dem unter anderem alle Investoren denselben Informationsstand haben und verzögerungsfrei handeln können, sollte es keine Möglichkeiten zur Arbitrage geben. Jeder Investor würde andernfalls versuchen, einen risikolosen Gewinn augenblicklich mitzunehmen. Die dadurch ausgelösten Transaktionen würden ebenfalls augenblicklich die Preise der involvierten Finanzinstrumente so ändern, dass die Arbitragemöglichkeit sofort verschwindet.

Zusätzlich zur Arbitragefreiheit fordern wir in diesem Kapitel, dass der Finanzmarkt einigen weiteren idealisierenden Annahmen genügt, die allerdings in diesem Zusammenhang einen geringeren Stellenwert haben und nur die Argumentation wesentlich vereinfachen. Sind diese Annahmen erfüllt, sprechen wir von einem perfekt funktionierenden Finanzmarkt.

ANNAHME (*perfekter Finanzmarkt*) :

Es gibt keine Arbitragemöglichkeiten, keine Transaktionskosten, keine Steuern und keine Einschränkungen beim Short-selling. Soll- und Habenzinsen sind gleich, und alle Wertpapiere sind beliebig teilbar.

Die Annahme eines perfekt funktionierenden Finanzmarktes reicht allein schon aus, um den Wert von Terminkontrakten und wichtige Beziehungen zwischen den Preisen verschiedener Optionen herzuleiten. Vor allem ist es nicht nötig, ein mathematisches Modell für den Kurs des Finanzinstrumentes, das Gegenstand der Optionen oder Terminkontrakte ist, zu formulieren. Um den Wert von Optionen zu bestimmen, genügen rein ökonomische Annahmen allerdings nicht mehr, sondern es ist eine detailliertere mathematische Modellierung nötig. Jeder mathematische Ansatz muss sich allerdings an den in diesem Kapitel hergeleiteten Arbitragebeziehungen messen lassen: wenn das Modell zu anderen Ergebnissen bei der Bewertung von Terminkontrakten führt und die aus ihm abgeleiteten Optionspreise nicht den folgenden Relationen genügen, dann beruht es auf falschen Voraussetzungen.

Eine wichtige Folgerung aus der Annahme eines perfekten Marktes und dabei vor allem aus der Arbitragefreiheit, die wir in den folgenden Beweisen immer wieder benutzen, ist die Tatsache, dass zwei Portfolios, die zu einem bestimmten Zeitpunkt T denselben Wert haben, auch zu jedem früheren Zeitpunkt $t < T$ wertgleich sind. Wegen ihrer Bedeutung stellen wir diese Argumentation hier in ihren Einzelheiten vor. Wir gehen von zwei Portfolios A und B , die sich aus beliebigen Anlageformen zusammensetzen, aus. Ihr Wert zur

Zeit t bezeichnen wir mit $W_A(t)$ bzw. $W_B(t)$. Wir nehmen für einen festen Zeitpunkt T an, dass sicher $W_A(T) = W_B(T)$ unabhängig von der Wertentwicklung der einzelnen, in A und B enthaltenen Anlagen. Für einen beliebigen früheren Zeitpunkt $t < T$ nehmen wir ohne Beschränkung der Allgemeinheit an, dass $W_A(t) \leq W_B(t)$. Ein Investor kann zur Zeit t ohne Eigenkapitaleinsatz ein kombiniertes Portfolio aufbauen, indem er eine Einheit aller Anlagen in A kauft, eine Einheit aller Anlagen in B verkauft (Short-selling) und den Differenzbetrag $\Delta(t) = W_B(t) - W_A(t) \geq 0$ festverzinslich zum Zinssatz r anlegt. Das kombinierte Portfolio hat zur Zeit t den Wert

$$W_A(t) - W_B(t) + \Delta(t) = 0,$$

kostet den Investor also nichts. Zur Zeit T hat der festverzinsliche Anteil des kombinierten Portfolios den aufgezinsten Wert $\Delta(T) = \Delta(t)e^{r(T-t)}$, das Portfolio als Ganzes also sicher den Wert

$$W_A(T) - W_B(T) + \Delta(T) = \Delta(t)e^{r(T-t)} > 0,$$

falls $\Delta(t) > 0$. Der Investor hat mit dem kombinierten Portfolio also einen sicheren Gewinn gemacht, ohne ein Risiko einzugehen. Dies ist ein Widerspruch zur angenommenen Arbitragefreiheit, und daher muss $\Delta(t) = 0$ gelten, d.h. $W_A(t) = W_B(t)$.

Diese Argumentation kann nun ausgenutzt werden, um den unbekanntem Wert eines Finanzderivats zu bestimmen. Hierzu wird ein Portfolio A zusammengestellt, das neben Anlagen mit bekanntem Preis auch eine Einheit des Derivats enthält. Diesem Portfolio wird ein zweites, sogenanntes *Duplikationsportfolio* B gegenübergestellt, das ausschließlich aus Anlagen mit bekanntem Preis besteht und so konstruiert worden ist, dass es zu einem festen Zeitpunkt T sicher denselben Wert wie A hat. Dann folgt aus der Arbitragefreiheit, dass beide Portfolios zu jedem früheren Zeitpunkt wertgleich sind, und daraus ergibt sich sofort der Wert des Finanzderivats zu jedem Zeitpunkt $t \leq T$. Wir illustrieren diese Vorgehensweise am Beispiel eines Terminkontrakts.

SATZ 2.1

Wir betrachten den Terminkauf eines Objekts, das zur Zeit t den Kurs S_t hat. K sei der Terminkurs und T der Zeitpunkt der Fälligkeit. $V(s, \tau)$ bezeichne den Wert des Terminkaufs zur Zeit t als Funktion des aktuellen Kurses $S_t = s$ und der verbleibenden Laufzeit $\tau = T - t$. Wir nehmen an, dass die Zinsen während der Laufzeit gleich einem konstanten Wert r sind.

1. *Wirft das zugrundeliegende Objekt während der Restlaufzeit τ keine Erträge ab und verursacht es auch keine Kosten, so gilt*

$$V(S_t, \tau) = V_{K,T}(S_t, \tau) = S_t - Ke^{-r\tau} \quad (2.1)$$

Der Forward-Preis ist in diesem Fall $F_t = S_t e^{r\tau}$.

2. *Fallen während der Restlaufzeit τ durch das Halten des zugrundeliegenden Objektes zu diskreten Zeitpunkten Erträge oder Kosten an, deren auf den aktuellen Zeitpunkt t abgezinster Gesamtwert D_t ist, so gilt*

$$V(S_t, \tau) = V_{K,T}(S_t, \tau) = S_t - D_t - Ke^{-r\tau} \quad (2.2)$$

Der Forward-Preis ist in diesem Fall $F_t = (S_t - D_t)e^{r\tau}$.

3. Sind mit dem Objekt stetige Bestandshaltkosten mit einer Rate b verbunden, so gilt

$$V(S_t, \tau) = V_{K,T}(S_t, \tau) = S_t e^{(b-r)\tau} - K e^{-r\tau} \quad (2.3)$$

Der Forward-Preis ist in diesem Fall $F_t = S_t e^{b\tau}$.

BEWEIS:

Wir nennen das zugrundeliegende Objekt der Einfachheit halber eine Aktie mit diskreten, auf den Zeitpunkt t abgezinsten Dividenden D_t bzw. mit stetigem Dividendenertrag d . Im letzteren Fall sind haben die stetigen Bestandshaltkosten der Aktie die Rate $b = r - d$: der Investor gewinnt durch Halten der Aktie Dividenden (als negative Kosten) mit der Rate d , verliert aber gleichzeitig Zinsen mit der Rate r , da er sein Kapital in die Aktie anstelle einer festverzinslichen Anleihe gesteckt hat. Statt Aktien können genausogut Anleihen, Devisen oder andere einfache Anlageformen betrachtet werden.

1. Wir betrachten zum Zeitpunkt t zwei Portfolios A und B , die folgendes Aussehen haben:

Portfolio A: Terminkauf der Aktie zum Terminkurs K , fällig zum Zeitpunkt T .

Kauf eines Zerobonds mit Nominalwert K , fällig zum Zeitpunkt T .

Portfolio B: Kauf einer Aktie

Am Ende der Laufzeit zur Zeit T wird in Portfolio A der Zerobond verkauft. Der Erlös K wird benutzt, um die mit dem Terminkauf eingegangene Verpflichtung zu erfüllen und eine Aktie zum Preis K zu kaufen. Danach enthält Portfolio A eine Aktie und hat somit zur Zeit T denselben Wert wie Portfolio B . Daher gilt wegen der vorausgesetzten Arbitragefreiheit zum Zeitpunkt t ebenfalls Gleichheit, also

$$V(S_t, \tau) + K e^{-r\tau} = S_t, \quad (2.4)$$

da sich durch Abzinsen als Wert des Zerobonds zur Zeit t gerade $K e^{-r\tau}$ ergibt. Der Forward-Preis ist nach Definition die Lösung von

$$0 = V_{F_t,T}(S_t, \tau) = S_t - F_t e^{-r\tau}.$$

2. Wir betrachten zur Zeit t die beiden Portfolios A und B aus dem ersten Teil des Beweises, ergänzen B aber jetzt um eine Position zu:

Portfolio B: Kauf einer Aktie und leihen eines Geldbetrages in Höhe D_t zum Zinssatz r (= Verkauf eines Zerobonds).

Zur Zeit T werden in Portfolio B die Dividendenerträge der Aktie - im auf die Zeit T abgezinsten Wert von $D_t e^{r\tau}$ - genutzt, um die Anleihe zurückzuzahlen. Damit besteht Portfolio B wie Portfolio A zur Zeit T wieder nur aus einer Aktie, und es ergibt sich wieder Wertgleichheit auch zum Zeitpunkt t :

$$V(S_t, \tau) + K e^{-r\tau} = S_t - D_t. \quad (2.5)$$

Der Forward-Preis ergibt sich dann wie in Teil 1. aus der Definition.

3. Besitzt die Aktie eine stetige Dividendenrendite d , kann ähnlich argumentiert werden. Wieder betrachten wir zwei Portfolios A und B zum Zeitpunkt t , wobei A wie im ersten Teil des Beweises gewählt wird. Portfolio B hat folgendes Aussehen

Portfolio B : Kauf von $e^{-d\tau}$ Aktien .

Wird der Dividendenenertrag in Portfolio B laufend direkt wieder in die Aktie reinvestiert, besteht Portfolio B auch in diesem Fall zum Zeitpunkt T aus genau einer Aktie. Heuristisch sieht man das folgendermaßen: Im Zeitintervall $[t, t + \delta]$ wirft für kleines δ eine Aktie eine Dividende von approximativ $d \cdot \delta \cdot S_t$ ab, der aktuelle Bestand von $e^{-d\tau} = e^{-d(T-t)}$ Aktien also eine Dividende von $d \cdot \delta \cdot S_t \cdot e^{-d(T-t)}$. Dieser Betrag wird in Aktien reinvestiert. Unter der Annahme, dass sich der Kurs in dem kleinen Zeitintervall $[t, t + \delta]$ nicht wesentlich ändert, d.h. $S_{t+\delta} \approx S_t$, enthält Portfolio B zur Zeit $t + \delta$ insgesamt

$$(1 + d \cdot \delta) \cdot e^{-d(T-t)} \approx e^{d\delta} \cdot e^{-d(T-t)} = e^{-d(T-t-\delta)}$$

Aktien. Diese Argumentation kann man mit dem Grenzübergang $\delta \rightarrow 0$ exakt machen und zeigen, dass Portfolio B zu jedem Zeitpunkt s zwischen t und T genau $e^{-d(T-s)}$, für $s = T$ also eine Aktie enthält. Die gleiche Argumentation wie im ersten Teil des Beweises liefert dann die Wertgleichheit der beiden Portfolios A und B zum Zeitpunkt t , also

$$V(S_t, \tau) + Ke^{-r\tau} = e^{-d\tau} S_t . \quad (2.6)$$

Mit $b = r - d$ erhalten wir die gewünschte Form. Der Forward-Preis ergibt sich dann wie in Teil 1. aus der Definition. \square

BEISPIEL 2.1

- *Wir betrachten den Terminkauf einer 5-Jahres Anleihe, die zum Kurs 900 EUR gehandelt wird. Der Terminkurs betrage 910 EUR, die Laufzeit des Terminkontraktes ein Jahr. Kuponzahlungen in Höhe von 60 EUR fallen in 6 bzw. 12 Monaten (letztere kurz vor Fälligkeit des Kontraktes) an. Der stetige Jahreszins für 6 bzw. 12 Monate betrage 9% bzw. 10%. In diesem Fall ist*

$$S_t = 900 , K = 910 , r = 0.10 , \tau = 1 , D_t = 60e^{-0.09 \cdot \frac{1}{2}} + 60e^{-0.10} = 111.65 \quad (2.7)$$

Der Wert des Terminkaufs ist dann

$$V(S_t, \tau) = 900 - 111.65 - 910e^{-0.10} = -35.05. \quad (2.8)$$

Der Wert der entsprechenden Short-Position ist dann +35.05. Der Forward-Preis F_t beträgt $F_t = (S_t - D_t)e^{r\tau} = 871.26$.

- Wir betrachten den Terminkauf von 1000 Dollar. In diesem Fall liegt ein stetiger Dividendenenertrag d in Höhe des amerikanischen Zinssatzes vor, wenn der Investor den Dollarbetrag erwirbt und in den USA anlegt. r sei der inländische Zinssatz. Die Bestandshaltkosten $b = r - d$ sind die Differenz zwischen Inlands- und Auslandszinssatz. Bezeichnet S_t den Dollarkurs, so ist der Forward-Preis dann

$$F_t = S_t e^{b\tau} = S_t e^{(r-d)\tau}. \quad (2.9)$$

Für $r > d$ ergibt sich ein Report $S_t < F_t$ (Zinsaufschlag), für $r < d$ ergibt sich ein Deport $S_t > F_t$ (Zinsabschlag). Ist $r > d$ und wählt man den Terminkurs gleich dem aktuellen Kurs, d.h. $K = S_t$, so ist der Wert des Terminkaufs

$$V_{S_t, T}(S_t, \tau) = S_t(e^{-d\tau} - e^{-r\tau}) > 0.$$

Der Terminkauf zum Preis S_t ist also teurer als der sofortige Kauf zum selben Preis, denn er ermöglicht dem Investor, noch bis zur Zeit T den aktuell höheren Inlandszinssatz auf seine heimische Währung zu nutzen.

Das folgende Resultat besagt, dass es keinen Unterschied zwischen Forward- und Future-Kontrakten mit gleichem Terminkurs und gleicher Laufzeit gibt, wenn der Zinssatz sich über die ganze Laufzeit hinweg nicht ändert. Wir benutzen dabei, dass nach Definition von Forward- und Future-Preis ein Terminkauf nichts kostet, wenn der Terminkurs gleich dem aktuellen Forward- bzw. Future-Preis gewählt wird.

SATZ 2.2

Ist der Zinssatz während der Laufzeit konstant, so stimmen Future- und Forward-Preis miteinander überein.

BEWEIS:

Wir gehen davon aus, dass der Future-Kontrakt am Tag 0 geschlossen wird und eine Gesamtlaufzeit von N Tagen hat. Wir nehmen an, dass der Gewinn-Verlust-Ausgleich täglich bei einem Tageszinssatz von ρ erfolgt. F bezeichne den Forward-Preis bei Kontraktbeginn, d.h. am Ende des 0-ten Tages, und F_t sei der Future-Preis am Ende des t -ten Tages, $t = 0, 1, \dots, N$. Unser Ziel ist es, $F = F_0$ zu zeigen. Dazu stellen wir wieder zwei Portfolios zusammen:

Portfolio A: Kauf von $e^{N\rho}$ Forward-Kontrakten mit Terminkurs F , Laufzeit N .
Kauf eines Zerobonds mit Nominalwert $F e^{N\rho}$ und Fälligkeit in N Tagen.

Portfolio B: Täglicher Zukauf von Future-Kontrakten mit jeweiligem Terminkurs F_t und Fälligkeit zur Zeit N derart, dass am Ende des t -ten Tages genau $e^{(t+1)\rho}$ Future-Kontrakte im Portfolio enthalten sind ($t = 0, 1, \dots, N$).
Kauf eines Zerobonds mit Nominalwert $F_0 e^{N\rho}$ und Fälligkeit in N Tagen.

Der Erwerb der Forward- und Future-Kontrakte ist kostenlos, da er jeweils mit dem Forward- bzw. dem tagesaktuellen Future-Preis als Terminkurs erfolgt. Der Halter von Portfolio B

erhält von Tag $t - 1$ auf den Tag t durch den Gewinn-Verlust-Ausgleich $F_t - F_{t-1}$ je Future-Kontrakt im Portfolio. Dies kann negativ sein, d.h. er selbst muss zahlen.

Am Ende der Laufzeit, d.h. des N -ten Tages, wird in Portfolio A der Ertrag $F e^{N\rho}$ des Zerobonds benutzt, um den Terminkontrakt zu erfüllen und $e^{N\rho}$ Aktien zum Preis von jeweils F zu kaufen. Das Portfolio A enthält dann nur noch diese Aktien und hat den Wert $S_N e^{N\rho}$. Wir zeigen nun, dass dies auch für Portfolio B gilt.

Zu Beginn des Tages t sind $e^{t\rho}$ Futures im Portfolio, und der Halter erhält durch den Gewinn-Verlust-Ausgleich den Betrag $(F_t - F_{t-1})e^{t\rho}$, der auch negativ sein kann. Während des Tages stockt er kostenfrei, da mit dem aktuellen Future-Preis als Terminkurs, sein Portfolio um weitere Futures auf, so dass es am Tagesende insgesamt $e^{(t+1)\rho}$ Futures enthält. Der am Tag t eingenommene Betrag hat auf den Fälligkeitstermin aufgezinst den Wert

$$(F_t - F_{t-1})e^{t\rho} \cdot e^{(N-t)\rho} = (F_t - F_{t-1})e^{N\rho}. \quad (2.10)$$

Am Ende der Laufzeit ist der Future-Kontrakt durch den laufenden Gewinn-Verlust-Ausgleich bereits erfüllt und verursacht keine weiteren Kosten. Die gesamten während der Laufzeit angefallenen Gewinne oder Kosten sind aufgezinst auf den Tag N :

$$\sum_{t=1}^N (F_t - F_{t-1})e^{N\rho} = (F_N - F_0)e^{N\rho}. \quad (2.11)$$

Zusammen mit dem Zerobond hat Portfolio B daher zur Zeit N den Wert

$$(F_N - F_0)e^{N\rho} + F_0 e^{N\rho} = F_N e^{N\rho} = S_N e^{N\rho},$$

da am Ende der Laufzeit Future-Preis F_N und aktueller Kurs S_N offensichtlich zusammenfallen.

Beide Portfolios haben also am Tag N denselben Wert und somit wegen der Arbitragefreiheit auch am Tag 0. Da der Forward-Kontrakt mit Terminkurs F am Tag 0 wegen der Definition des Forward-Preises den Wert 0 hat, ist der Wert von Portfolio A gleich dem Wert des Zerobonds, d.h. F (Nominalwert $F e^{N\rho}$ abgezinst auf den Tag 0). Entsprechend haben auch die am Ende von Tag 0 in Portfolio B enthaltenen e^ρ Futures wegen der Definition des Future-Preises den Wert 0, so dass auch hier nur der Zerobond zum Portfoliowert beiträgt. Er hat entsprechend den Wert F_0 ($F_0 e^{N\rho}$ abgezinst auf den Tag 0). Insgesamt schließen wir $F = F_0$.

□

Wir wollen nun mit ähnlichen Methoden Beziehungen zwischen den Preisen für Optionen herleiten. Die elementarsten Eigenschaften fassen wir in der folgenden Bemerkung ohne Beweis zusammen. Dazu benötigen wir den Begriff des inneren Wertes (*intrinsic value*).

DEFINITION 2.1 (Innerer Wert)

Der innere Wert einer Kaufoption zum Zeitpunkt t ist gegeben durch $\max(S_t - K, 0)$, der einer Verkaufsoption durch $\max(K - S_t, 0)$. Ist der innere Wert einer Option positiv, so sagen wir, dass sie im Geld ist. Ist $S_t = K$, so ist die Option am Geld; andernfalls ist die Option aus dem Geld.

BEMERKUNG 2.1

Für Optionen gelten folgende elementare Beziehungen. $C(s, \tau) = C_{K,T}(s, \tau)$ und $P(s, \tau) = P_{K,T}(s, \tau)$ bezeichnen dabei den Wert zur Zeit t eines Calls bzw. Puts mit Ausübungskurs K und Verfallszeitpunkt T , wenn $\tau = T - t$ die Restlaufzeit ist und der Kurs des zugrundeliegenden Finanzinstruments den Wert s hat: $S_t = s$.

1. Optionspreise sind nicht-negativ, da eine Ausübung nur stattfindet, wenn es im Interesse des Optionshalters liegt. Eine Option gewährt Rechte ohne Pflichten.
2. Zum Verfallszeitpunkt T besitzen amerikanische und europäische Optionen den gleichen Wert, da sie dann dem Besitzer identische Rechte einräumen. Zum Verfallszeitpunkt T stimmen der Wert der Option und der innere Wert überein:

$$C_{K,T}(S_T, 0) = \max(S_T - K, 0), \quad P_{K,T}(S_T, 0) = \max(K - S_T, 0).$$

3. Eine amerikanische Option muss mindestens zu ihrem inneren Wert gehandelt werden, da andernfalls durch Kauf einer Option und augenblickliche Ausübung Arbitrage möglich wäre. Diese Relation gilt für europäische Optionen im allgemeinen nicht. Grund hierfür ist, dass eine europäische Option nur indirekt über ein Termingeschäft zum heutigen Zeitpunkt vorzeitig ausgeübt werden kann. Der hierbei auftretende Abzinsungsfaktor kann den Wert unter den inneren Wert der Option drücken.
4. Der Wert zweier amerikanischer Optionen, die sich nur in den Laufzeiten unterscheiden (Verfallszeitpunkte $T_1 \leq T_2$), ist monoton in den Laufzeiten:

$$C_{K,T_1}(s, T_1 - t) \leq C_{K,T_2}(s, T_2 - t), \quad P_{K,T_1}(s, T_1 - t) \leq P_{K,T_2}(s, T_2 - t).$$

Dies folgt - zum Beispiel für Calls - unter Ausnutzung von 2. und 3. aus der zur Zeit $t = T_1$ gültigen Ungleichung mit $s = S_{T_1}$

$$C_{K,T_2}(s, T_2 - T_1) \geq \text{innerer Wert} = \max(s - K, 0) = C_{K,T_1}(s, 0) \quad (2.12)$$

Wegen der Arbitragefreiheit muss die Ungleichung auch zu jedem Zeitpunkt $t \leq T_1$ gelten. Für europäische Optionen ist diese Aussage im allgemeinen nicht erfüllt.

5. Eine amerikanische Option hat mindestens den gleichen Wert wie die ansonsten identische europäische Option, da die amerikanische Option dem Besitzer mehr Rechte einräumt.
6. Der Wert eines Calls ist als Funktion des Ausübungskurses monoton fallend, da das Kaufrecht umso wertvoller ist, je niedriger der vereinbarte Kaufpreis. Entsprechend ist der Wert eines Puts als Funktion des Ausübungskurses monoton wachsend.

$$C_{K_1,T}(s, \tau) \geq C_{K_2,T}(s, \tau), \quad P_{K_1,T}(s, \tau) \leq P_{K_2,T}(s, \tau)$$

für $K_1 \leq K_2$. Dies gilt für amerikanische und europäische Optionen.

Der Wert einer europäischen Kauf- und Verkaufsoption auf das gleiche Objekt mit gleicher Laufzeit und gleichem Ausübungskurs sind direkt miteinander verbunden, ohne dass die Annahme eines detaillierten mathematischen Modells nötig ist.

SATZ 2.3 (Put-Call-Parität für europäische Optionen)

Für den Wert einer europäischen Calls und eines europäischen Puts mit gleichem Verfallszeitpunkt T , gleichem Ausübungskurs K auf dasselbe zugrundeliegende Finanzinstrument gelten folgende Aussagen, wobei r den stetigen Zinssatz bezeichnet:

1. Fallen während der Restlaufzeit $\tau = T - t$ der Optionen Erträge mit dem auf den Zeitpunkt t abgezinsten Gesamtbarwert D_t an, so ist

$$C(S_t, \tau) = P(S_t, \tau) + S_t - D_t - Ke^{-r\tau} \tag{2.13}$$

 [SFMPutCall.xpl](#)

2. Fallen während der Restlaufzeit $\tau = T - t$ der Optionen stetige Bestandshaltkosten mit Rate b auf das Objekt an, so ist

$$C(S_t, \tau) = P(S_t, \tau) + S_t e^{(b-r)\tau} - Ke^{-r\tau} \tag{2.14}$$

BEWEIS:

Zur Vereinfachung der Sprechweise nehmen wir wieder speziell an, dass das Objekt eine Aktie ist. Wir betrachten ein Portfolio A , das nur aus einem Call besteht, und duplizieren es durch ein geeignetes Portfolio B , das unter anderem einen Put enthält.

1. Im Fall diskreter Dividenden betrachten wir folgendes Portfolio B zum Zeitpunkt t :

1. Kaufe den Put.
2. Verkaufe einen Zerobond mit Nominalwert K , fällig zum Zeitpunkt T .
3. Kaufe eine Aktie.
4. Verkaufe einen Zerobond zum (aktuellen) Wert D_t .

Da das Portfolio eine Aktie enthält, wirft es Dividenden ab, deren Wert abgezinst auf den Zeitpunkt t gerade D_t ist. Diese Erträge werden am Ende der Laufzeit benutzt, um die Anleihe in Position d) zurückzuzahlen, so dass der Wert dieser Position zu 0 wird. Tabelle 2.1 zeigt nun den Wert des Portfolios B zum Zeitpunkt T abhängig davon, ob der Put ausgeübt wird ($K \geq S_T$) oder nicht. Zur Zeit T hat Portfolio B damit denselben Wert $\max(S_T - K, 0)$ wie der Call. Um Arbitragemöglichkeiten zu verhindern, müssen daher beide Portfolios auch zum Zeitpunkt t denselben Wert haben, d.h. es gilt

$$C(S_t, \tau) = P(S_t, \tau) - Ke^{-r\tau} + S_t - D_t \tag{2.15}$$

2. Für den Fall einer Aktie mit stetiger Dividendenrendite d und entsprechenden Bestandshaltkosten $b = r - d$ betrachten wir dasselbe Portfolio B wie in Teil 1., jedoch ohne

Position	Wert zum Zeitpunkt T	
	$K < S_T$	$K \geq S_T$
a)	0	$K - S_T$
b)	$-K$	$-K$
c)	S_T	S_T
d)	0	0
Summe	$S_T - K$	0

Tabelle 2.1.: Wert von Portfolio B zum Zeitpunkt T (Satz 2.3)

die Position d). Stattdessen werden in Position c) $e^{-d\tau}$ Aktien anstelle von einer Aktie gekauft, deren Dividenden direkt wieder in die gleiche Aktie reinvestiert werden. Bei negativem d werden die auftretenden Kosten durch Verkäufe des Objektes finanziert. Hierdurch enthält Portfolio B zum Zeitpunkt T genau eine Aktie, und man kann wie in Teil 1. schließen, dass der Wert des Portfolios B zum Zeitpunkt t gleich dem Wert des Calls ist.

□

Der Beweis der Put-Call-Parität funktioniert nur für europäische Optionen. Für amerikanische Optionen kann es vorkommen, dass Put oder Call vorzeitig ausgeübt wird und die beiden Portfolios gar nicht bis zum Ende der Laufzeit gehalten werden.

Das folgende Resultat bietet eine Möglichkeit zur Überprüfung der Konsistenz von Optionspreisen auf dasselbe Objekt. Ist die hier formulierte Konvexität verletzt, ist Arbitrage möglich, wie wir in dem sich anschließenden Beispiel zeigen.

SATZ 2.4

Der Preis einer (amerikanischen oder europäischen) Option ist als Funktion des Ausübungskurses konvex.

BEWEIS:

Es genügt Calls zu betrachten. Für Puts läuft der Beweis analog; für europäische Optionen folgt dies direkt aus der Put-Call-Parität, da dort der explizit von K abhängige Term linear in K ist.

Für $0 \leq \lambda \leq 1$ und $K_1 < K_0$ definieren wir $K_\lambda := \lambda K_1 + (1 - \lambda)K_0$. Wir betrachten ein Portfolio A , das zur Zeit $t < T$ aus einem Call mit Ausübungskurs K_λ und Fälligkeitszeitpunkt T besteht. Wir duplizieren dieses Portfolio durch folgendes Portfolio B zum Zeitpunkt t :

1. Kaufe λ Calls mit Ausübungskurs K_1 , Fälligkeit T .
2. Kaufe $(1 - \lambda)$ Calls mit Ausübungskurs K_0 , Fälligkeit T .

Liquidieren wir das Portfolio zu einem beliebigen Zeitpunkt $t', t \leq t' \leq T$, so lässt sich die Differenz der Werte der Portfolios B und A zu diesem Zeitpunkt in der Tabelle 2.2 ablesen.

Position	Wert zum Zeitpunkt t'			
	$S_{t'} \leq K_1$	$K_1 < S_{t'} \leq K_\lambda$	$K_\lambda < S_{t'} \leq K_0$	$K_0 < S_{t'}$
B 1.	0	$\lambda(S_{t'} - K_1)$	$\lambda(S_{t'} - K_1)$	$\lambda(S_{t'} - K_1)$
B 2.	0	0	0	$(1 - \lambda)(S_{t'} - K_0)$
$-A$	0	0	$-(S_{t'} - K_\lambda)$	$-(S_{t'} - K_\lambda)$
Summe	0	$\lambda(S_{t'} - K_1)$	$(1 - \lambda)(K_0 - S_{t'})$	0

Tabelle 2.2.: Wertdifferenz der Portfolios B und A zum Zeitpunkt t' (Satz 2.4)

Ausübungskurs	Optionspreis
$K_1 = 190$	30.6 EUR
$K_\lambda = 200$	26.0 EUR
$K_0 = 220$	14.4 EUR

Tabelle 2.3.: Daten des Beispiels 2.2

Da in der letzten Zeile der Tabelle jeweils $\lambda(S_{t'} - K_1) \geq 0$ und $(1 - \lambda)(K_0 - S_{t'}) \geq 0$ gilt, ist die Differenz der Werte von Portfolio B und Portfolio A zum Zeitpunkt t' und damit wegen der Arbitragefreiheit auch zum Zeitpunkt $t < t'$ größer gleich Null. Es gilt also mit $\tau = T - t$

$$\lambda C_{K_1, T}(S_t, \tau) + (1 - \lambda) C_{K_0, T}(S_t, \tau) - C_{K_\lambda, T}(S_t, \tau) \geq 0 \quad (2.16)$$

□

BEISPIEL 2.2

Wir betrachten drei europäische Kaufoptionen auf die MD*TECH A.G. mit gleicher Laufzeit und den Ausübungskursen $K_1 = 190$, $K_\lambda = 200$, $K_0 = 220$, d.h. $\lambda = \frac{2}{3}$. Die Daten des Beispiels sind in der Tabelle 2.3 gegeben.

Nach dem letzten Satz muss gelten:

$$\frac{2}{3} C_{K_1, T}(S_t, \tau) + \frac{1}{3} C_{K_0, T}(S_t, \tau) \geq C_{K_\lambda, T}(S_t, \tau) \quad (2.17)$$

Diese Bedingung ist offensichtlich verletzt, und folgendes Portfolio ermöglicht Arbitrage:

1. Kaufe $\lambda = \frac{2}{3}$ Calls mit Ausübungskurs K_1 .
2. Kaufe $1 - \lambda = \frac{1}{3}$ Calls mit Ausübungskurs K_0 .
3. Verkaufe 1 Call mit Ausübungskurs $K_\lambda := \frac{2}{3}K_1 + \frac{1}{3}K_0$.

Zum jetzigen Zeitpunkt t liefert dieses Portfolio den sicheren Ertrag von +0.80 EUR. Der Wert des Portfolios zum Verfallszeitpunkt T der Optionen ist in der Tabelle 2.4 dargestellt. Das Portfolio liefert für Aktienkurse S_T zwischen 190 und 220 zusätzlich noch einen Ertrag von maximal $\frac{20}{3}$ EUR.

Position	Wert zum Zeitpunkt T			
	$S_T \leq 190$	$190 < S_T \leq 200$	$200 < S_T \leq 220$	$220 < S_T$
1.	0	$\frac{2}{3}(S_T - 190)$	$\frac{2}{3}(S_T - 190)$	$\frac{2}{3}(S_T - 190)$
2.	0	0	0	$\frac{1}{3}(S_T - 220)$
3.	0	0	$-(S_T - 200)$	$-(S_T - 200)$
Summe	0	$\frac{2}{3}(S_T - 190)$	$\frac{1}{3}(220 - S_T)$	0

Tabelle 2.4.: Wert des Portfolios zum Zeitpunkt T im Beispiel 2.2

Wir haben bereits angemerkt, dass Optionen monotone Funktionen des Ausübungskurses sind. Der folgende Satz liefert für europäische Optionen eine schärfere Aussage.

SATZ 2.5

Für zwei europäische Calls (bzw. Puts) mit gleichem Verfallsdatum T und den Ausübungskursen $K_1 \leq K_2$ gilt zur Zeit $t \leq T$:

$$0 \leq C_{K_1, T}(S_t, \tau) - C_{K_2, T}(S_t, \tau) \leq (K_2 - K_1)e^{-r\tau} \quad (2.18)$$

mit Restlaufzeit $\tau = T - t$ und Zinssatz r bzw.

$$0 \leq P_{K_2, T}(S_t, \tau) - P_{K_1, T}(S_t, \tau) \leq (K_2 - K_1)e^{-r\tau} \quad (2.19)$$

Sind die Call- bzw. Putwerte als Funktion des Ausübungskurses differenzierbar, folgt durch Grenzübergang $K_2 - K_1 \rightarrow 0$ auch

$$-1 \leq -e^{-r\tau} \leq \frac{\partial C}{\partial K} \leq 0 \quad \text{bzw.} \quad 0 \leq \frac{\partial P}{\partial K} \leq e^{-r\tau} \leq 1 \quad (2.20)$$

BEWEIS:

Wir führen den Beweis für Calls; für Puts ist die Argumentation analog. Hierzu betrachten wir ein Portfolio A , das nur einen Call mit Ausübungskurs K_1 enthält, und vergleichen es mit einem Duplikationsportfolio B , das zum Zeitpunkt t aus den beiden folgenden Positionen besteht:

1. Kaufe einen Call mit Ausübungskurs K_2 .
2. Kaufe einen Zerobond mit Nominalwert $(K_2 - K_1)$, fällig zum Zeitpunkt T .

Die Differenz der Werte der Portfolios B und A zum Zeitpunkt T lässt sich in der Tabelle 2.5 ablesen. Der Wert von Portfolio B zum Zeitpunkt T ist offensichtlich mindestens so groß wie der Wert von Portfolio A . Folglich muss dies wegen der Arbitragefreiheit auch für den Zeitpunkt t gelten und wir schließen:

$$C_{K_2, T}(S_t, \tau) + (K_2 - K_1)e^{-r\tau} \geq C_{K_1, T}(S_t, \tau).$$

□

Position	Wert zum Zeitpunkt T		
	$S_T \leq K_1$	$K < S_T < K_2$	$K_2 \leq S_T$
B 1.	0	0	$S_T - K_2$
B 2.	$K_2 - K_1$	$K_2 - K_1$	$K_2 - K_1$
$-A$	0	$-(S_T - K_1)$	$-(S_T - K_1)$
Summe	$K_2 - K_1$	$K_2 - S_T$	0

Tabelle 2.5.: Wertdifferenz der Portfolios B und A zum Zeitpunkt T (Satz 2.5)

2.2. Portfeuille-Versicherung

Ein wichtiger Verwendungszweck von Optionen ist das Hedgen, d.h. der Schutz von Investitionen gegen das Marktrisiko, das durch die zufälligen Kursentwicklungen hervorgerufen wird. Ein Beispiel für aktives Hedgen mit Optionen ist die Portfeuille-Versicherung (Portfolio Insurance). Darunter verstehen wir den Abschluss von Geschäften, welche zu einem bestimmten Zeitpunkt die Risikostruktur eines Portfolios so verändern, dass zu einem im voraus bestimmten Termin in der Zukunft

- die positiven Erträge durchgehend um einen kleinen Betrag gemindert werden (was als Zahlung einer Versicherungsprämie interpretiert werden kann) und dadurch
- ein Absinken des Werts des Portfolios unter ein bestimmtes Niveau, den *Floor*, verhindert wird.

Durch die Portfeuille-Versicherung wird eine Risikostruktur des Portfolios erzeugt, bei der extreme Verluste ausgeschlossen sind. Wir betrachten hierzu zuerst ein einfaches Beispiel.

BEISPIEL 2.3

Ein Kapitalbetrag von 10500 EUR steht zum Kauf von Aktien zur Verfügung, deren aktueller Kurs 100 EUR ist. Auf dem Markt erhältlich sind auch zu einem Stückpreis von 5 EUR Verkaufsoptionen auf die Aktie mit Ausübungskurs $K = 100$ und einer Laufzeit von einem Jahr. Wir betrachten zwei alternative Investitionsmöglichkeiten.

Portfolio A: *Kauf von 105 Aktien.*

Portfolio B: *Kauf von 100 Aktien für 10000 EUR und Kauf von 100 Verkaufsoptionen für 500 EUR.*

Der Preis der Verkaufsoptionen stellt in diesem Fall die Versicherungsprämie für die Aktien auf einen Versicherungsbetrag von 10000 EUR dar. Bezeichnet S_T den Aktienkurs nach einem Jahr, so ist der Wert des unversicherten Portfolios A gerade $105 \cdot S_T$ und leidet unter dem vollen Marktrisiko, dass der Kurs deutlich unter den Ausgangswert von 100 EUR fällt. Der Wert des versicherten Portfolios B beträgt dagegen nach einem Jahr mindestens 10000 EUR, da der Besitzer den Put ausüben und die 100 Aktien zu einem Stückpreis von 100 EUR verkaufen kann, wenn $S_T < 100$. Steigt der Aktienkurs über 100 EUR, wird der Put nicht ausgeübt und verfällt wertlos. Da bei Portfolio B ein Teil des Kapitals als Versicherungsprämie gegen hohe Verluste in den Put geht, werden die Wahrscheinlichkeiten für Gewinne

Aktienkurs S_T [EUR]	Unversichertes Portfolio		Versichertes Portfolio		Versichertes Portfolio in % des unversicherten Portfolios
	Wert [EUR]	Rendite % p.a.	Wert [EUR]	Rendite % p.a.	
50	5250	-50	10000	-4.8	190
60	6300	-40	10000	-4.8	159
70	7350	-30	10000	-4.8	136
80	8400	-20	10000	-4.8	119
90	9450	-10	10000	-4.8	106
100	10500	0	10000	-4.8	95
110	11550	+10	11000	+4.8	95
120	12600	+20	12000	+14.3	95
130	13650	+30	13000	+23.8	95
140	14700	+40	14000	+33.3	95

Tabelle 2.6.: Die Wirkung einer Portfeuille-Versicherung: Wert- und Renditeentwicklung

mäßig reduziert, wofür die Wahrscheinlichkeit großer Verluste auf 0 sinkt. Der Investor verliert mit Portfolio B maximal den Betrag von 500 EUR, den er für die Optionen bezahlt hat. Die Tabelle 2.6 zeigt die Wert- und Renditeentwicklung der beiden Portfolios nach einem Jahr in Abhängigkeit vom Aktienkurs S_T .

Die Vielzahl möglicher Strategien im Bereich der Portfeuille-Versicherung lassen sich nach der Häufigkeit der zur Risikominderung notwendigen Anpassungsmaßnahmen grob in zwei Gruppen unterteilen in

- statische Strategien, bei denen nach Bildung des Ausgangsportfolios höchstens eine weitere Umschichtung vor Ende des Anlagezeitraums erfolgt
- dynamische Strategien, bei denen laufend, im Idealfall sogar kontinuierlich Anpassungen im Bestand nach einer von vorneherein festgelegten Regel vorgenommen werden.

Die im vorausgehenden Beispiel geschilderte statische Strategie lässt sich neben dem geschilderten Hedgen mit Verkaufsoptionen auch noch auf eine andere Art durchführen, so dass der Anleger die Wahl zwischen zwei Strategien hat:

Strategie 1: Der Anleger kauft *Aktien und Puts* in gleicher Anzahl.

Strategie 2: Der Anleger kauft *Anleihen* im Nominalwert des gewünschten Floors und für den Rest des Geldes *Calls* auf das Objekt.

Alle üblichen Strategien beruhen auf einer Abwandlung dieser beiden Grundstrategien. Während bei der ersten Strategie die Ausübung der Puts bei sinkenden Aktienkursen dafür sorgt, dass das Kapital nicht unter den Floor sinkt, wird dies bei der zweiten Strategie durch die risikolose Anleihe garantiert. Die Aktien bzw. die Calls sorgen für entsprechende Gewinne

bei steigenden Kursen. Die Gleichwertigkeit beider Strategien folgt aus der Put-Call-Parität aus Satz 2.3.

Vor der Entscheidung über die Art einer Portfeuille-Versicherung müssen folgende Fragen geklärt werden:

1. Welche Finanzinstrumente stehen am Markt zur Verfügung, und welche Charakteristika haben sie (Zinssätze, Volatilitäten, Korrelation zum Index u.a.)?
2. Welche Vorstellungen hat der Anleger bezüglich
 - des Aufbaus des Portfolios (gewünschte Finanztitel),
 - des zu investierenden Kapitals,
 - des gewünschten zeitlichen Absicherungshorizonts,
 - des gewünschten Floors (sichere untere Schranke an den Portfoliowert) bzw. der gewünschten Minimalrendite nach Ablauf des Anlagezeitraums. Bei gegebenem Floor F und investiertem Kapital V ist die (negative) Minimalrendite ρ bei einem Zeithorizont von einem Jahr $\rho = \frac{F-V}{V}$.

Die oben beschriebenen Strategien 1 und 2 zur Portfeuille-Versicherung erläutern wir nochmals in einem Beispiel.

BEISPIEL 2.4

Wir gehen davon aus, dass der Anleger sich bereits entschieden hat, in welches Objekt er investieren will, und dass es sich dabei um eine Aktie handelt. Je nach Art des Ertrages, den das Objekt abwirft, unterscheiden wir zwei Fälle (für Kosten, d.h. negative Erträge wie z.B. Lagerhaltungskosten bei Sachwerten, funktioniert das Verfahren analog):

- i) eine stetige Ertragsrendite d
- ii) im voraus bekannte Erträge zu diskreten Zeitpunkten mit dem auf den Zeitpunkt 0 abgezinsten Gesamtbarwert D_0

Die Daten des Beispiels sind in der Tabelle 2.7 gegeben. Die Volatilität ist dabei ein Maß für die Schwankung des Aktienkurses, das wir später noch eingehend betrachten werden und das wesentlich in die Bewertung von Optionen eingeht. Da wir unsere Betrachtungen zum Beginn des Anlagezeitraums zur Zeit $t = 0$ durchführen, ist die Laufzeit $\tau = T - t = T$. Bei beiden Strategien stehen wir vor dem Problem, die Anzahl n der Aktien und/oder der (europäischen) Optionen und den Ausübungskurs K der Optionen zu bestimmen.

Fall i)

Die Aktie liefert eine stetige Dividendenrendite $d = 2\%$ p.a., die wir laufend automatisch wieder in die Aktie investieren. Am Ende der Laufzeit zur Zeit T sind dann aus anfangs n Aktien $ne^{d\tau}$ Aktien geworden, wobei $\tau = T - 0 = T$. Für Strategie 1 müssen wir daher zu Beginn des Anlagezeitraums entsprechend viele Puts kaufen. Da wir zur Zeit 0 gerade ein Kapital vom Umfang V investieren wollen, muss gelten

$$n \cdot S_0 + ne^{d\tau} \cdot P_{K,T}(S_0, \tau) = V \quad (2.21)$$

Die Daten des Beispiels:	
Aktueller Zeitpunkt t	0
Verfügbares Kapital V	100000 EUR
Gewünschter Floor F	95000 EUR
Anlagehorizont T	2 Jahre
Aktueller Aktienkurs S_0	100 EUR
Stetiger Jahreszinssatz r	10% p.a.
Jahresvolatilität σ der Aktie	30%
Dividenden während der Laufzeit	
Fall i): stetige Dividendenrendite d	2% p.a.
Fall ii): Dividenden im Barwert D_0	5 EUR

Tabelle 2.7.: Daten des Beispiels

Den Ausübungskurs K der Verkaufsoptionen wählen wir so, dass der gewünschte Floor F nach zwei Jahren garantiert ist. Dies bedeutet, dass die Ausübung der Puts zum Zeitpunkt T (bei einem Aktienkurs $S_T \leq K$) genau den Floor ergeben muss. Dies liefert die zweite Bedingung

$$ne^{d\tau} \cdot K = F \iff n = \frac{F}{K} e^{-d\tau} \quad (2.22)$$

Einsetzen von Gleichung (2.22) in Gleichung (2.21) liefert

$$e^{-d\tau} S_0 + P_{K,T}(S_0, \tau) - \frac{V}{F} \cdot K = 0 \quad (2.23)$$

Mit der Black-Scholes-Formel für europäische Optionen, die wir später in Abschnitt 6.2 herleiten, erhalten wir den Putpreis als Funktion des noch zu wählenden Parameters K . Der gesuchte Ausübungskurs K kann dann durch numerisches Lösen der Gleichung (2.23) berechnet werden, z.B. mit dem Newtonverfahren. Man erhält in diesem Fall den Wert $K = 99.58$. Um den Floor von 95000 EUR sicher zu erreichen, sind also $n = \frac{F}{K} e^{-d\tau} = 916.6$ Aktien und $n \cdot e^{d\tau} = 954$ Puts mit dem Ausübungskurs $K = 99.58$ zu kaufen. Der Preis der Verkaufsoptionen ist nach der Black-Scholes-Formel gleich 8.73 EUR/Put.  [SFM0ptStart.xpl](#)
Für die entsprechende Strategie 2 mit Anleihen und Calls werden zur Zeit 0 $Fe^{-r\tau} = 77779.42$ EUR festverzinslich angelegt, was aufgezinst auf den Zeitpunkt T gerade den Floor von 95000 EUR ergibt. Für den verbleibenden Rest des Kapitals $V - Fe^{-r\tau} = 22220.58$ EUR kaufen wir 954 Calls mit Ausübungskurs $K = 99.58$ im - nach der Black-Scholes-Formel - Wert von 23.28 EUR/Call. Aus der Put-Call-Parität für europäische Optionen folgt die Gleichwertigkeit beider Strategien, d.h. beide Portfolios aus Aktien und Puts bzw. aus Zerobonds und Calls haben für zu jedem Zeitpunkt $t, 0 \leq t \leq T$, denselben Wert:

$$n \cdot S_t + ne^{d\tau} P_{K,T}(S_t, \tau) = nKe^{-b\tau} + ne^{d\tau} C_{K,T}(S_0, \tau) \quad (2.24)$$

mit $\tau = T - t, b = r - d$ und $nKe^{-b\tau} = Fe^{-r\tau}$ wegen (2.22). Die risikobegrenzende Wirkung der Versicherung lässt sich in der Tabelle 2.8 ablesen.

Aktienkurs S_T [EUR]	Unversichertes Portfolio		Versichertes Portfolio		Versichertes Portfolio in % des unversicherten Portfolios
	Wert [EUR]	Rendite % p.a.	Wert [EUR]	Rendite % p.a.	
70	72857	-27	95000	-5	130
80	83265	-17	95000	-5	114
90	93673	-6	95000	-5	101
100	104081	+4	95400	-5	92
110	114489	+15	104940	+5	92
120	124897	+25	114480	+14	92
130	135305	+35	124020	+24	92
140	145714	+46	133560	+34	92

Tabelle 2.8.: Die Wirkung einer Portfeuille-Versicherung im Fall i): Wert- und Renditeentwicklung  [SFMOptman.xpl](#)

Fall ii)

Während der Laufzeit fallen Dividenden mit auf den Zeitpunkt 0 abgezinstem Gesamtwert $D_0 = 5$ EUR an, die bei Erhalt jeweils sofort festverzinslich angelegt werden. Zum Zeitpunkt T hat dieser Ertrag den aufgezinnten Wert: $D_T = D_0 e^{rT} = 6.107$ EUR, wobei $\tau = T$ wieder für die Gesamtlaufzeit steht. Der Kauf von n Aktien und n Puts führt dann unter Berücksichtigung der Dividende D_T analog zu den Überlegungen in Fall i) auf die Gleichungen

$$n \cdot S_0 + nP_{K,T}(S_0 - D_0, \tau) = V \quad (2.25)$$

und

$$nK + nD_T = F. \quad (2.26)$$

Die Subtraktion der Bardividende D_0 vom Kurs S_0 bei der Berechnung des Optionspreises kann erst im Zusammenhang mit dem später betrachteten Binomialmodell begründet werden. Grob gesagt hängt dies damit zusammen, dass in einem perfekten Markt der Kurs einer Aktie nach Zahlung der Dividende augenblicklich um den gezahlten Dividendenbetrag nach unten springt. Andernfalls wäre Arbitrage möglich. Einsetzen von Gleichung (2.26) in Gleichung (2.25) liefert:

$$S_0 + P_{K,T}(S_0 - D_0, \tau) - \frac{V}{F} \cdot (K + D_T) = 0 \quad (2.27)$$

Analog zum Fall i) ergeben sich aus den Gleichungen die Anzahl n der Aktien und Verkaufsoptionen und der Ausübungskurs K der Puts für die Strategie 1:

$$K = 96.42 \quad \text{und} \quad n = \frac{F}{K + D_T} = 926.55$$

Für die Strategie 2 mit Calls und Anleihen kaufen wir 926.55 Calls mit einem Wert von 23.99 EUR/Call und einem Ausübungskurs von $K = 96.42$ und legen $95000e^{-rT} = 77779.42$

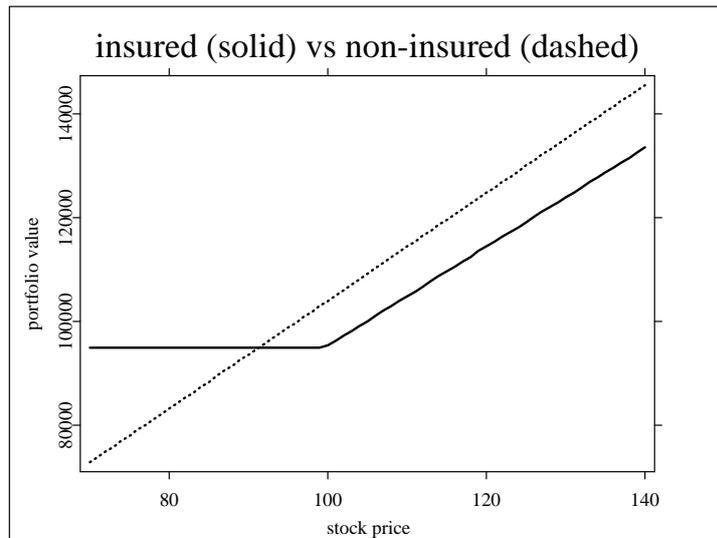


Abbildung 2.1.: Die Wirkung einer Portfeuille-Versicherung: Die durchgezogene Linie stellt den Werteverlauf eines versicherten, die gebrochene den eines unversicherten Portfeuillees als Funktion des Aktienkurses dar. [SFMoptman.xpl](#)

EUR festverzinslich an. Der für beide Strategien gleiche Effekt der Portfeuille-Versicherung im Fall ii) lässt sich in der Tabelle 2.9 ablesen, wobei die auf den Zeitpunkt T aufgezinste Gesamtdividende zu berücksichtigen ist.

Das Beispiel zeigt, wie eine Portfeuille-Versicherung mit Optionen im Prinzip durchgeführt werden kann. In der Praxis stößt man dabei aber auf Probleme:

- Die Anzahl n der benötigten Aktien und Optionen ist nicht ganzzahlig. In einem perfekten Markt sind Finanzinstrumente zwar beliebig teilbar, in der Realität allerdings nicht. Der durch Runden auf den nächsten ganzzahligen Wert entstehende Fehler ist nur in großen Portfolios vernachlässigbar.
- Normalerweise wird ein Put oder Call mit genau dem berechneten Ausübungskurs nicht am Markt erhältlich sein. Darüber hinaus betragen die Laufzeiten von Optionen in der Regel maximal ein Jahr, so dass bei längeren Anlagezeiträumen rein statische Strategien nur eingeschränkt einsetzbar sind.
- Schließlich werden am Markt meist amerikanische Optionen angeboten, die teurer als die für statische Portfeuille-Versicherung ausreichenden europäischen Optionen sind. Die zusätzlichen Möglichkeiten, die amerikanische Optionen bieten, nützen dem Investor nur, wenn er das Portfolio eventuell vorzeitig auflösen muss.

Die Tatsache, dass die gewünschte Option nicht angeboten wird, legt es nahe, sie mit dem in Kapitel 6.1 beschriebenen Delta-Hedge-Prozess zu erzeugen. Da es sich hierbei um eine dynamische Strategie handelt, treten durch die dann nicht mehr zu vernachlässigenden

Aktienkurs S_T [EUR]	Unversichertes Portfolio		Versichertes Portfolio		Versichertes Portfolio in % des unversicherten Portfolios
	Wert [EUR]	Rendite % p.a.	Wert [EUR]	Rendite % p.a.	
70	76107	-24	94996	-5	125
80	86107	-14	94996	-5	110
90	96107	-4	94996	-5	99
96,42	102527	+3	94996	-5	93
100	106107	+6	98313	-2	93
110	116107	+16	107579	+8	93
120	126107	+26	116844	+17	93
130	136107	+36	126110	+26	93
140	146107	+46	135375	+35	93

Tabelle 2.9.: Die Wirkung einer Portfeuille-Versicherung im Fall ii): Wert- und Renditeentwicklung

Transaktionskosten neue Probleme auf. Zum Schluss merken wir noch an, dass es bei der Versicherung großer Portfolios aus vielen verschiedenen Aktien schon aus Kostengründen sinnvoll ist, das Hedgen mit Indexoptionen, d.h. Puts oder Calls auf z.B. den DAX, durchzuführen. Solche Optionen ersetzen näherungsweise die nicht verfügbaren Optionen auf das Portfolio selbst. Um den genauen Effekt eines Hedgen mit Indexoptionen zu berechnen, benötigt man allerdings die Korrelation des Portfolios mit dem Index, die man aus den Korrelationen der einzelnen Aktien mit dem Index erhält. Außerdem sind detaillierte Modellannahmen wie das Capital Asset Pricing Model (CAPM - vgl. 10.4.1) nötig, die u.a. auch die Verteilung von Aktienrenditen betreffen.

3. Grundlegende Begriffe der Wahrscheinlichkeitstheorie

3.1. Reellwertige Zufallsgrößen

Lässt man einen Stein aus 10 m Höhe fallen, so ist aufgrund der Newtonschen Gesetze der Aufprallzeitpunkt bereits bekannt, bevor das Experiment ausgeführt wird. Messgrößen in komplexen Systemen (Aktienkurs an einem bestimmten Datum, Tageshöchsttemperatur an einem bestimmten Ort) lassen sich dagegen nicht exakt vorhersagen. Andererseits weiß man, mit welchen Werten man eher zu rechnen hat und welche gar nicht oder nur extrem selten auftreten. Solche Daten aus der Praxis, die - im Gegensatz zum fallenden Stein - nicht befriedigend durch einen einfachen deterministischen Mechanismus beschrieben werden können, lassen sich als Zufallsgrößen modellieren.

Sei X eine solche Zufallsgröße, die - z.B. als Modell für den Kurs einer Aktie - Werte in den reellen Zahlen annimmt. Die Einschätzung, mit welchen Werten von X eher gerechnet wird und welche weniger plausibel sind, wird durch Angabe der *Wahrscheinlichkeit* von Ereignissen der Form $\{a < X < b\}$ oder auch $\{X \leq b\}$ präzisiert. Das Set aller Wahrscheinlichkeiten

$$P(a \leq X \leq b), \quad -\infty < a \leq b < \infty,$$

bestimmt die komplette *Verteilung* der Zufallsgröße, d.h. alle Wahrscheinlichkeiten von Ereignissen, die in irgendeiner Weise vom Wert von X abhängen. Als Kurzschreibweise für die Verteilung von X benutzen wir $\mathcal{L}(X)$.

Die Verteilung wird bereits durch die *Verteilungsfunktion*

$$F(x) = P(X \leq x), \quad -\infty < x < \infty,$$

eindeutig bestimmt. $F(x)$ ist eine monoton wachsende Funktion, die für $x \rightarrow -\infty$ gegen 0 und für $x \rightarrow \infty$ gegen 1 konvergiert. Gibt es eine Funktion p , so dass sich die Wahrscheinlichkeiten

$$P(a < X < b) = \int_a^b p(x) dx$$

als Integrale ausrechnen lassen, so heißt p die *Wahrscheinlichkeitsdichte* oder kurz *Dichte* der Zufallsgröße X . Die Verteilungsfunktion ist dann eine Stammfunktion der Dichte:

$$F(x) = \int_{-\infty}^x p(y) dy.$$

Für sehr kleine h gilt:

$$P(x - h < X < x + h) \approx 2h \cdot p(x)$$

$p(x)$ ist also ein Maß dafür, wie groß die Chance ist, einen Wert von X in der Nähe von x zu beobachten.

Die wichtigste Klasse von Verteilungen mit Dichten sind die *Normalverteilungen*, die durch zwei Parameter μ, σ^2 charakterisiert werden. Die Wahrscheinlichkeitsdichten haben die Form

$$\varphi_{\mu, \sigma^2}(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} = \frac{1}{\sigma} \varphi\left(\frac{x-\mu}{\sigma}\right),$$

$$\varphi(x) = \varphi_{0,1}(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}.$$

Die Verteilung mit Dichte $\varphi(x)$, d.h. mit $\mu = 0, \sigma^2 = 1$, heißt *Standardnormalverteilung*. Als eine übliche Kurzschreibweise für "X ist eine normalverteilte Zufallsvariable mit Parametern μ, σ^2 " benutzen wir "X ist $N(\mu, \sigma^2)$ - verteilt". Mit Φ bezeichnen wir die Verteilungsfunktion der Standardnormalverteilung

$$\Phi(x) = \int_{-\infty}^x \varphi(y) dy.$$

Wenn X eine $N(\mu, \sigma^2)$ - verteilte Zufallsgröße ist, so ist die zentrierte und skalierte Zufallsgröße $(X - \mu)/\sigma$ standardnormalverteilt. Die Verteilungsfunktion von X ist daher

$$F(x) = P(X \leq x) = P\left(\frac{X - \mu}{\sigma} \leq \frac{x - \mu}{\sigma}\right) = \Phi\left(\frac{x - \mu}{\sigma}\right).$$

Die *Lognormalverteilung* ist eine von der Normalverteilung abgeleitete Verteilung, die in der Finanzmathematik als Modell für Aktienkurse eine wichtige Rolle spielt. Eine positive Zufallsgröße X ist lognormalverteilt mit den beiden Parametern μ, σ^2 , wenn die logarithmierte Variable $\ln(X)$ $N(\mu, \sigma^2)$ - verteilt ist. Die entsprechende Verteilungsfunktion erhält man sofort aus der Definition:

$$F(x) = P(X \leq x) = P(\ln X \leq \ln x) = \Phi\left(\frac{\ln x - \mu}{\sigma}\right), \quad x > 0.$$

Durch Ableiten von $F(x)$ erhalten wir

$$p(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \frac{1}{x} e^{-\frac{(\ln x - \mu)^2}{2\sigma^2}} = \frac{1}{\sigma x} \varphi\left(\frac{\ln x - \mu}{\sigma}\right), \quad x > 0,$$

als die Dichte der Lognormalverteilung mit den Parametern μ, σ^2 .

Die Verteilung einer Zufallsgröße X , die nur endlich viele verschiedene Werte x_1, \dots, x_n annehmen kann, wird vollständig durch die Wahrscheinlichkeiten

$$P(X = x_k), \quad k = 1, \dots, n$$

bestimmt. Die einfachsten Beispiele bilden Zufallsgrößen, die nur 2 bzw. 3 Werte annehmen können, z.B. ± 1 oder $-1, 0, +1$. Sie bilden die Basis der Binomial- bzw. Trinomialverfahren, mit denen auf dem Rechner diskrete Zufallsprozesse erzeugt werden können, die gute Approximationen für stetige Prozesse darstellen, wie sie z.B. bei der Modellierung des zeitlichen Verlaufs eines Aktienkurses vorkommen.

In diesem Zusammenhang treten binomialverteilte Zufallsgrößen auf. Seien Y_1, \dots, Y_n unabhängige Zufallsgrößen, die nur die beiden Werte 0 und 1 annehmen, mit

$$p = P(Y_k = 1) , \quad 1 - p = P(Y_k = 0) , \quad k = 1, \dots, n .$$

Solche Zufallsgrößen nennen wir Bernoulli-verteilt mit Parameter p . Ihre Summe X , d.h. die Anzahl der Einsen in der Stichprobe Y_1, \dots, Y_n , ist binomialverteilt mit Parametern n, p :

$$X = \sum_{k=1}^n Y_k , \quad P(X = m) = \binom{n}{m} p^m (1 - p)^{n-m} , \quad m = 0, \dots, n .$$

 [SFMBinomial.xpl](#)

Hierfür verwenden wir die Kurzschreibweise "X ist $B(n, p)$ -verteilt" anstelle von "X ist binomialverteilt mit Parametern n, p ." Bernoulli-verteilte Zufallsgrößen sind dann $B(1, p)$ -verteilt.

Wenn n groß genug ist, kann eine $B(n, p)$ -verteilte Zufallsgröße X näherungsweise durch eine $N(np, np(1 - p))$ -verteilte Zufallsgröße Z ersetzt werden in dem Sinn, dass

$$P(a < X < b) \approx P(a < Z < b) . \tag{3.1}$$

Eine genauere Aussage liefert der zentrale Grenzwertsatz. In der klassischen Statistik macht man sich dieses Ergebnis zunutze, um die für größere n aufwendige Berechnung von Binomialwahrscheinlichkeiten zu umgehen. Umgekehrt weist die approximative Austauschbarkeit von binomial- und normalverteilten Zufallsgrößen aber auch einen Weg, auf der Normalverteilung aufbauende stetige Zufallsprozesse durch auf dem Rechner leicht zu simulierende Binomialprozesse anzunähern.  [SFMclt.xpl](#)

3.2. Erwartungswert und Varianz

Der *Erwartungswert* oder Mittelwert $E[X]$ einer reellwertigen Zufallsgröße X ist ein Maß dafür, in welchem Bereich der reellen Zahlen die Werte von X zu finden sind. Addiert man zu dem zufälligen Wert X eine konstante Zahl c , so verschiebt sich der Erwartungswert entsprechend: $E[X + c] = E[X] + c$. Besitzt die Zufallsgröße X eine Dichte $p(x)$, so kann ihr Erwartungswert folgendermaßen berechnet werden:

$$E(X) = \int_{-\infty}^{\infty} xp(x)dx .$$

Existiert das Integral nicht, so existiert auch der Erwartungswert nicht. In der Praxis tritt dieser Fall selten auf.

Liegt eine Stichprobe von unabhängigen, identisch verteilten (u.i.v.) Zufallsgrößen X_1, \dots, X_n mit derselben Verteilung wie X vor, lässt sich $E[X]$ durch das Stichprobenmittel schätzen:

$$\hat{\mu} = \frac{1}{n} \sum_{t=1}^n X_t$$

Die *Varianz* $\text{Var}(X)$ von X ist ein Maß dafür, wie stark die Zufallsgröße um ihren Erwartungswert herum streut:

$$\begin{aligned} \text{Var}(X) &= E[(X - EX)^2] \\ \text{Varianz} &= \text{mittlere quadratische Abweichung der Zufallsgröße} \\ &\quad \text{vom eigenen Erwartungswert.} \end{aligned}$$

Besitzt die Zufallsgröße X eine Dichte $p(x)$, so kann ihre Varianz folgendermaßen berechnet werden:

$$\text{Var}(X) = \int_{-\infty}^{\infty} (x - EX)^2 p(x) dx.$$

Das Integral kann dabei auch unendlich sein. Empirische Befunde lassen gelegentlich Zweifel aufkommen, ob einige in der Finanz- und Versicherungsmathematik auftretende Zufallsgrößen, die Verluste in hochriskanten Geschäften modellieren, noch eine endliche Varianz haben.

Die Varianz besitzt als quadratische Größe eine andere Einheit als X selbst. Eine bessere Skalierung für die Variabilität von X liefert die *Standardabweichung*

$$\sigma(X) = \sqrt{\text{Var}(X)},$$

die wie $E[X]$ in derselben Einheit wie X gemessen wird.

Aus einer Stichprobe von u.i.v. Zufallsgrößen X_1, \dots, X_n mit derselben Verteilung wie X lässt sich $\text{Var}(X)$ durch die Stichprobenvarianz schätzen:

$$\hat{\sigma}^2 = \frac{1}{n} \sum_{t=1}^n (X_t - \hat{\mu})^2$$

Eine $N(\mu, \sigma^2)$ -verteilte Zufallsgröße X hat Mittelwert μ und Varianz σ^2 . Der 2σ -Bereich um μ enthält mit etwas mehr als 95 % Wahrscheinlichkeit Beobachtungen einer solchen Zufallsgröße

$$P(\mu - 2\sigma < X < \mu + 2\sigma) \approx 0.95.$$

Eine lognormalverteilte Zufallsgröße X mit den Parametern μ, σ^2 hat Erwartungswert und Varianz

$$E(X) = e^{\mu + \frac{1}{2}\sigma^2}, \quad \text{Var}(X) = e^{2\mu + \sigma^2}(e^{\sigma^2} - 1).$$

Eine $B(n, p)$ -verteilte Zufallsgröße X hat Erwartungswert np und Varianz $np(1-p)$. Die Approximation (3.1) ist also so gewählt, dass die näherungsweise einander ersetzenden binomial- und normalverteilten Zufallsgrößen identische Erwartungswerte und Varianzen besitzen.

3.3. Schiefe und Kurtosis

DEFINITION 3.1 (Schiefe)

Die Schiefe einer Zufallsgröße X mit Erwartungswert μ und Varianz σ^2 ist definiert als

$$S(X) = \frac{E[(X - \mu)^3]}{\sigma^3}.$$

Ist die Schiefe negativ (positiv), so ist die Verteilung linksschief (rechtsschief). Normalverteilte Zufallsgrößen haben eine Schiefe von Null, da die Verteilung symmetrisch um den Erwartungswert ist. Geschätzt wird die Schiefe aus einer Stichprobe von u.i.v. Zufallsgrößen X_1, \dots, X_n durch

$$\hat{S}(X) = \frac{\frac{1}{n} \sum_{t=1}^n (X_t - \hat{\mu})^3}{\hat{\sigma}^3} \quad (3.2)$$

wobei $\hat{\mu}, \hat{\sigma}^2$ die im letzten Abschnitt eingeführten Stichprobenmittel und -varianz sind.

DEFINITION 3.2 (Kurtosis)

Die Kurtosis (oder auch Wölbung) einer Zufallsgröße X mit Erwartungswert μ und Varianz σ^2 ist definiert als

$$\text{Kurt}(X) = \frac{E[(X - \mu)^4]}{\sigma^4}.$$

Normalverteilte Zufallsgrößen haben eine Kurtosis von 3. Eine größere Kurtosis ist bei Finanzdaten oft zu beobachten; dies ist ein Anzeichen dafür, dass Werte nahe bei 0 und vor allem sehr große positive und/oder negative Werte mit im Vergleich zur Normalverteilung hoher Wahrscheinlichkeit auftreten, mittelgroße Werte dagegen seltener. Geschätzt wird die Kurtosis durch

$$\widehat{\text{Kurt}}(X) = \frac{\frac{1}{n} \sum_{t=1}^n (X_t - \hat{\mu})^4}{\hat{\sigma}^4}. \quad (3.3)$$

3.4. Zufallsvektoren, Abhängigkeit, Korrelation

Interessiert man sich gleichzeitig für mehrere Zufallsgrößen X_1, \dots, X_n - z.B. die Kurse verschiedener Aktien - und für ihre wechselseitige Abhängigkeit, so betrachtet man (X_1, \dots, X_n) als zufälligen Punkt oder *Zufallsvektor* im \mathbb{R}^n . Die gemeinsame Verteilung der Zufallsgrößen X_1, \dots, X_n , d.h. die Verteilung des Zufallsvektors, wird wie im eindimensionalen eindeutig durch die Wahrscheinlichkeiten

$$P(a_1 \leq X_1 \leq b_1, \dots, a_n \leq X_n \leq b_n), \quad -\infty < a_i \leq b_i < \infty, \quad i = 1, \dots, n,$$

bestimmt. Besitzt der Zufallsvektor (X_1, \dots, X_n) eine Wahrscheinlichkeitsdichte $p(x_1, \dots, x_n)$, so können Wahrscheinlichkeiten wieder als Integrale von p berechnet werden:

$$P(a_1 \leq X_1 \leq b_1, \dots, a_n \leq X_n \leq b_n) = \int_{a_n}^{b_n} \dots \int_{a_1}^{b_1} p(x_1, \dots, x_n) dx_1 \dots dx_n.$$

Der intuitive Begriff der *Unabhängigkeit* zweier Zufallsgrößen X_1, X_2 wird präzisiert durch die Forderung, dass

$$P(a_1 \leq X_1 \leq b_1, a_2 \leq X_2 \leq b_2) = P(a_1 \leq X_1 \leq b_1) \cdot P(a_2 \leq X_2 \leq b_2),$$

d.h. Wahrscheinlichkeiten von Ereignissen, die von dem Wert des Zufallsvektors (X_1, X_2) abhängen, lassen sich faktorisieren, und es genügt, die eindimensionalen Verteilungen von X_1 und X_2 zu kennen. Besitzt der Zufallsvektor (X_1, X_2) eine Dichte $p(x_1, x_2)$, so besitzen X_1 und X_2 ebenfalls Dichten $p_1(x)$ und $p_2(x)$. Die Unabhängigkeit der beiden Zufallsgrößen ist dann gleichbedeutend mit der Faktorisierung der gemeinsamen Dichte:

$$p(x_1, x_2) = p_1(x_1)p_2(x_2)$$

Abhängigkeit zweier Zufallsgrößen X_1, X_2 kann in komplizierter Form vorliegen. Sind X_1, X_2 gemeinsam normalverteilt, lässt sich der Grad ihrer Abhängigkeit aber auf einfache Weise quantifizieren durch ihre Kovarianz

$$\text{Cov}(X_1, X_2) = E[(X_1 - E[X_1])(X_2 - E[X_2])]$$

oder Korrelation

$$\text{Corr}(X_1, X_2) = \frac{\text{Cov}(X_1, X_2)}{\sigma(X_1) \cdot \sigma(X_2)}.$$

Die Korrelation hat den Vorteil, immer zwischen -1 und +1 zu liegen und skaleninvariant zu sein. Für gemeinsam *normalverteilte* Zufallsgrößen ist Unabhängigkeit gleichbedeutend mit Korrelation gleich Null, während vollständige Abhängigkeit gleichbedeutend mit Korrelation +1 (X_1 ist groß, wenn X_2 groß ist) oder Korrelation -1 (X_1 ist groß, wenn X_2 klein ist) ist.

Allgemein gilt für *unabhängige* Zufallsgrößen X_1, \dots, X_n

$$\text{Cov}(X_i, X_j) = 0 \quad \text{für } i \neq j,$$

woraus eine nützliche Rechenregel folgt:

$$\text{Var}\left(\sum_{j=1}^n X_j\right) = \sum_{j=1}^n \text{Var}(X_j).$$

Sind X_1, \dots, X_n unabhängige Zufallsgrößen, die alle dieselbe Verteilung haben:

$$P(a \leq X_i \leq b) = P(a \leq X_j \leq b) \quad \text{für alle } i, j,$$

so nennen wir sie *unabhängig identisch verteilt* (*u.i.v.*).

3.5. Bedingte Wahrscheinlichkeiten und Erwartungswerte

Die *bedingte Wahrscheinlichkeit*, dass eine Zufallsgröße Y Werte zwischen a und b annimmt, gegeben, dass eine andere Zufallsgröße X Werte zwischen x und $x + \Delta_x$ annimmt, ist definiert als

$$P(a \leq Y \leq b | x \leq X \leq x + \Delta_x) = \frac{P(a \leq Y \leq b, x \leq X \leq x + \Delta_x)}{P(x \leq X \leq x + \Delta_x)}, \quad (3.4)$$

vorausgesetzt, dass der Nenner der rechten Seite nicht Null ist. Die bedingte Wahrscheinlichkeit von Ereignissen der Art $a \leq Y \leq b$ spiegelt unsere Einschätzung wider, mit welchen Werten von Y wir eher rechnen und welche wir für nicht plausibel halten, *wenn wir schon wissen*, dass eine andere Zufallsgröße X bestimmte Werte angenommen hat. Unsere Meinung über die Wahrscheinlichkeiten der Werte von Y wird durch unsere Vorabinformation über X nicht geändert, wenn Y gar nicht von X abhängt, wenn also Y und X *unabhängig* sind. Dann gilt:

$$P(a \leq Y \leq b | x \leq X \leq x + \Delta_x) = P(a \leq Y \leq b).$$

Wenn wir in (3.4) Δ_x gegen 0 gehen lassen, so konvergiert heuristisch die linke Seite gegen $P(a \leq Y \leq b | X = x)$. Wenn X eine Wahrscheinlichkeitsdichte p_X besitzt, so ist die rechte Seite von (3.4) allerdings nirgends definiert, da dann stets $P(X = x) = 0$. Es ist aber möglich, die bedingte Verteilung von Y gegeben $X = x$ sauber zu definieren. Wenn die beiden Zufallsgrößen Y und X eine gemeinsame Wahrscheinlichkeitsdichte $p(x, y)$ besitzen, so hat die bedingte Verteilung die Wahrscheinlichkeitsdichte

$$p_{Y|X}(y|x) = \frac{p(x, y)}{p_X(x)} \quad \text{für } p_X(x) \neq 0$$

und $p_{Y|X}(y|x) = 0$ sonst. Es gilt dann

$$P(a \leq Y \leq b | X = x) = \int_a^b p_{Y|X}(y|x) dy.$$

Der Erwartungswert bzgl. der bedingten Verteilung lässt sich dann folgendermaßen berechnen:

$$E(Y | X = x) = \int_{-\infty}^{\infty} y p_{Y|X}(y|x) dy \equiv \eta(x).$$

Die Funktion $\eta(x) = E(Y | X = x)$ heißt der *bedingte Erwartungswert von Y gegeben $X = x$* . Intuitiv handelt es sich um den Erwartungswert der Zufallsgröße Y , wenn bereits feststeht, dass X den Wert x angenommen hat.

Betrachtet man $\eta(x)$ als Funktion der Zufallsgröße X , so erhält man den bedingten Erwartungswert von Y gegeben X :

$$E(Y | X) = \eta(X).$$

$E(Y | X)$ ist eine Zufallsgröße, die sich als Funktion von X darstellen lässt und denselben Erwartungswert wie Y hat. Der bedingte Erwartungswert besitzt eine Reihe von nützlichen Eigenschaften, die wir in dem folgenden Satz zusammenfassen.

SATZ 3.1 X, Y, Z seien reellwertige Zufallsgrößen, deren gemeinsame Verteilung eine Dichte besitzt.

a) Wenn X, Y unabhängig sind, so gilt stets $E(Y|X = x) = E(Y)$

b) Wenn $Y = g(X)$ eine Funktion von X ist, so gilt

$$E[Y|X = x] = E[g(X)|X = x] = g(x).$$

Allgemeiner gilt für Zufallsgrößen der Form $Y = Zg(X)$:

$$E[Y|X = x] = E[Zg(X)|X = x] = g(x)E[Z|X = x].$$

c) Der bedingte Erwartungswert ist linear, d.h. für beliebige reelle Zahlen a, b gilt:

$$E(aY + bZ|X = x) = aE(Y|X = x) + bE(Z|X = x).$$

d) $E[E(Y|X)] = E(Y)$.

Der Begriff des bedingten Erwartungswert $E(Y|X = x)$ lässt sich entsprechend auch auf mehrdimensionale Zufallsvektoren Y und X verallgemeinern. Ist $S_t, t = 0, 1, 2, \dots$ eine Folge zeitlich geordneter Zufallsgrößen, z.B. als Modell für täglich notierte Kurse, so repräsentiert mit $Y = S_{t+1}$ und $X = (S_t, \dots, S_{t-p+1})^T$ der bedingte Erwartungswert

$$E(Y|X = x) = E(S_{t+1}|S_t = x_1, \dots, S_{t-p+1} = x_p)$$

den Erwartungswert des Kurses am nächsten Tag $t + 1$ gegeben die Kurse $x = (x_1, \dots, x_p)^T$ der letzten p Tage. Da die am Tag t verfügbare Information, die für die zukünftige Kursentwicklung relevant ist, mehr als nur ein paar vergangene Kurse enthalten kann, benutzen wir auch oft die Schreibweise $E(Y|\mathcal{F}_t)$ für den Erwartungswert von Y gegeben die bis zur Zeit t verfügbare Information. \mathcal{F}_t bezeichnet dabei für jedes $t \geq 0$ eine Familie von Ereignissen (mit der Struktur einer σ -Algebra, d.h. gewisse Kombinationen von Ereignissen aus \mathcal{F}_t liegen wieder in \mathcal{F}_t), die die bis zur Zeit t verfügbare Information repräsentiert. \mathcal{F}_t besteht aus solchen Ereignissen, von denen bis zur Zeit t feststeht, ob sie eintreffen oder nicht. Da die Information mit der Zeit wächst, gilt $\mathcal{F}_s \subset \mathcal{F}_t$ für $s < t$ - s. auch Definition 5.1. Auf die exakte Definition von $E(Y|\mathcal{F}_t)$ verzichten wir, weisen aber darauf hin, dass auch dieser allgemeine bedingte Erwartungswert die entsprechend umformulierten Rechenregeln von Satz 3.1 erfüllt.

3.6. Literaturhinweise

Eine Einführung in die Wahrscheinlichkeitstheorie findet man unter anderem in ?, ?, ? und ?. Anspruchsvoller, dafür aber gleichzeitig auch eine Einführung in die für die fortgeschrittene Finanzmathematik unerlässliche Martingaltheorie beinhaltend ist das Buch von ?.

4. Stochastische Prozesse in diskreter Zeit

Ein *stochastischer Prozess* oder Zufallsprozess besteht aus zeitlich angeordneten Zufallsgrößen $\{X_t; t \geq 0\}$. Der Einfachheit halber lassen wir den Prozess immer zur Zeit $t = 0$ anfangen. Werden die Beobachtungen nur in regelmäßigen Abständen erhoben, ist $t = 0, 1, 2, \dots$, und wir sprechen von einem Prozess in *diskreter Zeit*. In diesem Abschnitt werden nur solche Prozesse betrachtet. Typische Beispiele sind täglich, monatlich oder jährlich erhobene Wirtschaftsdaten (Aktienkurse, Arbeitslosenziffern, Absatzzahlen).

4.1. Binomialprozesse

Einer der einfachsten stochastischen Prozesse ist die *einfache Irrfahrt*, ein Prozess, dessen Zuwachs $Z_t = X_t - X_{t-1}$ von der Zeit $t - 1$ zur Zeit t nur die beiden Werte $+1$ und -1 annehmen kann. Zusätzlich nehmen wir an, dass die Zuwächse Z_1, Z_2, \dots u.i.v. und unabhängig von Anfangswert X_0 sind. Die einfache Irrfahrt hat daher die Form:

$$X_t = X_0 + \sum_{k=1}^t Z_k \quad , \quad t = 1, 2, \dots \quad (4.1)$$

X_0, Z_1, Z_2, \dots unabhängig und

$$P(Z_k = 1) = p \quad , \quad P(Z_k = -1) = 1 - p \quad \text{für alle } k .$$

Lassen wir stattdessen zu, dass der Prozess von der Zeit $t - 1$ zur Zeit t um u wächst oder um d fällt:

$$P(Z_k = u) = p \quad , \quad P(Z_k = -d) = 1 - p \quad \text{für alle } k ,$$

so erhalten wir die allgemeinere Klasse der *Binomialprozesse*. $u, d \geq 0$ sind dabei beliebige, aber feste Konstanten ($u = \text{up}$, $d = \text{down}$).

Der *Pfad* einer einfachen Irrfahrt ergibt sich durch lineare Interpolation der Punkte (t, X_t) und spiegelt den zeitlichen Verlauf des Prozesses anschaulich wider. Er kann sich - ausgehend von einem festen Anfangswert $X_0 = a$ - nur innerhalb eines Netzes von Verbindungen der Punkte (t, b_t) , $t = 0, 1, 2, \dots$, $b_t = a - t, a - t + 1, \dots, a + t$ bewegen, denn von a aus kann X_t bis zur Zeit t maximal auf $a + t$ wachsen ($Z_1 = \dots = Z_t = 1$) oder minimal auf $a - t$ fallen ($Z_1 = \dots = Z_t = -1$). Fünf generierte einfache Irrfahrten sind in den Abbildungen [4.1](#) ($p = 0.5$), [4.2](#) ($p = 0.4$) und [4.3](#) ($p = 0.6$) gezeigt.

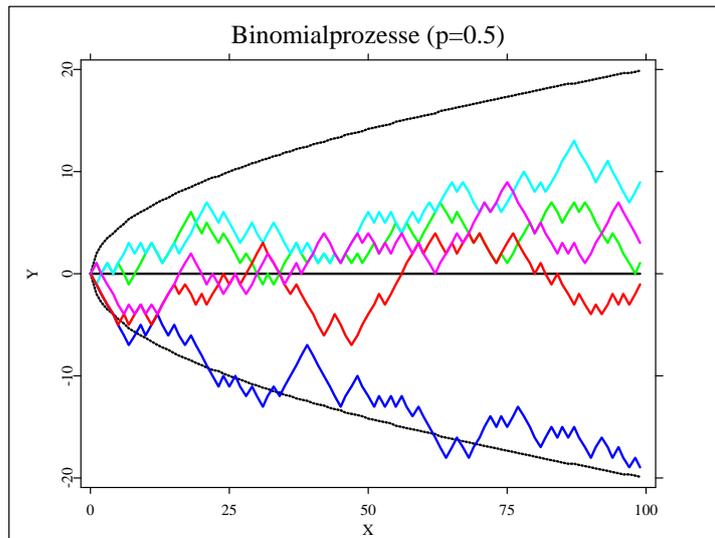


Abbildung 4.1.: Fünf Realisierungen einer symmetrischen einfachen Irrfahrt. Um den Trend (gleich Null) sind die (2σ) -Intervalle eingezeichnet. [SFMBinomp.xpl](#)

Für allgemeine Binomialprozesse sieht das Netz der möglichen Pfade komplizierter aus. Die möglichen Werte, die ausgehend von a zur Zeit t erreicht werden können sind von der Form

$$b_t = a + n \cdot u - m \cdot d, \text{ wobei } n, m \geq 0, \quad n + m = t.$$

$X_t = a + n \cdot u - m \cdot d$, wenn der Prozess von der Zeit 0 bis zur Zeit t n -mal gewachsen und m -mal gefallen ist, d.h. n der Zuwächse Z_1, \dots, Z_t haben den Wert u , m der Zuwächse haben den Wert $-d$ angenommen. Bei Aktienkursen nennt man das Netz, in dem sich die Pfade des Prozesses bewegen können, den *Kursbaum*.

Die *symmetrische einfache Irrfahrt* ($p = \frac{1}{2}$) mit Start in 0 ($X_0 = 0$) ist für alle Zeiten im Mittel 0 :

$$E[X_t] = 0 \quad \text{für alle } t.$$

Andernfalls hat die *Irrfahrt* einen *Trend* oder eine *Drift* nach oben ($p > \frac{1}{2}$) bzw. nach unten ($p < \frac{1}{2}$). Der Prozess wächst bzw. fällt im Mittel:

$$E[X_t] = t \cdot (2p - 1),$$

denn $E[Z_k] = 2p - 1$ für alle Zuwächse. Der Trend ist also linear in der Zeit. Er ist in [Abbildung 4.3](#) ($p = 0.6$) als steigende Gerade, in [Abbildung 4.2](#) ($p = 0.4$) als fallende Gerade eingezeichnet.

Für den allgemeinen *Binomialprozess* mit beliebigem Startwert X_0 gilt entsprechend $E[Z_k] = (u + d)p - d$ und somit:

$$E[X_t] = E[X_0] + t \cdot \{(u + d)p - d\}.$$

Mit wachsender Zeit t kann X_t immer mehr verschiedene Werte annehmen, und die Variabilität wächst. Da die einzelnen Summanden in [\(4.1\)](#) unabhängig sind und zudem $\text{Var}(Z_k) =$

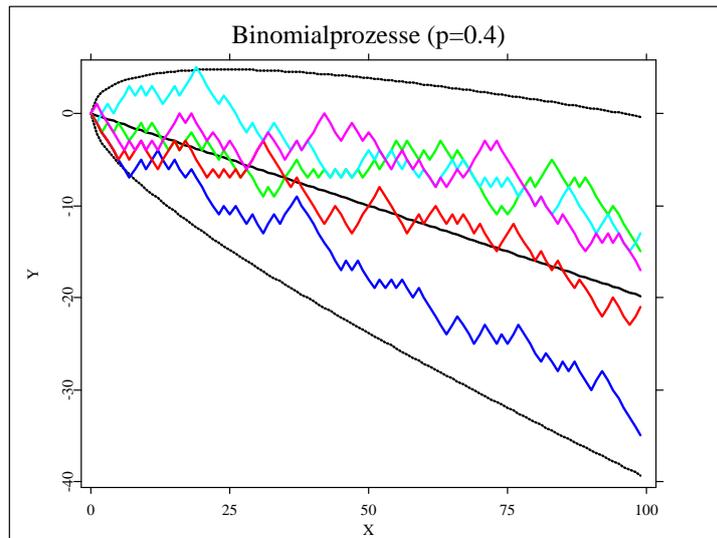


Abbildung 4.2.: Fünf Realisierungen einer einfachen Irrfahrt mit $p = 0.4$. Die fallende Gerade ist der Trend. Um den Trend sind die (2σ) -Intervalle eingezeichnet.
[SFMBinomp.xpl](#)

$\text{Var}(Z_1)$ für alle k , ist die Varianz von X_t (vgl. Abschnitt 3.3):

$$\text{Var}(X_t) = \text{Var}(X_0) + t \cdot \text{Var}(Z_1),$$

wächst also linear mit der Zeit, und die Standardabweichung $\sigma(X_t)$ wächst mit der Rate \sqrt{t} . Für die Irrfahrten in den Abbildungen 4.1 ($p = 0.5$), 4.2 ($p = 0.4$) und 4.3 ($p = 0.6$) sind die Intervalle $[E[X_t] - 2\sigma(X_t); E[X_t] + 2\sigma(X_t)]$ eingezeichnet. Für ein großes t sollten etwa 95% der Realisierungen des Prozesses in diesem Intervall liegen.

Die Varianz der Zuwächse Z_k lässt sich leicht ausrechnen, aber wir benutzen stattdessen die folgende einfache Beziehung zur Binomialverteilung. Setzen wir

$$Y_k = \frac{Z_k + d}{u + d} = \begin{cases} 1 & \text{wenn } Z_k = u \\ 0 & \text{wenn } Z_k = -d \end{cases}$$

bzw. umgekehrt

$$Z_k = (u + d) Y_k - d \tag{4.2}$$

so erhalten wir als Darstellung des Binomialprozesses

$$X_t = X_0 + (u + d) B_t - td \tag{4.3}$$

mit der $B(t, p)$ -verteilten Zufallsgröße

$$B_t = \sum_{k=1}^t Y_k \tag{4.4}$$

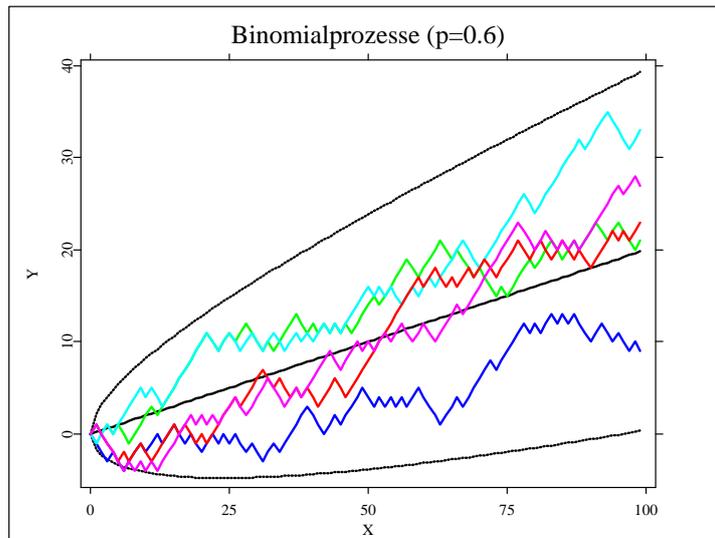


Abbildung 4.3.: Fünf Realisierungen einer einfachen Irrfahrt mit $p = 0.6$. Die steigende Gerade ist der Trend. Um den Trend sind die (2σ) -Intervalle eingezeichnet.
[SFMBinomp.xpl](#)

Bei gegebener Verteilung des Anfangswertes X_0 ist die Verteilung von X_t also für alle Zeiten bekannt und durch eine einfache Transformation aus der Binomialverteilung $B(t, p)$ ableitbar. Aus (4.2) – (4.4) erhalten wir, z.B. für den Fall $X_0 = 0$:

$$\text{Var}(X_t) = t(u + d)^2 p(1 - p)$$

und für große t erhalten wir eine Approximation der Verteilung von X_t durch

$$\mathcal{L}(X_t) \approx N(t\{(u + d)p - d\}, t(u + d)^2 p(1 - p)).$$

Für $p = \frac{1}{2}$, $u = d = \Delta x$ erhalten wir beispielsweise die folgende Annäherung an $\mathcal{L}(X_t)$:

$$N(0, t \cdot (\Delta x)^2).$$

Wie gut die Annäherung für $t = 100$ ist zeigt Abbildung 4.4.

4.2. Trinomialprozesse

Im Gegensatz zum Binomialprozess darf ein *Trinomialprozess* in der Zeiteinheit nicht nur um eine vorgegebene Größe wachsen oder fallen, sondern kann auch unverändert bleiben. Die Zuwächse Z_k haben hier die Eigenschaft:

$$P(Z_k = u) = p, P(Z_k = -d) = q, P(Z_k = 0) = r = 1 - p - q,$$

und der stochastische Prozess hat wieder die Gestalt:

$$X_t = X_0 + \sum_{k=1}^t Z_k$$

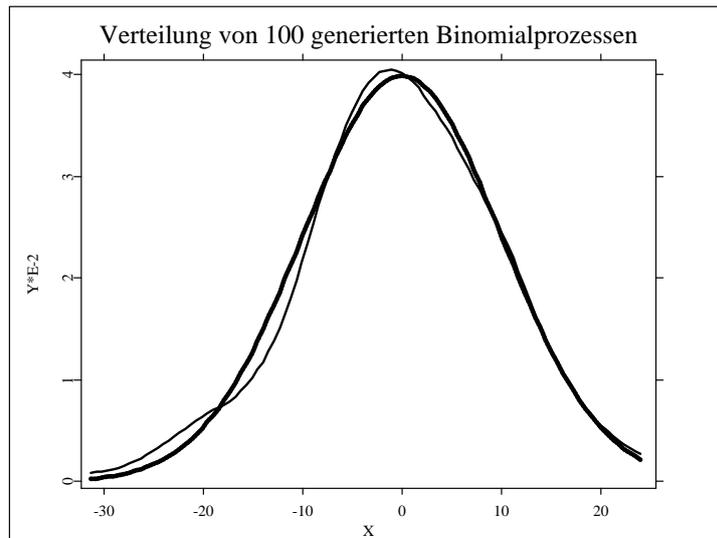


Abbildung 4.4.: Die Verteilung von 100 Realisierungen einer einfachen symmetrischen Irrfahrt der Länge 100 und eine Dichteschätzung von 100 normalverteilten Zufallsvariablen. [SFMbinomv.xpl](#)

mit unabhängigen X_0, Z_1, Z_2, \dots . Trinomialschemata, bei denen die Wahrscheinlichkeiten p, q, r allerdings zeit- und zustandsabhängig sein dürfen, werden von manchen Algorithmen angewendet, um die Black-Scholes Gleichung näherungsweise zu lösen. Abbildung 4.5 zeigt die Pfade einiger simulierter Trinomialprozesse mit $u = d = 1, p = q = 0.25$.

Die exakte Verteilung der X_t lässt sich nicht auf die Binomialverteilung zurückführen, aber es gilt ähnlich wie für den Binomialprozess:

$$\begin{aligned} E[X_t] &= E[X_0] + t \cdot E[Z_1] = E[X_0] + t \cdot (pu - qd) \\ \text{Var}(X_t) &= \text{Var}(X_0) + t \cdot \text{Var}(Z_1), \text{ wobei} \\ \text{Var}(Z_1) &= p(1-p)u^2 + q(1-q)d^2 + 2pq ud, \end{aligned}$$

und für große t ist X_t näherungsweise $N(E[X_t], \text{Var}(X_t))$ -verteilt.

4.3. Allgemeine Irrfahrten

Binomial- und Trinomialprozesse sind einfache Beispiele allgemeiner *Irrfahrten*, d.h. stochastischer Prozesse $\{X_t; t \geq 0\}$ der Form

$$X_t = X_0 + \sum_{k=1}^t Z_k, \quad t = 1, 2, \dots$$

wobei X_0 unabhängig von den u.i.v. Z_1, Z_2, \dots . Die Verteilung der Zuwächse ist irgendeine Verteilung reellwertiger Zufallsgrößen. Die einzelnen Z_k können endlich

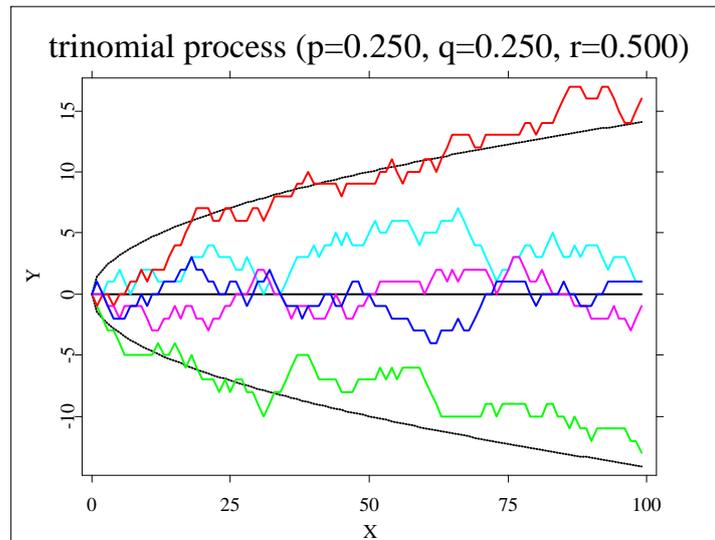


Abbildung 4.5.: Fünf Realisierungen eines Trinomialprozesses mit $p = q = 0.25$. Um den Trend (gleich Null) sind die (2σ) -Intervalle eingezeichnet.
[SFMTrinomp.xpl](#)

oder abzählbar viele Werte annehmen; sie können aber auch einen kontinuierlichen Wertebereich haben.

Als Beispiel betrachten wir eine *Gaußsche Irrfahrt*, mit Start in 0, d.h. eine Irrfahrt mit $X_0 = 0$, für die endlich viele Werte X_1, \dots, X_t jeweils gemeinsam normalverteilt sind. Eine solche Irrfahrt erhalten wir, wenn wir die Zuwächse als unabhängig und identisch $N(\mu, \sigma^2)$ -verteilt wählen. Aus den Eigenschaften der Normalverteilung folgt dann, dass zu jedem Zeitpunkt t der Wert des Prozesses $X_t \sim N(\mu t, \sigma^2 t)$ -verteilt ist.

Näherungsweise gilt dies für alle Irrfahrten für genügend große t :

$$\text{Verteilung von } X_t \approx N(t \cdot E[Z_1], t \cdot \text{Var}(Z_1)),$$

wenn $X_0 = 0$ und $\text{Var}(Z_1)$ endlich ist. Diese Approximation folgt aus dem zentralen Grenzwertsatz für u.i.v. Zufallsgrößen.

Irrfahrten sind *Prozesse mit unabhängigen Zuwächsen*, d.h. der Zuwachs Z_{t+1} des Prozesses von der Zeit t zur Zeit $t + 1$ ist unabhängig von den Werten X_0, \dots, X_t des Prozesses in der Vergangenheit bis zur Zeit t . Allgemeiner ist dann für beliebige $s > 0$ der Zuwachs

$$X_{t+s} - X_t = Z_{t+1} + \dots + Z_{t+s}$$

des Prozesses nach der Zeit t bis zur Zeit $t + s$ unabhängig von X_0, \dots, X_t . Als Folge ist die beste *Vorhersage* - im Sinne eines möglichst kleinen mittleren quadratischen Fehlers - für X_{t+1} gegeben die Kenntnis aller X_0, \dots, X_t einfach $X_t + E[Z_{t+1}]$. Solange der Kurs einer Aktie isoliert betrachtet wird, trifft diese Aussage für den Aktienkursprozess oft in guter Näherung zu. Schon vor rund hundert Jahren postulierte Bachelier (hier für den Fall

$E[Z_k] = 0$ für alle k): “Die beste Vorhersage für den Kurs von morgen ist der Kurs von heute.”

Prozesse mit unabhängigen Zuwächsen sind automatisch *Markoff-Prozesse*: die zukünftige Entwicklung des Prozesses von der Zeit t an hängt nur von X_t ab und - gegeben die Kenntnis von X_t - nicht mehr von den weiter zurückliegenden Werten X_0, \dots, X_{t-1} . Wenn die Zuwächse Z_k und der Anfangswert X_0 und damit auch die Werte X_t selbst nur endlich oder abzählbar viele Werte annehmen, dann wird diese *Markoff-Eigenschaft* formal ausgedrückt durch

$$\begin{aligned} P(a_{t+1} < X_{t+1} < b_{t+1} | X_t = c, a_{t-1} < X_{t-1} < b_{t-1}, \dots, a_0 < X_0 < b_0) \\ = P(a_{t+1} < X_{t+1} < b_{t+1} | X_t = c). \end{aligned}$$

Wenn $X_t = c$ bekannt ist, ändern zusätzliche Informationen über die Werte von X_0, \dots, X_{t-1} die Einschätzung, in welchem Bereich X_{t+1} vermutlich liegen wird, nicht.

4.4. Geometrische Irrfahrten

Die wesentliche Idee hinter dem Modell ”Irrfahrt” für reale Prozesse ist die Annahme, dass die Änderungen in einer Zeiteinheit unabhängig voneinander und immer von derselben Größenordnung sind. Letzteres ist gerade für wirtschaftliche Zeitreihen oft nicht erfüllt. Saisonale Schwankungen von z.B. monatlichen Absatzzahlen sind typischerweise *absolut* wesentlich größer, wenn der Absatz sich im Jahresmittel auf hohem Niveau bewegt. Nur die relativen oder prozentualen Änderungen sind über die Zeit stabil und hängen nicht davon ab, wie groß die Werte des Prozesses X_t gerade sind. Analog zur Irrfahrt, bei der die absoluten Zuwächse $Z_t = X_t - X_{t-1}$ als u.i.v. angenommen werden, spricht man von einer *geometrischen Irrfahrt* $\{X_t; t \geq 0\}$, wenn die *relativen Zuwächse*

$$R_t = \frac{X_t}{X_{t-1}}, \quad t = 1, 2, \dots$$

u.i.v. sind. Ein *geometrischer Binomialprozess* hat z.B. die Form

$$X_t = R_t \cdot X_{t-1} = X_0 \cdot \prod_{k=1}^t R_k \tag{4.5}$$

wobei X_0, R_1, R_2, \dots unabhängig und für $u > 1, d < 1$:

$$P(R_k = u) = p, \quad P(R_k = d) = 1 - p.$$

Da $E[R_k] = (u-d)p + d$, folgt aus (4.5) mit den Rechenregeln für unabhängige Zufallsgrößen, dass $E[X_t]$ geometrisch (d.h. mit exponentieller Rate) steigt oder fällt, je nachdem ob $E[R_k] > 1$ oder $E[R_k] < 1$ ist:

$$E[X_t] = E[X_0] \cdot (E[R_1])^t = E[X_0] \cdot \{(u-d)p + d\}^t.$$

Wenn $E[R_k] = 1$, bleibt der Prozess im Mittel stabil. Dies ist der Fall, wenn

$$p = \frac{1-d}{u-d}.$$

Für $d = \frac{1}{u}$, d.h. für einen Prozess, der nach zwei Zeiteinheiten wieder den Ausgangszustand erreicht haben kann, vereinfacht sich diese Beziehung zu:

$$p = \frac{1}{u + 1}.$$

Aus (4.5) folgt durch logarithmieren:

$$\ln X_t = \ln X_0 + \sum_{k=1}^t \ln R_k.$$

Der Prozess $\tilde{X}_t = \ln X_t$ zum geometrischen Binomialprozess ist also selbst ein gewöhnlicher Binomialprozess mit Anfangswert $\ln X_0$ und Zuwächsen $Z_k = \ln R_k$, für die gilt:

$$P(Z_k = \ln u) = p, \quad P(Z_k = \ln d) = 1 - p.$$

Für große t ist \tilde{X}_t näherungsweise normalverteilt. $X_t = \exp(\tilde{X}_t)$ ist damit selbst näherungsweise lognormalverteilt.

4.5. Binomialmodelle mit zustandsabhängigen Zuwächsen

Binomialprozesse und auch allgemeinere Irrfahrten beschreiben den Verlauf eines Aktienkurses bestenfalls lokal. Sie gehen davon aus, dass die Verteilung der Zuwächse $Z_t = X_t - X_{t-1}$ stets dieselbe ist - unabhängig davon, ob der Kurs mittlerweile deutlich über oder unter dem Anfangskurs X_0 liegt. Geometrische Irrfahrten lassen den Zuwachs $X_t - X_{t-1} = (R_t - 1) X_{t-1}$ vom erreichten Kursniveau X_{t-1} abhängen (und sind damit keine Prozesse mit unabhängigen Zuwächsen mehr). Will man den Aktienkursverlauf über einen größeren Bereich modellieren, sind jedoch auch diese Prozesse noch zu einfach, um den Einfluss des erreichten Kursniveaus auf die zukünftige Entwicklung zu beschreiben. Eine einfache Klasse von Prozessen, die diesen Effekt berücksichtigen, sind die Binomialprozesse mit zustandsabhängigen (und eventuell auch zeitabhängigen) Zuwächsen:

$$X_t = X_{t-1} + Z_t, \quad t = 1, 2, \dots \quad (4.6)$$

$$P(Z_t = u) = p(X_{t-1}, t), \quad P(Z_t = -d) = 1 - p(X_{t-1}, t).$$

Die Zuwächse sind jetzt weder unabhängig noch identisch verteilt, denn die Verteilung von Z_t hängt vom erreichten Kursniveau X_{t-1} und eventuell auch von der Zeit ab. Die festen Funktionen $p(x, t)$ ordnen jedem möglichen Wert des Prozesses zur Zeit t und jedem t eine Wahrscheinlichkeit zu. Stochastische Prozesse $\{X_t; t \geq 0\}$, die nach (4.6) konstruiert werden, sind also keine Prozesse mit unabhängigen Zuwächsen mehr, haben aber immer noch die *Markoff-Eigenschaft*.

Entsprechend lassen sich geometrische Binomialprozesse mit zustandsabhängigen relativen Zuwächsen definieren (für $u > 1$, $d < 1$):

$$X_t = R_t \cdot X_{t-1} \quad (4.7)$$

$$P(R_t = u) = p(X_{t-1}, t), \quad P(R_t = d) = 1 - p(X_{t-1}, t).$$

Prozesse der Form (4.6) und (4.7) sind in dieser Allgemeinheit in erster Linie für theoretische Untersuchungen geeignet, da es ohne weitere Annahmen schwierig ist, die Wahrscheinlichkeiten $p(x, t)$ aus Beobachtungen des tatsächlichen Aktienkursverlaufs zu schätzen. Diese allgemeinen Binomialmodelle (bzw. auch die analog definierten Trinomialmodelle) sind aber in einem ganz anderen Zusammenhang nützlich, und zwar für die numerische Lösung von Differentialgleichungen wie z.B. der Black-Scholes-Gleichung für amerikanische Optionen.

5. Stochastische Integrale und Differentialgleichungen

Dieser Abschnitt stellt das Handwerkzeug bereit, das für Bewertung von Optionen benötigt wird. Dabei spielen stochastische Prozesse in stetiger Zeit eine wesentliche Rolle, die als Lösungen stochastischer Differentialgleichungen definiert werden. Um diese Begriffe ohne langwierige Vorarbeiten verständlich zu machen, benutzen wir wiederholt Approximationen durch stochastische Prozesse in diskreter Zeit und bauen auf den in Abschnitt 4 erarbeiteten Grundlagen auf.

Ein stochastischer Prozess in stetiger Zeit $\{X_t; t \geq 0\}$ besteht wiederum aus zeitlich angeordneten Zufallsvariablen, wobei die Zeitvariable t jetzt aber kontinuierlich ist, d.h. alle reellen Zahlen durchläuft.

Aktienkurse sind eigentlich Prozesse in diskreter Zeit. Um die Black-Scholes-Formel herzuleiten, werden sie durch Prozesse in stetiger Zeit approximiert, da man mit diesen wesentlich leichter analytisch rechnen kann. Um anschließend solche Prozesse auf dem Rechner zu simulieren oder - für amerikanische Optionen - den Optionswert numerisch zu berechnen, werden die Prozesse in stetiger Zeit wieder durch Prozesse in diskreter Zeit angenähert. Wir wechseln also zwischen diskreter und stetiger Zeit, je nachdem, was für die aktuellen Rechnungen bequemer ist.

5.1. Der Wiener-Prozess

Wir beginnen mit einer einfachen symmetrischen Irrfahrt $\{X_n; n \geq 0\}$, die in 0 anfängt ($X_0 = 0$). Die u.i.v. Zuwächse seien wieder $Z_n = X_n - X_{n-1}$:

$$P(Z_n = 1) = P(Z_n = -1) = \frac{1}{2}.$$

Wir beschleunigen jetzt den Prozess, indem wir die Zeiteinheit, d.h. die Zeitspanne zwischen zwei aufeinanderfolgenden Beobachtungen, immer kleiner werden lassen. Gleichzeitig werden auch die Änderungen, die der Prozess in der Zeiteinheit erfahren kann, immer kleiner. Genauer gesagt betrachten wir einen stochastischen Prozess $\{X_t^\Delta; t \geq 0\}$ in stetiger Zeit, der nach einer Zeitspanne von jeweils Δt mit Wahrscheinlichkeit $\frac{1}{2}$ um Δx wächst oder um Δx fällt. Zwischen diesen Sprüngen nach oben oder unten ist der Prozess konstant (alternativ

könnten wir auch linear interpolieren). Zur Zeit $t = n \cdot \Delta t$ ist der Prozess dann:

$$X_t^\Delta = \sum_{k=1}^n Z_k \cdot \Delta x = X_n \cdot \Delta x$$

wobei die Zuwächse $Z_1 \Delta x, Z_2 \Delta x, \dots$ unabhängig voneinander mit Wahrscheinlichkeit jeweils $\frac{1}{2}$ die Werte Δx oder $-\Delta x$ annehmen. Aus Abschnitt 4.1 wissen wir:

$$E[X_t^\Delta] = 0, \quad \text{Var}(X_t^\Delta) = (\Delta x)^2 \cdot \text{Var}(X_n) = (\Delta x)^2 \cdot n = t \cdot \frac{(\Delta x)^2}{\Delta t}.$$

Wir lassen jetzt $\Delta t, \Delta x$ immer kleiner werden. Damit der Grenzprozess in vernünftigem Sinn existiert, muss $\text{Var}(X_t^\Delta)$ endlich bleiben. Andererseits soll $\text{Var}(X_t^\Delta)$ auch nicht gegen 0 konvergieren, da der Grenzprozess sonst nicht mehr zufällig und damit für uns uninteressant wäre. Also müssen wir wählen:

$$\Delta t \rightarrow 0, \quad \Delta x = c \cdot \sqrt{\Delta t}, \quad \text{so dass} \quad \text{Var}(X_t^\Delta) \rightarrow c^2 t.$$

Wenn Δt klein ist, dann ist $n = t/\Delta t$ groß, so dass die Zufallsvariable X_n der einfachen symmetrischen Irrfahrt ungefähr normalverteilt $N(0, n)$ ist, und somit für alle t (nicht nur für diejenigen der Form $t = n \Delta t$)

$$\mathcal{L}(X_t^\Delta) \approx N(0, n(\Delta x)^2) \approx N(0, c^2 t).$$

Der Grenzprozess $\{X_t; t \geq 0\}$, den wir aus $\{X_t^\Delta; t \geq 0\}$ für $\Delta t \rightarrow 0, \Delta x = c \sqrt{\Delta t}$ erhalten, hat also die Eigenschaften:

- (i) X_t ist $N(0, c^2 t)$ -verteilt für alle $t \geq 0$.
- (ii) $\{X_t; t \geq 0\}$ hat *unabhängige Zuwächse*, d.h. für $0 \leq s < t$ ist $X_t - X_s$ unabhängig von X_s (da die Irrfahrt $\{X_n; n \geq 0\}$, über die die $\{X_t^\Delta; t \geq 0\}$ definiert sind, unabhängige Zuwächse hat)
- (iii) Für $0 \leq s < t$ ist der Zuwachs $(X_t - X_s)$ $N(0, c^2 \cdot (t - s))$ -verteilt, d.h. seine Verteilung hängt nur von der Länge $t - s$ des Zeitintervalls, über das der Zuwachs betrachtet wird, ab (dies folgt aus (i) und (ii) und den Eigenschaften der Normalverteilung).

Ein stochastischer Prozess $\{X_t; t \geq 0\}$ in stetiger Zeit mit den Eigenschaften (i)-(iii) heißt *Wiener-Prozess* oder *Brownsche Bewegung* mit Start in 0 ($X_0 = 0$). Den Standard-Wiener-Prozess, der sich durch die Wahl $c = 1$ ergibt, bezeichnen wir im folgenden stets mit $\{W_t; t \geq 0\}$. Für ihn gilt für alle $0 \leq s < t$:

$$E[W_t] = 0, \quad \text{Var}(W_t) = t$$

$$\begin{aligned} \text{Cov}(W_t, W_s) &= \text{Cov}(W_t - W_s + W_s, W_s) = \text{Cov}(W_t - W_s, W_s) + \text{Cov}(W_s, W_s) \\ &= 0 + \text{Var}(W_s) = s \end{aligned}$$

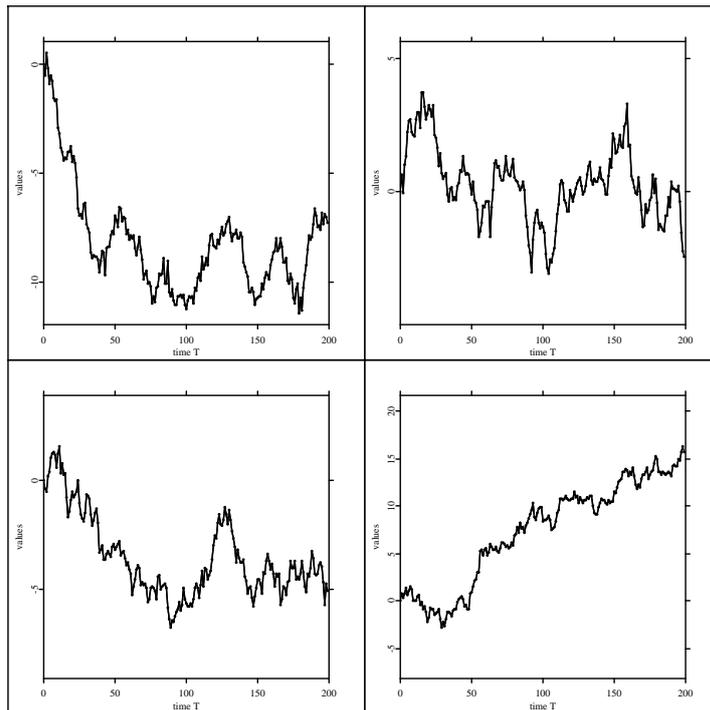


Abbildung 5.1.: Typische Pfade von Wiener-Prozessen [SFMWienerProcess.xpl](#)

Wie jeden stochastischen Prozess in stetiger Zeit können wir eine Realisation oder einen Pfad des Wiener-Prozesses als *zufällig ausgewählte Funktion* der Zeit auffassen. Man kann mit gewissem mathematischen Aufwand zeigen, dass die Pfade mit Wahrscheinlichkeit 1 *stetig* sind; der Wiener-Prozess hat keine Sprünge:

$$P(W_t \text{ ist stetig als Funktion von } t) = 1 .$$

Darüberhinaus fluktuiert W_t extrem stark; die Pfade sind zwar stetig, aber sehr erratisch. Man kann zeigen, dass die Pfade mit Wahrscheinlichkeit 1 nirgends differenzierbar sind.

Als Prozess mit unabhängigen Zuwächsen ist der Wiener-Prozess automatisch ein *Markoff-Prozess*. Für $0 \leq s < t$ ist $W_t = W_s + (W_t - W_s)$, d.h. W_t hängt nur von W_s und dem von der Vergangenheit bis zur Zeit s unabhängigen Zuwachs $W_t - W_s$ ab:

$$\begin{aligned} &P(a < W_t < b \mid W_s = x, \text{ Information über } W_{t'}, 0 \leq t' < s) \\ &= P(a < W_t < b \mid W_s = x) \end{aligned}$$

Aus den Eigenschaften (i)-(iii) lässt sich die bedingte Verteilung von W_t gegeben die Kenntnis von $W_s = x$ explizit angeben; sie ist $N(x, t-s)$, da der Zuwachs $(W_t - W_s)$ $N(0, t-s)$ -verteilt ist:

$$P(a < W_t < b \mid W_s = x) = \int_a^b \frac{1}{\sqrt{t-s}} \varphi\left(\frac{y-x}{\sqrt{t-s}}\right) dy.$$

Geht man bei den obigen Überlegungen statt von einer symmetrischen von einer beliebigen einfachen Irrfahrt $\{X_n; n \geq 0\}$ aus, die einen Trend

$$E[X_n] = n(2p - 1)$$

hat, so ist der Prozess X_t^Δ nicht mehr im Mittel 0, sondern

$$\begin{aligned} E[X_t^\Delta] &= n \cdot (2p - 1) \cdot \Delta x = (2p - 1) \cdot t \cdot \frac{\Delta x}{\Delta t} \\ \text{Var}(X_t^\Delta) &= n \cdot 4p(1 - p) \cdot (\Delta x)^2 = 4p(1 - p) \cdot t \cdot \frac{(\Delta x)^2}{\Delta t} \end{aligned}$$

Für $\Delta t \rightarrow 0$, $\Delta x = \sqrt{\Delta t}$, $p = \frac{1}{2}(1 + \mu\sqrt{\Delta t})$ erhalten wir für alle t :

$$E[X_t^\Delta] \rightarrow \mu t, \quad \text{Var}(X_t^\Delta) \rightarrow t$$

Der Grenzprozess ist der *Wiener-Prozess* $\{X_t; t \geq 0\}$ mit *Drift* oder *Trend* μt . Er ergibt sich auf einfache Weise aus dem Standard-Wiener-Prozess:

$$X_t = \mu t + W_t,$$

verhält sich also genauso wie der gewöhnliche Wiener-Prozess mit der einzigen Ausnahme, dass er nicht im Mittel um Null schwankt, sondern im Mittel linear wächst ($\mu > 0$) oder fällt ($\mu < 0$).

5.2. Stochastische Integration

Um anschließend einen stochastischen Prozess als Modell für Aktienkurse als Lösung einer stochastischen Differentialgleichung einführen zu können, benötigen wir den Begriff des Itô-Integrals, eines stochastischen Integrals bzgl. des Wiener-Prozesses. Formal ähnelt die Konstruktion des Itô-Integrals derjenigen eines deterministischen Stieltjes-Integrals, wobei die Funktion, bzgl. der integriert wird, zufällig - genauer: ein Pfad des Wiener-Prozesses - ist. Da der Integrand ebenfalls zufällig - d.h. ein Pfad eines stochastischen Prozesses - sein kann, muss man genaue Analysen bezüglich der wechselseitigen Abhängigkeit von Integrand und Wiener-Prozess miteinbeziehen.

Der Prozess, den wir integrieren wollen, sei $\{Y_t; t \geq 0\}$, und $\{W_t; t \geq 0\}$ sei der Standard-Wiener-Prozess. Für die Definition eines stochastischen Integrals benötigen wir die Annahme, dass $\{Y_t; t \geq 0\}$ *nicht antizipierend* (vorwegnehmend) ist. Intuitiv bedeutet dies, dass der Verlauf des Prozesses bis zu einem Zeitpunkt s keine Informationen über die zukünftigen Zuwächse $W_t - W_s$, $t > s$, des Wiener-Prozesses enthält. Insbesondere sind Y_s und $W_t - W_s$ unabhängig.

Wie stets bei der Einführung eines Integralbegriffs definieren wir das *Itô-Integral* bzgl. des Wiener-Prozesses als Grenzwert von (zufällig) gewichteten Summen der (zufälligen) Funktion $\{Y_t; t \geq 0\}$:

$$I_n = \sum_{k=1}^n Y_{(k-1)\Delta t} \cdot (W_{k\Delta t} - W_{(k-1)\Delta t}), \quad \Delta t = \frac{t}{n} \quad (5.1)$$

$$\int_0^t Y_s dW_s = \lim_{n \rightarrow \infty} I_n$$

Der asymptotische limes ist hier als Grenzwert von Zufallsgrößen im quadratischen Mittel zu verstehen, d.h. genauer gilt

$$E\left\{\left(\int_0^t Y_s dW_s - I_n\right)^2\right\} \rightarrow 0, \quad n \rightarrow \infty.$$

Entscheidend ist bei dieser Definition, dass die Summanden von I_n jeweils ein Produkt von zwei unabhängigen Zufallsgrößen sind: dem Wert $Y_{(k-1)\Delta t}$ des integrierten Prozesses am *linken Rand* des kleinen Zeitintervalls $[(k-1)\Delta t, k\Delta t]$ und dem Zuwachs $W_{k\Delta t} - W_{(k-1)\Delta t}$ des Wiener-Prozesses in diesem Intervall.

Die wesentliche Eigenschaft von $\{Y_t; t \geq 0\}$, nicht antizipierend zu sein, lässt sich etwas mehr präzisieren, ohne zu sehr in die technischen Details zu gehen.

DEFINITION 5.1 (Informationsmenge, nicht-antizipierend)

\mathcal{F}_t bezeichne für jedes $t \geq 0$ eine Familie von Ereignissen (mit der Struktur einer σ -Algebra, d.h. gewisse Kombinationen von Ereignissen aus \mathcal{F}_t sind wieder in \mathcal{F}_t), die die bis zur Zeit t verfügbare Information enthält. \mathcal{F}_t besteht aus solchen Ereignissen, von denen bis zur Zeit t feststeht, ob sie eintreffen oder nicht. Wir nehmen an:

$\mathcal{F}_s \subset \mathcal{F}_t$ für $s < t$ (die verfügbare Information wächst mit der Zeit)

$\{a < Y_t < b\} \in \mathcal{F}_t$ (Y_t enthält keine Information über Ereignisse nach der Zeit t) Wir nennen

$\{a < W_t < b\} \in \mathcal{F}_t$ (der Wiener-Prozess ist an die

$W_t - W_s$ unabhängig zu \mathcal{F}_s für $s < t$ Informationsentwicklung angepasst)

dann \mathcal{F}_t die Informationsmenge zur Zeit t und den Prozess $\{Y_t; t \geq 0\}$ nicht-antizipierend bzgl. der Informationsmengen $\mathcal{F}_t; t \geq 0$.

Der Prozess $\{Y_t\}$ heißt nicht-antizipierend, da er wegen der zweiten Forderung keine Informationen aus der Zukunft vorwegnimmt. Die zeitliche Entwicklung der Informationsmengen \mathcal{F}_t und der Zufallsvariablen Y_t, W_t sind aneinander angepasst.

Der Wert des stochastischen Integrals hängt entscheidend davon ab, an welcher Stelle des Intervalls $[(k-1)\Delta t, k\Delta t]$ der Wert der Zufallsfunktion Y_s in (5.1) eingesetzt wird. Als Beispiel betrachten wir $Y_t = W_t, t \geq 0$, d.h. wir integrieren den Wiener-Prozess bzgl. sich selbst. In (5.1) ersetzen wir $(k-1)\Delta t$ durch eine beliebige Stelle $t(n, k)$ im Intervall $[(k-1)\Delta t, k\Delta t]$. Würden wir definieren:

$$\int_0^t W_s dW_s = \lim_{n \rightarrow \infty} \sum_{k=1}^n W_{t(n,k)} (W_{k\Delta t} - W_{(k-1)\Delta t})$$

so müssten auch die Erwartungswerte konvergieren, also:

$$\begin{aligned} \mathbb{E} \left[\sum_{k=1}^n W_{t(n,k)} (W_{k\Delta t} - W_{(k-1)\Delta t}) \right] &= \sum_{k=1}^n \text{Cov}(W_{t(n,k)}, W_{k\Delta t} - W_{(k-1)\Delta t}) \\ &= \sum_{k=1}^n (t(n,k) - (k-1)\Delta t) \rightarrow \mathbb{E} \left[\int_0^t W_s dW_s \right] \end{aligned}$$

wegen der Rechenregeln für die Kovarianzen des Wiener-Prozesses (s. oben). Für $t(n,k) = (k-1)\Delta t$ - den Fall des Itô-Integrals - erhalten wir 0, für $t(n,k) = k\Delta t$ erhalten wir $n \cdot \Delta t = t$, und für passend gewählte Folgen $t(n,k)$ könnten wir jeden Wert zwischen 0 und t als Erwartungswert des stochastischen Integrals erhalten. Um $\int_0^t W_s dW_s$ einen eindeutigen Wert zuzuweisen, müssen wir uns also auf eine Folge $t(n,k)$ einigen.

Zur Illustration, wie Itô-Integrale berechnet werden und dass sie anderen Rechenregeln als gewöhnliche Integrale folgen, zeigen wir:

$$\int_0^t W_s dW_s = \frac{1}{2}(W_t^2 - W_0^2) - \frac{t}{2} = \frac{1}{2}(W_t^2 - t) \quad (5.2)$$

Da sich bei Summation der Differenzen $W_{k\Delta t}^2 - W_{(k-1)\Delta t}^2$ alle Terme bis auf den ersten und letzten wegheben und $n\Delta t = t$, haben wir

$$\begin{aligned} \frac{1}{2}(W_t^2 - W_0^2) &= \frac{1}{2} \sum_{k=1}^n (W_{k\Delta t}^2 - W_{(k-1)\Delta t}^2) \\ &= \frac{1}{2} \sum_{k=1}^n (W_{k\Delta t} - W_{(k-1)\Delta t})(W_{k\Delta t} + W_{(k-1)\Delta t}) \\ &= \frac{1}{2} \sum_{k=1}^n (W_{k\Delta t} - W_{(k-1)\Delta t})^2 + \sum_{k=1}^n (W_{k\Delta t} - W_{(k-1)\Delta t}) W_{(k-1)\Delta t} \end{aligned}$$

Der zweite Term konvergiert gegen $\int_0^t W_s dW_s$. Der erste Term ist eine Summe von n unabhängig, identisch verteilten Zufallsgrößen und stimmt daher nach dem Gesetz der großen Zahlen näherungsweise mit seinem Erwartungswert überein, d.h. mit

$$\frac{n}{2} \mathbb{E}[(W_{k\Delta t} - W_{(k-1)\Delta t})^2] = \frac{n}{2} \Delta t = \frac{t}{2}.$$

Für glatte, z.B. stetig differenzierbare, Funktionen f_s gilt $\int_0^t f_s df_s = \frac{1}{2}(f_t^2 - f_0^2)$. Das stochastische Integral (5.2) enthält dagegen noch den Zusatzterm $-\frac{t}{2}$, da der lokale Zuwachs des Wiener-Prozesses über ein Intervall der Länge Δt von der Größenordnung seiner Standardabweichung - also $\sqrt{\Delta t}$ - ist, während der entsprechende Zuwachs einer glatten Funktion f_s proportional zu Δt - also für $\Delta t \rightarrow 0$ wesentlich kleiner - ist.

5.3. Stochastische Differentialgleichungen

Der Wiener-Prozess fluktuiert um seinen Erwartungswert 0 und wird dementsprechend durch symmetrische Irrfahrten approximiert. Wie bei Irrfahrten ist aber auch bei stochastischen

Prozessen in stetiger Zeit der Fall interessant, dass ein Prozess im Mittel wächst oder fällt, d.h. dass er einen *Trend* oder eine *Drift* hat. Aus dem Wiener-Prozess mit beliebigem σ (vgl. Abschnitt 5.1) erhalten wir so den *allgemeinen Wiener-Prozess* $\{X_t; t \geq 0\}$ mit *Driftrate* μ und *Varianzrate* σ^2 :

$$X_t = \mu \cdot t + \sigma \cdot W_t \quad , \quad t \geq 0 \quad (5.3)$$

Der allgemeine Wienerprozess X_t ist zur Zeit t demnach $N(\mu t, \sigma^2 t)$ -verteilt. Für den Zuwachs des allgemeinen Wiener-Prozesses über ein kleines Zeitintervall Δt erhalten wir so

$$X_{t+\Delta t} - X_t = \mu \cdot \Delta t + \sigma(W_{t+\Delta t} - W_t)$$

Für $\Delta t \rightarrow 0$ geht dies in die differentielle Form über:

$$dX_t = \mu \cdot dt + \sigma \cdot dW_t \quad (5.4)$$

Dies ist nur eine andere Schreibweise für die Beziehung (5.3), die sich auch in Integralform schreiben lässt:

$$X_t = \int_0^t \mu ds + \int_0^t \sigma dW_s \quad (5.5)$$

da $\int_0^t dW_s = W_t - W_0 = W_t$, wie man sofort aus der Definition des stochastischen Integrals sieht.

In der differentiellen Form geht (5.4) nun davon aus, dass der lokale Trend, dessen Größe die Driftrate μ wiedergibt, und die lokale Variabilität, beschrieben durch den Parameter σ , stets konstant sind. Eine wesentlich größere und zum Modellieren vieler Vorgänge in Natur und Wirtschaft geeignete Klasse stochastischer Prozesse erhält man, wenn μ und σ in (5.4) von Zeit und vom erreichten Niveau abhängen dürfen. Solche Prozesse $\{X_t; t \geq 0\}$, die wir *Itô-Prozesse* nennen, sind als Lösungen *stochastischer Differentialgleichungen* gegeben:

$$dX_t = \mu(X_t, t)dt + \sigma(X_t, t)dW_t \quad (5.6)$$

Intuitiv bedeutet dies:

$$X_{t+\Delta t} - X_t = \mu(X_t, t)\Delta t + \sigma(X_t, t)(W_{t+\Delta t} - W_t),$$

d.h. der Zuwachs des Prozesses in einem kleinen Intervall der Länge Δt nach der Zeit t ist $\mu(X_t, t) \cdot \Delta t$ zuzüglich einer zufälligen Fluktuation, die $N(0, \sigma^2(X_t, t) \cdot \Delta t)$ -verteilt ist. Präzise wird eine Lösung von (5.6) als ein stochastischer Prozess definiert, der die Integralgleichung

$$X_t - X_0 = \int_0^t \mu(X_s, s)ds + \int_0^t \sigma(X_s, s)dW_s \quad (5.7)$$

erfüllt. (5.6) ist in diesem Sinn nur eine Kurzschreibweise für (5.7). Es folgt sofort für $0 \leq t' < t$:

$$X_t = X_{t'} + \int_{t'}^t \mu(X_s, s)ds + \int_{t'}^t \sigma(X_s, s)dW_s$$

Da der Zuwachs des Wiener-Prozesses zwischen t' und t nicht von den Ereignissen bis zur Zeit t' abhängt, folgt daraus, dass ein Itô-Prozess die *Markoff-Eigenschaft* besitzt.

Diskrete Approximationen von (5.6) und (5.7), die sich auch für Simulationen von Itô-Prozessen eignen, erhalten wir, wenn wir den Prozess zwischen 0 und t nur in regelmässigen Abständen $k\Delta t$, $k = 0, \dots, n$, $n\Delta t = t$, beobachten. Mit $X_k = X_{k\Delta t}$ und $Z_k = (W_{k\Delta t} - W_{(k-1)\Delta t})/\sqrt{\Delta t}$ erhalten wir

$$X_{k+1} - X_k = \mu(X_k, k) \cdot \Delta t + \sigma(X_k, k) \cdot Z_{k+1} \cdot \sqrt{\Delta t}$$

bzw. mit den Abkürzungen $\mu_k(X) = \mu(X, k)\Delta t$, $\sigma_k(X) = \sigma(X, k)\sqrt{\Delta t}$:

$$X_n - X_0 = \sum_{k=1}^n \mu_{k-1}(X_{k-1}) + \sum_{k=1}^n \sigma_{k-1}(X_{k-1}) \cdot Z_k$$

mit unabhängig, identisch verteilten $N(0, 1)$ -Zufallsgrößen Z_1, Z_2, \dots .

 [SFMSimCIR.xpl](#),  [SFMSimOU.xpl](#)

5.4. Der Aktienkurs als stochastischer Prozess

Aktienkurse sind an sich stochastische Prozesse in *diskreter Zeit*, die wegen der eingeschränkten Messgenauigkeit auch nur *diskrete Werte* annehmen. Dennoch werden oft stochastische Prozesse in *stetiger Zeit* als Modelle benutzt, da sie rechnerisch nicht so aufwendig wie diskrete Modelle - z.B. der Binomial- oder Trinomialprozess - sind. Letztere sind jedoch oft anschaulicher und sind besonders nützlich für Simulationen.

Der allgemeine Wiener-Prozess $dX_t = \mu dt + \sigma dW_t$ eignet sich selbst weniger als Aktienkursmodell, da er zum einen negative Aktienkurse zulassen würde, zum anderen die lokale Variabilität größer ist, wenn der Kurs selbst sich auf hohem Niveau bewegt. Daher wird in einem allgemeinen Ansatz der Börsenkurs S_t einer Aktie als Itô-Prozess modelliert:

$$dS_t = \mu(S_t, t)dt + \sigma(S_t, t)dW_t$$

Dieses Modell hängt von den nicht von vorneherein bekannten Funktionen $\mu(X, t)$ und $\sigma(X, t)$ ab. Eine brauchbare, einfache Variante, in der nur noch zwei reellwertige Modellparameter μ und σ unbekannt sind, erhält man durch folgende Überlegung: Die Rendite als prozentualer Zuwachs des eingesetzten Kapitals soll im Mittel nicht vom aktuellen Kurs bei Kauf der Aktien und schon gar nicht von der *Einheit* abhängen (EUR, USD, ...), in der der Aktienwert gemessen wird. Außerdem soll die mittlere Rendite wie bei anderen Anlageformen proportional zur Länge des Anlagezeitraums sein. Zusammen ergibt sich die Forderung

$$\frac{E[dS_t]}{S_t} = \frac{E[S_{t+dt} - S_t]}{S_t} = \mu \cdot dt.$$

Da $E[dW_t] = 0$, ist diese Bedingung bei gegebenem Ausgangskurs S_t erfüllt, wenn

$$\mu(S_t, t) = \mu \cdot S_t.$$

Darüberhinaus wird analog angesetzt:

$$\sigma(S_t, t) = \sigma \cdot S_t,$$

was die Tatsache berücksichtigt, dass die absolute Größe der Fluktuationen des Kurses sich proportional ändert, wenn wir den Kurs in einer anderen Einheit messen. Zusammengefasst modellieren wir den *Aktienkurs* S_t als Lösung der stochastischen Differentialgleichung

$$dS_t = \mu \cdot S_t dt + \sigma \cdot S_t \cdot dW_t$$

μ ist die zu *erwartende Aktienrendite*, σ die *Volatilität*. Ein solcher Prozess $\{S_t; t \geq 0\}$ heißt *geometrische Brownsche Bewegung*, da

$$\frac{dS_t}{S_t} = \mu dt + \sigma dW_t.$$

Mit Itô's Lemma, welches wir im Abschnitt 5.5 vorstellen, lässt sich zeigen, dass für einen passenden allgemeinen Wiener-Prozess $\{Y_t; t \geq 0\}$

$$S_t = e^{Y_t} \quad \text{bzw.} \quad Y_t = \ln S_t.$$

Da Y_t normalverteilt ist, ist S_t somit *lognormalverteilt*. So wie Irrfahrten diskrete Approximationen für den allgemeinen Wiener-Prozess liefern, lässt sich die geometrische Brownsche Bewegung, d.h. das einfache Modell für den Aktienkurs, durch geometrische Irrfahrten annähern.

5.5. Itô's Lemma

Ein entscheidendes Hilfsmittel beim Umgang mit stochastischen Differentialgleichungen ist das Lemma von Itô. Ist $\{X_t, t \geq 0\}$ ein Itô-Prozess:

$$dX_t = \mu(X_t, t)dt + \sigma(X_t, t) dW_t, \tag{5.8}$$

so interessiert man sich oft für den Verlauf anderer stochastischer Prozesse, die Funktionen von X_t sind: $Y_t = g(X_t)$. $\{Y_t; t \geq 0\}$ lässt sich dann ebenfalls als Lösung einer stochastischen Differentialgleichung beschreiben, und aus dieser Gleichung lassen sich interessierende Eigenschaften von Y_t wie z.B. das mittlere Wachstum mit der Zeit t ablesen.

Um die Gleichung für $\{Y_t, t \geq 0\}$ heuristisch herzuleiten, nehmen wir an, dass g genügend oft differenzierbar ist. Aus der Taylorreihenentwicklung folgt:

$$\begin{aligned} Y_{t+dt} - Y_t &= g(X_{t+dt}) - g(X_t) \\ &= g(X_t + dX_t) - g(X_t) \\ &= \frac{dg}{dX}(X_t) \cdot dX_t + \frac{1}{2} \frac{d^2g}{dX^2}(X_t) \cdot (dX_t)^2 + \dots \end{aligned} \tag{5.9}$$

wobei die Punkte für Terme vernachlässigbarer Größenordnung (für $dt \rightarrow 0$) stehen. Die für $dt \rightarrow 0$ dominanten Terme in dX_t sind aufgrund der Definition (5.8)

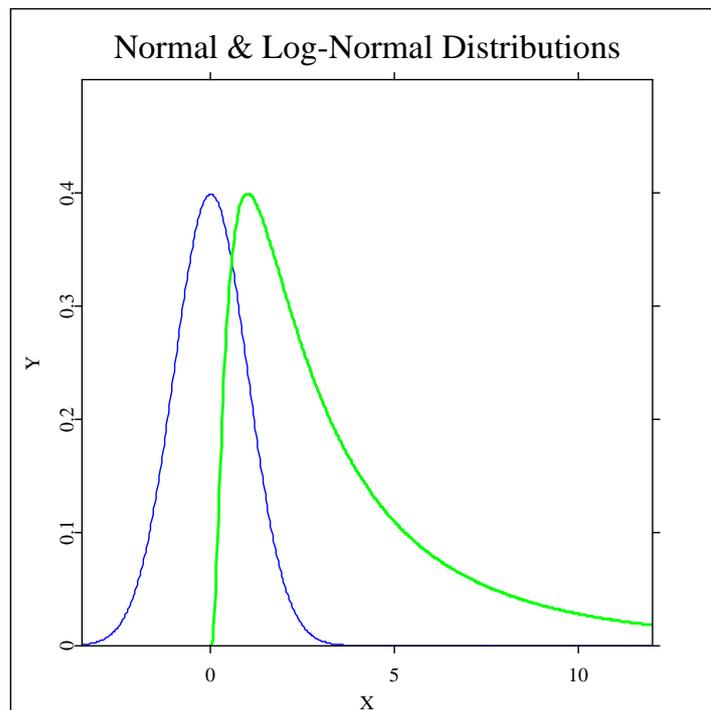


Abbildung 5.2.: Vergleich der Dichten von lognormal und normal verteilten Zufallsgrößen
[SFMLogNormal.xpl](#)

der Driftterm $\mu(X_t, t)dt$ von der Größenordnung dt und der Volatilitätsterm $\sigma(X_t, t)dW_t$ von der Größenordnung \sqrt{dt} .

Dabei berücksichtigen wir $E[(dW_t)^2] = dt$, so dass $dW_t = W_{t+dt} - W_t$ von der Größenordnung seiner Standardabweichung, also \sqrt{dt} , ist. Terme von kleinerer Größenordnung als dt vernachlässigen wir hier bei diesem heuristischen Argument. Daher können wir $(dX_t)^2$ durch einen einfacheren Ausdruck ersetzen:

$$\begin{aligned}(dX_t)^2 &= (\mu(X_t, t)dt + \sigma(X_t, t) dW_t)^2 \\ &= \mu^2(X_t, t)(dt)^2 + 2\mu(X_t, t) \sigma(X_t, t) dt dW_t + \sigma^2(X_t, t)(dW_t)^2\end{aligned}$$

Wir sehen, dass der erste Term von der Größenordnung $(dt)^2$ ist. Der zweite ist von der Größenordnung $dt \cdot \sqrt{dt}$, so dass beide vernachlässigt werden können. Der dritte Term hat allerdings die Größenordnung dt . Genauer kann man zeigen, dass für $dt \rightarrow 0$

$$(dW_t)^2 = dt.$$

Mit dieser einfachen Identität lassen sich Rechenregeln für stochastische Differentialgleichungen aus den bekannten Regeln für den Umgang mit deterministischen Funktionen (wie z.B. der Taylorentwicklung) ableiten. Unter Vernachlässigung von Termen kleinerer Größenordnung als dt erhalten wir so aus (5.9) die folgende Form von *Itô's Lemma*:

LEMMA 5.1 (Itô's Lemma)

$$\begin{aligned}dY_t &= dg(X_t) \\ &= \left(\frac{dg}{dX}(X_t) \cdot \mu(X_t, t) + \frac{1}{2} \frac{d^2g}{dX^2}(X_t) \cdot \sigma^2(X_t, t) \right) dt \\ &\quad + \frac{dg}{dX}(X_t) \cdot \sigma(X_t, t) dW_t\end{aligned}$$

bzw. - durch Weglassen des Zeitindex t und des Arguments X_t der Funktion g und ihrer Ableitungen:

$$dg = \left(\frac{dg}{dX}\mu(X, t) + \frac{1}{2} \frac{d^2g}{dX^2}\sigma^2(X, t) \right) dt + \frac{dg}{dX}\sigma(X, t)dW.$$

BEISPIEL 5.1

Als Beispiel betrachten wir $Y_t = \ln S_t$, den Logarithmus der geometrischen Brownschen Bewegung. Für $g(X) = \ln X$ haben wir $\frac{dg}{dX} = \frac{1}{X}$, $\frac{d^2g}{dX^2} = -\frac{1}{X^2}$, und für die geometrische Brownsche Bewegung ist $\mu(X, t) = \mu X$, $\sigma(X, t) = \sigma X$, so dass Itô's Lemma ergibt:

$$\begin{aligned}dY_t &= \left(\frac{1}{S_t}\mu S_t - \frac{1}{2} \frac{1}{S_t^2}\sigma^2 S_t^2 \right) dt + \frac{1}{S_t} \cdot \sigma S_t dW_t \\ &= \left(\mu - \frac{1}{2}\sigma^2 \right) dt + \sigma dW_t\end{aligned}$$

Der Logarithmus des Aktienkurses ist also ein allgemeiner Wiener-Prozess mit Driftrate $\mu^ = \mu - \frac{1}{2}\sigma^2$ und Varianzrate σ^2 . Da $Y_t \sim N(\mu^*t, \sigma^2t)$ -verteilt ist, ist S_t selbst lognormalverteilt mit Parametern μ^*t und σ^2t .*

Eine allgemeine Form von *Itô's Lemma* für Funktionen $g(X, t)$, die auch noch von der Zeit t abhängen dürfen, lautet :

LEMMA 5.2 (Itô's Lemma für explizit zeitabhängige Funktionen)

$$dg = \left(\frac{\partial g}{\partial X} \cdot \mu(X, t) + \frac{1}{2} \frac{\partial^2 g}{\partial X^2} \sigma^2(X, t) + \frac{\partial g}{\partial t} \right) dt + \frac{\partial g}{\partial X} \sigma(X, t) dW \quad (5.10)$$

$Y_t = g(X_t, t)$ ist also wieder ein Itô-Prozess, wobei als zusätzlicher Term in der Driftrate $\frac{\partial g}{\partial t}(X_t, t)$ auftaucht.

5.6. Literaturhinweise

Die in diesem Abschnitt kurz zusammengefassten Resultate gehören zu den zentralen Themen der stochastischen Analysis. Zahlreiche Bücher führen auf unterschiedlichen Niveaus in diesen Kalkül der stochastischen Integrale und stochastischen Differentialgleichungen ein, zum Beispiel ?, ? oder ?.

6. Black-Scholes-Optionsmodell

6.1. Die Black-Scholes-Differentialgleichung

Einfache, allgemein akzeptierte ökonomische Annahmen reichen nicht aus, um eine rationale Optionspreistheorie zu entwickeln. Die Voraussetzung eines perfekten Marktes hat in Kapitel 2.1 nur zum Herleiten elementarer Arbitragebeziehungen gereicht, denen Optionspreise genügen müssen und die sich daher als Test für fortgeschrittene Modellierungsansätze eignen. Die explizite Berechnung eines Optionspreises als Funktion von Zeit und Kurs sowie seiner zugrundeliegenden Parameter K, T ist damit aber nicht möglich. Hierfür wird ein Modell für den Kurs des zugrundeliegenden Finanzinstruments (Aktie, Devise, ...) gebraucht, bei dem es sich in aller Regel um einen stochastischen Prozess in diskreter oder stetiger Zeit handelt. Prozesse in stetiger Zeit haben den Vorteil, dass sich mit ihnen viele Problemstellungen leichter analytisch behandeln lassen. Prozesse in diskreter Zeit betrachten wir vorerst nur als Approximationen, mit denen sich leichter numerische Berechnungen durchführen lassen. Im zweiten Teil des Buches werden sie als eigenständige Modelle in der Finanzzeitreihenanalyse diskutiert.

Ein auch heute noch häufig verwendetes Modell für Aktienkurse, das auch die Grundlage des klassischen Black-Scholes-Ansatzes darstellt, ist die geometrische Brownsche Bewegung. Bei Annahme dieses Modells erfüllt der Aktienkurs S_t die einfache stochastische Differentialgleichung

$$dS_t = \mu S_t dt + \sigma S_t dW_t. \quad (6.1)$$

Äquivalent hierzu ist die Annahme, dass der Prozess der Renditen eine gewöhnliche Brownsche Bewegung ist, d.h.

$$\frac{dS_t}{S_t} = \mu dt + \sigma dW_t. \quad (6.2)$$

Die Drift μ gibt die erwartete Rendite des Objektes je Zeiteinheit. Die Volatilität σ ist ein Maß für die zu erwartende Größe der zufälligen Schwankungen der Rendite um ihren Erwartungswert. μ und σ hängen voneinander ab und sind für Anlageentscheidungen auf der Basis von Risikopräferenzen wesentlich: je größer die erwartete Rendite, also μ , sein soll, ein umso höheres, durch σ quantifiziertes Risiko muss in der Regel eingegangen werden.

Das Modell der geometrischen Brownschen Bewegung liefert eine brauchbare und in der Praxis bewährte Näherung für Aktienkurse über kurze und mittlere Zeiträume, zum Beispiel

über die Laufzeit typischer Aktienoptionen. Inzwischen sind zwar eine Reihe von Diskrepanzen zwischen Realität und Modell bekannt, z.B. Abweichungen der Volatilitätsfunktion $\sigma(x, t)$ im allgemeinen Modell (5.8) von der linearen Form $\sigma \cdot x$ der geometrischen Brownschen Bewegung in bestimmten Situationen, aber der Black-Scholes-Ansatz wird dennoch als erste einfache Näherung zur Optionsbewertung benutzt. Die prinzipielle Idee zur Ableitung von Optionspreisen lässt sich außerdem auf allgemeinere Aktienkursmodelle übertragen.

Der Ansatz von Black und Scholes beruht auf der bereits in Kapitel 2.1 benutzten Idee, den Wert eines Portfolios, das die interessierende Option enthält, durch ein zweites Portfolio zu duplizieren, das ausschließlich aus Finanzinstrumenten mit bekanntem Preis besteht. Dieses Duplikationsportfolio wird so gewählt, dass beide Portfolios zu einem Zeitpunkt T , typischerweise dem Fälligkeitstermin der Option, wertgleich sind. Dann folgt aus der Annahme eines perfekten Marktes und insbesondere aus der Arbitragefreiheit die Wertgleichheit zu jedem früheren Zeitpunkt. Hierfür gibt es zwei äquivalente Möglichkeiten, die wir am Beispiel eines Calls auf eine Aktie mit Kurs S_t erläutern:

1. Eines der Portfolios besteht nur aus dem Call, dessen Preis ermittelt werden soll. Das Duplikationsportfolio wird aus Aktien und risikolosen Zerobonds zusammengestellt, deren Anzahlen sich kontinuierlich an den sich ändernden Kurs anpassen müssen. Der Nominalwert der Zerobonds kann ohne Einschränkung als 1 gesetzt werden, da ihre Anzahl frei wählbar ist. Zur Zeit t bestehen die beiden Portfolios aus

Portfolio A: Eine Kaufoption auf die Aktie mit Terminkurs K und Fälligkeit T .

Portfolio B: $n_t = n(S_t, t)$ Aktien und $m_t = m(S_t, t)$ Zerobonds mit Nominalwert $B_T = 1$ und Fälligkeit T .

2. Eines der Portfolios ist ein perfektes Hedge-Portefeuille aus Aktien und verkauften Calls (mit Short Selling). Durch eine dynamische Hedge-Strategie wird dafür gesorgt, dass das Portfolio jederzeit risikolos ist, d.h. dass die Calls gerade die Gewinne und Verluste der Aktien neutralisieren. Das Duplikationsportfolio ist dementsprechend ebenfalls risikofrei und besteht ausschließlich aus Zerobonds. Auch dieser Ansatz erfordert wie im Fall 1. eine kursabhängige kontinuierliche Anpassung im Bestand der Portfolios. Zur Zeit t enthalten die beiden Portfolios

Portfolio A: Eine Aktie und $n_t = n(S_t, t)$ (über Short selling) verkaufte Kaufoptionen auf die Aktie mit Terminkurs K und Fälligkeit T .

Portfolio B: $m_t = m(S_t, t)$ Zerobonds mit Nominalwert $B_T = 1$ und Fälligkeit T .

Sei $T^* = T$, wenn der Besitzer die Kaufoption verfallen lässt, und sei T^* andernfalls der Zeitpunkt, zu dem die Option ausgeübt wird. Für einen europäischen Call ist z.B. stets $T^* = T$, während ein amerikanischer Call vorzeitig ausgeübt werden kann. Aus der Wertgleichheit

des Calls und des Duplikationsportfolios in Ansatz 1. oder aus der Wertgleichheit des Hedgeportfolios mit dem risikolosen Bondportfolio in Ansatz 2. zu jedem Zeitpunkt $t \leq T^*$ folgt jeweils dieselbe partielle Differentialgleichung für den Wert des Calls, die sogenannte *Black-Scholes-Gleichung*.

Der Ansatz von Black und Scholes ist nicht auf die Preisbestimmung einfacher Puts und Calls beschränkt, sondern lässt sich auf jedes derivative Finanzinstrument \mathcal{F} auf ein zugrundeliegendes Objekt mit Kurs S_t anwenden, wenn S_t eine geometrische Brownsche Bewegung ist und wenn der Preis F_t des Derivats bis zu seiner Ausübung nur eine Funktion des aktuellen Kurses und der Zeit ist: $F_t = F(S_t, t)$. Dann existiert nach dem folgenden Satz ein das Finanzinstrument duplizierendes Portfolio und der obige Ansatz 1. ist auf die Preisbestimmung anwendbar. Bei allgemeinen Derivaten muss das Duplikationsportfolio nicht nur zum Ausübungszeitpunkt T^* denselben Wert haben, sondern auch vorher schon denselben Cashflow, d.h. dieselben entnehmbaren Erträge oder zuzuschießenden Kosten, wie das Derivat generieren. Die Existenz eines perfekten Hedge-Portfolios als Grundlage von Ansatz 2. kann analog gezeigt werden.

SATZ 6.1

Sei der Kurs S_t eines Objektes eine geometrische Brownsche Bewegung (6.1). Sei \mathcal{F} ein von dem Objekt abhängiges derivatives Finanzinstrument mit Fälligkeit T . Sei $T^* \leq T$ der Ausübungszeitpunkt des Derivats bzw. $T^* = T$, falls \mathcal{F} nicht ausgeübt wird. Sei der Wert des Derivats zu jedem Zeitpunkt $t \leq T^*$ durch eine Funktion $F(S_t, t)$ von Kurs und Zeit gegeben.

- a) Es existiert ein Portfolio aus dem zugrundeliegenden Objekt und risikolosen Anleihen, das das Derivat in dem Sinn dupliziert, dass es bis zur Zeit T^* denselben Cashflow generiert und zur Zeit T^* denselben Wert wie \mathcal{F} hat.
- b) Die Wertfunktion $F(S, t)$ des Derivats erfüllt die Differentialgleichung von Black und Scholes

$$\frac{\partial F(S, t)}{\partial t} - rF(S, t) + bS \frac{\partial F(S, t)}{\partial S} + \frac{1}{2} \sigma^2 S^2 \frac{\partial^2 F(S, t)}{\partial S^2} = 0, \quad t \leq T^*, \quad (6.3)$$

BEWEIS:

Der Einfachheit halber gehen wir davon aus, dass das Objekt eine Aktie mit dem stetigen Dividendenertrag d und damit den Bestandshaltekosten $b = r - d$ ist, wobei r wieder den kontinuierlichen Zinssatz bezeichnet. Wir beschränken uns außerdem auf den Fall, dass \mathcal{F} ein Derivat mit der Aktie als zugrundeliegendem Objekt ist und dass \mathcal{F} vor der Zeit T^* keinen Cashflow erzeugt.

Wir bilden ein Portfolio aus $n_t = n(S_t, t)$ Aktien und aus $m_t = m(S_t, t)$ Zerobonds mit Fälligkeit T und Nominalwert $B_T = 1$.

$$B_t = B_T e^{-r(T-t)} = e^{r(t-T)}$$

sei der auf den Zeitpunkt t abgezinste Wert eines Bonds. Den Wert des Portfolios zur Zeit t bezeichnen wir mit

$$V_t \equiv V(S_t, t) = n(S_t, t) \cdot S_t + m(S_t, t) \cdot B_t.$$

Zu zeigen ist, dass n_t und m_t so gewählt werden können, dass bei Ausübung bzw. am Ende der Laufzeit von \mathcal{F} Portfolio und Derivat wertgleich sind, d.h. $V(S_{T^*}, T^*) = F(S_{T^*}, T^*)$, und dass das Portfolio vorher keinen Cashflow erzeugt, d.h. dem Portfolio darf vor der Zeit T^* weder Geld entnommen noch ihm Geld zugeschossen werden. Alle Änderungen im Bestand müssen aus den Dividenden bzw. durch Kauf und Verkauf von Aktien bzw. Bonds realisiert werden.

Wir untersuchen zuerst, wie sich der Wert V_t des Portfolios in einem kleinen Zeitraum dt ändert. Dabei schreiben wir $dV_t = V_{t+dt} - V_t$, $dn_t = n_{t+dt} - n_t$ usw.

$$\begin{aligned} dV_t &= n_{t+dt}S_{t+dt} + m_{t+dt}B_{t+dt} - n_tS_t - m_tB_t \\ &= dn_tS_{t+dt} + n_t dS_t + dm_tB_{t+dt} + m_t dB_t, \end{aligned}$$

so dass

$$dV_t = dn_t(S_t + dS_t) + n_t dS_t + dm_t(B_t + dB_t) + m_t dB_t. \quad (6.4)$$

Da der stochastische Prozess S_t eine geometrische Brownsche Bewegung und damit ein Itô-Prozess (5.8) mit $\mu(x, t) = \mu x$ und $\sigma(x, t) = \sigma x$ ist, folgt aus der allgemeinen Form (5.10) des Lemmas von Itô zusammen mit (6.1)

$$dn_t = \frac{\partial n_t}{\partial t} dt + \frac{\partial n_t}{\partial S} dS_t + \frac{1}{2} \frac{\partial^2 n_t}{\partial S^2} \sigma^2 S_t^2 dt. \quad (6.5)$$

und eine völlig analoge Beziehung für m_t . Daraus folgt unter Verwendung von

$$(dS_t)^2 = (\mu S_t dt + \sigma S_t dW_t)^2 = \sigma^2 S_t^2 (dW_t)^2 + o(dt) = \sigma^2 S_t^2 dt + o(dt),$$

$$dB_t = r B_t dt, \quad dS_t \cdot dt = o(dt) \quad \text{und} \quad dt^2 = o(dt)$$

und vernachlässigen der Terme, die klein gegenüber dt sind:

$$dn_t(S_t + dS_t) = \left(\frac{\partial n_t}{\partial t} dt + \frac{\partial n_t}{\partial S} dS_t + \frac{1}{2} \frac{\partial^2 n_t}{\partial S^2} \sigma^2 S_t^2 dt \right) S_t + \frac{\partial n_t}{\partial S} \sigma^2 S_t^2 dt, \quad (6.6)$$

$$dm_t(B_t + dB_t) = \left(\frac{\partial m_t}{\partial t} dt + \frac{\partial m_t}{\partial S} dS_t + \frac{1}{2} \frac{\partial^2 m_t}{\partial S^2} \sigma^2 S_t^2 dt \right) B_t. \quad (6.7)$$

Die entscheidende Forderung, dass das Portfolio wie das zu duplizierende Derivat vor der Zeit T^* keinen Cashflow generiert, bedeutet nun, dass die beiden Komponenten $dn_t(S_t + dS_t)$ und $dm_t(B_t + dB_t)$ von dV_t in (6.4), die dem Kauf/Verkauf von Aktien bzw. Zerobonds entsprechen, aus den Dividenden finanziert werden müssen. Da eine Aktie in einem kleinen Zeitintervall der Länge dt die Dividende $d \cdot S_t \cdot dt$ abwirft, bedeutet dies

$$d \cdot n_t S_t \cdot dt = (r - b) \cdot n_t S_t \cdot dt = dn_t(S_t + dS_t) + dm_t(B_t + dB_t)$$

oder, nach Einsetzen von (6.6) und (6.7),

$$0 = (b-r)n_t S_t dt + \left(\frac{\partial m_t}{\partial t} dt + \frac{\partial m_t}{\partial S} dS_t + \frac{1}{2} \frac{\partial^2 m_t}{\partial S^2} \sigma^2 S_t^2 dt \right) B_t \\ + \left(\frac{\partial n_t}{\partial t} dt + \frac{\partial n_t}{\partial S} dS_t + \frac{1}{2} \frac{\partial^2 n_t}{\partial S^2} \sigma^2 S_t^2 dt \right) S_t + \frac{\partial n_t}{\partial S} \sigma^2 S_t^2 dt$$

Einsetzen von (6.1) und Zusammenfassen der stochastischen Anteile mit Differential dW_t , der deterministischen Anteile mit Differential dt , die den Driftparameter μ enthalten, und der restlichen deterministischen Terme liefert

$$0 = \left(\frac{\partial n_t}{\partial S} S_t + \frac{\partial m_t}{\partial S} B_t \right) \mu S_t dt \\ + \left\{ \left(\frac{\partial n_t}{\partial t} + \frac{1}{2} \frac{\partial^2 n_t}{\partial S^2} \sigma^2 S_t^2 \right) S_t + \frac{\partial n_t}{\partial S} \sigma^2 S_t^2 + \left(\frac{\partial m_t}{\partial t} + \frac{1}{2} \frac{\partial^2 m_t}{\partial S^2} \sigma^2 S_t^2 \right) B_t + (b-r)n_t S_t \right\} dt \\ + \left(\frac{\partial n_t}{\partial S} S_t + \frac{\partial m_t}{\partial S} B_t \right) \sigma S_t dW_t \quad (6.8)$$

Dies ist nur möglich, wenn der stochastische Anteil verschwindet, d.h.

$$\frac{\partial n_t}{\partial S} S_t + \frac{\partial m_t}{\partial S} B_t = 0. \quad (6.9)$$

Damit verschwindet der erste Term in (6.8) ebenfalls, und der mittlere Term muss schließlich auch 0 sein:

$$\left(\frac{\partial n_t}{\partial t} + \frac{1}{2} \frac{\partial^2 n_t}{\partial S^2} \sigma^2 S_t^2 \right) S_t + \frac{\partial n_t}{\partial S} \sigma^2 S_t^2 + \left(\frac{\partial m_t}{\partial t} + \frac{1}{2} \frac{\partial^2 m_t}{\partial S^2} \sigma^2 S_t^2 \right) B_t + (b-r)n_t S_t = 0 \quad (6.10)$$

Wir können diese Gleichung weiter vereinfachen, indem wir Gleichung (6.9) partiell nach S differenzieren:

$$\frac{\partial^2 n_t}{\partial S^2} S_t + \frac{\partial n_t}{\partial S} + \frac{\partial^2 m_t}{\partial S^2} B_t = 0 \quad (6.11)$$

und das Ergebnis in (6.10) einsetzen. Wir erhalten dann

$$\frac{\partial n_t}{\partial t} S_t + \frac{\partial m_t}{\partial t} B_t + \frac{1}{2} \frac{\partial n_t}{\partial S} \sigma^2 S_t^2 + (b-r)n_t S_t = 0 \quad (6.12)$$

Da der Aktienkurs S_t nicht explizit von der Zeit abhängt und somit $\partial S_t / \partial t = 0$, liefert Differenzieren des Portfoliowerts $V_t = n_t S_t + m_t B_t$ nach der Zeit

$$\frac{\partial V_t}{\partial t} = \frac{\partial n_t}{\partial t} S_t + \frac{\partial m_t}{\partial t} B_t + m_t \frac{\partial B_t}{\partial t} = \frac{\partial n_t}{\partial t} S_t + \frac{\partial m_t}{\partial t} B_t + m_t r B_t.$$

Daraus folgt

$$\frac{\partial n_t}{\partial t} S_t + \frac{\partial m_t}{\partial t} B_t = \frac{\partial V_t}{\partial t} - r m_t B_t = \frac{\partial V_t}{\partial t} - r(V_t - n_t S_t).$$

Durch Einsetzen dieser Gleichung in (6.12) eliminieren wir den m_t enthaltenden Term und erhalten

$$\frac{1}{2}\sigma^2 S_t^2 \frac{\partial n}{\partial S} + \frac{\partial V_t}{\partial t} + b n_t S_t - r V_t = 0 \quad (6.13)$$

Da der Bondwert B_t nicht vom Kurs abhängt und somit $\partial B_t / \partial S = 0$, liefert Differenzieren des Portfoliowerts $V_t = n_t S_t + m_t B_t$ nach dem Kurs

$$\frac{\partial V_t}{\partial S} = \frac{\partial n_t}{\partial S} S_t + n_t + \frac{\partial m_t}{\partial S} B_t = n_t$$

wegen Gleichung (6.9), d.h.

$$n_t = \frac{\partial V_t}{\partial S}. \quad (6.14)$$

n_t ist somit gleich dem sogenannten *Delta* oder der Hedge-Rate des Portfolios (vgl. Abschnitt 6.3.1). Da

$$m_t = \frac{V_t - n_t S_t}{B_t},$$

können wir das gesuchte Duplikationsportfolio konstruieren, wenn wir $V_t = V(S_t, t)$ kennen. Diese Funktion von Kurs und Zeit erhalten wir als Lösung der Differentialgleichung von Black und Scholes

$$\frac{\partial V(S, t)}{\partial t} - rV(S, t) + bS \frac{\partial V(S, t)}{\partial S} + \frac{1}{2}\sigma^2 S^2 \frac{\partial^2 V(S, t)}{\partial S^2} = 0, \quad (6.15)$$

die sich durch Einsetzen von (6.14) in Gleichung (6.13) ergibt. Zur Bestimmung von V fehlt noch ein Randbedingung, die wir aus der Forderung erhalten, dass der Cashflow des Portfolios bei Ausübung bzw. Fälligkeit des Derivats, d.h. zur Zeit T^* , mit dem Cashflow des Derivats übereinstimmt, also denselben Pay-off hat:

$$V(S_{T^*}, T^*) = F(S_{T^*}, T^*). \quad (6.16)$$

Da das Derivat jederzeit denselben Cashflow wie das Portfolio hat, erfüllt $F(S, t)$ ebenfalls die Black-Scholes- Differentialgleichung, und zu jedem Zeitpunkt $t \leq T^*$ gilt $F_t = F(S_t, t) = V(S_t, t) = V_t$.

□

Die Differentialgleichung von Black und Scholes beruht wesentlich auf der Annahme, dass der Kurs durch eine geometrische Brownsche Bewegung modelliert werden kann. Eine derartige Annahme ist nur gerechtfertigt, wenn die darauf aufbauende Theorie auch die Arbitragebeziehungen aus Kapitel 2.1 reproduziert, Wir überprüfen dies am Beispiel eines Terminkontraktes mit Terminkurs K und Fälligkeit T auf ein Objekt, mit dem stetige Bestandshaltkosten mit Rate b verbunden sind. Der Wert $V(S_t, t)$ hängt nur von Kurs und Zeit ab, erfüllt also die Bedingung von Satz 6.1. Aus Satz 2.1 erhalten wir nach Einsetzen der Definition der Restlaufzeit $\tau = T - t$

$$V(S, t) = S e^{(r-b)(t-T)} - K e^{r(t-T)}.$$

Durch Einsetzen überprüft man leicht, dass diese Funktion die eindeutige Lösung der Differentialgleichung (6.3) mit der Randbedingung $V(S, T) = S - K$ ist. Der Black-Scholes-Ansatz liefert also denselben Preis des Terminkontrakts wie die modellfreie Arbitragebeziehung.

Zum Schluss bemerken wir noch, dass das Modell der geometrischen Brownschen Bewegung für Aktienkurse über mittelgroße Zeiträume eine brauchbare Näherung liefert. Die Übertragung des Modells auf andere grundlegende Finanzinstrumente wie Devisen oder Anleihen ist nicht ohne weiteres möglich. Bondoptionen haben in der Regel eine deutlich längere Laufzeit als Aktienoptionen. Ihre Wertentwicklung hängt nicht vom Kurs, sondern von dem dann ebenfalls als stochastischer Prozess aufzufassenden Zins ab. Zur adäquaten Modellierung der Zinsentwicklung braucht man aber andere Arten stochastischer Prozesse, die wir in späteren Kapiteln diskutieren werden.

Wechselkurse lassen sich eingeschränkt als geometrische Brownsche Bewegung modellieren. Empirische Vergleiche zeigen, dass dieses Modell für manche Währungen oder Laufzeiten nicht geeignet ist, so dass im Einzelfall anhand von Daten überprüft werden muss, ob der Black-Scholes-Ansatz für die Bewertung von Devisenoptionen benutzt werden kann. Ist dies der Fall, so entspricht die Fremdwährung, die das zugrundeliegende Objekt der Option ist, einer Aktie mit einem kontinuierlichen Dividendenertrag, der durch die kontinuierliche Zinsrate d im Land der Fremdwährung gegeben ist. Die Bestandshaltekosten $b = r - d$ entsprechen der Differenz der Zinsen im Inland und im Land der Fremdwährung: Kauft der Investor die Fremdwährung vorzeitig, so kann er sein Kapital nicht mehr im Inland anlegen und verliert den Zinssatz r , gewinnt dafür aber den Zinssatz d , den er für die Fremdwährung in deren Herkunftsland erhält. Der Wert der Devisenoption ergibt sich dann durch Lösen der Differentialgleichung (6.3) von Black und Scholes mit der passenden, durch den Typ der Option bestimmten Randbedingung.

6.2. Die Black-Scholes-Formel für europäische Optionen

In diesem Abschnitt verwenden wir die Black-Scholes-Gleichung zur Berechnung des Preises europäischer Optionen. Wir halten uns an die Notation des vorigen Kapitels und schreiben

$$C(S, t) = C_{K,T}(S, t), \quad P(S, t) = P_{K,T}(S, t)$$

für den Wert zur Zeit $t \leq T$ eines europäischen Calls bzw. Puts mit Ausübungskurs K und Fälligkeitstermin T , wenn das zugrundeliegende Objekt, z.B. eine Aktie, zur Zeit t den Kurs $S_t = S$ hat. Der Wert der Kaufoption erfüllt dann die Differentialgleichung

$$rC(S, t) - bS \frac{\partial C(S, t)}{\partial S} - \frac{1}{2} \sigma^2 S^2 \frac{\partial^2 C(S, t)}{\partial S^2} = \frac{\partial C(S, t)}{\partial t}, \quad 0 \leq t \leq T, \quad 0 < S < \infty \quad (6.17)$$

$$C(S, T) = \max\{0, S - K\}, \quad 0 < S < \infty, \quad (6.18)$$

$$C(0, t) = 0, \quad \lim_{S \rightarrow \infty} C(S, t) - S = 0, \quad 0 \leq t \leq T. \quad (6.19)$$

Die erste Randbedingung (6.18) ergibt sich sofort aus der Definition des Calls, der nur ausgeübt wird, wenn $S_T > K$, und der dann den Gewinn $S_T - K$ einbringt. Die erste Bedingung in (6.19) ergibt sich aus der Beobachtung, dass die geometrische Brownsche Bewegung S_t in

der Null bleibt, wenn sie erst einmal den Wert Null angenommen hat, d.h. aus $S_t = 0$ für ein $t < T$ folgt automatisch $S_T = 0$, und der Call wird nicht ausgeübt. Die zweite Bedingung in (6.19) ergibt sich aus der Überlegung, dass die geometrische Brownsche Bewegung S_t den Ausübungskurs nur mit sehr geringer Wahrscheinlichkeit wieder unterschreitet, wenn sie erst einmal ein Niveau weit oberhalb von K erreicht hat. Wenn $S_t \gg K$ für ein $t < T$ dann gilt auch mit hoher Wahrscheinlichkeit $S_T \gg K$. Der Call wird also ausgeübt und liefert den Ertrag $S_T - K \approx S_T$.

Die Differentialgleichung (6.17) mit den Randbedingungen (6.18), (6.19) kann analytisch gelöst werden. Wir transformieren sie dazu in eine aus der Literatur bekannte Form. Zuerst ersetzen wir die Zeitvariable t durch die Restlaufzeit $\tau = T - t$. Dadurch wird das Problem mit Endbedingung (6.18) bei $t = T$ zu einem Anfangswertproblem bei $\tau = 0$. Anschließend multiplizieren wir (6.17) mit $2/\sigma^2$ und ersetzen die Parameter r, b durch

$$\alpha = \frac{2r}{\sigma^2}, \quad \beta = \frac{2b}{\sigma^2},$$

sowie die Variablen τ, S durch

$$v = \sigma^2(\beta - 1)^2 \frac{\tau}{2}, \quad u = (\beta - 1) \ln \frac{S}{K} + v.$$

Während für die alten Parameter $0 \leq S < \infty, 0 \leq t \leq T$ gegolten hat, ergibt sich für den Wertebereich der neuen Parameter

$$-\infty < u < \infty, \quad 0 \leq v \leq \frac{1}{2}\sigma^2(\beta - 1)^2 T \equiv v_T.$$

Schließlich setzen wir

$$g(u, v) = e^{r\tau} C(S, T - \tau)$$

und erhalten die neue Differentialgleichung

$$\frac{\partial^2 g(u, v)}{\partial u^2} = \frac{\partial g(u, v)}{\partial v}. \quad (6.20)$$

mit der Anfangsbedingung

$$g(u, 0) = K \max\{0, e^{\frac{u}{\beta-1}} - 1\} \equiv g_0(u), \quad -\infty < u < \infty. \quad (6.21)$$

Anfangswertprobleme dieser Art sind aus der Literatur über partielle Differentialgleichungen bestens bekannt. Sie treten beispielsweise bei der mathematischen Modellierung von Wärmeleitungs- und Diffusionsprozessen auf. Die Lösung ist

$$g(u, v) = \int_{-\infty}^{\infty} \frac{1}{2\sqrt{\pi v}} g_0(\xi) e^{-\frac{(\xi-u)^2}{4v}} d\xi.$$

Den Optionspreis erhalten wir daraus, indem wir die Variablen- und Parametertransformationen rückgängig machen. Wir schreiben im folgenden wieder wie in Kapitel 2.1 $C(S, \tau)$ für

den Call-Preis als Funktion der Restlaufzeit $\tau = T - t$ anstelle der aktuellen Zeit t . Dann haben wir

$$C(S, \tau) = e^{-r\tau} g(u, v) = e^{-r\tau} \int_{-\infty}^{\infty} \frac{1}{2\sqrt{\pi v}} g_0(\xi) e^{-\frac{(\xi-u)^2}{4v}} d\xi$$

Durch Substitution $\xi = (\beta-1) \ln(x/K)$ erhalten wir die Anfangsbedingung in ihrer ursprünglichen Form $\max\{0, x - K\}$. Ersetzen wir außerdem u und v wieder durch die Variablen S und τ , erhalten wir

$$C(S, \tau) = e^{-r\tau} \int_0^{\infty} \max(0, x - K) \frac{1}{\sqrt{2\pi\sigma}\sqrt{\tau}x} e^{-\frac{[\ln x - \{\ln S + (b - \frac{1}{2}\sigma^2)\tau\}]^2}{2\sigma^2\tau}} dx. \quad (6.22)$$

Für die geometrische Brownsche Bewegung ist $S_T - S_t$ lognormalverteilt, d.h. $\ln(S_T - S_t)$ ist normalverteilt mit den Parametern $(b - \frac{1}{2}\sigma^2)\tau$ und $\sigma^2\tau$. Die bedingte Verteilung von S_T gegeben $S_t = S$ ist demnach ebenfalls eine Lognormalverteilung mit den Parametern $\ln S + (b - \frac{1}{2}\sigma^2)\tau$ und $\sigma^2\tau$. Der Integrand von (6.22) ist aber bis auf den Faktor $\max(0, x - K)$ gerade die Dichte dieser Verteilung. Damit erhalten wir die interessante Interpretation des Callwertes $C(S, \tau)$ als abgezinster Erwartungswert der Endbedingung $\max(0, S_T - K)$ gegeben den aktuellen Kurs $S_t = S$:

$$C(S, \tau) = e^{-r\tau} E\{\max(0, S_T - K) \mid S_t = S\}. \quad (6.23)$$

Diese Eigenschaft ist bei der Herleitung numerischer Verfahren zur Berechnung von Optionswerten hilfreich. Zuerst nutzen wir aber aus, dass (6.22) ein Integral bzgl. der Dichte einer Lognormalverteilung enthält, um den Ausdruck weiter zu vereinfachen. Durch geeignete Substitution formen wir ihn in ein Integral bzgl. der Dichte einer Normalverteilung um und erhalten

$$C(S, \tau) = e^{(b-r)\tau} S \Phi(y + \sigma\sqrt{\tau}) - e^{-r\tau} K \Phi(y), \quad (6.24)$$

wobei y abkürzend für

$$y = \frac{\ln \frac{S}{K} + (b - \frac{1}{2}\sigma^2)\tau}{\sigma\sqrt{\tau}}, \quad (6.25)$$

steht und Φ die Verteilungsfunktion der Standardnormalverteilung bezeichnet:

$$\Phi(y) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^y e^{-\frac{z^2}{2}} dz.$$

(6.24) ist zusammen mit (6.25) die *Black-Scholes-Formel* für den Preis einer europäischen Kaufoption. In den Grenzfällen $S \gg K$ und $S = 0$ erhalten wir:

- Ist $S \gg K$, so ist $y \gg 0$ und damit $\Phi(y + \sigma\sqrt{\tau}) \approx \Phi(y) \approx 1$, so dass sich der Wert der Kaufoption im dividendenfreien Fall $b = r$ etwa wie $S - e^{-r\tau}K$ also wie der aktuelle Kurs abzüglich des auf den aktuellen Zeitpunkt abgezinnten Ausübungskurses verhält.

- Ist $S = 0$, so ist $y = -\infty$ und somit $\Phi(y) = 0$, so dass die Option wertlos ist: ($C(0, \tau) = 0$).

Die entsprechende Black-Scholes-Formel für den Preis $P(S, \tau)$ einer europäischen Verkaufsoption könnten wir genauso durch Lösen der Black-Scholes-Differentialgleichung mit den passenden Randbedingungen herleiten. Einfacher ist aber das Ausnutzen der Put-Call-Parität (Satz 2.3)

$$P(S, \tau) = C(S, \tau) - Se^{(b-r)\tau} + Ke^{-r\tau}.$$

Hieraus erhalten wir zusammen mit (6.24) sofort

$$P(S, \tau) = e^{-r\tau} K \Phi(-y) - e^{(b-r)\tau} S \Phi(-y - \sigma\sqrt{\tau}). \quad (6.26)$$

Für den Wert europäischer Kauf- und Verkaufsoptionen gibt es also explizite Formeln. Die beiden Terme in der Formel (6.24) für z.B. den Wert der Kaufoption lassen sich grob folgendermaßen interpretieren, wobei wir uns auf den dividendenfreien Fall $b = r$ beschränken. Der erste Term ist dann $S\Phi(y + \sigma\sqrt{\tau})$ und repräsentiert in erster Linie den Wert der Aktie, die der Käufer der Kaufoption im Fall der Ausübung seines Rechts beziehen kann. Der andere Term $e^{-r\tau} K\Phi(y)$ steht in erster Linie für den Wert des Ausübungskurses. Über die Variable y beeinflusst natürlich das Verhältnis S/K beide Terme.

Bei der Herleitung der Black-Scholes-Differentialgleichung haben wir gesehen, dass insbesondere der Wert einer Kaufoption durch Aktien und Anleihen dupliziert werden kann. Der in Aktien investierte Betrag ergibt sich dabei als $\frac{\partial C}{\partial S} S$, wobei $\frac{\partial C}{\partial S}$ die Hedge-Rate ist. Sie bestimmt das zum Hedgen notwendige, auch Delta genannte Verhältnis zwischen Aktien und Optionen. Leitet man die Black-Scholes-Formel (6.24) nach S ab, so erhält man

$$\frac{\partial C(S, t)}{\partial S} = \Phi(y + \sigma\sqrt{\tau}).$$

Der erste Term in (6.24) steht somit für den im Duplikationsportfolio in Aktien und der zweite für den in Anleihen investierten Betrag.

Da die analytischen Lösungsformeln die Auswertung der Standardnormalverteilungsfunktion erfordert, was nur numerisch erfolgen kann, muss man für die praktische Anwendung der Black-Scholes-Formel auf Näherungen der Normalverteilung zurückgreifen. Davon kann in geringem Maße der berechnete Optionswert abhängen. Zur Illustration betrachten wir verschiedene in der Literatur (beispielsweise [Hastings \(1955\)](#)) beschriebene Näherungsformeln.

a.) Die Normalverteilung kann folgendermaßen approximiert werden:

$$\Phi(y) \approx 1 - (a_1 t + a_2 t^2 + a_3 t^3) e^{-\frac{y^2}{2}}, \quad \text{mit}$$

$$\begin{aligned} t &= (1 + by)^{-1}, & b &= 0.332672527, \\ a_1 &= 0.17401209, & a_2 &= -0.04793922, \\ a_3 &= 0.373927817. \end{aligned}$$

Der Approximationsfehler hat unabhängig von y die Größenordnung $O(10^{-5})$.

 [SFMNormalApprox1.xpl](#)

b.)

$$\Phi(y) \approx 1 - (a_1 t + a_2 t^2 + a_3 t^3 + a_4 t^4 + a_5 t^5) e^{-\frac{y^2}{2}}, \quad \text{mit}$$

$$\begin{aligned} t &= (1 + by)^{-1}, & b &= 0.231641888, \\ a_1 &= 0.127414796, & a_2 &= -0.142248368, & a_3 &= 0.71070687, \\ a_4 &= -0.726576013, & a_5 &= 0.530702714. \end{aligned}$$

Der Fehler dieser Approximation liegt bei $O(10^{-7})$.

[SFMMNormalApprox2.xpl](#)

c.) Eine andere Approximation der Normalverteilung, deren Fehler von der Größenordnung $O(10^{-5})$ ist, lautet:

$$\Phi(y) \approx 1 - \frac{1}{2(a_1 t + a_2 t^2 + a_3 t^3 + a_4 t^4 + a_5 t^5)^8}, \quad \text{mit}$$

$$\begin{aligned} a_1 &= 0.099792714, & a_2 &= -0.044320135, & a_3 &= 0.009699203, \\ a_4 &= -0.000098615, & a_5 &= 0.00581551. \end{aligned}$$

[SFMMNormalApprox3.xpl](#)

d.) Als letzte Möglichkeit wollen wir hier noch die Taylorreihenzerlegung angeben:

$$\begin{aligned} \Phi(y) &\approx \frac{1}{2} + \frac{1}{\sqrt{2\pi}} \left(y - \frac{y^3}{1!2^1 3} + \frac{y^5}{2!2^2 5} - \frac{y^7}{3!2^3 7} + \dots \right) = \\ &= \frac{1}{2} + \frac{1}{\sqrt{2\pi}} \sum_{n=0}^{\infty} (-1)^n \frac{y^{2n+1}}{n! 2^n (2n+1)}. \end{aligned}$$

Mit Hilfe dieser Reihe kann die Normalverteilung beliebig genau angenähert werden, wobei natürlich mit der Genauigkeit auch die Anzahl der benötigten Summanden und somit der arithmetischen Operationen steigt.

[SFMMNormalApprox4.xpl](#)

Ein Vergleich aller vier Näherungsformeln ist in der Tabelle 6.1 enthalten. Die Taylorreihe wurde beim ersten Summanden, der betragsmäßig kleiner als 10^{-5} ist, abgebrochen. Die letzte Spalte gibt dabei die Nummer des letzten Summanden an.

Die Tabelle 6.2 zeigt den Optionspreis, der mit Hilfe der Formel (6.24) für eine konkrete europäische Kaufoption auf der Grundlage verschiedener Approximationen der Normalverteilung berechnet wurde.

6.3. Risikomanagement mit Hedge-Strategien

Das Handeln mit Optionen ist wegen der starken zufälligen Komponente mit einem vergleichsweise hohen Risiko verbunden. Fortgeschrittene Strategien zur Verringerung und Kontrolle dieses Risikos können aus der Black-Scholes-Formel (6.24) abgeleitet werden. Um die Problematik zu verdeutlichen, betrachten wir zuerst ein Beispiel und einige traditionelle Strategien.

x	norm-a	norm-b	norm-c	norm-d	iter
1.0000	0.8413517179	0.8413447362	0.8413516627	0.8413441191	6
1.1000	0.8643435425	0.8643338948	0.8643375717	0.8643341004	7
1.2000	0.8849409364	0.8849302650	0.8849298369	0.8849309179	7
1.3000	0.9032095757	0.9031994476	0.9031951398	0.9031993341	8
1.4000	0.9192515822	0.9192432862	0.9192361959	0.9192427095	8
1.5000	0.9331983332	0.9331927690	0.9331845052	0.9331930259	9
1.6000	0.9452030611	0.9452007087	0.9451929907	0.9452014728	9
1.7000	0.9554336171	0.9554345667	0.9554288709	0.9554342221	10
1.8000	0.9640657107	0.9640697332	0.9640670474	0.9640686479	10
1.9000	0.9712768696	0.9712835061	0.9712842148	0.9712839202	11
2.0000	0.9772412821	0.9772499371	0.9772538334	0.9772496294	12

Tabelle 6.1.: Verschiedene Approximationen der Normalverteilung

Aktienkurs S_t	230.00	EUR/Aktie		
Ausübungskurs K	210.00	EUR/Aktie		
Restlaufzeit $\tau = T - t$	0.50000	a		
stetiger Zinssatz r	0.04545			
Volatilität σ	0.25000	p.a.		
keine Dividende				
	norm-a	norm-b	norm-c	norm-d
Optionspreise	30.74261	30.74158	30.74352	30.74176

Tabelle 6.2.: Bestimmung des europäischen Callpreises mit verschiedenen Approximationen der Normalverteilung [SFMBSCopt1.xpl](#)

Aktueller Zeitpunkt t	6 Wochen
Laufzeit T	26 Wochen
Restlaufzeit $\tau = T - t$	20 Wochen = 0.3846 a
stetiger Jahreszinssatz r	5% p.a.
Jahresvolatilität der Aktie σ	20%
aktueller Aktienkurs S_t	98 EUR
Ausübungskurs K	100 EUR

Tabelle 6.3.: Daten des Beispiels

BEISPIEL 6.1

Eine Bank verkauft für 600000 EUR einen europäischen Call auf 100000 Stück einer dividendenfreien Aktie. Die Parameter der Option sind in der Tabelle 6.3 dargestellt.

Die Black-Scholes-Formel (6.24) mit $b = r$ wegen der Dividendenfreiheit ergibt als Wert der Option 480105 EUR, also rund 480000 EUR. Die Bank hat die Option um 120000 EUR teurer als ihren theoretischen Wert verkauft, setzt sich dafür aber dem Risiko von nennenswerten Verlusten aus.

Eine Risikomanagementstrategie ist es, nichts zu tun, d.h. eine ungedeckte Position (naked position) einzunehmen. Wenn die Option nach 20 Wochen ausgeübt wird, muss die Bank die Aktie zum dann aktuellen Kurs kaufen. Steigt der Kurs z.B. auf $S_T = 120$ EUR, so kostet dies effektiv $100000 \cdot (S_T - K) = 2000000$ EUR, also deutlich mehr als die durch den Verkauf der Option erzielten Einnahmen. Steigt andererseits der Kurs nicht über den Ausübungskurs K , so wird die Option nicht ausgeübt, und die Bank kann 600000 EUR als Gewinn verbuchen.  SFMBSCopt2.xpl

Das andere Extrem ist das Einnehmen einer gedeckten Position (covered position). Sofort nach dem Verkauf der Option kauft die Bank die bei Ausübung fälligen Aktien zum Preis von $100000 \cdot S_t = 9800000$ EUR. Wird im Falle $S_T > K$ die Option ausgeübt, so werden die Aktien zum Preis von $100000 \cdot K = 10000000$ EUR geliefert. Da dies abgezinst auf den Zeitpunkt t ungefähr dem ursprünglichen Kaufpreis der Aktien entspricht, beträgt der Gewinn der Bank in diesem Fall rund 600000 EUR, d.h. dem für den Verkauf der Option erzielten Betrag. Fällt dagegen der Aktienkurs auf z.B. $S_T = 80$ EUR, so wird die Option nicht ausgeübt. Die Bank verliert aber aus dem Kursverlust der von ihr gehaltenen Aktien rund 2000000 EUR, also wieder wesentlich mehr als der für die Option erzielte Erlös. Aus der Put-Call-Parität für europäische Optionen (Satz 2.3) folgt übrigens, dass das Risiko aus dem Verkauf eines gedeckten Calls identisch zum Risiko aus dem Verkauf eines ungedeckten Puts ist.

Beide Risikomanagementstrategien sind unbefriedigend, da die Kosten sehr stark zwischen 0 und großen Werten variieren. Nach Black-Scholes entstehen im Mittel Kosten von rund 480000 EUR, und ein perfekter Hedge sollte dafür sorgen, dass der Einfluss des Zufalls ausgeschaltet wird und genau diese Kosten entstehen.

Ein aufwendigeres Verfahren zum Hedgen, d.h. zum Beschränken des Risikos, das mit dem Verkauf eines Calls verbunden ist, ist die *Stop-Loss-Strategie*: Die die Option herausgebende Bank nimmt eine ungedeckte Position ein, solange der Aktienkurs unterhalb des Ausübungskurses ist ($S_t < K$), und wechselt in eine gedeckte Position, sobald der Call im Geld ist ($S_t > K$).

Die beim Ausüben der Option zu liefernden Aktien werden gekauft, sobald S_t K überschreitet, und wieder abgestoßen, sobald S_t unter K fällt.

Da sämtliche Käufe und Verkäufe nach der Zeit 0 zum Kurs K getätigt wurden, und da bei Fälligkeit T entweder keine Aktien gehalten werden ($S_T < K$) oder sie zum Kurs K an den Besitzer des Calls abgegeben werden ($S_T > K$), fallen nach der Zeit 0 nominell keine Kosten an.

Kosten entstehen durch diese Hedge-Strategie nur, wenn $S_0 > K$ ist, zur Zeit 0 also Aktien zum Kurs S_0 gekauft werden, die später zum Kurs K veräußert werden:

$$\text{Kosten für Stop-Loss-Hedgen: } \max[S_0 - K, 0].$$

Δt (Wochen)	5	4	2	1	$\frac{1}{2}$	$\frac{1}{4}$
L	1.02	0.93	0.82	0.77	0.76	0.76

Tabelle 6.4.: Leistung der Stop-Loss-Strategie

Dies liegt unter dem Black-Scholes-Preis $C(S_0, T)$, so dass Arbitrage durch Verkaufen einer Option und Hedgen mit der Stop-Loss-Strategie möglich wäre. Bei dieser Argumentation sind aber mehrere Dinge übersehen worden:

- Kauf und Verkauf der Aktien verursachen Transaktionskosten,
- durch die Kapitalbindung beim Kauf von Aktien vor der Zeit T entstehen Zinsverluste,
- praktisch sind Käufe und Verkäufe zum exakten Kurs K nicht möglich; bei steigendem Kurs werden die Aktien für $K + \delta$ gekauft, bei fallendem Kurs für $K - \delta$ verkauft für ein $\delta > 0$.

Praktisch werden Käufe und Verkäufe nur nach jeweils Δt Zeiteinheiten getätigt. Je größer Δt , desto größer ist in der Regel δ , desto seltener fallen aber die Transaktionskosten an. Hull (2000) hat in einer Monte Carlo-Studie mit $M = 1000$ simulierten Kursverläufen untersucht, wie gut sich die Stop-Loss-Strategie als Verfahren zur Verringerung des Risikos, das mit der Ausgabe einer Kaufoption verbunden ist, eignet. In jedem Simulationslauf werden die durch Stop-Loss-Verfahren verursachten Kosten $\Lambda_m, m = 1, \dots, M$, registriert und ihre Stichprobenvarianz

$$\hat{v}_\Lambda^2 = \frac{1}{M} \sum_{m=1}^M (\Lambda_m - \frac{1}{M} \sum_{j=1}^M \Lambda_j)^2$$

berechnet. Die Größe des nach Anwendung des Stop-Loss-Hedgens verbleibenden Risikos misst die zur Standardisierung durch den Preis des gehedgten Calls geteilte Stichprobenstandardabweichung

$$L = \frac{\sqrt{\hat{v}_\Lambda^2}}{C(S_0, T)}.$$

Das Ergebnis zeigt Tabelle 6.4. Ein perfekter Hedge würde das Risiko, d.h. die Variabilität der Kosten, völlig ausschalten und hätte den Wert $L = 0$.

6.3.1. Delta-Hedgen

Zur Begrenzung des Risikos beim Optionshandel werden komplexere Hedge-Strategien als die bisher betrachteten eingesetzt. Wir betrachten als Beispiel wieder den Verkauf eines Calls auf Aktien. Zuerst wird versucht, den Wert des Portfolios über kleine Zeitintervalle hinweg unempfindlich gegen kleine Schwankungen des Kurses der der Option zugrunde liegenden Aktie zu machen. Dies nennt man Delta-Hedgen. Anschließend werden andere der "Greek letters" (Gamma, Theta, Vega, Rho) betrachtet, um die Feinabstimmung des Hedge-Portfolios zu erreichen.

Unter dem *Delta* oder der *Hedge-Rate* einer Kaufoption versteht man die Ableitung des Optionspreises nach dem Aktienkurs oder - als diskrete Version - den Differenzenquotienten der Änderung ΔC des Optionspreises bei Änderung ΔS des Aktienkurses:

$$\Delta = \frac{\partial C}{\partial S} \quad \text{oder} \quad \Delta = \frac{\Delta C}{\Delta S}$$

Entsprechend ist das Delta anderer Anlageformen definiert. Die Aktie selbst hat den Wert S , also $\Delta = \partial S / \partial S = 1$. Ein Terminkontrakt auf eine dividendenfreie Aktie hat den Forward Price $V = S - K \cdot e^{-r\tau}$ (siehe Satz 2.1) und daher ebenfalls $\Delta = \partial V / \partial S = 1$. Aktien und Terminkontrakte sind daher in Delta-Hedge-Portfolios gegeneinander austauschbar. Letztere werden wegen der niedrigeren Transaktionskosten vorgezogen, wenn sie verfügbar sind.

BEISPIEL 6.2

Eine Bank verkauft Calls auf 2000 Aktien zum Preis von $C = 10$ EUR/Aktie bei einem Kurs $S_0 = 100$ EUR/Aktie. Das Delta des Calls sei $\Delta = 0.4$. Zum Hedgen werden die verkauften Calls im Portfolio durch Kauf von $\Delta \cdot 2000 = 800$ Aktien ergänzt. Kleine Änderungen des Optionswertes werden durch entsprechende Änderungen des Wertes der Aktien im Portfolio ausgeglichen. Steigt der Aktienkurs um 1 EUR, d.h. der Wert der Aktien im Portfolio um 800 EUR, so steigt der Optionswert eines Calls auf 1 Aktie um $\Delta C = \Delta \cdot \Delta S = 0.4$ EUR, der Wert aller Calls im Portfolio also um 800 EUR, was einen Verlust für den Verkäufer des Calls darstellt. Gewinn und Verlust gleichen sich aus, da das Delta der Optionsposition durch das Delta der Aktienposition ausgeglichen wird. Das Gesamtportfolio hat $\Delta = 0$; die Bank nimmt eine deltaneutrale Position ein.

Da das Delta von Optionen u.a. von Aktienkurs und Zeit abhängt, bleibt eine deltaneutrale Position nur kurzfristig erhalten. In der Praxis muss beim Delta-Hedgen das Portfolio regelmäßig den sich ändernden Gegebenheiten angepasst werden. Dies nennt man *Rebalanzieren*. Strategien zum Risikomanagement von Portfolios, die häufiges Umschichten beinhalten, sind als *dynamisches Hedgen* bekannt. Wir merken an, dass die Black-Scholes-Differentialgleichung (6.3) mit Hilfe eines dynamischen Hedgeportfolios hergeleitet werden kann, dessen Position durch kontinuierliches Umschichten deltaneutral gehalten wird. Dieser Ansatz ist analog zur Betrachtung eines die Option duplizierenden Portfolios.

BEISPIEL 6.3

Der Kurs der dem obigen Call zugrunde liegenden Aktie steigt innerhalb einer Woche auf 110 EUR. Durch die verstrichene Zeit und den geänderten Kurs erhöht sich das Delta der Option auf $\Delta = 0.5$. Um erneut eine deltaneutrale Position zu erreichen, müssen $(0.5 - 0.4) \cdot 2000 = 200$ Aktien zum Portfolio hinzugekauft werden.

Aus den Black-Scholes-Formeln für den Wert europäischer Calls und Puts auf dividendenfreie Aktien ergibt sich als Delta:

$$\begin{aligned} \Delta &= \frac{\partial C}{\partial S} = \Phi(y + \sigma\sqrt{\tau}) & (6.27) \\ \text{bzw. } \Delta &= \frac{\partial P}{\partial S} = \Phi(y + \sigma\sqrt{\tau}) - 1, \end{aligned}$$

wobei y wie in (6.25) definiert ist.

Δt (Wochen)	5	4	2	1	$\frac{1}{2}$	$\frac{1}{4}$
L	0.43	0.39	0.26	0.19	0.14	0.09

Tabelle 6.5.: Leistung des Delta-Hedgens

Das Delta (6.27) ist in Abbildung 6.1 als Funktion von Aktienkurs und Laufzeit dargestellt. Für steigende Aktienkurse geht das Delta gegen Eins, für fallende gegen Null. Mit anderen Worten, ist die Option sehr weit im Geld, so wird sie sehr wahrscheinlich am Laufzeitende ausgeübt werden, weshalb man als Stillhalter dieser Option zur Deckung des Risikos die Aktie (oder einen sehr großen Teil) halten sollte. Ist die Option andererseits sehr weit aus dem Geld, so wird sie sehr wahrscheinlich nicht ausgeübt werden, so dass man die Aktie nicht (oder nur einen kleinen Teil) zu halten braucht.

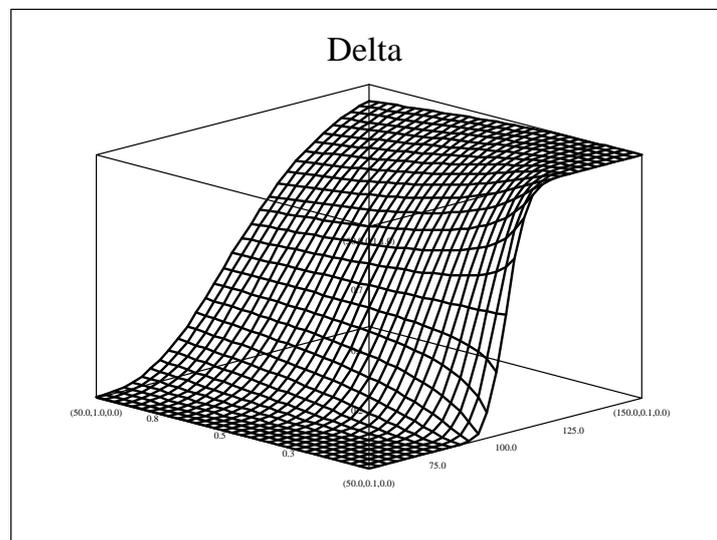


Abbildung 6.1.: Das Delta als Funktion von Aktienkurs (rechte Achse) und Laufzeit (linke Achse). [SFMdelta.xpl](#)

Weiterhin ist die Wahrscheinlichkeit p , dass eine Option aus dem Geld am Laufzeitende ausgeübt und eine Option im Geld nicht ausgeübt wird größer, je länger die Laufzeit ist. Das erklärt, weshalb das Delta für längere Laufzeiten flacher (linearer) wird.

Tabelle 6.5 nach Hull (2000) zeigt in derselben Weise wie Tabelle 6.4 für die Stop-Loss-Strategie die Leistung des Delta-Hedgens in Abhängigkeit von den Abständen Δt zwischen den Rebalanzierungen des Portfolios. Wenn Δt klein genug ist, wird das Risiko aus dem Verkauf des Calls sehr gut eingegrenzt. Für $\Delta t \rightarrow 0$ erhält man als Grenzfall die der Ableitung der Black-Scholes-Formel zugrunde liegende kontinuierliche Rebalanzierung und die perfekte Ausschaltung des Risikos ($L = 0$).

Wegen der Linearität der Ableitung ist das Delta Δ_p eines Portfolios, das aus w_1, \dots, w_m

Stück von Derivaten $1, \dots, m$ mit Deltas $\Delta_1, \dots, \Delta_m$ besteht, einfach:

$$\Delta_p = \sum_{j=1}^m w_j \Delta_j$$

BEISPIEL 6.4

Ein Portfolio aus verschiedenen USD-Optionen besteht aus

1. 200000 gekauften Calls (long position) mit Ausübungskurs 1.70 EUR und Fälligkeit in 4 Monaten. Das Delta einer auf einen einzelnen Dollar bezogenen Option ist $\Delta_1 = 0.54$.
2. 100000 verkaufte Calls (short position) mit Ausübungskurs 1.75 EUR und Fälligkeit in 6 Monaten mit einem Delta von $\Delta_2 = 0.48$.
3. 100000 verkauften Puts (short position) mit Ausübungskurs 1.75 EUR und Fälligkeit in 3 Monaten mit $\Delta_3 = -0.51$.

Das Delta des Portfolios ist (Wertsteigerungen verkaufter Optionen wirken sich negativ auf den Wert des Portfolios aus):

$$\begin{aligned} \Delta_p &= 200000 \cdot \Delta_1 - 100000 \cdot \Delta_2 - 100000 \cdot \Delta_3 \\ &= 111000 \end{aligned}$$

Das Portfolio kann durch Verkauf von 111000 USD oder eines passenden Terminkontrakts auf USD (beides mit $\Delta = 1$) deltaneutral gemacht werden.

6.3.2. Gamma und Theta

Beim Delta-Hedgen wird der Optionspreis C lokal durch eine im Aktienkurs S lineare Funktion approximiert. Wenn die Zeit Δt bis zum nächsten Rebalanzieren des Portfolios nicht sehr kurz ist, ist die Näherung unbefriedigend (vgl. Tabelle 6.5). Daher betrachtet man eine feinere Approximation, d.h. die Taylor-Entwicklung von C als Funktion von S und t :

$$\Delta C = C(S + \Delta S, t + \Delta t) - C(S, t) = \frac{\partial C}{\partial S} \cdot \Delta S + \frac{\partial C}{\partial t} \cdot \Delta t + \frac{1}{2} \frac{\partial^2 C}{\partial S^2} \Delta S^2 + o(\Delta t)$$

Dabei ist (wie wir schon im Beweis von Satz 6.1 gesehen haben) ΔS von der Größenordnung $\sqrt{\Delta t}$, und die in $o(\Delta t)$ zusammengefassten Terme sind von kleinerer Größenordnung als Δt . Vernachlässigt man alle Terme bis auf den ersten (von der Größenordnung $\sqrt{\Delta t}$), so erhält man die dem Delta-Hedgen zugrunde liegende Näherung

$$\Delta C \approx \Delta \cdot \Delta S$$

Berücksichtigt man auch die Terme der Größenordnung Δt , so folgt

$$\Delta C \approx \Delta \cdot \Delta S + \Theta \cdot \Delta t + \frac{1}{2} \Gamma \Delta S^2$$

wobei $\Theta = \partial C / \partial t$ das *Theta* und $\Gamma = \partial^2 C / \partial S^2$ das *Gamma* der Option ist. Θ heißt auch Zeitverfall (*time decay*) der Option. Für einen dividendenfreien Call erhält man aus der Black-Scholes-Formel (6.24):

$$\Theta = -\frac{\sigma S}{2\sqrt{\tau}} \varphi(y + \sigma\sqrt{\tau}) - rKe^{-r\tau}\Phi(y)$$

und

$$\Gamma = \frac{1}{\sigma S\sqrt{\tau}} \varphi(y + \sigma\sqrt{\tau}) \quad (6.28)$$

mit dem in (6.25) definierten y .

Das Gamma in (6.28) ist in Abbildung (6.2) geplottet als Funktion von Aktienkurs und Laufzeit. Optionen, deren Deltas am sensitivsten auf Kursänderungen reagieren, sind am Geld und haben kurze Laufzeiten. Bei diesen Optionen sind also häufige Umschichtungen erforderlich, wenn man Delta-Hedgen betreibt.

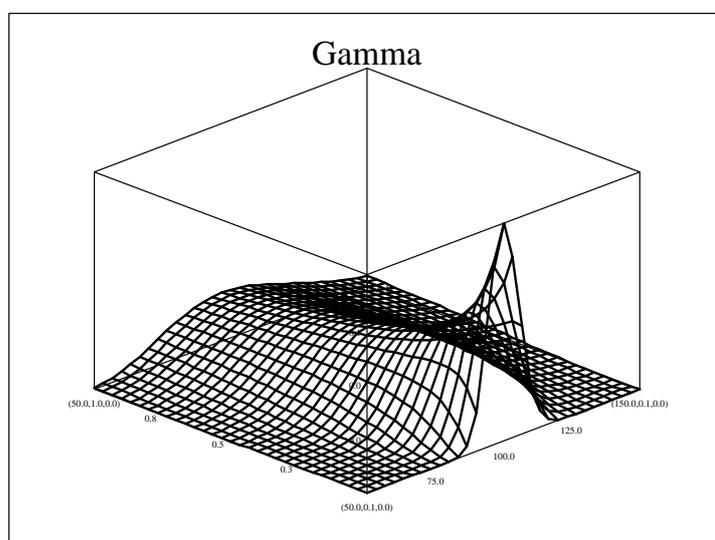


Abbildung 6.2.: Das Gamma als Funktion von Aktienkurs (rechte Achse) und Laufzeit (linke Achse). [SFMgamma.xpl](#)

Beim *Gamma-Hedgen* geht man von einem bereits deltaneutralen Portfolio aus und versucht es, durch Hinzukauf oder Verkauf von Derivaten gammaneutral zu machen, d.h. $\Gamma = 0$ zu erreichen und den Wert des Portfolios noch stabiler gegen Änderungen des Aktienkurses zu machen. Für das Gamma-Hedgen kann man allerdings weder Aktien noch Terminkontrakte benutzen, da beide konstantes Δ und daher $\Gamma = 0$ haben. Andererseits ist dies auch von Vorteil, da man diese Instrumente einsetzen kann, um ein bereits gammaneutrales Portfolio deltaneutral zu machen, ohne das Gamma zu ändern. Zum Gamma-Hedgen einer Position aus Optionen mit einem Gammawert Γ kann man z.B. w Stück einer in der Börse gehandelten Option mit einem Gamma von Γ_B einsetzen. Das Gamma des Portfolios ist dann $\Gamma + w\Gamma_B$, also 0 für $w = -\Gamma/\Gamma_B$.

BEISPIEL 6.5

Ein Portfolio aus USD-Optionen und US-Dollars sei bereits *deltaneutral* mit $\Gamma = -150000$. An der Terminbörse wird ein USD-Call mit $\Delta_B = 0.52$ und $\Gamma_B = 1.20$ angeboten. Das Portfolio wird durch Hinzufügen von $-\Gamma/\Gamma_B = 125000$ dieser Calls *gammaneutral*. Sein Δ ist danach aber $125000 \cdot \Delta_B = 65000$. Die *Deltaneutralität* kann z.B. durch Verkauf von 65000 USD aus dem Portfolio wiedergestellt werden, ohne das Gamma zu ändern.

Im Gegensatz zur Entwicklung des Aktienkurses ist der Ablauf der Zeit deterministisch und mit keinen risikosteigernden Unwägbarkeiten verbunden. Wenn Δ und Γ beide 0 sind, so ändert sich der Wert der Option (näherungsweise) risikofrei mit der Rate $\Theta = \Delta C/\Delta t$. Der Parameter Θ ist für die meisten Optionen negativ; eine Option verliert an Wert, je näher der Fälligkeitsterm rückt.

Aus der Black-Scholes-Gleichung (6.24) folgt für ein *deltaneutrales* Portfolio aus Aktienoptionen

$$rV = \Theta + \frac{1}{2}\sigma^2 S^2 \Gamma,$$

wobei V der Wert des Portfolios ist. Θ und Γ hängen auf einfache Weise voneinander ab, so dass Θ anstelle von Γ zum Gamma-Hedgen eines *deltaneutralen* Portfolios benutzt werden kann.

6.3.3. Rho und Vega

Der Black-Scholes-Ansatz geht davon aus, dass die Volatilität σ konstant ist. Das Auftreten von Smiles zeigt, dass diese Annahme in der Praxis nicht gerechtfertigt ist. Daher kann es sinnvoll sein, den Wert eines Portfolios auch unempfindlich gegenüber Änderungen der Volatilität zu machen. Dabei benutzt man das *Vega* eines Portfolios (in der Literatur manchmal auch Lambda oder Kappa genannt), das für z.B. einen Call definiert ist als $\mathcal{V} = \frac{\partial C}{\partial \sigma}$.

Aktien und Terminkontrakte haben $\mathcal{V} = 0$, so dass man beim Vega-Hedgen wieder auf an der Börse gehandelte Optionen zurückgreifen muss. Da ein *veganeutrales* Portfolio nicht automatisch *gammaneutral* ist, muss man beim Hedgen zwei verschiedene Optionen einsetzen, um gleichzeitig $\mathcal{V} = 0$ und $\Gamma = 0$ zu erreichen.

Aus der Black-Scholes-Formel (6.24) mit dem dort definierten y ergibt sich als Vega eines Calls auf eine dividendenfreie Aktie

$$\mathcal{V} = S\sqrt{\tau}\varphi(y + \sigma\sqrt{\tau}). \quad (6.29)$$

Da die Black-Scholes-Formel unter der Annahme konstanter Volatilität hergeleitet worden ist, ist es eigentlich nicht gerechtfertigt, \mathcal{V} durch Ableiten von (6.24) nach σ zu berechnen. Die obige Formel für \mathcal{V} ist aber sehr ähnlich zu einer Gleichung für $\partial C/\partial \sigma$, die aus einem allgemeineren Modell mit stochastischer Volatilität folgt, so dass sie näherungsweise benutzt werden kann.

Das Vega in (6.29) ist in Abbildung 6.3 dargestellt als Funktion des Aktienkurses und der Laufzeit. Optionen am Geld mit langer Laufzeit reagieren am sensitivsten auf Änderungen der Volatilität.

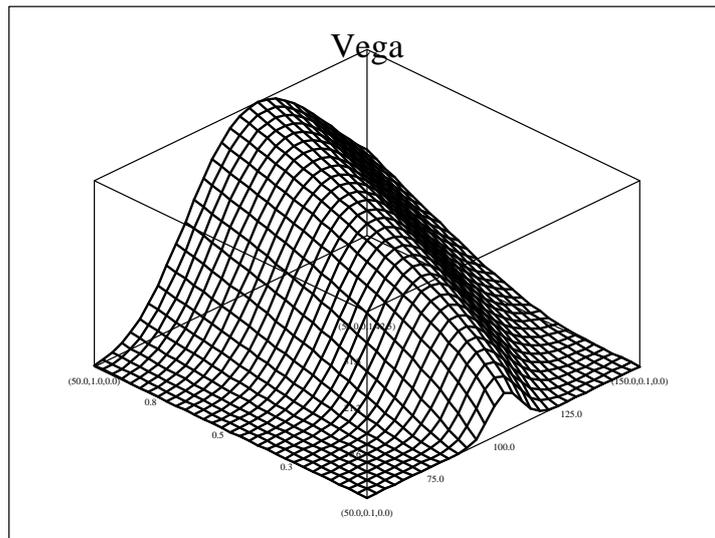


Abbildung 6.3.: Das Vega als Funktion von Aktienkurs (rechte Achse) und Laufzeit (linke Achse). [SFMvega.xpl](#)

Schließlich kann man das mit einem Call verbundene Risiko aufgrund von Zinsschwankungen begrenzen, indem man das *Rho* der Option zum Hedgen benutzt:

$$\rho = \frac{\partial C}{\partial r}$$

Für einen Call auf eine dividendenfreie Aktie ergibt sich aus (6.24)

$$\rho = K \tau e^{-r\tau} \Phi(y).$$

Bei Währungsoptionen müssen dabei neben den inländischen auch die Zinsschwankungen im Land der Fremdwährung berücksichtigt werden, so dass beim *Rho*-Hedgen zwei verschiedene Werte ρ_1, ρ_2 betrachtet werden.

6.3.4. Historische und implizierte Volatilität

In den Black-Scholes-Formeln (6.22), (6.24) sind alle Optionsparameter bekannt mit Ausnahme der Volatilität σ . In der Praxis schätzt man σ aus vorhandenen Beobachtungen des Aktienkurses oder der an der Börse notierten Preise ähnlicher Optionen.

Die *historische Volatilität* (*historic volatility*) ist ein Schätzer für σ auf der Grundlage des Schwankungsverhaltens der Aktie in der Vergangenheit. Sind z.B. S_0, \dots, S_n die Aktienkurse zu den Zeitpunkten $0, \Delta t, 2\Delta t, \dots, n\Delta t$, so sind die logarithmierten relativen Zuwächse

$$R_t = \ln \frac{S_t}{S_{t-1}}, \quad t = 1, \dots, n$$

unabhängige identisch normalverteilte Zufallsgrößen, wenn man den Kurs S_t als geometrische Brownsche Bewegung modelliert. R_t ist der Zuwachs $Y_t - Y_{t-1}$ des logarithmierten Aktienkurses $Y_t = \ln S_t$, der nach Abschnitt 5.4 ein Wiener-Prozess mit Varianzrate σ^2 ist, über ein Zeitintervall der Länge Δt . Die Varianz von R_t ist daher

$$v \equiv \text{var}(R_t) = \sigma^2 \cdot \Delta t.$$

Ein guter Schätzer für $\text{var } R_t$ ist die Stichprobenvarianz

$$\hat{v} = \frac{1}{n-1} \sum_{t=1}^n (R_t - \bar{R}_n)^2$$

wobei $\bar{R}_n = \frac{1}{n} \sum_{t=1}^n R_t$ das Stichprobenmittel ist. \hat{v} ist erwartungstreu, d.h. $E[\hat{v}] = v$, und die Zufallsgröße

$$(n-1) \frac{\hat{v}}{v}$$

hat eine bekannte Verteilung, eine χ_{n-1}^2 -Verteilung (Chi-Quadrat mit $n-1$ Freiheitsgraden). Daraus folgt insbesondere, dass der mittlere relative quadratische Schätzfehler von \hat{v}

$$E \left(\frac{\hat{v} - v}{v} \right)^2 = \frac{1}{(n-1)^2} \text{Var} \left((n-1) \frac{\hat{v}}{v} \right) = \frac{2}{n-1}$$

ist. Da $v = \sigma^2 \Delta t$, erhält man auf der Basis der historischen Aktienkurse S_0, \dots, S_n als Schätzer für die Volatilität σ

$$\hat{\sigma} = \sqrt{\hat{v}/\Delta t}.$$

Durch Taylorentwicklung der Wurzel und mit den bekannten Größen $E[\hat{v}]$ und $\text{Var}(\frac{\hat{v}}{v})$ folgt, dass $\hat{\sigma}$ erwartungstreu bis auf Terme der Größenordnung $\frac{1}{n}$ ist:

$$E\hat{\sigma} = \sigma + O\left(\frac{1}{n}\right),$$

und dass der mittlere relative quadratische Schätzfehler von $\hat{\sigma}$ bis auf Terme kleinerer Größenordnung als $\frac{1}{n}$ gegeben ist durch

$$E \left(\frac{\hat{\sigma} - \sigma}{\sigma} \right)^2 = \frac{1}{2(n-1)} + o\left(\frac{1}{n}\right).$$

Mit dieser Beziehung lässt sich die Zuverlässigkeit des Schätzers $\hat{\sigma}$ beurteilen.

Wahl der Stichprobenparameter:

- a) Als Ausgangsdaten werden meist Tagesschlusskurse S_0, \dots, S_n betrachtet. Δt entspricht einem Tag bezogen auf 1 Jahr, da σ in der Regel die Jahresvolatilität ist. Für Kalendertage wäre also $\Delta t = \frac{1}{365}$. Für Wochenenden und Feiertage liegen allerdings keine Kurse vor. Die folgende, empirisch begründete Argumentation legt aber nahe, Wochenenden und Feiertage einfach zu ignorieren: Würde sich der Kursprozess über das Wochenende genauso entwickeln wie unter der Woche, auch wenn er nicht

beobachtet wird, so müsste die Standardabweichung der Kursänderung zwischen Freitag und Montag das Dreifache der Standardabweichung der Kursänderung zwischen zwei Werktagen, z.B. Montag und Dienstag, betragen. Dies folgt aus der Beobachtung, dass für den Wiener-Prozess $Y_t = \ln S_t$ die Standardabweichung der Änderung $Y_{t+\delta} - Y_t$ gerade $\sigma \cdot \delta$ ist. Empirische Untersuchungen der Aktienmärkte haben aber gezeigt, dass das Verhältnis der Standardabweichungen nur wenig größer als 1 und auf jeden Fall deutlich kleiner als 3 ist. Die Volatilität sinkt also über das Wochenende, woraus man schließen kann, dass der Handel selbst die Volatilität erhöht und sie nicht nur durch externe Wirtschaftsentwicklungen bestimmt ist. Beim Schätzen der Volatilität ist es daher ratsam, nur Handelstage zu betrachten, an denen Kursnotierungen vorliegen. Üblicherweise rechnet man mit 252 Handelstagen pro Jahr, so dass $\Delta t = \frac{1}{252}$.

- b) $\hat{\sigma}$ wird theoretisch immer zuverlässiger, je größer n . In der Praxis ist die Volatilität aber über längere Zeiträume nicht konstant, so dass lange zurückliegende Kurse wenig aussagekräftig für das gegenwärtige σ sind. Als Kompromiss benutzt man oft die Schlusskurse der letzten 90 oder 180 Tage. Manche Autoren empfehlen auch, um dieselbe Zeitspanne in die Vergangenheit zurückzugehen, in der man die Volatilität in der Zukunft benutzen will, also z.B. Kurse der Handelstage aus den letzten 9 Monaten zu benutzen, wenn der Wert eines Calls mit Laufzeit 9 Monaten berechnet werden soll.

Die *implizierte Volatilität* (*implied volatility* - im deutschen Sprachgebrauch auch *implizite Volatilität* genannt) einer Option wird nicht aus den Aktienkursen, sondern aus den an der Terminbörse notierten Optionen auf dieselbe Aktie berechnet. Betrachtet man einen europäischen Call auf eine dividendenfreie Aktie ($d = 0$, $b = r$), der an der Börse mit C_B notiert wird, so ist seine implizierte Volatilität σ_I die Lösung der Gleichung

$$S \Phi(y + \sigma_I \sqrt{\tau}) - e^{-r\tau} K \Phi(y) = C_B$$

$$\text{mit} \quad y = \frac{1}{\sigma_I \sqrt{\tau}} \left\{ \ln \frac{S}{K} + \left(r - \frac{1}{2} \sigma_I^2 \right) \tau \right\}.$$

σ_I ist derjenige Wert der Volatilität σ , für den der nach der Black-Scholes-Formel (6.24) berechnete Optionswert $C(S, \tau)$ mit dem beobachteten Marktwert übereinstimmt. σ_I ist implizit als Lösung der obigen Gleichung definiert und muss numerisch bestimmt werden, da die Gleichung sich nicht nach σ auflösen lässt.

Die implizierte Volatilität kann benutzt werden, um sich einen Eindruck über die Einschätzung des Marktes hinsichtlich der Volatilität einer Aktie zu verschaffen. Man kann die implizierten Volatilitäten aus (hinsichtlich Restlaufzeit τ und Ausübungskurs K) verschiedenen Optionen auf dieselbe Aktie auch zu einem Schätzer von σ kombinieren. Dabei sollte man durch Gewichte die Stärke der Abhängigkeit des Optionspreises von der Volatilität berücksichtigen.

BEISPIEL 6.6

Zwei Optionen auf eine Aktie werden gehandelt. Eine von ihnen ist am Geld und hat als implizierte Volatilität $\sigma_{I1} = 0.25$; die andere ist weit im Geld und hat die implizierte Volatilität $\sigma_{I2} = 0.21$. Am Geld ist die Abhängigkeit zwischen Optionspreis und Volatilität besonders stark, so dass der Marktpreis der ersten Option mehr Information über das σ der Aktie

enthält und σ_{I1} eine zuverlässigere Approximation von σ ist. In der Kombination der beiden implizierten Volatilitäten zu einem Schätzer $\tilde{\sigma}$ für σ sollte σ_{I1} ein größeres Gewicht erhalten, z.B.

$$\tilde{\sigma} = 0.8 \cdot \sigma_{I1} + 0.2 \cdot \sigma_{I2}.$$

Andere Autoren empfehlen, $\tilde{\sigma} = \sigma_{Im}$ zu setzen, wobei σ_{Im} die implizierte Volatilität derjenigen Option auf die Aktie ist, deren Wert am empfindlichsten auf Änderungen von σ reagiert, d.h. der Option mit dem betragsgrößten Vega = $\partial C / \partial \sigma$.

Soll die Methode risiko-neutraler Bewertung angewandt werden (vgl. [Cox and Ross \(1976\)](#)), muß das Maß derart transformiert werden, dass der resultierende Preisprozeß ein Martingal ist. Dies garantiert die Abwesenheit von Arbitragemöglichkeiten (siehe [Harrison and Kreps \(1979\)](#)). Bei unvollständigen Märkten gibt es jedoch eine Vielzahl solcher Transformationen (siehe [Harrison and Pliska \(1981\)](#)). Im Gegensatz zu vollständigen Märkten kann der Händler bei unvollständigen Märkten kein selbstfinanzierendes Portfolio aufbauen, das die Auszahlung der Option exakt repliziert. Daher ist das Hedgen nicht vollkommen risikolos, und Optionspreise hängen infolgedessen von den Risikopräferenzen ab. Es sollte an dieser Stelle betont werden, dass das Fehlen einer perfekten Hedge-Strategie in der Praxis von eminenter Bedeutung ist.

6.4. Literaturhinweise

Die beiden klassischen Arbeiten [Black and Scholes \(1973\)](#) und ?, die die moderne Theorie der Derivatebewertung begründet haben, sind auch heute noch lesenswert. Eine wie das vorliegende Buch ohne Martingaltheorie auskommende, aber ausführlichere Einführung in die Mathematik der Finanzderivate bieten ?. Zwei einflussreiche Arbeiten zur modernen Finanzmathematik, die tieferliegende Resultate aus der Theorie stochastischer Prozesse benutzt, sind [Harrison and Pliska \(1981\)](#) und ?. Eine übersichtliche Diskussion der mathematischen Grundlagen der Arbitragefreiheit findet man bei ?. ? und ? geben eine kompakte Einführung in die moderne Finanzmathematik. Einen für mathematisch vorgebildete Leser gut lesbaren Einstieg in die tieferliegende martingalbasierte Theorie der Derivatebewertung bieten die Bücher von ? und ?. In ? stehen die mit dem Hedgen und Optimieren von Portfolios verbundenen Probleme im Vordergrund.

7. Das Binomialmodell für europäische Optionen

Bei vielen Optionen wie schon bei einfachen amerikanischen Optionen sind die Randbedingungen der Black-Scholes-Differentialgleichung so komplex, dass es keine analytische Lösung mehr gibt. Daher muss man sich mit der numerischen Berechnung des Optionspreises zufrieden geben. Am bekanntesten sind dabei Methoden, die den Aktienkurs durch einen stochastischen Prozess in diskreter Zeit approximieren oder ihn - wie beim Ansatz von Cox, Ross und Rubinstein - gleich als solchen modellieren. Dabei wird zunächst die Laufzeit T der Option in n äquidistante Zeitschritte zerlegt mit der Länge

$$\Delta t = \frac{T}{n}.$$

Wir betrachten nur noch die diskreten Zeitpunkte

$$t_j = j\Delta t, \quad j = 0, \dots, n.$$

Mit S_j bezeichnen wir den Kurs zur Zeit t_j . Gleichzeitig diskretisieren wir auch den Wertebereich des Kursprozesses, so dass der Kurs zu jedem festen Zeitpunkt t_j nur noch endlich viele Werte $S_j^k, k = 1, \dots, m_j$, annehmen kann, wobei j den Zeitpunkt und k den Wert des Kurses repräsentiert. Hat der Kurs zur Zeit t_j den Wert S_j^k , so kann er im nächsten Zeitschritt in einen von m_{j+1} neuen Zuständen $S_{j+1}^l, l = 1, \dots, m_{j+1}$, springen. Die Wahrscheinlichkeiten dafür nennen wir p_{kl}^j :

$$p_{kl}^j = P(S_{j+1} = S_{j+1}^l | S_j = S_j^k),$$

wobei

$$\sum_{l=1}^{m_{j+1}} p_{kl}^j = 1, \quad 0 \leq p_{kl}^j \leq 1.$$

Wenn wir den Kurs zum gerade aktuellen Zeitpunkt kennen, können wir einen Baum möglicher Kurse bis hin zu einem gewünschten Zeitpunkt (z.B. dem Verfallsdatum $T = t_n$) bestimmen. Einen solchen Baum nennt man auch *Kursbaum*. Ist für den Endzeitpunkt t_n des Kursbaumes der Optionspreis bekannt (z.B. durch den inneren Wert der Option), so lässt sich der Wert der Option zum Zeitpunkt t_{n-1} (entsprechend (6.23)) als abgezinster bedingter Erwartungswert der zugehörigen Optionspreise des Zeitpunkts t_n gegeben den Kurs zum Zeitpunkt t_{n-1} bestimmen:

$$V(S_{n-1}^k, t_{n-1}) = e^{-r\Delta t} \mathbb{E}\{V(S_n, t_n) | S_{n-1} = S_{n-1}^k\} = e^{-r\Delta t} \sum_{l=1}^{m_n} p_{kl}^{n-1} V(S_n^l, t_n). \quad (7.1)$$

$V(S, t)$ bezeichnet hier wieder den Wert einer Option zur Zeit t , wenn der Kurs des zugrundeliegenden Objekts gerade den Wert S hat. Diesen Prozess kann man für alle weiteren Zeitschritte t_j , $j = n - 2, n - 3, \dots, 0$, fortsetzen und so näherungsweise die Preise bis hin zum Anfangszeitpunkt $t = 0$ berechnen.

7.1. Der Cox-Ross-Rubinstein-Ansatz zur Optionsbewertung

Wir betrachten als einfachstes Beispiel den Ansatz von Cox, Ross und Rubinstein (CRR) zur Optionsbewertung, der auf der Annahme eines *Binomialmodells* beruht und der sich als numerisches Verfahren zur Lösung der Black-Scholes-Gleichung interpretieren lässt. Wir beschränken uns auf europäische Optionen und nehmen vorerst an, dass das der Option zugrundeliegende Objekt während der Laufzeit keine Dividenden abwirft. Wir diskretisieren wieder die Zeit und betrachten nur die Zeitpunkte $t_0 = 0, t_1 = \Delta t, t_2 = 2\Delta t, \dots, t_n = n\Delta t = T$ mit $\Delta t = \frac{T}{n}$. Das Binomialmodell geht davon aus, dass der Kursprozess in diskreter Zeit S_j eine geometrische Irrfahrt ist (vgl. Kapitel 4), d.h. das diskrete Analogon zur geometrischen Brownschen Bewegung. Charakteristisch für dieses Modell ist es, dass der Kurs sich zu jedem Zeitpunkt nur in eine von zwei möglichen Richtungen bewegen kann:

- entweder bewegt sich der Kurs mit einer Wahrscheinlichkeit p und einer Änderungsrate u in die eine Richtung (z.B. nach oben),
- oder er bewegt sich mit der Wahrscheinlichkeit q und der Änderungsrate d in die andere Richtung (z.B. nach unten).

Mit der oben eingeführten allgemeinen Notation kann der Kursprozess ausgehend von einem Wert S_j^k zur Zeit t_j im nächsten Zeitschritt zur Zeit t_{j+1} nur die beiden Werte $u \cdot S_j^k$ und $d \cdot S_j^k$ erreichen, und die Wahrscheinlichkeiten dafür sind p bzw. q unabhängig von j . Alle anderen Wahrscheinlichkeiten p_{kl}^j für $S_{j+1}^l \neq u \cdot S_j^k$ und $\neq d \cdot S_j^k$ haben den Wert 0.

Um die Lösung der Black-Scholes-Differentialgleichung mit dem Cox-Ross-Rubinstein-Ansatz zu approximieren, müssen die Wahrscheinlichkeiten p, q und die Änderungsraten u, d so gewählt werden, dass im Grenzfalle $\Delta t \rightarrow 0$ eine geometrische Brownsche Bewegung entsteht, d.h. gemäß den Überlegungen im Anschluss an (6.22) muß die bedingte Verteilung von $\ln S_{j+1}$ gegeben S_j asymptotisch eine Normalverteilung mit dem Mittelwert $\ln S_j + (b - \frac{1}{2}\sigma^2)\Delta t$ und der Varianz $\sigma^2\Delta t$ sein. Die exakte bedingte Verteilung von $\ln S_{j+1}$ gegeben S_j ist auf die beiden Werte $\ln(u \cdot S_j), \ln(d \cdot S_j)$ mit den Wahrscheinlichkeitsgewichten p, q konzentriert. Wir wählen die Parameter der geometrischen Irrfahrt so, dass die exakten bedingten Erwartungswerte und Varianzen mit ihren asymptotischen Werten für $\Delta t \rightarrow 0$ übereinstimmen. Berücksichtigt man zusätzlich, dass sich die beiden Wahrscheinlichkeiten p, q zu 1 aufsummieren müssen, erhalten wir drei Gleichungen für die insgesamt vier Parameter p, q, u, d :

$$\begin{aligned} p + q &= 1, \\ E \equiv p \ln(u \cdot S_j) + q \ln(d \cdot S_j) &= \ln(S_j) + (b - \frac{1}{2}\sigma^2)\Delta t, \\ p\{\ln(u \cdot S_j) - E\}^2 + q\{\ln(d \cdot S_j) - E\}^2 &= \sigma^2\Delta t. \end{aligned}$$

Wegen der ersten Gleichung hebt sich der aktuelle Kurswert S_j in den beiden anderen Gleichungen weg. Setzen wir $q = 1 - p$ in die anderen Gleichungen ein, so bleiben nach einigen Umformungen noch zwei Gleichungen mit drei Unbekannten übrig:

$$\begin{aligned} p \ln\left(\frac{u}{d}\right) + \ln d &= \left(b - \frac{1}{2}\sigma^2\right)\Delta t, \\ (1-p)p\left\{\ln\left(\frac{u}{d}\right)\right\}^2 &= \sigma^2\Delta t. \end{aligned}$$

Zur eindeutigen Lösung dieses nichtlinearen Gleichungssystems können wir eine zusätzliche Bedingung einführen. Wir fordern

$$u \cdot d = 1,$$

d.h. wenn der Kurs sich direkt hintereinander einmal nach oben und einmal nach unten bewegt, gelangt er nach den zwei Zeitschritten wieder zum Ausgangswert zurück. Diese Rekombinationseigenschaft ist nicht nur intuitiv ansprechend, sondern sie vereinfacht den Kursbaum beträchtlich. Zur Zeit t_j gibt es dann nur noch $m_j = j + 1$ mögliche Werte für den Kurs S_j zur Zeit t_j , und zwar ausgehend von dem Startwert S_0 zur Zeit $t_0 = 0$ nur die Werte

$$S_j^k = S_0 u^k d^{j-k}, \quad k = 0, \dots, j,$$

denn es gilt jeweils $S_{j+1}^{k+1} = u \cdot S_j^k$ und $S_{j+1}^k = S_j^k / u$. Im allgemeinen Fall kämen $m_j = 2^j$ Werte in Frage, da dann die Reihenfolge der Aufwärts- und Abwärtsbewegungen und nicht nur ihre Anzahl den Zielzustand bestimmen.

Durch Lösen der nun insgesamt drei Gleichungen für p, u, d folgt näherungsweise durch Vernachlässigen von Termen, die klein gegenüber Δt sind:

$$p = \frac{1}{2} + \frac{1}{2}\left(b - \frac{1}{2}\sigma^2\right)\frac{\sqrt{\Delta t}}{\sigma}, \quad u = e^{\sigma\sqrt{\Delta t}}, \quad d = \frac{1}{u}. \quad (7.2)$$

Abkürzend schreiben wir $V_j^k = V(S_j^k, t_j)$ für den Optionswert zur Zeit t_j , wenn der Kurs $S_j = S_j^k$ ist. Analog zu Gleichung (7.1) erhalten wir den Optionswert zur Zeit t_j als abgezinste bedingten Erwartungswert des Optionswertes zur Zeit t_{j+1} :

$$V_j^k = e^{-r\Delta t}\{pV_{j+1}^{k+1} + (1-p)V_{j+1}^k\}. \quad (7.3)$$

Am Ende der Laufzeit zur Zeit $T = t_n$ ist der Optionspreis bekannt. Im Fall eines europäischen Calls haben wir zum Beispiel

$$V_n^k = \max\{0, S_n^k - K\}, \quad k = 0, \dots, n. \quad (7.4)$$

Davon ausgehend kann man mit der Rekursion (7.3) sukzessive alle Optionswerte V_j^k , $k = 0, \dots, j$, $j = n - 1, n - 2, \dots, 0$, bestimmen.

BEISPIEL 7.1

Ein Beispiel für eine Kaufoption ist in Tabelle 7.1 dargestellt. Zuerst werden die innerhalb der 5 Zeitschritte erreichbaren Kurse berechnet. Mit $\Delta t = \tau/n = 0.1$ erhalten wir aus (7.2) $u = 1.0823$. Ausgehend von dem aktuellen Kurs $S_0 = 230$ kann der Aktienkurs S_1 sich somit nach einem Zeitschritt in $S_1^1 = uS_0 = 248.92$ oder in $S_1^0 = S_0/u = 212.52$ befinden. Im

zweiten Zeitschritt erreichen wir von $S_1 = S_1^1$ aus die beiden Werte $S_2^2 = uS_1 = 269.40$ und $S_2^1 = S_1/u = 230$, von $S_1 = S_1^0$ dementsprechend die beiden Werte $S_2^1 = 230$ und $S_2^0 = 196.36$ usw. Am Ende der Laufzeit kann der Kurs S_5 nach 5 Schritten die 6 Werte $S_5^5 = u^5 S_0 = 341.51$, $S_5^4 = u^4 S_0 = 291.56$, ..., $S_5^0 = S_0/u^5 = 154.90$ annehmen.

Nachdem die im Kursbaum vertretenen Kurswerte feststehen, berechnen wir mit (7.4) die Optionswerte am Ende der Laufzeit, z.B. $V_5^4 = V(S_5^4, t_5) = S_5^4 - K = 81.561$ oder $V_5^1 = 0$, da $S_5^1 = 181.44 < K$. Aus (7.2) erhalten wir $p = 0.50898$, da die Bestandshaltekosten b im dividendenfreien Fall mit dem Zinssatz r zusammenfallen. Ausgehend von den Callwerten am Ende berechnen wir jetzt rekursiv mit (7.3) die Werte des Calls zu früheren Zeitpunkten im Kursbaum. Mit $V_5^4 = 81.561$, $V_5^3 = 38.921$ erhalten wir durch Einsetzen der bekannten Werte von $p, r, \Delta t$ den Callwert $V_4^3 = 60.349$ zur Zeit $t_4 = 0.4a$, der einem Aktienkurs $S_4 = S_4^3 = 269.40$ entspricht. Analog ergibt sich der Optionswert $V_0^0 = 30.378$ zum aktuellen Zeitpunkt $t_0 = 0$ und zum aktuellen Kurs $S_0 = 230$ mit (7.3) aus den beiden Werten $V_1^1 = 44.328$, $V_1^0 = 16.200$.

Für 5 Zeitschritte ist 30.378 nur eine grobe Näherung an den tatsächlichen Wert des Calls. Vergleicht man allerdings die Black-Scholes-Lösung (6.24) dieses Beispiels mit dem Ergebnis des Cox-Ross-Rubinstein-Verfahrens für unterschiedliche Anzahlen von Zeitschritten n , so sieht man, dass die numerische Lösung mit Hilfe des Binomialmodells mit wachsendem n gegen die Black-Scholes-Lösung konvergiert (Tabelle 7.2).

Wenn während der Laufzeit der Option stetige Dividendenerträge mit der Rate d anfallen, so ändert sich das numerische Verfahren zur Preisbestimmung nicht. Es reicht, die Bestandshaltekosten $b = r$ durch $b = r - d$ zu ersetzen. Dividendenzahlungen zu diskreten Zeitpunkten erfordern allerdings größere Modifikationen bei der rekursiven Berechnung des Optionspreises, die wir im folgenden Abschnitt diskutieren.

BEISPIEL 7.2

Wir betrachten einen Call auf US-Dollar mit einer Restlaufzeit von 4 Monaten, d.h. $\tau = 1/3$ Jahr, einem Kurs von $S=1.50$ EUR/USD und $K=1.50$ EUR/USD. Als Rate für den stetigen Dividendenertrag, der dem US-Zins entspricht, nehmen wir 1% an. Die Bestandshaltekosten als Differenz zwischen Inlands- und Auslandszins haben damit die Rate $b = r - d = 8\%$. Die Optionspreise, die wir mit Hilfe des Binomialmodells genauso wie im vorherigen Beispiel erhalten, sind in der Tabelle 7.3 dargestellt.

7.2. Diskrete Dividendenerträge

Werden zu wenigen festen Zeitpunkten Dividenden gezahlt, so ändert sich der Kursbaum. Dabei müssen wir zwischen den beiden Fällen unterscheiden, dass die Dividende als fester Prozentsatz des aktuellen Kurses bzw. als vom Kurs unabhängiger Festbetrag gezahlt wird. Wir beschränken uns auf den Fall, dass nur einmal während der Laufzeit eine Dividende anfällt, und zwar zur Zeit t^* , $0 < t^* \leq T$. Erträge, die an mehreren Zeitpunkten anfallen, lassen sich analog berücksichtigen. Als das der Option zugrundeliegende Objekt betrachten wir eine Aktie.

Wir nutzen aus, dass wegen der Arbitragefreiheit der Kurs im Augenblick der Dividendenausschüttung um den gezahlten Dividendenbetrag nach unten springt. Dies sieht man

aktueller Aktienkurs S_t	230.00
Ausübungskurs K	210.00
Restlaufzeit τ	0.50 a
Volatilität σ	0.25 p.a.
Zinssatz r	0.04545
Dividende	keine
Schritte	5
Optionstyp	europ. Call

Kurse	Optionspreise					
	0.00	0.10	0.20	0.30	0.40	0.50
341.50558						131.506
315.54682					106.497	
291.56126				83.457		81.561
269.39890			62.237		60.349	
248.92117		44.328		40.818		38.921
230.00000	30.378		26.175		20.951	
212.51708		16.200		11.238		2.517
196.36309			6.010		1.275	
181.43700				0.646		0.000
167.64549					0.000	
154.90230						0.000
Zeit	0.00	0.10	0.20	0.30	0.40	0.50

Tabelle 7.1.: Entwicklung der Optionspreise (keine Dividenden)

 SFMBiTree1.xpl

Schritte	5	10	20	50	100	150	Black-Scholes
Optionswert	30.378	30.817	30.724	30.751	30.769	30.740	30.741

Tabelle 7.2.: Konvergenz des Binomial-Optionspreises gegen den Black-Scholes-Preis

aktueller EUR/USD-Kurs S_t	1.50						
Ausübungskurs K	1.50						
Restlaufzeit τ	0.33 a						
Volatilität σ	0.20 p.a.						
Zinssatz r	0.09						
stetige Dividende d	0.01						
Schritte	6						
Optionstyp	europ. Call						

Kurse	Optionspreise						
1.99034							0.490
1.89869						0.405	
1.81127					0.324		0.311
1.72786				0.247		0.234	
1.64830			0.180		0.161		0.148
1.57240		0.127		0.105		0.079	
1.50000	0.087		0.067		0.042		0.000
1.43093		0.041		0.022		0.000	
1.36504			0.012		0.000		0.000
1.30219				0.000		0.000	
1.24223					0.000		0.000
1.18503						0.000	
1.13046							0.000
Zeit	0.00	0.06	0.11	0.17	0.22	0.28	0.33

Tabelle 7.3.: Entwicklung der Optionspreise (mit stetigen Dividenden)

Q SFMBiTree2.xpl

mit folgendem Argument: Wir erwerben zur Zeit $t^* - dt$ unmittelbar vor Dividendenzahlung die Aktie, streichen die Dividende in der Höhe D ein und verkaufen die Aktie umgehend zur Zeit $t^* + dt$ wieder. Dabei machen wir den Gewinn $D + S_{t^*+dt} - S_{t^*-dt}$, der für $dt \rightarrow 0$ nur dann nicht risikolos von 0 verschieden ist, wenn S_t zur Zeit t^* um D nach unten springt.

7.2.1. Dividenden als Prozentsatz des Aktienkurses

Nach Diskretisierung der Zeit liege t^* im i -ten Zeitintervall, d.h. $t_{i-1} < t^* \leq t_i$. Die dann zur Zeit t_i gezahlte Dividende sei ein fester Prozentsatz δ des Aktienkurses, hat also den Betrag δS_i . Sie hat den Effekt, dass der Kurs zur Zeit t_i um den Dividendenbetrag niedriger liegt als ohne die Dividendenzahlung. Entsprechend ändern sich auch alle Kurse nach der Zeit t_i . Im Kursbaum werden daher alle Kurswerte S_j^k , $j \geq i$, mit dem Faktor $(1 - \delta)$ multipliziert. Nach dieser Korrektur werden die Optionswerte wie im dividendenfreien Fall durch Rückwärtsrekursion bestimmt.

BEISPIEL 7.3

Als Beispiel betrachten wir eine Kaufoption mit Dividendenzahlung von $\delta = 1\%$ des Aktienkurses zum Zeitpunkt 0.15. Abgesehen davon entsprechen die Parameter des Beispiels exakt denen in Tabelle 7.1. Die Ergebnisse sind in der Tabelle 7.4 dargestellt. Zuerst wird die Dividende ignoriert und der Kursbaum wie in Tabelle 7.1 erstellt. Danach werden alle Kurse ab Ausschüttung der Dividende, d.h. ab dem Zeitpunkt $t_2 = 0.2$ (bei 5 betrachteten Unterteilungen des Zeitintervalls $0 \leq t \leq 0.5$), mit dem Faktor $(1 - \delta)$ multipliziert. Die Zahlen in Klammern in Tabelle 7.1) entsprechen diesen dividendenbereinigten Kursen, d.h. S_j^k , $j < i = 2$ bzw. $0.99 \cdot S_j^k$, $j \geq i = 2$. Dadurch ändern sich nach (7.4) die Optionswerte am Ende der Laufzeit entsprechend, z.B. zu $V_5^4 = V(0.99 \cdot S_5^4, t_5) = 0.99 \cdot 291.56 - K = 78.646$. Nachdem die Randwerte V_5^5, \dots, V_5^0 feststehen, werden die früheren Optionswerte wieder mit (7.2) berechnet. Dabei ist zu beachten, dass V_j^k nicht der dem Kurs S_j^k , sondern der dem Kurs $0.99 \cdot S_j^k$ entsprechende Callpreis ist, wenn $j \geq 2$, d.h. $t_j \geq t^*$. Der aktuelle Zeitpunkt $t_0 = 0 < t^* = 0.15$ ist davon nicht betroffen, d.h. $V_0^0 = 28.384$ ist der aktuelle Preis des Calls bei aktuellem Kurs $S_0 = 230$.

7.2.2. Dividenden als feste Beträge

Wir nehmen jetzt an, dass zum vorher bekannten Zeitpunkt t^* ein fester Betrag (z.B. 5.00 EUR) gezahlt wird. Der Kurs springt jetzt um einen kursunabhängigen Betrag nach unten, was zur Folge hat, dass eine Auf- und sofortige Abwärtsbewegung des Kurses nicht mehr auf den Ausgangszustand zurückführt. Dadurch spaltet sich der Kursbaum auf. Diese Tatsache kann man sich an einem einfachen Beispiel veranschaulichen. Angenommen zum Zeitpunkt t^* , $t_1 < t^* \leq t_2 < T$, wird eine feste Dividende D gezahlt. Der Kursbaum für diesen Fall ist in der Abbildung 7.1 dargestellt. Vor Ausschüttung der Dividende zur Zeit t_1 entsprechen die Knoten des Kursbaumes den möglichen Aktienkursen uS_0 und S_0/u . Nach der Dividendenausschüttung sind die möglichen Kurswerte zur Zeit t_2 : $u^2S_0 - D$, $S_0 - D$ und $S_0/u^2 - D$. Von diesen 3 Werten ausgehend erreicht der Kursbaum im nächsten Schritt zur Zeit t_3 6 mögliche Werte, zur Zeit t_4 9 Werte usw. Der Kursbaum wird im weiteren Verlauf sehr unübersichtlich und ist für praktische Rechnungen wenig geeignet. Um dieses Problem zu umgehen, nutzen

aktueller Aktienkurs S_t	230.00
Ausübungskurs K	210.00
Restlaufzeit τ	0.50 a
Volatilität σ	0.25 p.a.
Zinssatz r	0.04545
diskrete Dividende δ	0.01
Ausschüttungstermin t^*	0.15 a
Schritte	5
Optionstyp	europ. Call

Kurse	Optionspreise					
341.50558						128.091
315.54682					103.341	(338.09)
291.56126				80.542	(312.39)	78.646
269.39890			59.543	(288.65)	57.655	(288.65)
248.92117		41.942	(266.70)	38.329	(266.70)	36.432
230.00000	28.384	(248.92)	24.087	(246.43)	18.651	(246.43)
212.51708	(230.00)	14.592	(227.70)	9.547	(227.70)	0.392
196.36309		(212.52)	4.886	(210.39)	0.199	(210.39)
181.43700			(194.40)	0.101	(194.40)	0.000
167.64549				(179.62)	0.000	(179.62)
154.90230					(165.97)	0.000
						(153.35)
Zeit	0.00	0.10	0.20	0.30	0.40	0.50
Dividende	1.00	1.00	0.99	0.99	0.99	0.99

Tabelle 7.4.: Entwicklung der Optionspreise (Dividenden als fester Prozentsatz des Aktienkurses)

 SFMBiTree3.xpl

wir aus, dass die Dividende in diesem Fall kursunabhängig und damit nicht mehr zufällig ist. Wir zerlegen den Kurs S_j in eine zufällige und eine deterministische Komponente:

$$S_j = \tilde{S}_j + D_j,$$

D_j ist dabei der aktuelle Barwert der Dividendenzahlung, d.h. vor Dividendenausschüttung der auf den Zeitpunkt $t_j \leq t^*$ abgezinste Betrag D und nach der Ausschüttung 0:

$$D_j = \begin{cases} De^{-r(t^*-t_j)} & , \text{für } t_j \leq t^*, \\ 0 & , \text{für } t^* < t_j. \end{cases} \quad (7.5)$$

Insbesondere gilt am Ende der Laufzeit $D_n = 0$ und $S_n = \tilde{S}_n$.

Zur Berechnung des Optionspreises erzeugen wir zunächst einen Kursbaum der zufallsabhängigen Kurskomponente \tilde{S}_j , der im Anfangszustand $\tilde{S}_0 = S_0 - D_0$ beginnt. Von dem Fälligkeitsstermin $T = t_n$ ausgehend erhalten wir:

$$\tilde{V}_{n-1} = e^{-r\Delta t} \mathbf{E}[\max(0, \tilde{S}_n - K) | \tilde{S}_{n-1}]$$

Die weiteren Optionspreise ergeben sich dann wie im dividendenfreien Fall als

$$\tilde{V}_{j-1} = e^{-r\Delta t} \mathbf{E}[\tilde{V}_j | \tilde{S}_{j-1}].$$

Die eigentlichen Optionspreise stimmen dann mit den so berechneten Werten \tilde{V}_j^k überein, entsprechen aber nicht einem Kurs von \tilde{S}_j^k , sondern dem tatsächlichen Aktienkurs

$$S_k^j = \tilde{S}_k^j + D_j.$$

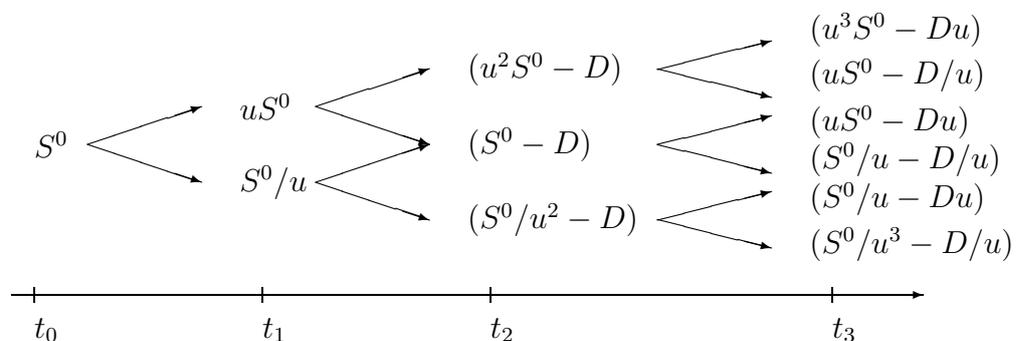


Abbildung 7.1.: Entwicklung des Kursbaumes bei festen Dividendenbeträgen.

BEISPIEL 7.4

Wir betrachten ein Beispiel, bei dem zu zwei Zeitpunkten $t_1^* = 0.25$ und $t_2^* = 0.75$ jeweils eine Dividende in Höhe von $D^{(1)} = D^{(2)} = 1.00$ EUR ausgeschüttet wird. Die Parameter und Ergebnisse des Beispiels sind in der Tabelle 7.5 dargestellt. Zuerst wird nach Gleichung (??) der gesamte, auf den Zeitpunkt t_j abgezinste Dividendenwert berechnet: $D_j = D^{(1)}e^{-r(t_1^*-t_j)} + D^{(2)}e^{-r(t_2^*-t_j)}$ für $t_j \leq t_1^*$, $D_j = D^{(2)}e^{-r(t_2^*-t_j)}$ für $t_1^* < t_j \leq t_2^*$ und

$D_j = 0$ für $t_2^* < t_j$. Insbesondere gilt $D_j = 0$ für $t_j > t_2^*$. Anschließend erstellen wir wie in Tabelle 7.1 den Kursbaum, allerdings nicht von dem aktuellen Kurs $S_0 = 100$, sondern von dem dividendenbereinigten Kurs $\tilde{S}_0 = S_0 - D_0 = 98.10$ aus. In diesem Baum berechnen wir ausgehend von den Randwerten $\tilde{V}_6^k = K - \tilde{S}_6^k$, $k = 0, \dots, 3$, $\tilde{V}_6^k = 0$, $k = 4, \dots, 6$ mit der Rekursion (7.3) die Putpreise zu früheren Zeitpunkten. Wir müssen nur berücksichtigen, dass z.B. der Optionspreis $\tilde{V}_3^2 = 2.810$ nicht dem dividendenbereinigten Kurs $\tilde{S}_3^2 = 110.88$, sondern zu dem Kurs $S_3^2 = \tilde{S}_3^2 + D_3 = 111.85$ gehört. Der berechnete Putpreis zum Ausgangskurs $S_0 = 100$ ist dann $\tilde{V}_0^0 = 7.631$.

7.3. Literaturhinweise

Ausgangspunkt für die Verwendung von Binomialprozessen bei der Optionsbewertung ist die klassische Arbeit von ?, die diesen Ansatz hier als eigenständiges Verfahren und nicht nur als numerische Approximation für die Black-Scholes Lösungen vorstellen. ? bieten eine ausführliche und moderne Darstellung der auf binomialen Kursbäumen aufbauenden Preisbestimmung. Die numerischen Aspekte der Binomialapproximation werden eingehend von [Dewynne, Howison and Wilmott \(1993\)](#) diskutiert.

aktueller Aktienkurs S_t	100.00
Ausübungskurs K	100.00
Restlaufzeit τ	1.00 a
Volatilität σ	0.30 p.a.
Zinssatz r	0.10
diskrete Dividende $D^{(1)}$	1.00
Ausschüttungstermin t_1^*	0.25 a
diskrete Dividende $D^{(2)}$	1.00
Ausschüttungstermin t_2^*	0.75 a
Schritte	6
Optionstyp	europ. Put

Kurse	Optionspreise						
204.54810							0.000
180.96953						0.000	(204.55)
160.10890					0.000	(180.97)	0.000
141.65290				0.179	(161.10)	0.000	(160.11)
125.32435			1.373	(142.63)	0.394	(141.65)	0.000
110.87803		3.906	(126.28)	2.810	(126.32)	0.866	(125.32)
98.09695	7.631	(112.81)	6.900	(111.85)	5.720	(110.88)	1.903
86.78916	(100)	12.335	(99.06)	12.100	(99.09)	11.567	(98.10)
76.78484		(88.72)	18.774	(87.76)	19.953	(86.79)	23.215
67.93374			(77.74)	27.211	(77.78)	30.420	(76.78)
60.10291				(68.91)	36.631	(67.93)	39.897
53.17476					(61.09)	45.178	(60.10)
47.04522						(53.17)	52.955
							(47.05)
Zeit	0.00	0.17	0.33	0.50	0.67	0.83	1.00
Dividende D_j	1.903	1.935	0.960	0.975	0.992	0.00	0.00

Tabelle 7.5.: Entwicklung der Optionspreise (diskrete feste Dividenden)

 SFMBiTree4.xpl

8. Amerikanische Optionen

8.1. Arbitragebeziehungen für amerikanische Optionen

Das mit amerikanischen Optionen verbundene Recht zur vorzeitigen Ausübung bereitet bei ihrer Bewertung große Schwierigkeiten. Ob der Besitzer der Option sie zur Zeit t ausübt, hängt vom aktuellen Kurs S_t des zugrundeliegenden Objekts ab. Der Ausübungszeitpunkt ist damit selbst zufällig. Die Black-Scholes-Differentialgleichung gilt zwar weiterhin, solange die Option nicht ausgeübt wird, aber die Randbedingung wird so kompliziert, dass eine analytische Lösung nicht länger möglich ist. In diesem Kapitel untersuchen wir amerikanische Optionen eingehender, während wir die numerische Bestimmung ihres Preises im nächsten Kapitel diskutieren.

Wie wir in Kapitel 2.1 gesehen haben, bewirkt das Recht zur vorzeitigen Ausübung, dass der Wert amerikanischer Optionen nicht unter den inneren Wert fallen darf, d.h. beispielsweise nicht unter $\max(K - S_t, 0)$ bei einem Put mit Ausübungskurs K . Diese Bedingung wird von europäischen Optionen im allgemeinen nicht erfüllt, was sich bei Verkaufsoptionen besonders einfach zeigen lässt. Daraus folgt dann sofort, dass es zur vorzeitigen Ausübung eines amerikanischen Puts kommen kann, wenn andernfalls der Optionswert unter den inneren Wert fallen würde.

Betrachten wir hierzu einen amerikanischen Put auf eine Aktie mit Fälligkeit T . Ist der Kurs S_{t^*} der Aktie zur Zeit t^* gleich Null, so bleibt er als geometrische Brownsche Bewegung auch anschließend in der Null: $S_t = 0, t \geq t^*$. Für den Besitzer eines amerikanischen Puts rentiert es sich daher nicht, noch länger mit der Ausübung seines Verkaufsrechts zu warten; er würde dadurch nur den möglichen Zinsertrag des Verkaufserlöses K verlieren, den er z.B. in Anleihen investieren kann. Wenn $S_{t^*} = 0$, so ist daher der Wert des Puts zu diesem Zeitpunkt gerade K , und dies stimmt mit dem inneren Wert überein. Einen entsprechenden europäischen Put können wir aber nicht zur Zeit t^* ausüben. Wir müssen bis zur Fälligkeit T warten und erhalten erst dann den Erlös K für die Aktie. Abgezinst auf den Zeitpunkt t^* ist dies - mit $\tau^* = T - t^*$ - nur $Ke^{-r\tau^*}$, was in diesem Fall auch der Wert des europäischen Puts zur Zeit t^* ist. Dieser liegt demnach unter dem Wert des amerikanischen Puts und damit auch unter dem inneren Wert. Die Abbildung 8.1 zeigt dies für die Putwerte bei stetigen Bestandshaltekosten b .

Wir bemerken, dass eine vorzeitige Ausübung des Put nötig sein kann, schon bevor der Kurs $S_t = 0$ erreicht ist. Heuristisch gesehen ist nämlich ab einem bestimmten kritischen Aktienkurs S^{**} der Zinsverlust auf den inneren Wert der Option, den der Besitzer bei sofortiger Ausübung erhalten würde, größer als die durch eventuelle Kursänderungen noch bis zur Fälligkeit zu erwartende Wertsteigerung der Option. Dies ist einer der Gründe für die

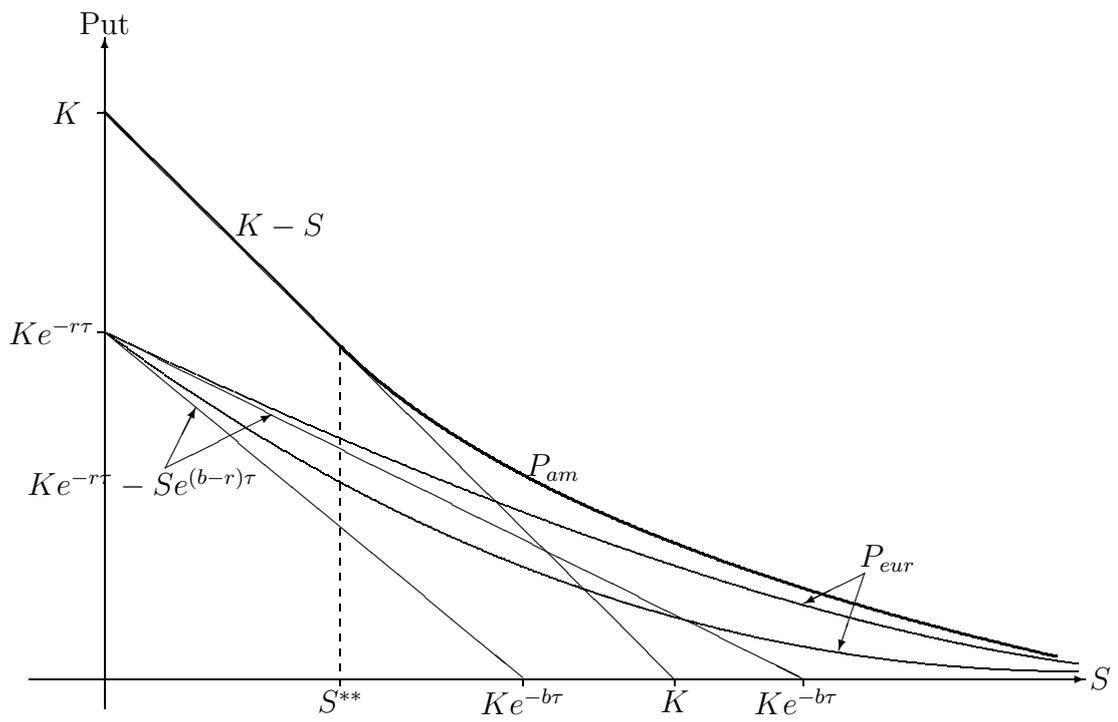


Abbildung 8.1.: Europäischer Put und vorzeitige Ausübung eines amerikanischen Puts

Abhängigkeit des kritischen Kurses von der Zeit: $S^{**} = S^{**}(t)$

Aus ihrer Herleitung folgt, dass die Differentialgleichung von Black und Scholes nur gilt, solange die Option nicht ausgeübt wird. Da in einem perfekten Markt keine Transaktionskosten zu berücksichtigen sind, stimmt der Erlös bei vorzeitiger Ausübung der Option und damit auch der Wert der Option zu diesem Zeitpunkt mit dem inneren Wert überein. Man sagt daher, dass die Option bei vorzeitiger Ausübung auf ihren inneren Wert zurückfällt. Dies bedeutet, dass die Bewertung amerikanischer Optionen auf ein freies Randwertproblem führt, bei dem die Black-Scholes-Differentialgleichung nur im Bereich der Kurse S gilt, die größer als der kritische Put-Kurs $S^{**} = S^{**}(t)$ bzw. kleiner als der kritische Call-Kurs $S^* = S^*(t)$ sind, wobei die durch $S^{**}(t), S^*(t)$ bestimmten Ränder des Gebiets nicht bekannt sind.

Die Abbildung 8.2 zeigt das Gebiet, in dem der Optionswert $C = C(S, t)$ die Black-Scholes-Differentialgleichung erfüllt, für einen amerikanischen Call.

- Im Innenbereich $\{(S, t) | 0 \leq S < S^*(t), t < T\}$ gilt die Black-Scholes Differentialgleichung.
- Am Rand $\{(S, t) | S = S^*(t), t < T\}$ bzw. $\{(S, t) | 0 \leq S, t = T\}$ fällt der Call auf den inneren Wert $\max(S - K, 0)$ zurück.
- Dabei sind $C(S, t)$ und $\frac{\partial C(S, t)}{\partial S}$ im gesamten Bereich einschließlich des Randes stetig.

Die numerische Lösung solcher freien Randwertprobleme beschreiben wir im nächsten Kapitel. Zuerst leiten wir aber einige Eigenschaften amerikanischer Optionen her, die wie in Kapitel 2.1 schon aus der Annahme eines perfekten Marktes und insbesondere der Arbitragefreiheit folgen, ohne dass ein spezifisches mathematisches Modell für den Kurs S_t angenommen werden muss.

SATZ 8.1

1. *Ein amerikanischer Call auf ein Objekt, das während der Laufzeit der Option keine positiven Erträge abwirft, wird nicht vorzeitig ausgeübt und hat daher den gleichen Wert wie der ansonsten gleiche europäische Call.*
2. *Für einen amerikanischen Call auf ein Objekt, das während der Laufzeit der Option positive Erträge nur zu bestimmten Zeitpunkten t_1, \dots, t_n abwirft, ist eine vorzeitige Ausübung nur kurz vor einem dieser Zeitpunkte t_1, \dots, t_n optimal. Dementsprechend muss bei stetigen positiven Erträgen jederzeit mit einer vorzeitigen Ausübung gerechnet werden.*

BEWEIS:

Sei K der Ausübungskurs, T der Fälligkeitstermin, $\tau = T - t$ die Restlaufzeit des Calls, und bezeichne S_t den Kurs des zugrundeliegenden Objektes. $C_{am}(S, \tau), C_{eur}(S, \tau)$ seien der Wert des amerikanischen bzw. des entsprechenden europäischen Calls zur Zeit t bei Restlaufzeit $\tau = T - t$ und aktuellem Kurs $S_t = S$. Wir benutzen die Put-Call-Parität für europäische Optionen und erhalten

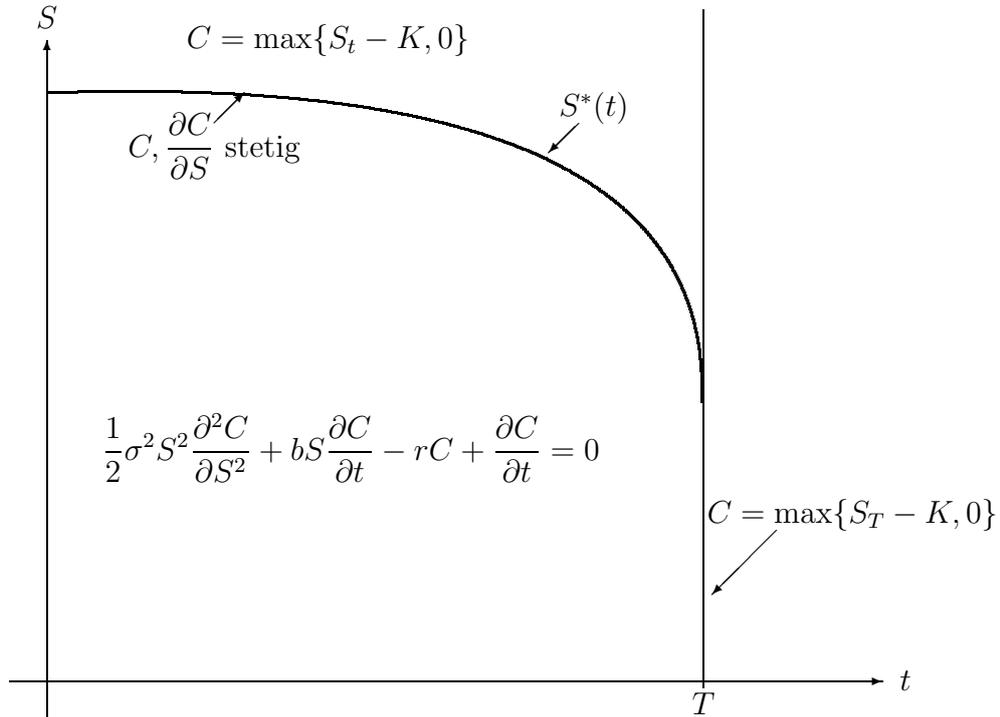


Abbildung 8.2.: Die Ausübungsgrenze $S^*(t)$ bei einem amerikanischen Call

1. Fallen diskrete Erträge D an, deren auf den Zeitpunkt t abgezinster Wert D_t ist, gilt nach Voraussetzung $D_t \leq 0$, d.h. es handelt sich um Kosten. Daraus folgt mit Satz 2.3, wobei wir zur Notationsvereinfachung $C = C_{eur}$ schreiben und P den entsprechenden Putpreis bezeichnen:

$$C(S_t, \tau) = P(S_t, \tau) + S_t - D_t - Ke^{-r\tau} \geq S_t - Ke^{-r\tau} > S_t - K \quad (8.1)$$

Fallen stetige Erträge mit Rate d an, gilt nach Voraussetzung wieder $d \leq 0$. Daraus folgt wegen $b - r = -d \geq 0$:

$$C(S_t, \tau) = P(S_t, \tau) + S_t e^{(b-r)\tau} - Ke^{-r\tau} \geq S_t e^{(b-r)\tau} - Ke^{-r\tau} > S_t - K \quad (8.2)$$

In beiden Fällen gilt also für den europäischen Call $C(S_t, \tau) > S_t - K$. Da aber $C_{am} \geq C$ ist, schließen wir

$$C_{am}(S_t, \tau) > S_t - K,$$

d.h. der Wert des amerikanischen Calls ist während der ganzen Laufzeit immer echt größer als der innere Wert, so dass eine vorzeitige Ausübung nicht erfolgt.

2. Ohne Einschränkung betrachten wir den Fall, dass t_1 der nächstgelegene Zahlungszeitpunkt ist. $\tilde{t} < t_1$ bezeichnet einen beliebigen früheren Zeitpunkt. $\tilde{C}(S_t, \tilde{\tau})$ mit $\tilde{\tau} = \tilde{t} - t$ sei der Wert eines europäischen Calls mit gleichem Ausübungskurs K , jedoch dem Verfallszeitpunkt \tilde{t} . Da vor \tilde{t} keine Zahlungen anfallen, folgt aus Teil 1. $\tilde{C}(S_t, \tilde{\tau}) > S_t - K$

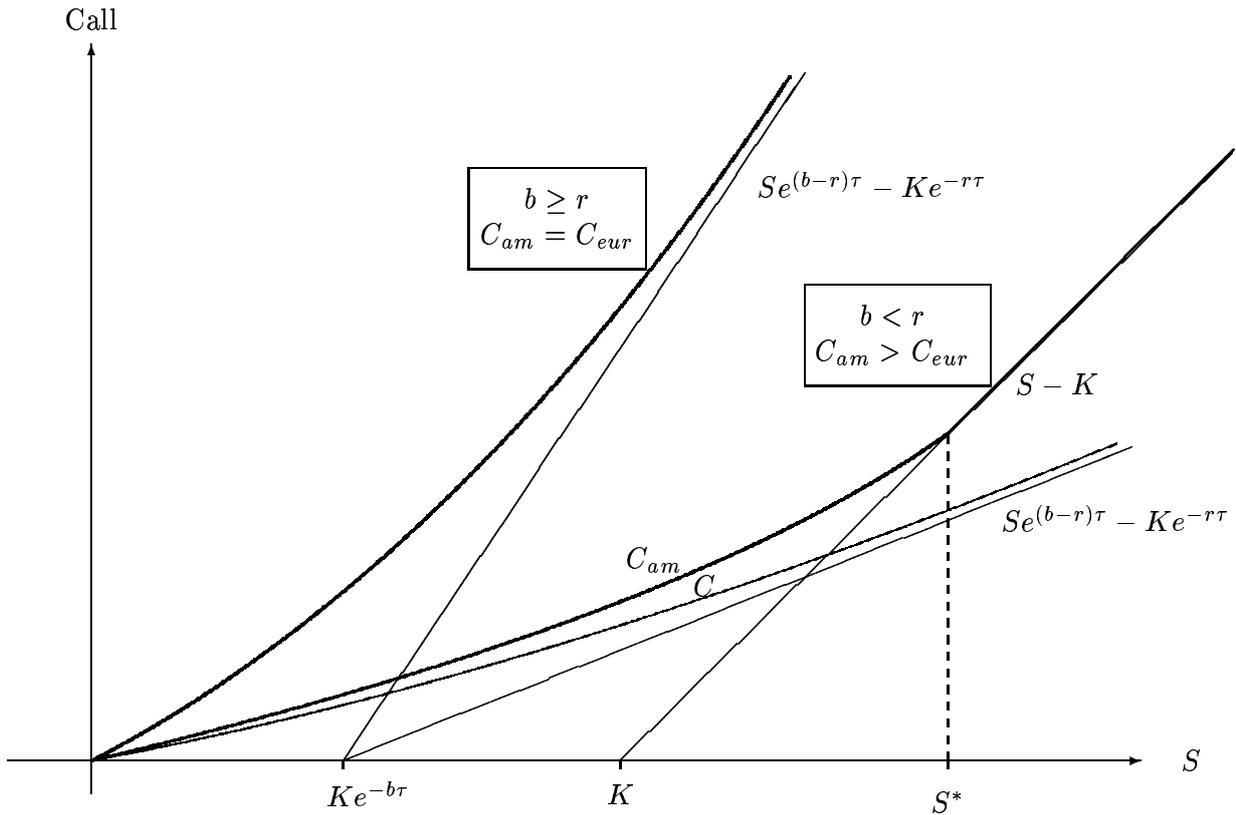


Abbildung 8.3.: Europäischer Call und vorzeitige Ausübung eines amerikanischen Calls

für $t < \tilde{t}$. Wegen der längeren Laufzeit und der vorzeitigen Ausübbarkeit des amerikanischen Calls gilt dann

$$C_{am}(S_t, \tau) \geq \tilde{C}(S_t, \tilde{\tau}) > S_t - K \quad (8.3)$$

Wie in Teil 1. liegt der Wert des amerikanischen Calls zu jedem Zeitpunkt $t < \tilde{t}$ strikt über dem inneren Wert, was eine vorzeitige Ausübung ausschließt. Da $\tilde{t} < t_1$ beliebig war, kann $C_{am}(S_t, \tau)$ nur zur Zeit t_1 (oder zu einem entsprechenden späteren Zahlungstermin) auf den inneren Wert zurückfallen.

□

Die Abbildung 8.3 stellt Teil 1. des Satzes nochmals graphisch dar.

- Im Fall $b \geq r$, was äquivalent zu $d \leq 0$ ist, gilt $C_{am} = C_{eur}$.
- Im Fall $b < r$, was äquivalent zu $d > 0$ ist, gilt $C_{am} > C_{eur}$.

Eine zur Put-Call-Parität analoge Beziehung lässt sich auch für amerikanische Optionen herleiten. Da über die kritischen Kurse $S^*(t)$ und $S^{**}(t)$ und damit über mögliche Zeitpunkte einer vorzeitigen Optionsausübung ohne Modell wenig bekannt ist, müssen wir uns bei amerikanischen Optionen mit Ungleichungen zufrieden geben.

SATZ 8.2 (Put-Call-Parität für amerikanische Optionen)

Wir betrachten einen amerikanischen Call und einen amerikanischen Put mit gleichem Fälligkeitszeitpunkt T und gleichem Ausübungskurs K auf dasselbe Objekt. $C_{am}(S, \tau), P_{am}(S, \tau)$ seien die Optionspreise zur Zeit t bei aktuellem Kurs $S_t = S$ und Restlaufzeit $\tau = T - t$. Dann gilt

1. Fallen während der Laufzeit $\tau = T - t$ der Optionen Erträge bzw. Kosten mit dem auf den Zeitpunkt t abgezinstem Wert D_t an, so ist

$$P_{am}(S_t, \tau) + S_t - Ke^{-r\tau} \geq C_{am}(S_t, \tau) \geq P_{am}(S_t, \tau) + S_t - D_t - K \quad (8.4)$$

2. Fallen stetige Bestandshaltkosten mit Rate b auf das Objekt an, so ist

$$\begin{aligned} P_{am}(S_t, \tau) + S_t - Ke^{-r\tau} \geq C_{am}(S_t, \tau) \geq P_{am}(S_t, \tau) + S_t e^{(b-r)\tau} - K & \text{ falls } b < r \\ P_{am}(S_t, \tau) + S_t e^{(b-r)\tau} - Ke^{-r\tau} \geq C_{am}(S_t, \tau) \geq P_{am}(S_t) + S_t - K & \text{ falls } b \geq r \end{aligned} \quad (8.5)$$

BEWEIS:

Wir nehmen ohne Einschränkung an, dass das zugrundeliegende Objekt eine Aktie ist, die zu einem Zeitpunkt t_1 während der Laufzeit eine Dividende D_1 abwirft.

1. Wir zeigen zunächst die linke Ungleichung. Hierzu betrachten wir ein Portfolio, das aus den folgenden vier Positionen besteht:

1. Kaufe den amerikanischen Put
2. Kaufe eine Aktie
3. Verkaufe Anleihen zum Nominalwert K (d.h. leihe Geld), fällig zum Zeitpunkt T
4. Verkaufe den amerikanischen Call

Im Portfolio wird Position 1. trotz möglicher Suboptimalität bis zum Zeitpunkt T gehalten, d.h. wir üben den Put nicht vorzeitig aus, auch wenn der Käufer des Calls ihn vorzeitig ausübt. Zu beachten ist, dass nach Satz 8.1 eine vorzeitige Ausübung des Calls nur zum Zeitpunkt $\tilde{t} := t_1 - \Delta t$ mit $\Delta t \approx 0$ direkt vor dem Zahlungstermin t_1 stattfinden kann. In diesem Fall liefern wir die Aktie aus dem Portfolio. Der Wert des Portfolios zum Zeitpunkt T ergibt sich dann aus der Tabelle 8.1. Daher gilt für den Wert des Portfolios zu jedem Zeitpunkt t wie behauptet:

$$P_{am}(S_t, \tau) + S_t - Ke^{-r\tau} - C_{am}(S_t, \tau) \geq 0 \quad (8.6)$$

Die zweite Ungleichung beweist man analog, wobei hier zu berücksichtigen ist, dass der Put vorzeitig ausgeübt werden kann.

2. Bei stetigen Bestandshaltkosten betrachten wir zuerst den Fall $b \geq r \iff d \leq 0$. Wir zeigen zunächst die linke Ungleichung

$$P_{am}(S_t, \tau) + S_t e^{(b-r)\tau} - Ke^{-r\tau} \geq C_{am}(S_t, \tau) \quad (8.7)$$

Hierzu betrachten wir folgendes Portfolio zur Zeit t :

Position	$\tilde{t} = t_1 - \Delta t$ vorzeitige Ausübung des Call	t_1	T			
			Call vorzeitig ausgeübt		Call nicht vorzeitig ausgeübt	
			$S_T \leq K$	$S_T > K$	$S_T \leq K$	$S_T > K$
1.	≥ 0	–	$K - S_T$	0	$K - S_T$	0
2.	$S_{\tilde{t}}$	D_1	–	–	S_T	S_T
3.	$-Ke^{-r(T-\tilde{t})}$	–	$-K$	$-K$	$-K$	$-K$
4.	$-(S_{\tilde{t}} - K)$	–	$Ke^{r(T-\tilde{t})}$	$Ke^{r(T-\tilde{t})}$	0	$-(S_T - K)$
Summe	≥ 0	≥ 0	≥ 0	≥ 0	0	0

Tabelle 8.1.: Wert des Portfolios zu den Zeitpunkten \tilde{t}, t_1 und T

1. Kaufe den amerikanischen Put
2. Kaufe $e^{(b-r)\tau}$ Aktien
3. Verkaufe Anleihen zum Nominalwert K , fällig zum Zeitpunkt T
4. Verkaufe den amerikanischen Call

Wie in Teil 1. folgt, dass der Wert des Portfolios zum Zeitpunkt T gleich Null ist, wenn der Call *nicht* vorzeitig ausgeübt wird. Hierbei werden die vorausgesetzten kontinuierlich anfallenden Kosten ($d \leq 0$) durch den Verkauf eines Teils der Aktien finanziert, so dass zum Zeitpunkt T genau eine Aktie vorhanden ist.

Wird andererseits der Call zwischenzeitlich zum Zeitpunkt \tilde{t} ausgeübt, wird das gesamte Portfolio liquidiert, und wir erhalten:

$$\begin{aligned} P_{am}(S_{\tilde{t}}, \tau) - (S_{\tilde{t}} - K) + S_{\tilde{t}}e^{(b-r)(T-\tilde{t})} - Ke^{-r(T-\tilde{t})} &= \\ P_{am}(S_{\tilde{t}}, \tau) + K(1 - e^{-r(T-\tilde{t})}) + S_{\tilde{t}}(e^{(b-r)(T-\tilde{t})} - 1) &\geq 0 \end{aligned} \quad (8.8)$$

Also gilt für den Wert des Portfolios zum Zeitpunkt t wie behauptet:

$$P_{am}(S_t, \tau) + S_t e^{(b-r)\tau} - Ke^{-r\tau} - C_{am}(S_t, \tau) \geq 0 \quad (8.9)$$

Im Fall $b < r \iff d < 0$ beweist man die linke Ungleichung

$$P_{am}(S_t, \tau) + S_t - Ke^{-r\tau} \geq C_{am}(S_t, \tau) \quad (8.10)$$

völlig analog, wobei es wegen $d > 0$ genügt, eine Aktie im Portfolio zu halten.

Wir zeigen nun für den Fall $b \geq r$ die rechte Ungleichung

$$C_{am}(S_t, \tau) \geq P_{am}(S_t, \tau) + S_t - K \quad (8.11)$$

Wir betrachten hierzu folgendes Portfolio zum Zeitpunkt t :

1. Kaufe den amerikanischen Call
2. Verkaufe den amerikanischen Put

3. Verkauf einer Aktie (mit Short selling)

4. Kaufe Anleihen zum Nominalwert $Ke^{r\tau}$, fällig zum Zeitpunkt T

Wird der Put nicht vorzeitig ausgeübt, gilt zum Zeitpunkt T :

$$\begin{array}{rcccl} 0 & - & (K - S_T) & - S_T e^{-(b-r)\tau} + Ke^{r\tau} & \geq 0 & \text{falls } S_T < K \\ (S_T - K) & + & 0 & - S_T e^{-(b-r)\tau} + Ke^{r\tau} & \geq 0 & \text{falls } S_T \geq K \end{array} \quad (8.12)$$

Wird der Put zum Zeitpunkt \tilde{t} vorzeitig ausgeübt, wird das gesamte Portfolio liquidiert und man erhält:

$$C_{am}(S_{\tilde{t}}, \tau) - (K - S_{\tilde{t}}) - S_{\tilde{t}} e^{-(b-r)(\tilde{t}-t)} + Ke^{r(\tilde{t}-t)} \geq 0 \quad (8.13)$$

Damit gilt für den Wert des Portfolios zum Zeitpunkt t wie behauptet:

$$C_{am}(S_t, \tau) - P_{am}(S_t, \tau) - S_t + K \geq 0 \quad (8.14)$$

Analog verfährt man im Fall $b < r$ für die rechte Ungleichung

$$C_{am}(S_t, \tau) \geq P_{am}(S_t, \tau) + S_t e^{(b-r)\tau} - K \quad (8.15)$$

wobei die Aktienposition auf $e^{(b-r)\tau}$ Aktien reduziert wird. \square

8.2. Das Trinomialmodell für amerikanische Optionen

Der Optionspreis amerikanischer Optionen lässt sich nur numerisch bestimmen. Dabei kann wie bei europäischen Optionen das Binomialmodell nach Cox-Ross-Rubinstein eingesetzt werden. In diesem Kapitel stellen wir einen konzeptionell geringfügig komplexeren, dafür aber meist numerisch effizienteren Ansatz vor, der auf dem Trinomialmodell aufbaut (siehe [Dewynne et al. \(1993\)](#)). Er steht in Beziehung zu klassischen numerischen Differenzenverfahren zur Lösung partieller Differentialgleichungen, die natürlich auch zur Lösung der Black-Scholes-Differentialgleichung eingesetzt werden können.

Das Trinomialmodell folgt weitgehend der Vorgehensweise des Binomialmodells, wobei sich allerdings der Kurs zu jedem diskreten Zeitpunkt $t_j = j\Delta t$, $j = 0, \dots, n$ mit $\Delta t = T/n$ in drei statt in zwei Richtungen bewegen kann (Abbildung 8.4). Von einem Wert S_j^k zur Zeit t_j aus können zur Zeit t_{j+1} die Werte $u_1 \cdot S_j^k$, $u_2 \cdot S_j^k$, $u_3 \cdot S_j^k$ erreicht werden, wobei $u_i > 0$, $i = 1, 2, 3$, passend zu wählende Modellparameter sind. Die Wahrscheinlichkeit, dass der Kurs sich von S_j^k nach $u_i \cdot S_j^k$ bewegt, bezeichnen wir mit p_i , $i = 1, 2, 3$. Der Kursprozess S_j , $j = 0, \dots, n$, in diskreter Zeit ist also ein geometrischer Trinomialprozess, d.h. der logarithmierte Kurs $Z_j = \ln S_j$, $j = 0, \dots, n$, ist ein gewöhnlicher Trinomialprozess mit möglichen Zuwächsen $\ln u_i$, $i = 1, 2, 3$.

Wie beim Binomialmodell müssen drei Bedingungen erfüllt sein: Die Summe der Wahrscheinlichkeiten p_i , $i = 1, 2, 3$, ist Eins, und Erwartungswert und Varianz des logarithmierten Zuwachses Z_j müssen mit dem Erwartungswert und der Varianz des Zuwachses der

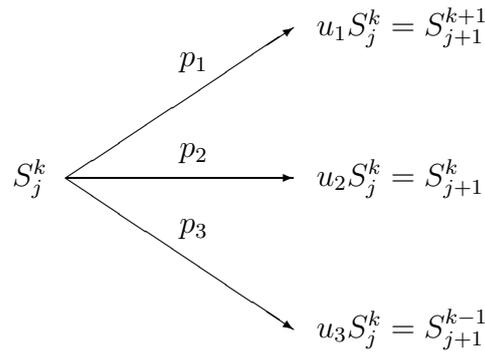


Abbildung 8.4.: Mögliche Kursbewegungen beim Trinomialmodell

logarithmierten geometrischen Brownschen Bewegung über ein Zeitintervall der Länge Δt übereinstimmen. Daraus ergeben sich drei Gleichungen:

$$p_1 + p_2 + p_3 = 1, \quad (8.16)$$

$$p_1 \ln u_1 + p_2 \ln u_2 + p_3 \ln u_3 = (b - \frac{1}{2}\sigma^2)\Delta t, \quad (8.17)$$

$$p_1(\ln u_1)^2 + p_2(\ln u_2)^2 + p_3(\ln u_3)^2 = \sigma^2\Delta t + (b - \frac{1}{2}\sigma^2)^2\Delta t^2. \quad (8.18)$$

wobei wir bei der letzten Gleichung $E(Z_j^2) = \text{var}(Z_j) + (EZ_j)^2$ benutzt haben. Die Anzahl der unbekannt Parameter ist im Fall des Trinomialmodells jedoch sechs, so dass wir die Freiheit haben, drei weitere Bedingungen einzuführen, um zu eindeutigen und möglichst einfachen Lösungen des Gleichungssystems zu gelangen. Um einen übersichtlichen Kursbaum konstruieren zu können, fordern wir die Rekombinationseigenschaft

$$u_1 u_3 = u_2^2.$$

Dadurch wird die Anzahl der zum Zeitpunkt t_n möglichen Kurse von maximal 3^n auf $2n + 1$ reduziert und somit Speicherplatz und Rechenzeit gespart. Zur Bestimmung der Parameter brauchen wir aber zwei weitere Bedingungen. Wir diskutieren zwei Ansätze, von denen der eine durch Bezüge zum Binomialmodell, der andere durch Ergebnisse der Numerik partieller Differentialgleichungen begründet werden kann.

a.) Der erste Ansatz fordert, dass ein Zeitschritt der Länge Δt des Trinomialmodells zwei Zeitschritten des Binomialmodells entspricht: u_1 steht für zwei Aufwärtsbewegungen, u_3 für zwei Abwärtsbewegungen und u_2 für eine Auf- und eine Abwärtsbewegung (oder umgekehrt). Das Binomialmodell erfülle die Rekombinationsbedingung $d = 1/u$. Da die Länge des Zeitschritts $\Delta t/2$ ist, folgt aus Abschnitt ??

$$u = e^{\sigma\sqrt{\Delta t/2}}$$

und für die Wahrscheinlichkeit p , dass der Kurs sich im Binomialmodell nach oben bewegt:

$$p = \frac{1}{2} + \frac{1}{2}(b - \frac{1}{2}\sigma^2)\frac{\sqrt{\Delta t/2}}{\sigma}.$$

Dann erhalten wir für die Parameter des Trinomialmodells die Bedingungen

$$\begin{aligned} u_1 &= u^2, u_2 = 1, u_3 = u^{-2}, \\ p_1 &= p^2, p_2 = 2p(1-p), p_3 = (1-p)^2. \end{aligned}$$

Mit diesen Parametern leistet das Trinomialmodell für europäische Optionen dasselbe wie das entsprechende Binomialmodell, benötigt dafür aber nur die Hälfte der Zeitschritte. Es konvergiert somit schneller als das Binomialmodell gegen die Black-Scholes-Lösung.

BEISPIEL 8.1

Mit den Parametern aus Tabelle 7.1 für eine europäische Kaufoption liefert das Trinomialmodell nach $n = 50$ Schritten einen Callpreis von 30.769. Dies entspricht exakt dem Wert, den das Binomialmodell nach $n = 100$ Schritten erreicht - s. Tabelle 7.2.

Amerikanische Optionen unterscheiden sich von den europäischen dadurch, dass die Ausübung zu einem beliebigen Zeitpunkt t^* , $0 < t^* \leq T$ erfolgen kann. Dabei fällt der Wert z.B. eines Calls bei vorzeitiger Ausübung auf den inneren Wert zurück:

$$C(S, t^*) = \max\{0, S(t^*) - K\}.$$

Mathematisch haben wir damit ein freies Randwertproblem zu lösen, was nur numerisch möglich ist.

V_j^k bezeichne wieder den Optionswert zur Zeit t_j , wenn für den aktuellen Kurs $S_j = S_j^k$ gilt. Wie bei der Anwendung des Binomialmodells auf europäische Optionen bestimmen wir V_j^k als abgezinste Erwartungswert der zu den im nächsten Zeitschritt erreichbaren Kursen gehörenden Optionspreise, also von V_{j+1}^{k+1} , V_{j+1}^k und V_{j+1}^{k-1} . Im Unterschied zu europäischen Optionen darf dieser Erwartungswert nicht unter den inneren Wert fallen. Damit lautet die Rekursion für amerikanische Callpreise:

$$C_j^k = \max\{S_j^k - K, e^{-r\Delta t}[p_1 C_{j+1}^{k+1} + p_2 C_{j+1}^k + p_3 C_{j+1}^{k-1}]\}.$$

BEISPIEL 8.2

Tabelle 8.2 gibt die Parameter und den mit dem Trinomialmodell in $n = 50$ Schritten bestimmten Wert einer amerikanischen Kaufoption. Dabei ergibt sich im Einklang mit Satz 8.1 derselbe Wert 30.769 wie bei der ansonsten gleichen europäischen Option, da die betrachtete Aktie während der Laufzeit keine Dividende abwirft.

Der amerikanische Put ist dagegen wertvoller als der europäische. Mit den Parametern des Beispiels von Tabelle 8.2 erhält man $P_{eur} = 6.05140$ und $P_{am} = 6.21159$).

b.) Beim zweiten Ansatz, die Trinomialparameter p_k und u_k durch Zusatzbedingungen eindeutig zu machen, zeigt der gesamte Kursbaum einen gewissen Aufwärtstrend, denn wir ersetzen die Bedingung $u_2 = 1$ durch

$$u_2 = u \equiv e^{(b - \frac{1}{2}\sigma^2)\Delta t}.$$

aktueller Aktienkurs S_t	230.00
Ausübungskurs K	210.00
Restlaufzeit τ	0.50 a
Volatilität σ	0.25 p.a.
Zinssatz r	0.04545
Dividende	keine
Schritte	50
Optionstyp	amerik. Call
Optionspreis	30.769

Tabelle 8.2.: Wert einer amerikanischen Kaufoption

Desweiteren fordern wir $p_1 = p_3$ und erhalten damit zusammen mit den vier obigen Bedingungen:

$$p_1 = p_3 = p, \quad p_2 = 1 - 2p, \quad \text{mit } p = \frac{\Delta t}{Th^2},$$

$$u_1 = ue^{\sigma h\sqrt{T/2}}, \quad u_2 = u, \quad u_3 = ue^{-\sigma h\sqrt{T/2}},$$

wobei h vorerst noch ein frei wählbarer Parameter ist. Diese p_k und u_k erfüllen das Gleichungssystem (8.16) - (8.18) exakt. Damit p_1, p_2, p_3 nicht triviale Wahrscheinlichkeiten sind, d.h. zwischen 0 und 1 liegen, muss $0 < p < 1/2$ sein, d.h. h muss die folgende Bedingung erfüllen:

$$h > \sqrt{\frac{2\Delta t}{T}}. \quad (8.19)$$

Wir betrachten nun europäische Optionen. Hier liefert das Trinomialmodell die folgende Rekursion für die möglichen Optionswerte in Abhängigkeit von den Wahrscheinlichkeiten p_i und den Änderungsraten u_i :

$$V_j^k = e^{-r\Delta t} \left(\frac{\Delta t}{Th^2} V_{j+1}^{k+1} + (1 - 2\frac{\Delta t}{Th^2}) V_{j+1}^k + \frac{\Delta t}{Th^2} V_{j+1}^{k-1} \right). \quad (8.20)$$

Wir berücksichtigen $\Delta t = -(T - t_{j+1}) + (T - t_j)$ für alle $j = 0, \dots, n - 1$, und wir setzen $h^* = \Delta t/T$ sowie

$$Z_j^k = V_j^k e^{-r(T-t_j)}, \quad Z_{j+1}^k = V_{j+1}^k e^{-r(T-t_{j+1})}.$$

Die Rekursion (8.20) für die Optionswerte V_j^k wird dann zu

$$\frac{Z_j^k - Z_{j+1}^k}{h^*} = \frac{Z_{j+1}^{k+1} - 2Z_{j+1}^k + Z_{j+1}^{k-1}}{h^2}. \quad (8.21)$$

Dies ist die explizite Differenzenapproximation der parabolischen Differentialgleichung (??), Samaskij (1984). Die Bedingung (8.19) entspricht dabei der bekannten Stabilitätsforderung für explizite Differenzenverfahren. Im Vergleich zu den bisher diskutierten Ansätzen sind

bei dieser Variante des Trinomialmodells die Wahrscheinlichkeiten p_k , und somit auch die Berechnungsvorschrift (8.20) nicht von der Volatilität abhängig. Die Rekursion (8.20) hängt nur über die Anfangsbedingung, d.h. über den Kurs $S_n = S_T$ zum Ausübungszeitpunkt als Beobachtung aus einer geometrischen Brownschen Bewegung, von σ ab.

8.3. Literaturhinweise

Amerikanische Optionen sind klassische derivative Finanzinstrumente und spielen als solche eine zentrale Rolle in der am Ende von Kapitel 6 genannten Literatur. Trinomialmodelle als numerische Verfahren zur Approximation von Optionspreisen werden ausführlich von [Dewynne et al. \(1993\)](#) und ? behandelt.

9. Exotische Optionen und Zinsderivate

Es existieren eine Reihe komplexer Optionen, sogenannte exotische Optionen, die vorwiegend im OTC-Geschäft (over the counter) eingesetzt werden, um besonderen Wünschen der Firmenkunden entgegenzukommen. Die wichtigsten Typen sind:

- Zusammengesetzte Optionen (*Compound Options*)
- Chooser-Optionen
- Barrier-Optionen
- Asiatische Optionen
- Lookback-Optionen

9.1. Beispiele für exotische Optionen

9.1.1. Zusammengesetzte Optionen, Optionen auf Optionen

Mit einer zusammengesetzten Option erwirbt man das Recht, zu einem späteren Zeitpunkt eine gewöhnliche Option zu erwerben. Als Beispiel betrachten wir einen *Call-on-a-Call* mit den Parametern:

Ausübungszeiten $T_1 < T_2$
Ausübungskurse K_1, K_2 .

Eine solche zusammengesetzte Option besteht aus dem Recht, zur Zeit T_1 zum Preis von K_1 einen Call zu kaufen, dessen Ausübungszeit T_2 und dessen Ausübungskurs K_2 ist.

Der Wert $V(S, t)$ dieser Option zur Zeit t bei aktuellem Kurs $S_t = S$ lässt sich durch zweifaches Anwenden der Black-Scholes-Formel berechnen:

- 1) Ausgehend von der Zeit $t = T_2$, berechne (explizit oder numerisch) den Wert des Calls, der zur Zeit T_1 gekauft werden kann, zur Zeit $t = T_1$. Dieser Wert sei $C(S, T_1)$.
- 2) Die Kaufoption auf den Call wird zur Zeit T_1 nur ausgeübt, wenn

$$C(S, T_1) > K_1$$

Daher gilt

$$V(S, T_1) = \max(C(S, T_1) - K_1, 0).$$

Berechne $V(S, t)$ für $t < T_1$ aus der Black-Scholes-Gleichung mit dieser Randbedingung bei $t = T_1$ analog zum gewöhnlichen Call.

BEMERKUNG 9.1 Da $V(S, t)$ nur von t und vom Kurs S_t zur Zeit t abhängt, erfüllt diese Wertfunktion die Black-Scholes-Gleichung so dass unsere Vorgehensweise gerechtfertigt ist.

9.1.2. Chooser Optionen oder “Wie es euch gefällt”-Optionen

Eine Chooser-Option ist eine Form der zusammengesetzten Option, bei der der Käufer zu einem späteren Zeitpunkt entscheiden kann, welchen Typ von Option er erwerben möchte. Als Beispiel betrachten wir eine reguläre Chooser-Option mit den Parametern:

Ausübungszeiten $T_1 < T_2$
 Ausübungskurse K_1, K_2 .

Diese Option gewährt das Recht, zur Zeit T_1 zum Preis von K_1 einen Call oder Put (nach eigener Wahl) zu kaufen, dessen Ausübungszeit T_2 und dessen Ausübungskurs K_2 ist: es handelt sich in der Sprache der zusammengesetzten Optionen um einen *Call-on-a Call* oder *Put*.

Der Wert $V(S, t)$ kann durch dreifaches Anwenden der Black-Scholes-Formel ermittelt werden:

- 1) Bestimme die Werte $C(S, T_1)$ und $P(S, T_1)$ von Call und Put mit Ausübungszeit T_2 , Ausübungskurs K_2 .
- 2) Löse die Black-Scholes-Gleichung für $t < T_1$ mit der Randbedingung

$$V(S, T_1) = \max\{C(S, T_1) - K_1, P(S, T_1) - K_1, 0\}$$

9.1.3. Barrier-Optionen

Eine Barrier-Option ändert ihren Wert sprunghaft, sobald der Aktienkurs eine vorgegebene Schranke, die i.a. auch zeitabhängig sein darf, erreicht. Als Beispiel betrachten wir einfache europäische Barrier-Optionen, die bei

Ausübungszeit T , Ausübungskurs K und
 Barriere B

das Recht geben, eine Aktie zur Zeit T zum Kurs K zu kaufen, vorausgesetzt, dass

- *down-and-out*: $S_t > B$ für alle $0 \leq t \leq T$
- *up-and-out*: $S_t < B$ für alle $0 \leq t \leq T$

Eine solche Knock-out-Option wird wertlos, sobald der Kurs S_t die Barriere erreicht. Eine Knock-in-Option ist umgekehrt so lange wertlos, bis die Barriere erreicht wird.

Zum Beispiel besteht ein europäischer *Knock-in-Call* aus dem Recht zum Aktienkauf, vorausgesetzt, dass

- *down-and-in*: $S_t \leq B$ für ein $0 \leq t \leq T$
- *up-and-in*: $S_t \geq B$ für ein $0 \leq t \leq T$

Der Wert einer Barrier-Option hängt nicht mehr von dem Aktienkurs zu festen Zeitpunkten, sondern von dem gesamten Aktienpfad während der Laufzeit ab. Im Prinzip erfüllt er daher nicht die Black-Scholes-Differentialgleichung. Die Abhängigkeit ist aber noch einfach genug, um im wesentlichen doch mit dem üblichen Black-Scholes-Ansatz arbeiten zu können. Als Beispiel betrachten wir einen *Down-and-out-Call* mit $K > B$. Solange $S_t > B$ ist, erfüllt $V(S, t)$ die Black-Scholes-Gleichung mit der Randbedingung:

$$V(S, T) = \max(S_T - K, 0)$$

Falls der Kurs die Barriere B erreicht, wird die Option allerdings wertlos:

$$V(B, t) = 0, \quad 0 \leq t \leq T,$$

ist daher als zusätzliche Randbedingung beim Lösen der Differentialgleichung (6.22) zu berücksichtigen. Als explizite Lösung ergibt sich:

$$V(S, t) = C(S, t) - \left(\frac{B}{S}\right)^\alpha C\left(\frac{B^2}{S}, t\right)$$

mit $\alpha = \frac{2r}{\sigma^2} - 1$, wobei $C(S, t)$ der Wert eines gewöhnlichen europäischen Calls auf die betreffende Aktien darstellt. Den Wert $\bar{V}(S, t)$ eines europäischen Down-and-in-Calls kann man analog berechnen. Wenn man $V(S, t)$ bereits kennt, kann man aber auch die Gleichung

$$\bar{V}(S, t) + V(S, t) = C(S, t),$$

benutzen. Sie ist erfüllt, da ein Down-and-in- und ein Down-and-out-Call zusammen gerade die Wirkung eines gewöhnlichen Calls haben.

9.1.4. Asiatische Optionen

Bei asiatischen Optionen hängt der Wert von dem über die ganze Laufzeit gemittelten Aktienkurs ab, Bei einer *Average Strike Option* über den Zeitraum $0 \leq t \leq T$ hat z.B. der Payoff am Ende der Laufzeit die Form

$$\max\left(S_t - \frac{1}{t} \int_0^t S_s ds, 0\right), t = T.$$

Bei einer amerikanischen Average Strike Option ist dies auch die Auszahlung bei vorzeitiger Ausübung der Option zu einem beliebigen Zeitpunkt $t \leq T$.

Zur Berechnung des Wertes asiatischer Optionen betrachten wir eine allgemeinere Klasse europäischer Optionen mit Payoff zur Zeit T , der von S_T und I_T abhängt mit

$$I_t = \int_0^t f(S_s, s) ds.$$

Analog zur Black-Scholes-Gleichung leiten wir eine Gleichung für den Wert zur Zeit t einer solchen *pfadabhängigen Option* $V(S, I, t)$ her. Zur Zeit t ergibt sich bei einem Aktienkurs S_t

$$\begin{aligned} I_t + dI_t &\equiv I_{t+dt} = \int_0^{t+dt} f(S_s, s) ds \\ &= I_t + \int_t^{t+dt} f(S_s, s) dt \\ &= I_t + f(S_t, t) dt + o(dt). \end{aligned}$$

Also ist das Differential von I_t als $dI_t = f(S_t, t)dt$.

Aus Itô's Lemma folgt für $V_t = V(S_t, I_t, t)$:

$$dV_t = \sigma S_t \frac{\partial V_t}{\partial S} dW_t + f(S_t, t) \frac{\partial V_t}{\partial I} dt + \left(\frac{1}{2} \sigma^2 S_t^2 \frac{\partial^2 V_t}{\partial S^2} + \frac{\partial V_t}{\partial t} \right) dt$$

Analog zur Herleitung der Black-Scholes-Formel liefert kontinuierliches Delta-Hedgen ein risikofreies Portfolio aus einer Option und $\Delta_t = \partial V_t / \partial S$ verkauften Aktien. Zusammen mit der Bedingung der Arbitrage-Freiheit folgt für den dividendenfreien Fall ($b = r$):

$$\frac{\partial V_t}{\partial t} + f(S_t, t) \frac{\partial V_t}{\partial I} + \frac{1}{2} \sigma^2 S_t^2 \frac{\partial^2 V_t}{\partial S^2} + r S_t \frac{\partial V_t}{\partial S} - r V_t = 0$$

Dies ist die Black-Scholes-Gleichung mit einem Zusatzterm $f(S_t, t) \frac{\partial V_t}{\partial I}$. Als Randbedingung haben wir in diesem Fall

$$V(S, I, T) = g(S, I, T).$$

Für einen Average Strike Call ist dabei speziell:

$$g(S, I, t) = \max\left(S - \frac{1}{t}I, 0\right) \quad \text{und} \quad f(S, t) = S.$$

Für europäische Optionen existiert eine explizite analytische Lösung der Differentialgleichung, in der komplizierte spezielle Funktionen, sogenannte konfluente hypergeometrische Funktionen, auftreten. Die numerische Lösung erhält man aber einfacher und schneller.

$\int_0^t S_s ds$ wird in der Praxis als Summe über alle notierten Kurse, z.B. im Zeittakt von 30 sec, berechnet. Zeitdiskrete asiatische Optionen benutzen stattdessen eine wesentliche größere Zeitskala. I_t ändert sich nur einmal täglich oder wöchentlich:

$$I_t = \sum_{j=1}^{n(t)} S_{t_j}, \quad t_{n(t)} \leq t < t_{n(t)+1}$$

mit $t_{j+1} - t_j = 1$ Tag oder = 1 Woche und Schlusskursen S_{t_j} .

Eine solche zeitdiskrete asiatische Option entspricht weitgehend einer gewöhnlichen Option mit diskreten Dividendenzahlungen zu den Zeitpunkten t_1, t_2, \dots . Aus der Bedingung der Arbitragefreiheit folgt als Stetigkeitsbedingung bei t_j :

$$V(S, I, t_j-) = V(S, I + S, t_j^+)$$

Zur Bestimmung des Optionswertes beginnt man wie üblich am Ende der Laufzeit, wo der Optionswert bekannt ist:

1) $T = t_n$

$$V(S, I, T) = \max\left(S - \frac{1}{T}I_T, 0\right)$$

$$I_T = \sum_{j=1}^n S_{t_j}$$

Löse Black-Scholes-Gleichung rückwärts bis zur Zeit t_{n-1} und erhalte

$$V(S, I + S, t_{n-1}^+)$$

2) Berechne aus der Stetigkeitsbedingung den neuen Endwert $V(S, I, t_{n-1}^-)$. Löse Black-Scholes rückwärts bis zur Zeit t_{n-2} und erhalte

$$V(S, I + S, t_{n-2}^+)$$

usw.

9.1.5. Lookback-Optionen

Der Wert einer Lookback-Option hängt vom Maximum oder Minimum des Aktienkurses über die gesamte Laufzeit ab, z.B. hat ein *Lookback-Put* über den Zeitraum $0 \leq t \leq T$ am Ende der Laufzeit den Payoff

$$\max(M_T - S_T, 0)$$

mit $M_t = \max_{0 \leq s \leq t} S_s$.

Zur Berechnung des Wertes einer solchen Option betrachten wir zuerst eine pfadabhängige Option mit

$$I_t(n) = \int_0^t S_s^n ds, \quad \text{d.h.} \quad f(S, t) = S^n.$$

Mit $M_t(n) = (I_t(n))^{\frac{1}{n}}$ gilt:

$$M_t = \lim_{n \rightarrow \infty} M_t(n).$$

Aus der Differentialgleichung für $I_t(n)$ und $n \rightarrow \infty$ folgt, dass der Wert $V_t = V(S_t, M_t, t)$ eines europäischen Lookback-Puts folgende Gleichung erfüllt:

$$\frac{\partial V_t}{\partial t} + \frac{1}{2}\sigma^2 S_t^2 \frac{\partial^2 V_t}{\partial S^2} + r S_t \frac{\partial V_t}{\partial S} - r V_t = 0$$

Dies ist die übliche Black-Scholes-Gleichung. M tritt nur als Argument von V und in der Randbedingung auf:

$$V(S, M, T) = \max(M - S, 0)$$

Die Lösung ist mit der Restlaufzeit $\tau = T - t$, $\alpha = 2r/\sigma^2$:

$$V(S, M, t) = S \left(\Phi(y_1) \cdot \left(1 + \frac{1}{\alpha}\right) - 1 \right)$$

$$+M e^{-r\tau} \left(\Phi(y_3) - \frac{1}{\alpha} \left(\frac{M}{S} \right)^{\alpha-1} \Phi(y_2) \right)$$

$$\text{mit } y_1 = \frac{1}{\sigma\sqrt{\tau}} \left\{ \ln \frac{S}{M} + \left(r + \frac{1}{2}\sigma^2 \right) \tau \right\}$$

$$y_2 = \frac{1}{\sigma\sqrt{\tau}} \left\{ \ln \frac{S}{M} - \left(r + \frac{1}{2}\sigma^2 \right) \tau \right\}$$

$$y_3 = \frac{1}{\sigma\sqrt{\tau}} \left\{ \ln \frac{M}{S} - \left(r + \frac{1}{2}\sigma^2 \right) \tau \right\}$$

9.2. Modelle für den Zinssatz und Zinsderivate

Wie wir in den vorangegangenen Kapiteln gesehen haben, sind wichtige Grundlagen des Ansatzes von Black und Scholes

- konstante risikoloser inländischer Zinssatz r (näherungsweise erfüllt für Aktienoptionen mit Laufzeit ≤ 9 Monate)
- Unabhängigkeit des Kurses des optierbaren Objekts vom Zinssatz r (empirische Untersuchungen zeigen, dass dies für Aktien näherungsweise erfüllt ist).

Beide Bedingungen sind verletzt für Rentenoptionen (*bond options*) und die damit typischerweise verbundenen längeren Laufzeiten.

Ein *Bond* liefert zum Fälligkeitstermin T einen festen Geldbetrag Z , den Nominalwert, und evtl. zu festen Zeiten vor T Dividendenzahlungen (*Coupon*). Bei Fehlen von Coupons spricht man von einem *Zero-coupon Bond* oder kurz *Zerobond*.

Coupons können bei der Bewertung von Bond-Optionen wie diskrete Dividendenzahlungen bei der Bewertung von Aktienoptionen behandelt werden.

9.2.1. Wert eines Bonds bei bekannten zeitabhängigem Zinssatz

Als Einstieg berechnen wir den Bond-Wert $V(t)$ zur Zeit t bei zeitabhängigem, aber bekanntem Zinssatz $r(t)$.

Aus der Bedingung der Arbitragefreiheit schließen wird, dass die Wertänderung eines Bonds über das Zeitintervall $[t, t + dt]$ mit eventueller Couponzahlung $K(t)dt$ übereinstimmt mit der Wertänderung eines Bargeldkontos der Größe $V(t)$ bei Verzinsung mit $r(t)$:

$$V(t + dt) - V(t) = \left(\frac{dV}{dt} + K(t) \right) dt = r(t)V(t) dt$$

Zusammen mit der Randbedingung $V(T) = Z$ folgt:

$$V(t) = e^{I(t)} \left\{ Z + \int_t^T K(s) e^{-I(s)} ds \right\} \quad (9.1)$$

mit $I(t) = -\int_t^T r(s) ds$, einer Stammfunktion von $r(t)$.

Für einen Zerobond vereinfacht sich dies zu: $V(t) = Z \cdot e^{I(t)}$

9.2.2. Stochastische Zinsmodelle

Wegen der Unsicherheit über die zukünftige Zinsentwicklung wird $r(t)$ als Zufallsgröße modelliert. Um einen eindeutig festgelegten Zins zu haben, betrachtet man dabei üblicherweise den Zinssatz einer Geldanlage über den kürzestmöglichen Zeitraum:

$r(t) = \text{spot rate} =$ Zinssatz für kürzestmögliche Anlage.

$r(t)$ folgt keiner geometrischen Brownschen Bewegung, so dass der Black-Scholes-Ansatz nicht anwendbar ist. Es gibt eine Reihe von Modellen für $r(t)$, die Spezialfälle des folgenden allgemeinen Ansatzes sind, der den Zins als allgemeinen Itô-Prozess modelliert:

$$dr(t) = \mu(r(t), t)dt + \sigma(r(t), t)dW_t \quad (9.2)$$

$\{W_t\}$ bezeichnet wie üblich einen Standard-Wienerprozess.

Drei der am häufigsten benutzten Modelle sind einfache Spezialfälle, und zwar sind die Koeffizientenfunktionen des Modells von

- *Vasicek*:

$$\mu(r, t) = a(b - r), \quad \sigma(r, t) = \sigma \quad (9.3)$$

a, b, σ Konstanten

- *Cox, Ingersoll, Ross*:

$$\mu(r, t) = a(b - r), \quad \sigma(r, t) = \sigma\sqrt{r} \quad (9.4)$$

a, b, σ Konstanten

- *Hull, White*:

$$\mu(r, t) = \delta(t) - ar, \quad \sigma(r, t) = \sigma \quad (9.5)$$

a, σ Konstanten, $\delta(t)$ aus Marktdaten
erschließbare bekannte Funktion

Allgemein können $\mu(r, t)$ und $\sigma(r, t)$ passend gewählt werden, um empirisch beobachtete Phänomene im Modell zu replizieren. Im folgendem schreiben wir $w(r, t)$ für $\sigma(r, t)$, um deutlicher zwischen der Funktion $w \equiv \sigma$ und der Konstanten σ , die als Parameter in den obigen drei Modellen auftaucht, zu unterscheiden.

9.2.3. Die Bondbewertungs-Gleichung

Eine Aktienoption kann mit Aktien gehedgt werden, und Black und Scholes nutzen dies bei der Herleitung der Optionspreisformel aus. Da es kein entsprechendes einem Bond zugrundeliegendes Finanzinstrument gibt, müssen Bonds mit unterschiedlichen Laufzeiten gegenseitig gegeneinander gehedgt werden, um eine Gleichung für Bondwerte herzuleiten.

Wir betrachten ein Portfolio aus einem Zerobond mit Restlaufzeit τ_1 und $-\Delta$ Zerobonds (d.h. Δ verkauften Zerobonds) mit Restlaufzeit τ_2 . Der Portfoliowert zur Zeit t bei aktuellem Zins $r(t) = r$ ist:

$$\Pi(r, t) = V_1(r, t) - \Delta \cdot V_2(r, t).$$

wobei $V_i, i = 1, 2$ für die Wertfunktionen der beiden Bonds stehen. Wir schreiben $\Pi_t = \Pi(r(t), t), V_{it} = V_i(r(t), t), i = 1, 2, \mu_t = \mu(r(t), t), w_t = w(r(t), t)$ Aus Itô's Lemma folgt

$$d\Pi_t = \frac{\partial V_{1t}}{\partial t} dt + \frac{\partial V_{1t}}{\partial r} dr(t) + \frac{1}{2} w_t^2 \cdot \frac{\partial^2 V_{1t}}{\partial r^2} dt \\ - \Delta \left(\frac{\partial V_{2t}}{\partial t} dt + \frac{\partial V_{2t}}{\partial r} dr(t) + \frac{1}{2} w_t^2 \frac{\partial^2 V_{2t}}{\partial r^2} dt \right)$$

Durch Hedgen des Risikos wird die Zufallskomponente zum Verschwinden gebracht. Dies erreicht man mit der Wahl

$$\Delta = \frac{\partial V_{1t}}{\partial r} \bigg/ \frac{\partial V_{2t}}{\partial r}.$$

Einsetzen und Vergleichen des Portfolios mit einer risikofreien Anleihe und Ausnutzen der Arbitragefreiheit, d.h. der Gleichheit der Wertänderung von Portfolio und Anleihe:

$$d\Pi_t = r(t) \cdot \Pi_t dt,$$

ergibt insgesamt

$$\left(\frac{\partial V_{1t}}{\partial t} + \frac{1}{2} w_t^2 \frac{\partial^2 V_{1t}}{\partial r^2} - r(t) V_{1t} \right) \bigg/ \frac{\partial V_{1t}}{\partial r} = \left(\frac{\partial V_{2t}}{\partial t} + \frac{1}{2} w_t^2 \frac{\partial^2 V_{2t}}{\partial r^2} - r(t) V_{2t} \right) \bigg/ \frac{\partial V_{2t}}{\partial r}.$$

Dies kann nur richtig sein, wenn beide Seiten unabhängig von den Restlaufzeiten τ_1, τ_2 sind. V_{1t}, V_{2t} genügen daher beide der folgenden Differentialgleichung

$$\frac{\partial V_t}{\partial t} + \frac{1}{2} w_t^2 \frac{\partial^2 V_t}{\partial r^2} - r(t) V_t = -a(r(t), t) \cdot \frac{\partial V_t}{\partial r}$$

für eine von der Restlaufzeit unabhängige Funktion $a(r, t)$. Mit der ökonomisch interpretierbaren Größe

$$\lambda(r, t) = \frac{\mu(r, t) - a(r, t)}{w(r, t)}$$

ergibt sich mit der Abkürzung $\lambda_t = \lambda(r(t), t)$ die *Zerobond-Bewertungsgleichung* für $V(r, t)$:

$$\frac{\partial V(r, t)}{\partial t} + \frac{1}{2} w_t^2 \frac{\partial^2 V(r, t)}{\partial r^2} - (\mu_t - \lambda_t w_t) \frac{\partial V(r, t)}{\partial r} - r(t) V(r, t) = 0 \quad (9.6)$$

mit der Randbedingung $V(r, T) = Z$ am Ende der Laufzeit und mit weiteren von μ, w abhängenden Randbedingungen. Man beachte, dass in der Gleichung μ_t, w_t, λ_t für Funktionen von r und t stehen.

Die Größe $\lambda(r, t)$ besitzt die folgende Interpretation. Wir betrachten ein ungehedgtes, risikobehaftetes Portfolio aus einem Bond mit Wert $V_t = V(r(t), t)$. Als Änderung des Wertes in der Zeit dt ergibt sich mit Itô's Lemma und der Zerobond-Bewertungsgleichung:

$$dV_t = r(t) V_t dt + w_t \cdot \frac{\partial V_t}{\partial r} (dW_t + \lambda(r(t), t) dt)$$

Da $E[dW_t] = 0$, liegt die mittlere Wertänderung $E[dV_t]$ um

$$\left(w_t \frac{\partial V_t}{\partial r} \right) \cdot \lambda(r(t), t) dt$$

über der Wertsteigerung $r(t) V_t dt$ einer risikofreien Anleihe. $\lambda(r, t) dt$ ist also der Bonus an Wertsteigerung, den man zur Zeit t bei aktuellem Zins $r(t) = r$ für die Inkaufnahme des Risikos erhält. $\lambda(r, t)$ wird daher als *Marktpreis des Risikos* (*market price of risk*) interpretiert.

9.2.4. Lösung der Zerobond-Bewertungsgleichung

Wir betrachten speziell:

$$w(r, t) = \sqrt{\alpha(t)r - \beta(t)}$$

$$\mu(r, t) = -\gamma(t) \cdot r + \eta(t) + \lambda(r, t)w(r, t).$$

Einsetzen des Lösungsansatzes

$$V(r, t) = Z \cdot e^{A(t) - rB(t)}$$

in die Zerobond-Bewertungsgleichung ergibt die beiden Gleichungen

$$\frac{\partial A(t)}{\partial t} = \eta(t) B(t) + \frac{1}{2} \beta(t) B^2(t)$$

$$\frac{\partial B(t)}{\partial t} = \frac{1}{2} \alpha(t) B^2(t) + \gamma(t) B(t) - 1$$

mit Randbedingungen $A(T) = B(T) = 0$ (da $V(r, T) = Z$).

Für zeitunabhängige $\alpha, \beta, \gamma, \eta$ gibt es eine explizite Lösung, die mit der Restlaufzeit $\tau = T - t$ die Form hat

$$B(t) = \frac{2(e^{\psi_1 \tau} - 1)}{(\gamma + \psi_1)(e^{\psi_1 \tau} - 1) + 2\psi_1}, \quad \psi_1 = \sqrt{\gamma^2 + \alpha}$$

$$\frac{2}{\alpha} A(t) = b_2 \psi_2 \ln(b_2 - B) + \left(\psi_2 - \frac{\beta}{2}\right) b_1 \ln\left(\frac{B}{b_1} + 1\right) + \frac{1}{2} B \beta - b_2 \psi_2 \ln b_2$$

mit

$$b_{1/2} = \frac{\pm \gamma + \sqrt{\gamma^2 + 2\alpha}}{\alpha}, \quad \psi_2 = \frac{\eta + b_2 \beta / 2}{b_1 + b_2}$$

Wahl der Parameter:

- 1) Die Spotrate-Volatilität ist $\sqrt{\alpha r(t) - \beta}$. Mit dieser Beziehung lassen sich α, β aus historischen Daten schätzen ähnlich wie die historische Volatilität bei Aktien.
- 2) Durch Betrachtung der Zinsstrukturkurve (*yield curve* - s. Abschnitt 10.4.3), deren Diskussion hier zu weit führen würde, lassen sich Schätzer für γ und η herleiten.

9.2.5. Bewertung von Bond-Optionen

Als Beispiel betrachten wir einen europäischen Call mit Ausübungskurs K und Fälligkeit T_C auf einen Zerobond mit Fälligkeit $T_B > T_C$, d.h. das Recht, den Bond zur Zeit T_C zum Preis K zu kaufen.

$$V_B(r, t) = \text{Wert des Bonds zur Zeit } t \text{ bei aktuellem Zins } r(t) = r$$

$$C_B(r, t) = \text{Wert des Calls zur Zeit } t \text{ bei aktuellem Zins } r(t) = r$$

C_B hängt nur von der Zufallsvariable $r(t)$ und der Zeit t ab und erfüllt daher selbst auch die Zerobond-Bewertungsgleichung, aber mit der Randbedingung

$$C_B(r, T_C) = \max(V_B(r, T_C) - K, 0).$$

Diese Gleichung lässt sich analog zu den entsprechenden Black-Scholes-Gleichungen numerisch lösen.

9.3. Literaturhinweise

Exotische Optionen wie auch Bond-Optionen werden ausführlich von ? vorgestellt. Die für die Bewertung komplexer pfadabhängiger Derivate notwendigen numerischen Verfahren werden im Überblick in ? und in größerem Detail in [Dewynne et al. \(1993\)](#) diskutiert. Die klassischen stochastischen Zinsmodelle werden in ?, [Cox, Ingersoll and Ross \(1985\)](#) und ? eingeführt. Eine für die moderne Betrachtungsweise des Zinssatzes in der Finanzmathematik grundlegende Arbeit ist ?.

Teil II.

Statistische Modellierung von Finanzzeitreihen

10. Einführung: Definitionen und Konzepte

Finanzmärkte lassen sich unter verschiedenen Gesichtspunkten betrachten. Zunächst einmal gibt es ökonomische Theorien, die Aussagen über die Preisbildung an den verschiedenen Märkten liefern. Für die unterschiedlichen Märkte (Devisen, Zinsen, Wertpapiere, Derivative, etc.) existieren verschiedene ökonomische Theorien. Als bekannte Beispiele seien genannt die Kaufkraftparitätentheorie für Wechselkurse, Theorien über Zinsstrukturkurven, das *capital asset pricing model* (CAPM) und das Optionspreismodell nach Black und Scholes. Vielen dieser Modelle liegen Denkkonzepte zugrunde, die z.B. die Bildung von Erwartungen, Nutzenfunktionen und Risikoeinstellungen betreffen. Es wird grundlegend angenommen, dass sich Wirtschaftssubjekte 'rational' verhalten, rationale Erwartungen bilden und risikoavers sind. Auf diese Weise lassen sich in den Gleichgewichtsmodellen (wie z.B. dem CAPM) Preise oder Renditen bestimmen, die den Markt räumen, d.h., dass das Angebot gleich der aggregierten Nachfrage ist. Einen etwas anderen Ansatz verfolgt die Arbitragetheorie (z.B. das Black und Scholes Modell), in der davon ausgegangen wird, dass risikolose Gewinne sofort von Marktteilnehmern erkannt und über Anpassung der Preise eliminiert werden. Arbitragetheorie und Gleichgewichtstheorie hängen oft eng miteinander zusammen. Die Arbitragetheorie kommt meistens mit weniger Annahmen aus, dafür kommen die Gleichgewichtstheorien in komplexen Situationen zu genauer festgelegten Lösungen.

Die traditionellen ökonomischen Modelle sind in der Mehrzahl so formuliert, dass sich die involvierten Parameter ökonomisch interpretieren lassen. Man ist an den folgenden empirischen Fragestellungen interessiert:

1. Wie gut beschreibt ein bestimmtes Modell einen gegebenen Datensatz (Querschnitts- oder Zeitreihendaten)?
2. Hilft das Modell den Marktteilnehmern, Aussagen über die zukünftige Entwicklung der relevanten Größen zu treffen?
3. Was implizieren die empirischen Befunde für das ökonomische Modell? Muss man es eventuell modifizieren? Kann man sogar Vorschläge machen, die die Funktionsweise und strukturelle Organisation des Marktes betreffen?

Um diese empirischen Fragestellungen wissenschaftlich zu behandeln, bedarf es einer statistischen Untersuchung. Da bei Finanzmarktdaten in der Regel die dynamischen Eigenschaften im Vordergrund stehen, werden wir uns hauptsächlich auf die Zeitreihenanalyse konzentrieren. Zunächst werden wir Konzepte der univariaten Analyse vorstellen, später aber auch

multivariate Modelle betrachten, die Interdependenzen der Finanztitel explizit als System modellieren können.

Um einige Begriffe, die in der Zeitreihenanalyse und der Analyse finanzieller Zeitreihen oft gebraucht werden, in kompakter Form vorliegen zu haben, definieren wir sie kurz im nächsten Abschnitt.

10.1. Einige Definitionen

Zunächst benötigen wir den Begriff des stochastischen Prozesses, das grundlegende Konstruktions- und Analyseobjekt in der Zeitreihenanalyse.

DEFINITION 10.1 (Stochastischer Prozess)

Ein stochastischer Prozess X_t , $t \in \mathbb{Z}$, ist eine Familie von Zufallsvariablen, definiert auf einem Wahrscheinlichkeitsraum (Ω, \mathcal{F}, P) .

Zu einem bestimmten Zeitpunkt t ist X_t also eine Zufallsvariable mit einer bestimmten Verteilungsfunktion, und für ein bestimmtes $\omega \in \Omega$ ist $X(\omega) = \{X_t(\omega), t \in \mathbb{Z}\}$ eine Realisierung oder ein Pfad des Prozesses.

DEFINITION 10.2 (Verteilungsfunktion eines stochastischen Prozesses)

Die gemeinsame Verteilungsfunktion eines stochastischen Prozesses X_t wird definiert durch

$$F_{t_1, \dots, t_n}(x_1, \dots, x_n) = P(X_{t_1} \leq x_1, \dots, X_{t_n} \leq x_n).$$

Der stochastische Prozess X_t ist bereits eindeutig identifiziert, wenn man das System seiner Verteilungsfunktionen kennt. Ist also für beliebige $t_1, \dots, t_n \in \mathbb{Z}$ die gemeinsame Verteilungsfunktion $F_{t_1, \dots, t_n}(x_1, \dots, x_n)$ bekannt, so bestimmt dies eindeutig den zugrundeliegenden stochastischen Prozess.

DEFINITION 10.3 (Bedingte Verteilungsfunktion)

Die bedingte Verteilungsfunktion eines stochastischen Prozesses X_t für beliebige $t_1, \dots, t_n \in \mathbb{Z}$ mit $t_1 < t_2 < \dots < t_n$ wird definiert durch

$$F_{t_n | t_{n-1}, \dots, t_1}(x_n | x_{n-1}, \dots, x_1) = P(X_{t_n} \leq x_n | X_{t_{n-1}} = x_{n-1}, \dots, X_{t_1} = x_1).$$

Als nächstes definieren wir Momentfunktionen eines reellen stochastischen Prozesses. Hierfür gehen wir davon aus, dass die Momente existieren. Ist dies nicht der Fall, so ist die entsprechende Funktion nicht definiert.

DEFINITION 10.4 (Mittelwertfunktion)

Die Mittelwertfunktion μ_t eines stochastischen Prozesses X_t ist definiert durch

$$\mu_t = E[X_t] = \int_{\mathbb{R}} x dF_t(x). \quad (10.1)$$

Im allgemeinen hängt μ_t von der Zeit t ab, z.B. bei Prozessen mit einer saisonalen oder periodischen Struktur oder bei Prozessen mit einem deterministischen Trend.

DEFINITION 10.5 (Autokovarianzfunktion)

Die Autokovarianzfunktion eines stochastischen Prozesses X ist definiert durch

$$\gamma(t, \tau) = E[(X_t - \mu_t)(X_{t-\tau} - \mu_{t-\tau})] = \int_{\mathbb{R}^2} (x_1 - \mu_t)(x_2 - \mu_{t-\tau}) dF_{t,t-\tau}(x_1, x_2) \quad (10.2)$$

für $\tau \in \mathbb{Z}$.

Die Autokovarianzfunktion ist symmetrisch, d.h. $\gamma(t, \tau) = \gamma(t - \tau, -\tau)$. Für den Spezialfall $\tau = 0$ ergibt sich die Varianzfunktion $\gamma(t, 0) = \text{Var}(X_t)$. Im allgemeinen hängt $\gamma(t, \tau)$ sowohl von t als auch von τ ab. Im folgenden definieren wir das wichtige Konzept der Stationarität, das es ermöglicht, die Momentfunktionen in vielen Fällen zu vereinfachen.

DEFINITION 10.6 (Stationarität)

Ein stochastischer Prozess X ist kovarianzstationär, wenn

1. $\mu_t = \mu$, und
2. $\gamma(t, \tau) = \gamma_\tau$.

Ein stochastischer Prozess X_t ist streng stationär, wenn für beliebige t_1, \dots, t_n und für alle $n, s \in \mathbb{Z}$ gilt

$$F_{t_1, \dots, t_n}(x_1, \dots, x_n) = F_{t_1+s, \dots, t_n+s}(x_1, \dots, x_n).$$

Häufig wird für Kovarianzstationarität auch der Begriff der schwachen Stationarität verwendet. Man beachte jedoch, dass ein stochastischer Prozess streng stationär sein kann, ohne kovarianzstationär zu sein, nämlich dann, wenn die Varianz (oder die Kovarianz) nicht existiert. Existieren die ersten beiden Momentfunktionen, so folgt aus der strengen Stationarität die Kovarianzstationarität.

DEFINITION 10.7 (Autokorrelationsfunktion (ACF))

Die Autokorrelationsfunktion ρ eines kovarianzstationären stochastischen Prozesses ist definiert als

$$\rho_\tau = \frac{\gamma_\tau}{\gamma_0}.$$

Die ACF ist normiert auf das Intervall $[-1, 1]$ und erleichtert somit die Interpretation von Autokovarianzstrukturen verschiedener stochastischer Prozesse. Da gefordert wurde, dass der Prozess kovarianzstationär ist, hängt die ACF nur von einem Parameter, der Verzögerung τ ab. Häufig wird die ACF als Funktion von τ geplottet, und man erhält das sogenannte *Korrelogramm*. Dies ist ein wichtiges graphisches Instrument, um lineare Abhängigkeitsstrukturen des Prozesses aufzudecken.

Als nächstes definieren wir zwei wichtige stochastische Prozesse, die die Grundbausteine für die weitere Modellierung bilden.

DEFINITION 10.8 (weißes Rauschen (WR))

Der stochastische Prozess X_t ist weißes Rauschen, wenn gilt

1. $\mu_t = 0$, und

$$2. \quad \gamma_\tau = \begin{cases} \sigma^2 & \text{wenn } \tau = 0 \\ 0 & \text{wenn } \tau \neq 0. \end{cases}$$

 SFMtimewr.xpl

Ist X_t ein Prozess von u.i.v. Zufallsgrößen mit Erwartungswert 0 und endlicher Varianz, so ist er weißes Rauschen. Diesen Spezialfall nennen wir *unabhängiges weißes Rauschen*. Umgekehrt kann es sein, dass weißes Rauschen Abhängigkeiten in dritten oder höheren Momenten hat, also nicht unabhängig ist.

DEFINITION 10.9 (Irrfahrt)

Der stochastische Prozess X_t folgt einer Irrfahrt (oder Random Walk), wenn er sich re-präsentieren lässt als

$$X_t = c + X_{t-1} + \varepsilon_t$$

mit einer Konstanten c und weißem Rauschen ε_t .

Ist c ungleich Null, so haben die Zuwächse $Z_t = X_t - X_{t-1} = c + \varepsilon_t$ einen von Null verschiedenen Erwartungswert, und man spricht von einer Irrfahrt mit Drift (vgl. Abschnitt 4.1). Im Gegensatz zu Abschnitt ?? verlangen wir hier nicht, dass die Zuwächse unabhängig sind, damit die hier definierte Irrfahrt für $\alpha \rightarrow 1$ Grenzfall des im folgenden Beispiel 10.1 eingeführten AR(1)-Prozesses ist. Verlangen wir wie in Abschnitt ??, dass ε_t unabhängiges weißes Rauschen ist, so nennen wir X_t eine Irrfahrt mit unabhängigen Zuwächsen. Historisch nimmt die Irrfahrt eine besondere Rolle ein, da sie Anfang dieses Jahrhunderts das erste stochastische Modell für die Entwicklung von Aktienkursen darstellte. Bis in die heutige Zeit wurde die Irrfahrt häufig als Grundhypothese angenommen, überprüft und zumindest in ihrer strengen Fassung mit unabhängigen Zuwächsen meistens verworfen.

Um die Momentfunktionen einer Irrfahrt zu bestimmen, gehen wir der Einfachheit halber davon aus, dass die Konstante c und der Startwert X_0 auf Null gesetzt werden. Dann erhalten wir durch sukzessive Substitution die Repräsentation

$$X_t = \varepsilon_t + \varepsilon_{t-1} + \dots + \varepsilon_1.$$

Die Mittelwertfunktion ist einfach

$$\mu_t = E[X_t] = 0, \tag{10.3}$$

und für die Varianzfunktion erhalten wir wegen der Unkorreliertheit der ε_t

$$\text{Var}(X_t) = \text{Var}\left(\sum_{i=1}^t \varepsilon_i\right) = \sum_{i=1}^t \text{Var}(\varepsilon_i) = t\sigma^2. \tag{10.4}$$

Die Varianz der Irrfahrt steigt also linear mit der Zeit an. Für die Autokovarianzfunktion gilt für $\tau < t$:

$$\begin{aligned}\gamma(t, \tau) &= \text{Cov}(X_t, X_{t-\tau}) \\ &= \text{Cov}\left(\sum_{i=1}^t \varepsilon_i, \sum_{j=1}^{t-\tau} \varepsilon_j\right) \\ &= \sum_{j=1}^{t-\tau} \sum_{i=1}^t \text{Cov}(\varepsilon_i, \varepsilon_j) \\ &= \sum_{j=1}^{t-\tau} \sigma^2 = (t - \tau)\sigma^2.\end{aligned}$$

Für $\tau < t$ ist die Autokovarianz also stets positiv. Da die Kovarianzfunktion von der Zeit t abhängt (und nicht nur von der Zeitdifferenz τ) ist die Irrfahrt nicht kovarianzstationär. Für die Autokorrelationsfunktion ρ erhalten wir

$$\rho(t, \tau) = \frac{(t - \tau)\sigma^2}{\sqrt{t\sigma^2(t - \tau)\sigma^2}} = \frac{(t - \tau)}{\sqrt{t(t - \tau)}} = \sqrt{1 - \frac{\tau}{t}}.$$

Wieder hängt ρ sowohl von t als auch von τ ab, da die Irrfahrt nicht kovarianzstationär ist.

Zur weiteren Illustration betrachten wir als nächstes einen einfachen, aber wichtigen stochastischen Prozess.

BEISPIEL 10.1 (AR(1) Prozess)

Der stochastische Prozess X_t folgt einem autoregressiven Prozess erster Ordnung, kurz AR(1) Prozess, wenn

$$X_t = c + \alpha X_{t-1} + \varepsilon_t$$

mit dem konstanten Parameter α , $|\alpha| < 1$. Der Prozess X_t lässt sich durch iterierte Substitution auch schreiben als

$$\begin{aligned}X_t &= c(1 + \alpha + \alpha^2 + \dots + \alpha^{k-1}) + \alpha^k X_{t-k} + \varepsilon_t + \alpha\varepsilon_{t-1} + \dots + \alpha^{k-1}\varepsilon_{t-k+1} \\ &= c\left(\sum_{i=0}^{k-1} \alpha^i\right) + \alpha^k X_{t-k} + \sum_{i=0}^{k-1} \alpha^i \varepsilon_{t-i} \\ &= c \frac{1 - \alpha^k}{1 - \alpha} + \alpha^k X_{t-k} + \sum_{i=0}^{k-1} \alpha^i \varepsilon_{t-i}\end{aligned}$$

Ist X_{t-k} für ein bestimmtes k gegeben (z.B. der Startwert des Prozesses), so hängen offensichtlich die Eigenschaften des Prozesses von diesem Wert ab. Dieser Einfluss verschwindet aber mit wachsender Zeit, da wir angenommen haben, dass $|\alpha| < 1$ und somit $\alpha^k \rightarrow 0$ für $k \rightarrow \infty$. Für $k \rightarrow \infty$ existiert ein Grenzwert im Sinne der quadratischen Abweichung, also können wir den Prozess X_t auch schreiben als

$$X_t = c \frac{1}{1 - \alpha} + \sum_{i=0}^{\infty} \alpha^i \varepsilon_{t-i}.$$

Für die Momentfunktionen ergibt sich dann

$$\mu_t = c \frac{1}{1 - \alpha},$$

und

$$\gamma_\tau = \frac{\sigma^2}{1 - \alpha^2} \alpha^\tau.$$

Die ACF ist also einfach $\rho_\tau = \alpha^\tau$. Für positive α ist diese Funktion stets positiv, für negative α alterniert sie um Null. In jedem Fall konvergiert sie gegen Null, wobei die Konvergenz bei z.B. $\alpha = 0.5$ sehr schnell, bei $\alpha = 0.99$ recht langsam ist.  SFMacfar1.xpl

DEFINITION 10.10 (Markoff Prozess)

Ein stochastischer Prozess hat die Markoff-Eigenschaft, wenn für alle $t \in \mathbb{Z}$ und $k \geq 1$

$$F_{t|t-1, \dots, t-k}(x_t | x_{t-1}, \dots, x_{t-k}) = F_{t|t-1}(x_t | x_{t-1}).$$

Mit anderen Worten, die bedingte Verteilung eines Markoff-Prozesses zu einem bestimmten Zeitpunkt ist vollständig durch den Zustand des Systems zu dem vorherigen Zeitpunkt bestimmt. Man kann auch Markoff-Prozesse höherer Ordnung definieren, bei denen die bedingte Verteilung nur von einer endlichen Anzahl von vergangenen Zuständen abhängt. Zwei Beispiele für den Markoff-Prozess erster Ordnung sind die oben besprochene Irrfahrt mit unabhängigen Zuwächsen und der AR(1) Prozess mit unabhängigem weißem Rauschen.

DEFINITION 10.11 (Martingal)

Der stochastische Prozess X_t ist ein Martingal, wenn gilt

$$E[X_t | X_{t-1} = x_{t-1}, \dots, X_{t-k} = x_{t-k}] = x_{t-1}$$

für jedes $k > 0$.

Das Martingal ist ein ebenfalls häufig benutztes Instrument zur Beschreibung von Preisen auf Finanzmärkten. Man beachte, dass man nur eine Aussage über den bedingten Erwartungswert macht und nicht wie beim Markoff-Prozess über die gesamte bedingte Verteilung. Ein Beispiel für ein Martingal ist die Irrfahrt ohne Drift. Der AR(1) Prozess mit $0 < \alpha < 1$ ist kein Martingal, da $E[X_t | x_{t-1}, \dots, x_{t-k}] = c + \alpha X_{t-1}$.

DEFINITION 10.12 (Faires Spiel)

Der stochastische Prozess X_t ist ein faires Spiel, wenn gilt

$$E[X_t | X_{t-1} = x_{t-1}, \dots, X_{t-k} = x_{t-k}] = 0$$

für jedes $k > 0$.

Manchmal wird das faire Spiel auch als Martingal-Differenz bezeichnet. Ist nämlich X_t ein Martingal, so ist $Z_t = X_t - X_{t-1}$ ein faires Spiel.

DEFINITION 10.13 (Lag-Operator)

Der Operator L verschiebt den Prozess X_t um eine Zeiteinheit nach hinten, d.h. $LX_t = X_{t-1}$ und $L^k X_t = X_{t-k}$. Außerdem definieren wir den Differenzenoperator Δ als $\Delta = 1 - L$, d.h. $\Delta X_t = X_t - X_{t-1}$, und $\Delta^k = (1 - L)^k$.

Nach diesen mathematischen Definitionen kommen wir nun zu eher ökonomischen Definitionen, und hier insbesondere auf den Begriff der Rendite. Wir gehen dafür von einer Zeitreihe von Preisen P_1, \dots, P_n aus und wollen die Rendite zwischen zwei Perioden berechnen.

DEFINITION 10.14 (Einfache Rendite)

Die einfache Rendite R_t wird definiert als

$$R_t = \frac{P_t - P_{t-1}}{P_{t-1}}.$$

Soll die Durchschnittsrendite $R_t(k)$ über k Perioden berechnet werden, so wird das geometrische Mittel aus den einfachen Bruttorenditen gezogen, d.h.

$$R_t(k) = \left(\prod_{j=0}^{k-1} (1 + R_{t-j}) \right)^{1/k} - 1.$$

Im allgemeinen ist dies natürlich ungleich dem arithmetischen Mittel $k^{-1} \sum_{j=0}^{k-1} R_{t-j}$.

DEFINITION 10.15 (log Rendite)

Die log Rendite r_t wird definiert als

$$r_t = \ln \frac{P_t}{P_{t-1}} = \ln(1 + R_t).$$

 [SFMContDiscRet.xpl](#)

Die log Rendite ist definiert für den Fall der stetigen Verzinsung. Für die Durchschnittsrendite über mehrere Perioden ergibt sich

$$\begin{aligned} r_t(k) &= \ln\{1 + R_t(k)\} = \frac{1}{k} \ln \prod_{j=0}^{k-1} (1 + R_{t-j}) \\ &= \frac{1}{k} \sum_{j=0}^{k-1} \ln(1 + R_{t-j}) \\ &= \frac{1}{k} \sum_{j=0}^{k-1} r_{t-j}, \end{aligned}$$

d.h., für die log Rendite ergibt sich die Durchschnittsrendite als einfaches arithmetisches Mittel der log Renditen über die einzelnen Perioden.

Für kleine Preisänderungen ist die unterschiedliche Definition von Renditen nicht sehr wichtig. Nach einer Taylor-Reihenentwicklung folgt, dass

$$\begin{aligned}\ln(1+x) &= \ln(1) + \frac{\partial \ln x}{\partial x}(1)x + \frac{\partial^2 \ln x}{\partial x^2}(1)\frac{x^2}{2!} + \dots \\ &= x - \frac{x^2}{2!} + \frac{x^3}{3!} + \dots\end{aligned}$$

Für x nahe bei Null ist eine Approximation erster Ordnung ausreichend, d.h. $\ln(1+x) \approx x$. Als Faustregel kann man sagen, dass bei Renditen unter 10% es keinen großen Unterschied macht, ob man einfache oder log Renditen untersucht. Dies ist vor allem dann der Fall, wenn man Finanzzeitreihen von hoher Frequenz untersucht, also z.B. Tagesdaten.

10.2. Statistische Analyse deutscher Aktienrenditen

Im folgenden beschreiben wir einige typische Eigenschaften von Finanzzeitreihen am Beispiel der täglichen deutschen Aktienrenditen von 1974 bis 1996. Wir konzentrieren uns dabei zum einen auf die lineare zeitliche (Un-)abhängigkeit der Renditen, zum anderen auf die Verteilungseigenschaften. Tabelle 10.1 stellt die deskriptiven Statistiken zusammenfassend dar. Die Autokorrelation erster Ordnung ist bei allen Aktienrenditen nahe bei Null. Die größte positive Autokorrelation hat Preussag (0.08), die größte negative Autokorrelation hat Allianz (-0.06). Die Mehrzahl der Autokorrelationen ist positiv (14 gegenüber 6 negativen). Dies ist ein empirisches Phänomen, das auch für den amerikanischen Markt dokumentiert ist.

Während die Autokorrelation erster Ordnung der Renditen bei allen Aktienrenditen nahe bei Null ist, ist diejenige der quadrierten und absoluten Renditen bei allen Aktien positiv und deutlich größer als Null. Offenbar gibt es einen linearen Zusammenhang der absoluten Größe zeitlich aufeinanderfolgender Renditen. Da die Autokorrelation positiv ist, lässt sich schliessen, dass auf kleine (positive oder negative) Renditen tendenziell wieder kleine Renditen und auf große Renditen wieder große folgen. Mit anderen Worten, es gibt ruhige Perioden mit kleinen Kursausschlägen und turbulente Perioden mit großen Ausschlägen. In der Tat lässt sich weiter feststellen, dass diese Perioden von relativ langer Dauer sind, d.h., die Autokorrelationen quadrierter Renditen von meist sehr hoher Ordnung sind immer noch positiv. Diese Effekte wurden von Mandelbrot und Fama bereits in den sechziger Jahren untersucht. Sie lassen sich u.a. mit den in Kapitel 17 untersuchten ARCH Modellen modellieren.

Weiterhin betrachten wir Schätzer für Schiefe und Kurtosis. Während die Schiefe in den meisten Fällen nahe bei Null und mal positiv, mal negativ ist, ist in allen Fällen die Kurtosis deutlich größer als 3. Die kleinste geschätzte Kurtosis hat Thyssen ($\widehat{\text{Kurt}} = 6.1$), die größte Allianz ($\widehat{\text{Kurt}} = 32.4$). Unter der Nullhypothese der Normalverteilung sind die Schätzer in (3.2) und (3.3) unabhängig und asymptotisch normalverteilt mit

$$\sqrt{n}\hat{S} \xrightarrow{\mathcal{L}} N(0, 6)$$

und

$$\sqrt{n}(\widehat{\text{Kurt}} - 3) \xrightarrow{\mathcal{L}} N(0, 24).$$

	$\rho_1(r_t)$	$\rho_1(r_t^2)$	$\rho_1(r_t)$	S	K	BJ
ALLIANZ	-0.063191	0.3699	0.33492	0.078121	32.409	207116
BASF	-0.027967	0.2461	0.22837	-0.17267	8.6577	7693.5
BAYER	-0.033263	0.3356	0.2487	0.049908	9.6044	10447
BMW	-0.013375	0.34494	0.25604	-0.010743	17.029	47128
COMMERZBANK	0.048356	0.13101	0.21408	-0.24489	10.033	11902
DAIMLER	-0.027337	0.40502	0.3195	0.038118	26.673	134201
DEUTSCHE BANK	0.030428	0.28809	0.24078	-0.30991	13.773	27881
DEGUSSA	0.025011	0.31491	0.23487	-0.39493	19.127	62427
DRESDNER	0.063609	0.18462	0.22135	0.12226	8.8289	8150.2
HOECHST	0.01175	0.20284	0.19769	-0.12052	9.9881	11708
KARSTADT	0.0060163	0.29633	0.19635	-0.40415	20.436	72958
LINDE	-0.034012	0.19072	0.23075	-0.24326	14.565	32086
MAN	0.028026	0.28241	0.25065	-0.59105	18.034	54454
MANNESMANN	0.058174	0.17367	0.20481	-0.27015	13.692	27442
PREUSSAG	0.082669	0.14192	0.19323	0.13862	10.341	12923
RWE	0.040789	0.16416	0.23851	-0.19264	16.727	45154
SCHERING	0.069593	0.24929	0.22165	-0.035854	9.5773	10360
SIEMENS	0.064762	0.15746	0.18026	-0.5474	10.306	13070
THYSSEN	0.042602	0.15897	0.15531	-0.05009	6.103	2308
VOLKSWAGEN	0.059504	0.18904	0.16868	-0.32751	10.235	12637

Tabelle 10.1.: Autokorrelation erster Ordnung der Renditen ($\rho_1(r_t)$), der quadrierten Renditen ($\rho_1(r_t^2)$) und der absoluten Renditen ($\rho_1(|r_t|)$) sowie Schiefe (S), Kurtosis (K) und Bera-Jarque Test Statistik (BJ) für die täglichen Renditen deutscher Aktienkurse 1974-1996.

 [SFMReturns.xpl](#)

Daraus lässt sich der gemeinsame Test auf Normalverteilung von Bera und Jarque (BJ) ableiten:

$$BJ = n \left(\frac{\hat{S}^2}{6} + \frac{(\widehat{\text{Kurt}} - 3)^2}{24} \right).$$

BJ ist asymptotisch χ^2 verteilt mit zwei Freiheitsgraden. Die letzte Spalte von Tabelle 10.1 zeigt, dass in allen Fällen die Normalverteilungshypothese bei einem Testniveau von 1% (kritischer Wert 9.21) deutlich abgelehnt wird. Dies ist vor allem der größeren Kurtosis zuzuschreiben. Typischerweise ist bei Finanzzeitreihen die Kurtosis deutlich größer als 3, was durch das häufige Auftreten von 'Ausreißern' verursacht wird. Weiterhin gibt es ein häufigeres Auftreten von ganz kleinen Renditen als man es unter der Normalverteilungshypothese erwarten würde. Man sagt, die empirische Verteilung der Renditen ist *leptokurtisch*, was bedeutet, dass die Verteilung stärker gewölbt ist als die Normalverteilung. Das Gegenteil, eine schwächer gewölbte oder *platykurtische* Verteilung ist bei Finanzdaten selten anzutreffen.

10.3. Erwartungsbildung und Markteffizienz

Markteffizienz ist ein sehr allgemeines Konzept in der ökonomischen Theorie. Ein Markt wird effizient genannt, wenn zu jedem Zeitpunkt jede für die Bewertung der gehandelten Objekte

relevante Information durch die Preise vollständig widerspiegelt wird. Diese allgemeine Definition muss konkretisiert werden, um sagen zu können, was ‘vollständig widerspiegeln’ heißt. Hierfür benötigen wir das Konzept der *rationalen Erwartungen*. Allgemein spricht man von einer rationalen Erwartungsbildung, wenn bei der Prognose eines stochastischen Prozesses P_t sämtliche relevante und verfügbare Information \mathcal{F}_{t-1} (siehe Definition 5.1) ‘optimal’ ausgenutzt wird. Optimal heißt, dass der mittlere quadratische Prognosefehler minimiert wird. Dies ist der Fall, wenn zur Prognose der bedingte Erwartungswert (s. Abschnitt 3.5) $E[P_t | \mathcal{F}_{t-1}]$ verwendet wird:

SATZ 10.1

Für jedes $h > 0$ minimiert der bedingte Erwartungswert $E[P_{t+h} | \mathcal{F}_t]$ als Prognose den mittleren quadratischen Prognosefehler $E[(P_{t+h} - P_{t+h|t}^*)^2]$ unter allen bzgl. der zur Zeit t gegebenen Informationsmenge \mathcal{F}_t messbaren Prognosen $P_{t+h|t}^*$.

BEWEIS:

Für eine beliebige Prognose $P_{t+h|t}^*$, die man sich als (im allgemeinen nichtlineare) Funktion der zur Zeit t bekannten Zufallsvariablen, die die Informationsmenge \mathcal{F}_t bestimmen, vorstellen kann, lässt sich der mittlere quadratische Prognosefehler schreiben als

$$\begin{aligned} E[(P_{t+h} - P_{t+h|t}^*)^2] &= E[(P_{t+h} - E[P_{t+h} | \mathcal{F}_t] + E[P_{t+h} | \mathcal{F}_t] - P_{t+h|t}^*)^2] \\ &= E[(P_{t+h} - E[P_{t+h} | \mathcal{F}_t])^2] + E[\{E[P_{t+h} | \mathcal{F}_t] - P_{t+h|t}^*\}^2], \end{aligned} \quad (10.5)$$

da der Kreuzterm gleich null ist:

$$\begin{aligned} 2E\{(P_{t+h} - E[P_{t+h} | \mathcal{F}_t]) (E[P_{t+h} | \mathcal{F}_t] - P_{t+h|t}^*)\} &= \\ 2E[E\{P_{t+h} - E[P_{t+h} | \mathcal{F}_t] | \mathcal{F}_t\} (E[P_{t+h} | \mathcal{F}_t] - P_{t+h|t}^*)] &= \\ 2E\{0 \cdot (E[P_{t+h} | \mathcal{F}_t] - P_{t+h|t}^*)\} &= 0. \end{aligned}$$

Der zweite Term auf der rechten Seite von (10.5) ist nichtnegativ und nimmt den Wert Null an gerade für $E[P_{t+h} | \mathcal{F}_t] = P_{t+h|t}^*$. □

Man kann nun zunächst nicht davon ausgehen, dass alle Wirtschaftssubjekte über ausreichende Information verfügen, um $E[P_t | \mathcal{F}_{t-1}]$ schätzen zu können. Dies betrifft zum einen die Art des zugrundeliegenden Prozesses, der P_t treibt, zum anderen die für die Prognose benötigte Information relevanter Größen. Um diesem konzeptuellen Problem zu begegnen, entwickelten sich insbesondere in der makroökonomischen Theorie Hypothesen über die Mechanismen der Erwartungsbildung, die ohne den mathematischen Erwartungswert $E[P_t | \mathcal{F}_{t-1}]$ auskommen. Die Hypothese der *adaptiven Erwartungen* unterstellt beispielsweise, dass die Prognose zum Zeitpunkt $t - 1$ von P_t , $E_{t-1}^a[P_t]$, nach folgendem Mechanismus generiert wird:

$$E_{t-1}^a[P_t] - E_{t-2}^a[P_{t-1}] = \theta(P_{t-1} - E_{t-2}^a[P_{t-1}]) \quad (10.6)$$

mit konstantem Parameter θ , $0 < \theta < 1$. Änderungen der Prognosen ergeben sich also durch den mit θ gewichteten letzten Prognosefehler.

SATZ 10.2

Die adaptive Erwartungsbildung in (10.6) ist optimal im Sinne des mittleren quadratischen Prognosefehlers genau dann, wenn P_t dem Prozess

$$P_t = P_{t-1} + \varepsilon_t - (1 - \theta)\varepsilon_{t-1} \quad (10.7)$$

folgt, wobei ε_t weißes Rauschen sei.

BEWEIS:

Mit dem Lag-Operator L (siehe Definition 10.13) lässt sich (10.6) repräsentieren als

$$\{1 - (1 - \theta)L\} E_{t-1}^a[P_t] = \theta P_{t-1}.$$

Da $0 < \theta < 1$ und somit $\{1 - (1 - \theta)z\}^{-1} = \sum_{i=0}^{\infty} (1 - \theta)^i z^i$ lässt sich weiter schreiben

$$E_{t-1}^a[P_t] = \theta \sum_{i=0}^{\infty} (1 - \theta)^i P_{t-i-1}.$$

Der Prozess (10.7) lässt sich umformen zu

$$\{1 - (1 - \theta)L\} \varepsilon_t = P_t - P_{t-1}$$

und

$$\begin{aligned} \varepsilon_t &= \sum_{j=0}^{\infty} (1 - \theta)^j (P_{t-j} - P_{t-j-1}) \\ &= P_t - \theta \sum_{j=0}^{\infty} (1 - \theta)^j P_{t-j-1}, \end{aligned}$$

so dass $P_t - E_{t-1}^a[P_t]$ weißes Rauschen ist. Damit ist $E_{t-1}^a[P_t]$ in diesem Fall die beste Prognose für P_t im Sinne des mittleren quadratischen Prognosefehlers. □

Der Prozess (10.7) wird auch als integrierter autoregressiver Moving-Average Prozess (ARIMA) der Ordnung (0,1,1) bezeichnet. Die ARIMA Modellklasse wird näher in Kapitel 11 behandelt. Im allgemeinen können natürlich auch exogene Faktoren, z.B. Angebotschocks, bei der Bestimmung des Gleichgewichtspreises eine Rolle spielen. In diesem Fall wären adaptive Erwartungen suboptimal. Bezeichne X_t den stochastischen exogenen Faktor, von dem das Angebot abhängt, und \mathcal{F}_t eine Familie von Ereignissen, die sich aus den zur Zeit t verfügbaren Beobachtungen $\{p_t, p_{t-1}, \dots, x_t, x_{t-1}, \dots\}$ zusammensetzt. Dann ist die optimale Prognose $E[P_t \mid \mathcal{F}_{t-1}]$ im allgemeinen eine Funktion von $\{p_t, p_{t-1}, \dots\}$ und von $\{x_t, x_{t-1}, \dots\}$. Zwar gibt es Spezialfälle, in denen adaptive Erwartungen mit rationalen Erwartungen zusammenfallen, z.B. bei einem linearen Nachfrage/Angebotssystem mit X_t als exogenem Angebotschock, der einer Irrfahrt folgt. Ist X_t in so einem System aber z.B. ein AR(1) Prozess, so haben Prognosen aufgrund adaptiver Erwartungen einen größeren mittleren quadratischen Fehler als Prognosen mit rationalen Erwartungen. Ist der Faktor X_t

Typ	P1	P2
I	300	50
II	50	300
III	150	250

Tabelle 10.2.: Auszahlungen in Perioden P1 und P2 in Abhängigkeit vom Typ des Investors

common knowledge, d.h. öffentlich verfügbar, so bedeuten rationale Erwartungen in diesem Beispiel also, dass der Preis unter Ausnutzung dieser Information optimal prognostiziert wird.

Ist der Faktor X_t hingegen nicht für jeden beobachtbar, so können prinzipiell die nicht-informierten Subjekte durch die Preisgebote der informierten Subjekte *lernen*. Das heißt, sie können durch Beobachtung des Preises an Information über den Zustand ω , die über ihre private Informationsmenge F_i hinausgeht, gewinnen. Voraussetzung hierfür ist, dass die Informationsfunktion des Preises korrekt beurteilt wird.

Um zu illustrieren, welche Rolle der Preis bei der Bildung von Erwartungen spielt, stelle man sich den Kauf einer Flasche Wein vor. Im Geschäft steht man vor der Wahl zwischen drei Flaschen mit den Preisen EUR 300, EUR 30 und EUR 3. Da die Flasche zu EUR 300 das Budget übersteigt, bleiben noch die zu EUR 3 und EUR 30. Nun nehmen wir an, dass man als Nicht-Weinkenner aus den Angaben des Etiketts keine oder nur wenige Folgerungen über die Güte des Weines ableiten kann. Da man unter Zeitdruck steht, ist eine Informationsbeschaffung über andere Personen zu aufwendig. Was bleibt ist die Information über den Preis. Nehmen wir weiter an, dass man aus vorherigen Käufen bereits gelernt hat, dass teurere Weine tendenziell besser sind als billigere. Man hat also eine Vorstellung von der Funktion des Preises in Abhängigkeit von dem Zustand, d.h. der Güte des Weines. Man wird sich für den Wein zu EUR 30 entscheiden, wenn man den Gütevorteil des teureren Weines unter seiner Nutzenfunktion, die eine Funktion ist der Güte und des Preises, höher bewertet als den Preisvorteil des billigen Weines. Der Käufer hat *rational* gehandelt, da er sein Entscheidungsproblem (hier: Maximierung seiner Nutzenfunktion) optimiert hat unter Zuhilfenahme der ihm zur Verfügung stehenden Information und der Preisfunktion, vorausgesetzt seine Vermutung über die Preisfunktion erweist sich als richtig.

Zur Illustration noch ein Beispiel eines experimentellen Marktes, das der Literatur entnommen ist. Wir haben ein Wertpapier, das in zwei Perioden P1 und P2 gehandelt wird und in jeder Periode unterschiedliche Dividenden je nach Typ des Investors auszahlt. Der Handelsmechanismus ist ein Auktionssystem, in dem man bei einem Gebot gleichzeitig Kauf- und Verkaufspreis mündlich mitteilt. Es gibt drei Investortypen und von jedem Typ gibt es drei Investoren, d.h. insgesamt gibt es neun Investoren, die untereinander das Wertpapier handeln können. Jeder Investor ist mit einem Anfangskapital von 10.000 Franken (1 ‘Franken’ = 0.002 USD) und zwei Wertpapieren ausgestattet. Das Anfangskapital von 10.000 Franken muss am Ende der zweiten Periode wieder zurückgezahlt werden. Jeder Gewinn, der aus dem Handeln des Wertpapiers erfolgt, darf behalten werden. Hält ein Investor am Ende von P1 oder P2 das Wertpapier, so bekommt er eine Dividende in Abhängigkeit von seinem Typ. Tabelle 10.2 gibt Auskunft über die Dividendenzahlungen.

Jeder Investor kennt nur seine eigenen Auszahlungen, nicht die der anderen. Die Frage ist, ob und wenn ja wie schnell die Investoren über die Preisbildung ‘lernen’, d.h. Information über den Wert der Aktie für die anderen Investoren gewinnen. Es gibt zunächst zwei grundsätzliche Hypothesen:

1. Investoren richten sich in ihren Preisgeboten nur nach den eigenen Auszahlungen in P1 und P2 (‘naives Verhalten’).
2. Investoren ziehen über den beobachteten Preis Rückschlüsse auf den Wert der Aktie für die anderen Investoren und berücksichtigen diese Information bei ihren eigenen Geboten (‘rationales Verhalten’).

Da das Experiment nach der Periode P2 beendet ist, spielen in dieser Periode für jeden Investor nur noch die eigenen Auszahlungen eine Rolle, so dass in P2 beide Hypothesen zusammentreffen: Der Gleichgewichtspreis wird 300 Franken betragen, da Typ II gerade noch bereit ist, zu diesem Preis zu kaufen und es eine Konkurrenz gibt zwischen Investoren vom Typ II. Zu Beginn von P1, bevor gehandelt wurde, besteht für jeden Investor nur Information über seine eigenen Zahlungen, so dass man auch hier zunächst ein naives Verhalten annehmen muss: Typ I und Typ II bieten maximal 350, Typ III bietet maximal 400, also ist der Gleichgewichtspreis nach der Hypothese naiven Verhaltens 400 Franken. Dies hat sich auch in dem Experiment empirisch als gute Hypothese herausgestellt. Wird nun jedoch das Experiment mit der gleichen Auszahlungsmatrix wiederholt, so können die Investoren aus vergangenen Experimenten über den Preis lernen, welchen Wert die Aktie in P2 für andere Typen hat. Insbesondere wird unter der Hypothese rationalen Verhaltens Typ I lernen, dass der Gleichgewichtspreis in P2 höher ist, als es seiner eigenen Auszahlung entspricht, dass er also die Aktie zu einem höheren Preis verkaufen kann. Der Gleichgewichtspreis in P1 wird nach der rationalen Hypothese also 600 Franken betragen. Typ I kauft zu diesem Preis in P1 und verkauft in P2 an Typ II zum Preis von 300.

Im wiederholten Experiment hat sich herausgestellt, dass die Teilnehmer tatsächlich vom naiven Erwartungsgleichgewicht zum rationalen Erwartungsgleichgewicht tendieren, dass der Übergang jedoch nicht sofort nach dem ersten Experiment erfolgte, sondern graduell war und etwa acht Wiederholungen in Anspruch nahm. Es wurden noch andere Experimente durchgeführt, unter anderem ein *Terminmarkt*, in dem in P1 bereits der Preis der Aktie in P2 ausgehandelt werden konnte. Hier zeigte sich, dass durch die sofortige Transparenz des Wertes der Aktie in zukünftigen Perioden eine deutlich schnellere Anpassung an das rationale Erwartungsgleichgewicht erfolgte.

Der beobachtete Marktpreis bewirkt durch Interaktion verschiedener Anbieter und Nachfrager eine Aggregation der einzelnen heterogenen Informationsmengen. Wir nehmen an, dass der Preis zum Zeitpunkt t eine Funktion des Zustands ist und nennen dies die *Preisfunktion* $p_t(\omega)$, $\omega \in \Omega$. Wir definieren im folgenden ein Gleichgewicht mit rationalen Erwartungen.

DEFINITION 10.16 (RE-Gleichgewicht)

Ein Gleichgewicht in t mit rationalen Erwartungen (RE-Gleichgewicht) ist ein Gleichgewicht, in dem jedes Wirtschaftssubjekt i sein Entscheidungsproblem optimiert, gegeben seine Informationsmenge $\mathcal{F}_{i,t}$ und die Preisfunktion $p_t(\omega)$.

Definition 10.16 setzt also insbesondere voraus, dass jedes Subjekt die Informationsfunktion des Preises korrekt in sein Entscheidungsproblem einbezieht.

Das Konzept der *Markteffizienz* ist eng mit dem Konzept der rationalen Erwartungen verknüpft. Die ursprüngliche und allgemeine Definition sagt, dass ein Markt effizient ist, wenn zu jedem Zeitpunkt jede relevante Information sich in den Preisen widerspiegelt. Das heißt z.B., dass neue Information sich sofort in den Preisen niederschlägt. Im folgenden definieren wir Markteffizienz bezüglich einer Informationsmenge \mathcal{G} .

DEFINITION 10.17 (Markteffizienz)

Ein Markt ist effizient bezüglich $\mathcal{G} = (G_t)$, $t \in \mathbb{N}$, $G_t \subset \mathcal{F}_t$, wenn der Markt zu jedem Zeitpunkt t im RE-Gleichgewicht mit Preisfunktion $p_t(\omega)$ ist und für jedes Subjekt i und für jeden Zeitpunkt t gilt

$$G_t \subset \{\mathcal{F}_{i,t} \cup p_t(\omega)\}.$$

Üblicherweise werden drei typische Fälle unterschieden, die als schwache, semi-starke und starke Effizienz bezeichnet werden.

1. Der Markt ist *schwach effizient*, wenn er effizient ist bezüglich der Menge der historischen Preise, also der Menge $\mathcal{G} = (G_t)$, $G_t = \{p_t, p_{t-1}, p_{t-2}, \dots\}$. Dies ist z.B. dann erfüllt, wenn für alle i gilt $\{p_t, p_{t-1}, p_{t-2}, \dots\} \subset F_{i,t}$, wenn also die historischen Preise in jeder privaten Informationsmenge enthalten sind.
2. Der Markt ist *semi-stark effizient*, wenn er effizient ist bezüglich einer Menge $\mathcal{G} = (G_t)$, $(\cap_i F_{i,t}) \subset G_t \subset (\cup_i F_{i,t})$, die alle öffentlich zugängliche Information umfasst.
3. Der Markt ist *stark effizient*, wenn er effizient ist bezüglich der Menge $\mathcal{G} = (G_t)$, $G_t = \cup_i F_{i,t}$, wenn also durch die Preisfunktion sämtliche private Information öffentlich wird. Man spricht in diesem Fall auch von einem *fully revealing* RE-Gleichgewicht.

Eine äquivalente Definition besagt, dass bei Markteffizienz keine anormalen Renditen erzielt werden können. Um dies zu untersuchen, muss man jedoch erst festlegen, was ‘normale’ Renditen sind, d.h., man muss ein ökonomisches Modell angeben. Markteffizienz kann man dann nur im Zusammenhang mit diesem Modell testen. Wird diese gemeinsame Hypothese verworfen, so kann es sein, dass der Markt ineffizient ist, oder dass das ökonomische Modell nicht adäquat ist.

Wir geben im folgenden einen kurzen Überblick über typische ökonomische Modelle, die als Erklärung von finanziellen Größen dienen. Dabei stellen wir für die in erster Linie interessierenden finanziellen Werte – Aktien, Wechselkurse, Zinsen und Optionen – jeweils eine Theorie dar, die in ihrem jeweiligen Bereich als eine Standardtheorie oder klassische Theorie angesehen werden kann. Auf diese Theorien greifen wir in späteren Kapiteln dann zurück, wenn wir ökonometrische Modelle spezifizieren wollen und über empirisch motivierte Erweiterungen nachdenken.

10.4. Ökonomische Modelle: Ein kurzer Überblick

10.4.1. Aktienkurse: Das CAPM

Das *capital asset pricing model* (CAPM), von verschiedenen Autoren unabhängig voneinander in den sechziger Jahren entwickelt, ist das klassische Gleichgewichtsmodell für die Bewertung von risikobehafteten Wertpapieren (Aktien). Es basiert auf folgenden Annahmen:

1. Es besteht homogene Information unter den Marktteilnehmern. Diese Annahme kann abgeschwächt werden, indem man annimmt, dass bei heterogener Information ein RE-Gleichgewicht *fully revealing* ist (siehe die starke Version von Definition 10.17).
2. Der Markt ist friktionslos, d.h., es gibt keine Transaktionskosten, keine Steuern, keine Restriktionen bezüglich Leerverkauf und Teilbarkeit von Aktien.
3. Es besteht vollständige Konkurrenz.
4. Es gibt keine Arbitragegelegenheiten.
5. Es gibt endlich (K) viele Aktien und ein risikoloses Wertpapier mit der Rendite r .
6. Jeder Investor hat eine strikt konkave Nutzenfunktion als Funktion der unsicheren zukünftigen Auszahlung. Dies ist gleichbedeutend mit der Annahme, dass jeder Investor risikoavers ist.
7. Jeder Investor maximiert seinen Erwartungsnutzen, der nur abhängt von dem Erwartungswert und der Varianz der unsicheren zukünftigen Auszahlung. Dies ist die entscheidende Annahme des CAPM. Hinreichende Bedingungen für dieses $(\mu - \sigma)$ -Kriterium sind die beiden folgenden:
 - a) Jeder Investor hat eine quadratische Nutzenfunktion.
 - b) Die Aktienrenditen sind normalverteilt.

Im Folgenden bezeichnen $X_{i,t}$ und $\alpha_{i,t}$ den Kurs und die im Gleichgewicht angebotene Stückzahl der i -ten Aktie zum Zeitpunkt t . Wir definieren das *Marktportefeuille* $X_{m,t}$ als

$$X_{m,t} = \sum_{i=1}^K \alpha_{i,t} X_{i,t}. \quad (10.8)$$

Das relative Gewicht $w_{i,t}$ der i -ten Aktie in diesem Portefeuille ist folglich

$$w_{i,t} = \frac{\alpha_{i,t} X_{i,t}}{\sum_k \alpha_{k,t} X_{k,t}}.$$

Die meisten der bekannten Aktienindizes sind derartige wertgewichtete Indizes, allerdings sind auch häufig nur die größten Aktien des Marktes im Index enthalten (im DAX z.B. nur die 30 größten). Nun definieren wir wie in Definition 10.15 die Aktienrendite als $R_{i,t} =$

$\ln\{X_{i,t}/X_{i,t-1}\}$ und die Marktrendite als $R_{m,t} = \{\ln X_{m,t}/X_{m,t-1}\}$. Wir nehmen an, dass die Renditeprozesse kovarianzstationär sind. Im Gleichgewicht gilt nun nach dem CAPM für jede Aktie i

$$E[R_{i,t}] = r + \beta_i(E[R_{m,t}] - r), \quad (10.9)$$

mit dem ‘Beta’-Faktor

$$\beta_i = \frac{\text{Cov}(R_{i,t}, R_{m,t})}{\text{Var}(R_{m,t})}.$$

Gleichung (10.9) besagt, dass im Gleichgewicht die erwartete Rendite der i -ten Aktie sich aus zwei Komponenten zusammensetzt: aus der Rendite des risikolosen Wertpapiers und aus einer Risikoprämie, die für jede Aktie spezifisch durch ihren Beta-Faktor bestimmt wird. Aktien, die positiv mit dem Marktportefeuille korreliert sind, haben eine positive Risikoprämie. Je höher die Korrelation einer Aktie mit dem Marktportefeuille, desto größer der Beitrag im CAPM zum Portefeuillerisiko.

Da sich das CAPM nutzentheoretisch herleiten lässt, wird es manchmal als nachfrageorientiertes Gleichgewichtsmodell aufgefasst. Dem gegenüber gibt es Modelle, die die Aktienrenditen aufgrund verschiedener aggregierter Größen, sogenannter *Faktoren*, zu erklären versuchen und die eher als angebotsorientiert gelten können.

Wir werden in späteren Kapiteln insbesondere auf die Restriktion in Gleichung (10.9) auf zeitkonstante Betas eingehen und argumentieren, dass man anstelle der unbedingten die bedingte Varianz und Kovarianz explizit modellieren kann und somit ein Modell mit zeitabhängiger Risikoprämie erhält.

10.4.2. Wechselkurse: Die Zinsparitätentheorie

Wenn man bei Aktien von einer großen Vielfalt ökonomischer Modelle sprechen kann, so ist dies bei Wechselkursen noch stärker der Fall. Es gibt zwei grundlegende und sehr einfache Theorien, die aber bei weitem nicht ausreichen, die z.T. beachtlichen Kursbewegungen auf Devisenmärkten vor allem im kurzfristigen Bereich erklären zu können. Die *Kaufkraftparitätentheorie* (PPP) setzt voraus, dass gleiche Güter in verschiedenen Ländern den selben relativen Preis, d.h. ausgedrückt in Einheiten eines Numeraire, haben müssen. Man hat empirisch festgestellt, dass langfristig diese Theorie in guter Näherung die Realität beschreibt, dass aber kurzfristig die Kursbewegungen nicht erklärt werden können. Die zweite einfache Theorie, die *Zinsparitätentheorie*, vermag hier etwas mehr zu leisten, da Kapitalströme schneller fließen als Güterströme. Zinsdifferenzen können somit die Verschiebung von Kapital in andere Devisen und damit Wechselkursänderungen hervorrufen. Die Grundannahme der Zinsparitätentheorie ist, dass in- und ausländische Wertpapiere vollständige Substitute sind im Hinblick auf Laufzeit und Risikostruktur.

Wir nehmen an, dass neben dem Kassamarkt ein Terminmarkt besteht, auf dem Devisen auf Zeit gehandelt werden können. Den Kassakurs (*spot*) bezeichnen wir mit W_t^K , den Terminkurs mit W_t^T , jeweils ausgedrückt in Einheiten der ausländischen Währung, also z.B. EUR/USD. Ein international agierender Investor steht vor der Wahl einer Kapitalanlage im Inland zum inländischen Zins r_t^i und einer Anlage im Ausland zum ausländischen Zins r_t^a . Wählt er die Auslandsanlage, so muss er zunächst sein Kapital zum Kassakurs in

ausländische Devisen umtauschen und am Ende der Laufzeit dann wieder zurücktauschen. Die Unsicherheit über die zukünftige Wechselkursentwicklung kann er ausschalten, indem er per Termin kauft. Seine Rendite im Fall der Auslandsanlage ist also $(1/W_t^K)(1+r_t^a)W_t^T - 1$. Ist diese Rendite ungleich dem inländischen Zinssatz, so besteht ein Ungleichgewicht, das von sogenannten Zinsarbitrageuren sofort ausgenutzt wird, wodurch sich die Kurse anpassen. Im Gleichgewicht muss also gelten

$$\frac{W_t^T}{W_t^K} = \frac{1+r_t^i}{1+r_t^a}, \quad (10.10)$$

d.h., das Verhältnis von Terminkurs zum Kassakurs entspricht genau dem Verhältnis von in- und ausländischem Bruttozins. Die Beziehung (10.10) wird auch als *gesicherte Zinsparität* bezeichnet, da sie zum Zeitpunkt der Anlageentscheidung auf sicheren Kursen und Zinsen beruht.

Neben den Zinsarbitrageuren gibt es sogenannte Terminspekulanten, die den erwarteten zukünftigen Kassakurs mit dem Terminkurs vergleichen und entsprechend ihrer Risikoeinstellung bei einem Ungleichgewicht Devisen kaufen beziehungsweise verkaufen. Nehmen wir der Einfachheit halber Risikoneutralität der Terminspekulanten an. Dann ist im Gleichgewicht der erwartete Kassakurs genau gleich dem Terminkurs, d.h.

$$W_t^T = E[W_{t+1}^K | \mathcal{F}_t], \quad (10.11)$$

mit der Informationsmenge \mathcal{F}_t , die alle relevanten und verfügbaren Informationen enthält. Wir gehen hier davon aus, dass die Terminspekulanten rationale Erwartungen haben, d.h., dass sie den wahren datengenerierenden Prozess kennen und auf dieser Grundlage die optimale Prognose bilden. Dies lässt sich auch durch die Beziehung

$$W_{t+1}^K = E[W_{t+1}^K | \mathcal{F}_t] + \varepsilon_{t+1} \quad (10.12)$$

beschreiben, die besagt, dass die Abweichung der von den Spekulanten verwendeten Prognose $E[W_{t+1}^K | \mathcal{F}_t]$ von dem realisierten Kassakurs ein weißes Rauschen ε_t ist (vergleiche die Definition des weißen Rauschens in Definition 10.8). Der Markt ist ineffizient, wenn die Spekulanten tatsächlich risikoneutral sind, aber ε_t kein weißes Rauschen ist. In diesem Fall spiegelt die Menge \mathcal{F}_t in der Erwartungsbildung der Spekulanten nicht die gesamte relevante Information wider – sie haben keine rationalen Erwartungen. Um Markteffizienz zu testen (beziehungsweise um zu testen, ob ε_t weißes Rauschen ist) brauchen wir jedoch erst ein Modell für $E[W_{t+1}^K | \mathcal{F}_t]$. Dieses bekommen wir nun aus (10.11) und (10.10).

Setzt man (10.11) in (10.10) ein, so erhält man die sogenannte *ungesicherte Zinsparität*,

$$\frac{E[W_{t+1}^K | \mathcal{F}_t]}{W_t^K} = \frac{1+r_t^i}{1+r_t^a}. \quad (10.13)$$

Ungesichert ist diese Zinsparität deswegen, weil der zukünftige Kassakurs unsicher ist und als Erwartungswert in die Beziehung eingeht.

Zusammen mit (10.12) impliziert die ungesicherte Zinsparität (10.13), dass gilt

$$W_{t+1}^K = \frac{1+r_t^i}{1+r_t^a} W_t^K + \varepsilon_{t+1}. \quad (10.14)$$

Ist die Zinsdifferenz über lange Zeiträume im Mittel Null, so ist (10.14) eine Irrfahrt (siehe Definition 10.9). Die Irrfahrt ist also nach diesem einfachen Ansatz ein erstes Modell, um Wechselkurszeitreihen zu beschreiben.

Betont werden sollte, dass die Herleitung dieses einfachen Modells unter der Annahme der Risikoneutralität der Termispekulanten erfolgte. Im Fall der Risikoaversion müsste man eine Risikoprämie berücksichtigen. Wir können also z.B., wenn wir die Effizienz des Devisenmarktes testen wollen, nur die gemeinsame Hypothese der Effizienz und der ungesicherten Zinsparität mit Risikoneutralität testen. Eine Ablehnung dieser Hypothese kann auf Marktineffizienz beruhen, oder aber darauf, dass das Zinsparitätenmodell ein ungenügendes Modell für den Devisenmarkt ist.

10.4.3. Zinsstruktur: Das Cox-Ingersoll-Ross Modell

Mit Zinsstrukturmodellen versucht man, die zeitliche Entwicklung der Renditen festverzinslicher Wertpapiere in Abhängigkeit von der Laufzeit zu modellieren. Der klassische Ansatz ist, dass man einen oder mehrere Faktoren angibt, von denen man meint, dass sie die Zinsstruktur bestimmen. Durch Spezifizierung ihrer zeitlichen Dynamik und über bestimmte Erwartungshypothesen gelangt man zu einer expliziten Lösung für die Renditen.

Als ein typisches Beispiel stellen wir kurz das Modell von Cox, Ingersoll und Ross (CIR) dar, das wir bereits in Abschnitt 9.2.2 erwähnt haben. Wir bezeichnen mit $P_T(t)$ den Preis zum Zeitpunkt t eines *Zero Bonds* mit Nominalwert 1 EUR, d.h. eines dividendenfreien Wertpapiers, das zum Laufzeitende T genau einen EURO auszahlt. Mit $Y_T(t)$ bezeichnen wir die log Rendite dieses *Zero Bonds*, d.h. wir unterstellen kontinuierliche Verzinsung. Der Prozess $Y_T(t)$ wird häufig als *Yield to Maturity* bezeichnet. Die Beziehung zwischen Preis und Rendite des *Zero Bonds* ist

$$P_T(t) = \exp\{-Y_T(t)\tau\}$$

mit der Restlaufzeit $\tau = T - t$, wie man sich leicht aus der Definition der log Rendite (Definition 10.15) überlegt. Für sehr kurze Laufzeiten definieren wir die *Short rate* $r(t)$ als

$$r(t) = \lim_{T \rightarrow t} Y_T(t).$$

Praktisch entspricht die Short rate der Spot rate, d.h. dem Zinssatz für die kürzestmögliche Anlage (s. Abschnitt 9.2.2). Man hat, anschaulich gesprochen, die Wahl zwischen einer Anlage im *Zero Bond* mit der Rendite $Y_T(t)$ und einer wiederholten Anlage in zukünftigen Perioden zum (unsicheren) kurzfristigen Zins. Eine wichtige Erwartungshypothese sagt aus, dass gilt

$$P_T(t) = \mathbb{E} \left[\exp\left(-\int_t^T r(s) ds\right) \middle| \mathcal{F}_t \right] \quad (10.15)$$

(s. auch Gleichung (9.1) für variablen, aber deterministischen Zins). Die Short rate wird häufig als wichtigster Bestimmungsfaktor der Zinsstruktur angesehen. Wie das CIR-Modell benutzen die meisten Ein-Faktorenmodelle die Short rate als Faktor. Das CIR-Modell spezifiziert die Dynamik der Short rate als zeitstetigen stochastischen Prozess

TO DO:
CIR Q zur
Gl. (10.16)
hinzufügen

$$dr(t) = a(b - r(t))dt + \sigma\sqrt{r(t)}dW_t \quad (10.16)$$

mit Wiener Prozess W_t und konstanten Parametern a, b und σ - s. auch Abschnitt 9.2.2. Der Prozess (10.16) hat ein sogenanntes *mean reversion* Verhalten, d.h., dass bei Abweichungen von dem stationären Mittelwert b der Prozess durch ein positives a wieder zu diesem Wert gezogen wird. Die Volatilität, ausgedrückt durch $\sigma\sqrt{r(t)}$, ist umso größer, je höher das Zinsniveau ist, was auch empirisch feststellbar ist.

Da Gleichung (10.16) $r(t)$ als Markoff-Prozess spezifiziert, ist $P_T(t)$ dann wegen Gleichung (10.15) eine Funktion der aktuellen Short rate, d.h.

$$P_T(t) = V(r(t), t).$$

Mit Itô's Lemma (5.10) und (9.6) erhalten wir dann aus (10.16) die Differentialgleichung

$$a(b - r)\frac{\partial V(r, t)}{\partial r} + \frac{1}{2}\sigma^2 r\frac{\partial^2 V(r, t)}{\partial r^2} + \frac{\partial V(r, t)}{\partial t} - rV(r, t) = 0.$$

Mit der Randbedingung $V(r, T) = P_T(T) = 1$ führt dies zur Lösung

$$P_T(t) = V(r(t), t) = \exp\{A(T - t) + B(T - t)r(t)\}, \quad (10.17)$$

wobei (s. Abschnitt 9.2.4)

$$\begin{aligned} A(\tau) &= \frac{2ab}{\sigma^2} \ln \frac{2\psi \exp\{(a + \psi)\tau/2\}}{g(\tau)}, \\ B(\tau) &= \frac{2\{1 - \exp(\psi\tau)\}}{g(\tau)} \\ \psi &= \sqrt{a^2 + 2\sigma^2} \\ g(\tau) &= 2\psi + (a + \psi)\{\exp(\psi\tau) - 1\}. \end{aligned}$$

Für wachsende Laufzeiten $T - t$ konvergiert die Zinsstrukturkurve $Y_T(t)$ gegen den Wert

$$Y_{lim} = \frac{2ab}{\psi + a}.$$

Liegt der kurzfristige Zins über b , so ist die Zinsstruktur fallend (siehe das Beispiel in Abbildung 10.1), liegt er unter Y_{lim} , so ist sie steigend (Abbildung 10.2). Liegt er zwischen b und Y_{lim} , so kann sich eine zunächst steigende, dann fallende Kurve ergeben.

10.4.4. Derivate: Das Black-Scholes Modell

Da wir im ersten Teil des Buches ausführlich das Black-Scholes Modell zur Optionspreisbestimmung behandelt haben, geben wir hier nur eine kurze Zusammenfassung des Modells. Optionen sind nicht nur von theoretischem Interesse für die Finanzmathematik, sondern auch aus einem empirischen Blickwinkel interessant. Gerade in letzter Zeit gibt es Anzeichen für systematische Abweichungen realer Marktpreise vom Black-Scholes Modell. Über diese Abweichungen werden wir in späteren Kapiteln, insbesondere bei der Behandlung von ARCH Modellen, genauer aussagen.

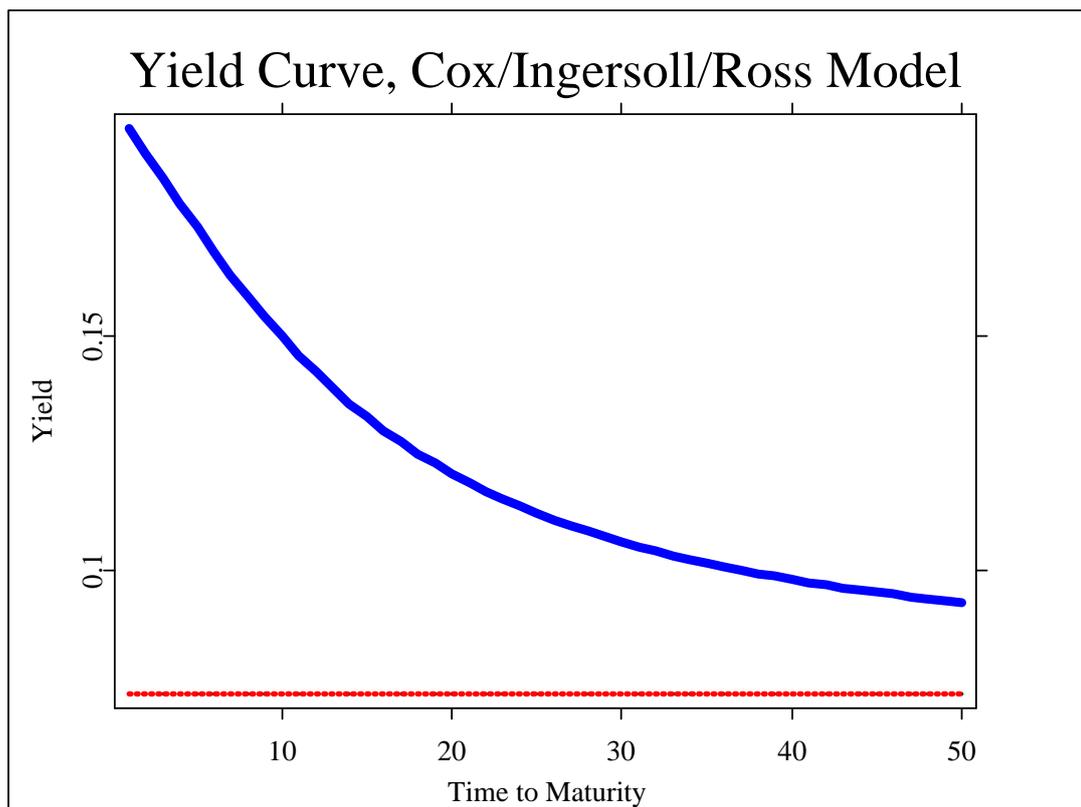


Abbildung 10.1.: Zinsstrukturkurve nach dem Modell von Cox-Ingersoll-Ross mit einer Short rate von $r_t=0.2$, $a = b = \sigma = 0.1$ und $Y_{lim} = 0.073$ (gestrichelte Gerade). [SFMcir.xpl](#)

Wir betrachten beispielhaft eine europäische Kaufoption auf eine im fraglichen Zeitraum dividendenfreie Aktie mit Kurs S_t zur Zeit t . $C(S, t)$ sei der Optionspreis zur Zeit t , wenn der aktuelle Kurs $S_t = S$ ist. Die Auszahlung am Laufzeitende T ist $C(S_T, T) = \max(0, S_T - K)$, wobei K der Ausübungspreis ist. Der Optionspreis bestimmt sich aus allgemeinen Arbitrageüberlegungen als

$$C(S_t, t) = E[e^{-r(T-t)} C(S_T, T) | \mathcal{F}_t],$$

wobei der Erwartungswert über eine geeignete risikoneutrale Verteilung gebildet wird - s. auch (6.23). r ist die als fest angenommene risikolose kontinuierliche Zinsrate.

Spezielle Ergebnisse können nur hergeleitet werden, wenn die Dynamik des Aktienkurses festgelegt wird. Die Grundannahme von Black und Scholes ist, dass der Aktienkurs S_t einer geometrischen Brownschen Bewegung folgt, d.h.,

$$dS_t = \mu S_t dt + \sigma S_t dW_t. \tag{10.18}$$

Der Optionspreis $C(S, t)$ erfüllt dann als Funktion von Zeit und Aktienkurs die Black-Scholes-

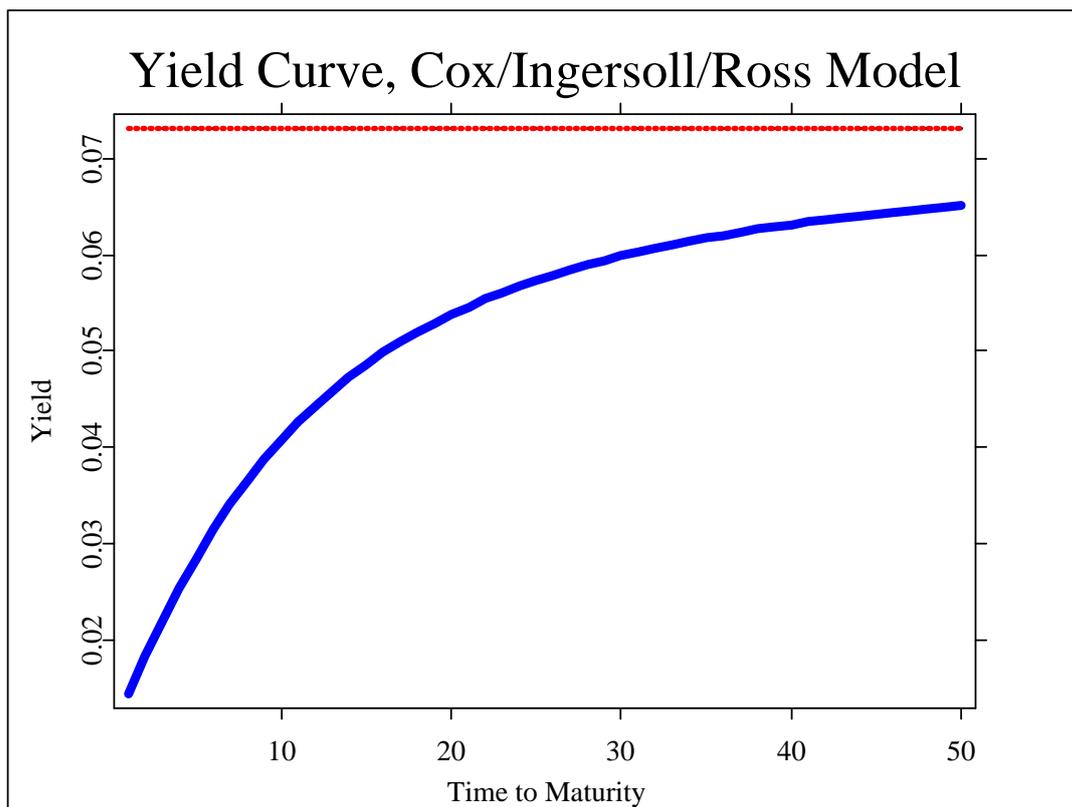


Abbildung 10.2.: Zinsstrukturkurve nach dem Modell von Cox-Ingersoll-Ross mit einer Short rate von $r_t=0.01$, $a = b = \sigma = 0.1$ und $Y_{lim} = 0.073$ (gestrichelte Gerade). [SFMcir.xpl](#)

Differentialgleichung (6.3)

$$\frac{1}{2}\sigma^2 S^2 \frac{\partial^2 C}{\partial S^2} + rS \frac{\partial C}{\partial S} + \frac{\partial C}{\partial t} = rC$$

Black und Scholes leiteten folgende Lösung her (s. Abschnitt 6.2):

$$C(S, t) = S\Phi(y + \sigma\sqrt{\tau}) - e^{-r\tau} K\Phi(y), \quad (10.19)$$

wobei $\tau = T - t$ für die Restlaufzeit der Option steht und y abkürzend für

$$y = \frac{\ln \frac{S}{K} + (r - \frac{1}{2}\sigma^2)\tau}{\sigma\sqrt{\tau}}.$$

10.4.5. Der Marktpreis des Risikos

In einer risiko-neutralen Welt ist der Marktpreis des Risikos - s. Abschnitt 9.2.3 - gleich Null. Im folgenden Abschnitt, der der Darstellung von Hull (2000) folgt, wollen wir auf den Marktpreis des Risikos (siehe 9.2.3) noch einmal eingehen und daraus eine heuristische Herleitung

der Black-Scholes Formel entwickeln. Hierzu betrachten wir derivative Finanzinstrumente, die zunächst von einem einzigen Zufallsprozess θ_t getrieben werden. Über den Prozess θ_t nehmen wir an, dass er einer geometrischen Brownschen Bewegung folgt:

$$d\theta_t = m\theta_t dt + s\theta_t dW_t. \quad (10.20)$$

Die Variable θ_t muss nicht notwendigerweise selbst eine finanzielle Größe darstellen. Sie könnte so etwas wie die Marktstimmung, ein Maß für die Popularität eines Politikers oder auch die Häufigkeit von ad-hoc Meldungen zur Zeit t sein. Nehmen wir nun an, dass V_{1t} und V_{2t} die Preise zweier derivativer Finanzinstrumente, die nur von θ_t und t abhängen, sind. Der Einfachheit halber lassen wir keine Zahlungen während der Beobachtungsperiode zu. Auch die Prozesse $V_{jt} = V_j(\theta, t)$, $j = 1, 2$ folgen dem Schema (10.20) mit dem gleichen Wienerprozess W_t

$$dV_{jt} = \mu_{jt}V_{jt}dt + \sigma_{jt}V_{jt}dW_t, \quad j = 1, 2 \quad (10.21)$$

wobei μ_{jt}, σ_{jt} Funktionen von θ_t und t sein können. Der Zufallsprozess W_t in (10.20) und (10.21) ist stets derselbe, da wir annehmen, dass dies die einzige Quelle der Unsicherheit darstellt.

Die Betrachtung von (10.21) in diskreter Zeit führt zu:

$$\Delta V_{1t} = \mu_{1t}V_{1t}\Delta t + \sigma_{1t}V_{1t}\Delta W_t \quad (10.22)$$

$$\Delta V_{2t} = \mu_{2t}V_{2t}\Delta t + \sigma_{2t}V_{2t}\Delta W_t \quad (10.23)$$

Wir können nun den ‘‘Zufall ΔW_t eliminieren’’ indem wir ein kontinuierlich sich veränderndes risikoloses Portfolio konstruieren. Dazu nehmen wir $\sigma_{2t}V_{2t}$ Einheiten des ersten Instrumentes und $-\sigma_{1t}V_{1t}$ des zweiten Instrumentes, d.h. wir verkaufen das zweite Instrument über Short selling. Falls Π_t den Gesamtwert des Portfolios zur Zeit t bezeichnet, ergibt sich:

$$\Pi_t = (\sigma_{2t}V_{2t})V_{1t} - (\sigma_{1t}V_{1t})V_{2t} \quad (10.24)$$

und

$$\Delta \Pi_t = (\sigma_{2t}V_{2t})\Delta V_{1t} - (\sigma_{1t}V_{1t})\Delta V_{2t} \quad (10.25)$$

Setzen wir (10.22) und (10.23) ein, ergibt sich:

$$\Delta \Pi_t = (\mu_{1t}\sigma_{2t}V_{1t}V_{2t} - \mu_{2t}\sigma_{1t}V_{1t}V_{2t})\Delta t. \quad (10.26)$$

Dieses Portfolio soll risikolos sein; daher muss es in der Zeitspanne Δt den risikolosen Ertrag $r\Delta t$ erbringen:

$$\frac{\Delta \Pi_t}{\Pi_t} = r\Delta t. \quad (10.27)$$

Setzen wir nun (10.24) und (10.26) in diese Gleichung ein, so folgt:

$$(\mu_{1t}\sigma_{2t}V_{1t}V_{2t} - \mu_{2t}\sigma_{1t}V_{1t}V_{2t})\Delta t = (\sigma_{2t}V_{1t}V_{2t} - \sigma_{1t}V_{1t}V_{2t})r\Delta t \quad (10.28)$$

$$\mu_{1t}\sigma_{2t} - \mu_{2t}\sigma_{1t} = r\sigma_{2t} - r\sigma_{1t} \quad (10.29)$$

$$\frac{\mu_{1t} - r}{\sigma_{1t}} = \frac{\mu_{2t} - r}{\sigma_{2t}} \quad (10.30)$$

Setzen wir dies wie in (9.2.3) gleich λ_t so sehen wir, dass bei einem Preis V_t eines derivativen Instrumentes, welches nur von θ_t und t abhängt und der Dynamik

$$dV_t = \mu_t V_t dt + \sigma_t V_t dW_t. \quad (10.31)$$

folgt, die Größe

$$\lambda_t = \frac{\mu_t - r}{\sigma_t} = \frac{\mu(\theta_t, t) - r}{\sigma(\theta_t, t)} \quad (10.32)$$

den *Marktpreis des Risikos* darstellt. Dieser Marktpreis des Risikos kann (durch μ_t, σ_t) von θ_t abhängen, jedoch nicht vom aktuellen Preis des Instrumentes V_t ! Wir können Gleichung (10.32) umschreiben zu:

$$\mu_t - r = \lambda_t \sigma_t \quad (10.33)$$

und können σ_t , welches bei diesen Überlegungen auch negativ werden kann, als die Größe des θ_t -Risikos in V_t interpretieren. Die Gleichung (10.33) hat starke Bezüge zum CAPM-Modell, welches wir in Sektion (10.4.1) behandelt haben - für weitere Details siehe auch (Hafner and Herwartz, 1998).

BEISPIEL 10.2 *Nehmen wir an, dass wir zwei Objekte haben, die beide von einer 90-tägigen Zinsrate abhängen. Das erste Instrument habe einen erwarteten Return von 3% p.a. und eine Volatilität von 20% p.a.. Für das zweite Instrument wird eine Volatilität von 30% p.a. angenommen. Weiterhin sei $r = 6\%$ p.a.. Der Marktpreis des Risikos für das erste Instrument ist nach (10.32):*

$$\frac{0.03 - 0.06}{0.2} = -0.15 \quad (10.34)$$

Durch Einsetzen in Gleichung (10.33) erhalten wir für das zweite Objekt:

$$0.06 - 0.15 \cdot 0.3 = 0.015 \quad (10.35)$$

oder 1.5% erwarteten Ertrag.

Da V_t eine Funktion von θ_t und t ist, können wir die Abhängigkeit von θ_t durch Itô's Lemma bestimmen. Die direkte Anwendung von Itô's Lemma (5.10) auf $V(\theta, t)$ ergibt durch Vergleich mit (10.31) für die Parameter in dieser Gleichung

$$\begin{aligned} \mu_t V_t &= m \theta_t \frac{\partial V_t}{\partial \theta_t} + \frac{\partial V_t}{\partial t} + \frac{1}{2} s^2 \theta_t^2 \frac{\partial^2 V_t}{\partial \theta^2} \\ \sigma_t V_t &= s \theta_t \frac{\partial V_t}{\partial \theta}. \end{aligned}$$

Wegen Gleichung (10.33) haben wir $\mu_t V_t - \lambda_t \sigma_t V_t = r V_t$, so dass wir die folgende Differentialgleichung für V_t erhalten:

$$\frac{\partial V_t}{\partial t} + (m - \lambda_t s) \theta_t \frac{\partial V_t}{\partial \theta} + \frac{1}{2} s^2 \theta_t^2 \frac{\partial^2 V_t}{\partial \theta^2} = r \cdot V_t \quad (10.36)$$

Diese Gleichung (10.36) ist der Black-Scholes Differentialgleichung sehr ähnlich und ist in der Tat identisch mit (6.3) für $\theta_t = S_t$, wenn S_t einen dividendenfreien Aktienkurs bezeichnet. In diesem Fall ist θ_t selbst der Preis eines Risikoträgers, nämlich der Aktie selbst, und muss daher wie der Preis V_t eines beliebigen auf der Aktie aufbauenden Derivats (10.33) erfüllen, d.h. wir erhalten

$$m - r = \lambda_t s, \quad (10.37)$$

so dass der zweite Term in (10.36) gleich

$$r\theta_t \frac{\partial V_t}{\partial \theta} \quad (10.38)$$

ist. Insgesamt erhalten wir also eine Differentialgleichung:

$$\frac{1}{2}s^2\theta_t^2 \frac{\partial^2 V_t}{\partial \theta^2} + r\theta_t \frac{\partial V_t}{\partial \theta} - rV_t + \frac{\partial V_t}{\partial t} = 0 \quad (10.39)$$

welche nach Umbenennen der Variablen identisch ist mit (6.3). Genauer setzt man $S_t = \theta_t$, $b = r$ (da es keine Dividendenerträge gibt) und $\sigma = s$ in der Notation des Abschnitts 6.1.

10.5. Die Irrfahrt-Hypothesen

Wir haben gesehen, dass ökonomische Modelle zumindest bei Aktienkursen und Wechselkursen eine Irrfahrt als statistisches Modell motivieren. Bei Wechselkursen sahen wir, dass als Konsequenz aus der ungesicherten Zinsparität und der angenommenen Risikoneutralität der Termispekulanten die Irrfahrt in (10.14) folgte. Nimmt man wie bei dem Modell von Black und Scholes eine geometrische Brownsche Bewegung für den Aktienkurs an wie in (10.18), so folgt aus Itô's Lemma, dass der Logarithmus des Aktienkurses einem Wiener Prozess mit konstanter Driftrate folgt:

$$d \ln S_t = \mu^* dt + \sigma dW_t \quad (10.40)$$

mit $\mu^* = \mu - \sigma^2/2$. Beobachtet man (10.40) in Zeitintervallen der Länge $\Delta > 0$, d.h. zu diskreten Zeitpunkten $0, \Delta, 2\Delta, \dots$, so ergibt sich

$$\ln S_{t\Delta} = \ln S_{(t-1)\Delta} + \Delta\mu^* + \sqrt{\Delta}\sigma\xi_t \quad (10.41)$$

mit unabhängig standardnormalverteilten ξ_t , $t = 1, 2, \dots$. Der Prozess (10.41) ist gerade eine Irrfahrt mit Drift für den logarithmierten Aktienkurs. Die log Renditen (siehe Definition 10.15) über Zeitintervalle der Länge Δ sind also unabhängig normalverteilt mit Erwartungswert $\Delta\mu^*$ und Varianz $\Delta\sigma^2$.

Bei langen Zinszeitreihen erscheint die Irrfahrt hingegen weniger plausibel, da man vermutet, dass es langfristig ein stationäres Niveau gibt, um das die Zinsen kurzfristig schwanken. Betrachten wir noch einmal den Prozess für die short rate in (10.16), der dem Cox-Ingersoll-Ross (CIR) Modell zugrundeliegt. Eine diskrete Approximation ist

$$r_t - r_{t-1} = \alpha + \beta r_{t-1} + \sigma\sqrt{r_{t-1}}\xi_t$$

oder

$$r_t = \alpha + (1 + \beta)r_{t-1} + \sigma\sqrt{r_{t-1}}\xi_t. \quad (10.42)$$

Ist β in (10.42) negativ (und größer als -2), so ist der Prozess ein stationärer AR(1) Prozess mit heteroskedastischen Fehlern. In Beispiel 10.1 haben wir einen solchen Prozess mit homoskedastischen Fehlern behandelt.

Es gibt auch die Auffassung, dass Zinsen zumindest auf kurze Sicht gut durch eine Irrfahrt zu beschreiben sind. Es ist deshalb allgemein von Interesse, zu testen, ob eine Irrfahrt vorliegt. Wir geben im folgenden noch eine wichtige Abgrenzung zwischen drei Versionen der Irrfahrt Hypothese. Allgemein betrachten wir die Irrfahrt mit Drift

$$P_t = \mu + P_{t-1} + \varepsilon_t. \quad (10.43)$$

1. Die stochastischen Fehler in (10.43) sind unabhängig identisch verteilt (u.i.v.) mit Erwartungswert Null und Varianz σ^2 . Diese Hypothese wurde schon in den sechziger Jahren an vielen Datensätzen überprüft und als empirisch nicht haltbar befunden. Man entdeckte z.B. bestimmte Volatilitätscluster, die unter der u.i.v. Hypothese statistisch nicht zu erwarten sind.
2. Die stochastischen Fehler in (10.43) sind unabhängig, aber nicht notwendig identisch verteilt mit Erwartungswert Null. Diese Hypothese ist schwächer als die erste, da sie z.B. Heteroskedastizität zulässt. Allerdings sind auch hier die empirischen Befunde derart, dass eine Abhängigkeit der Fehler unterstellt werden muss.
3. Die stochastischen Fehler in (10.43) sind unkorreliert, d.h. $\gamma_\tau(\varepsilon_t) = 0$ für alle $\tau \neq 0$. Dies ist die schwächste und am meisten diskutierte Irrfahrt Hypothese. Empirisch wird sie am häufigsten durch statistische (Nicht-)Signifikanz der geschätzten Autokorrelationen von ε_t getestet.

Die Diskussion der Irrfahrt Hypothesen behandelt vor allem die Prognostizierbarkeit von Finanzzeitreihen. Eine andere Diskussion behandelt die Frage, ob das Modell (10.43) mit unabhängigen bzw. unkorrelierten Fehlern überhaupt ein vernünftiges Modell für Finanzzeitreihen darstellt oder ob man nicht vielmehr ein Modell mit einem deterministischen Trend unterstellen sollte. Solch ein *trendstationäres* Modell hat die Form

$$P_t = \nu + \mu t + \varepsilon_t \quad (10.44)$$

mit konstanten Parametern ν und μ . Der Prozess (10.44) ist instationär, da z.B. gilt $E[P_t] = \nu + \mu t$, d.h., der Erwartungswert ist zeitabhängig. Bereinigt man P_t um den linearen Zeittrend, so erhält man den stationären Prozess $P_t - \mu t$.

Zum Vergleich zwischen der differenzenstationären Irrfahrt mit Drift und dem trendstationären Prozess (10.44) schreiben wir die Irrfahrt aus (10.43) durch sukzessives Einsetzen als

$$P_t = P_0 + \mu t + \sum_{i=1}^t \varepsilon_i, \quad (10.45)$$

mit festem Startwert P_0 . Man sieht, dass die Irrfahrt mit Drift auch einen linearen Zeittrend impliziert, dass aber die stochastischen Zuwächse nicht wie beim trendstationären Prozess stationär (ε_t) sondern instationär ($\sum_{i=1}^t \varepsilon_i$) sind. Aufgrund der Darstellung (10.45) wird die Irrfahrt mit oder ohne Drift auch als *integriert* bezeichnet, da die Abweichung von einem deterministischem Trend eine Summe von Fehlern ist. Außerdem hat jeder Fehler ε_t einen permanenten Einfluss auf alle zukünftigen Werte des Prozesses. Für die beste Prognose im Sinne des mittleren quadratischen Fehlers gilt für alle $k > 0$,

$$E[P_{t+k} | \mathcal{F}_t] = P_0 + \mu(t+k) + \sum_{i=1}^t \varepsilon_i.$$

Im Gegensatz dazu ist der Einfluss eines Schocks ε_t auf die Prognosen des trendstationären Prozesses (10.44) gleich Null, d.h.

$$E[P_{t+k} | \mathcal{F}_t] = \nu + \mu(t+k).$$

Es ist also prinzipiell wichtig, zwischen einem differenzen- und einem trendstationären Prozess zu unterscheiden. Betont sei hierbei, dass die Irrfahrt nur ein spezieller Fall eines differenzenstationären Prozesses ist. Sind z.B. die Zuwächse in (10.43) zwar stationär, jedoch autokorreliert, so liegt ein allgemeinerer differenzenstationärer Prozess vor. Es gibt nun viele statistische Tests, die darauf testen, ob der Prozess differenzenstationär ist oder nicht. Zwei dieser Tests wollen wir im nächsten Abschnitt besprechen.

10.6. Einheitswurzeltests

Wir haben in Beispiel 10.1 den AR(1) Prozess

$$X_t = c + \alpha X_{t-1} + \varepsilon_t \tag{10.46}$$

besprochen. Gilt $|\alpha| < 1$, so ist der Prozess stationär falls $E[X_0] = \frac{c}{1-\alpha}$ oder nach einer “Einschwingphase”. Der Fall $\alpha = 1$ entspricht der Irrfahrt, die nichtstationär ist. Das Verhalten eines stationären AR(1) Prozesses mit α nahe bei Eins ist dem der Irrfahrt recht ähnlich, so dass man häufig testen möchte, ob der Fall $\alpha = 1$ oder $\alpha < 1$ vorliegt. Zu diesem Zweck sind sogenannte *Einheitswurzeltests* entwickelt worden.

10.6.1. Dickey-Fuller Tests

Die von Dickey und Fuller entwickelten Einheitswurzeltests testen die Nullhypothese einer *Einheitswurzel*, d.h. einer Nullstelle der charakteristischen Gleichung (11.6) des AR(1)-Prozesses bei $z = 1$, gegen die Alternativhypothese eines Prozesses ohne Einheitswurzel. Grundlage des Tests bildet die Regression

$$\Delta X_t = c + (\alpha - 1)X_{t-1} + \varepsilon_t, \tag{10.47}$$

die sich aus Umformung von (10.46) ergibt. Ist X_t eine Irrfahrt, so ist der Koeffizient von X_{t-1} gleich Null. Ist X_t hingegen ein stationärer AR(1) Prozess, so ist der Koeffizient negativ.

Man bildet nun die Standard t-Statistik

$$\hat{t}_n = \frac{\hat{\alpha} - 1}{\sqrt{\hat{\sigma}^2 \sum_{t=2}^n X_{t-1}^2}}, \quad (10.48)$$

wobei $\hat{\alpha}$ und $\hat{\sigma}^2$ die Kleinste-Quadrate Schätzer für α und die Varianz σ^2 von ε_t sind. Die Statistik (10.48) konvergiert für wachsendes n nicht gegen eine Standardnormalverteilung, sondern gegen die Verteilung einer Funktion von Wienerprozessen,

$$\hat{t}_n \xrightarrow{\mathcal{L}} \frac{1 - W^2(1)}{2 \left\{ \int_0^1 W^2(u) du \right\}^{1/2}},$$

wobei W ein Standard-Wienerprozess ist. Die kritischen Werte der Verteilung sind z.B. zum 1%, 5% und 10% Niveau -2.58, -1.95, beziehungsweise -1.62.

Ein Problem des Tests ist, dass das nominale Testniveau (z.B. 5%) nicht eingehalten wird, wenn die Fehler ε_t in (10.47) autokorreliert sind. Je größer die Autokorrelation von ε_t , desto größer im allgemeinen auch die Verzerrung des Testniveaus. Vernachlässigt man also diese Autokorrelation, so kann es sein, dass man die Nullhypothese einer Einheitswurzel zum vermeintlichen Testniveau von 5% ablehnt, in Wirklichkeit aber das Niveau bei z.B. 30% liegt. Um diese negativen Effekte zu vermeiden, schlagen Dickey und Fuller eine erweiterte Regression vor, die verzögerte Differenzen enthält. Bei dem *augmented Dickey Fuller Test* (ADF) wird also aus der Regression

$$\Delta X_t = c + (\alpha - 1)X_{t-1} + \sum_{i=1}^p \alpha_i \Delta X_{t-i} + \varepsilon_t \quad (10.49)$$

wie beim einfachen Dickey-Fuller Test die Nullhypothese einer Einheitswurzel abgelehnt, wenn die Test-Statistik (10.48) kleiner ist als der kritische Wert, der wieder tabelliert ist. Problematisch ist natürlich die Wahl von p . Im allgemeinen gilt, dass das Testniveau besser eingehalten wird, wenn p größer wird, dass dann jedoch der Test an Güte (*Macht*) verliert. Dies sei an einem simulierten Beispielprozess illustriert. Die Fehler ε_t seien autokorreliert durch die Beziehung

$$\varepsilon_t = \beta \xi_{t-1} + \xi_t$$

wobei ξ_t u.i.v. $(0, \sigma^2)$ sei. Im nächsten Kapitel werden wir diesen Prozess als *moving average* Prozess der Ordnung 1, MA(1), bezeichnen. Es gilt offensichtlich $\text{Var}(\varepsilon_t) = \sigma^2(1 + \beta^2)$ und $\gamma_1(\varepsilon_t) = \text{Cov}(\varepsilon_t, \varepsilon_{t-1}) = \beta\sigma^2$, und $\gamma_\tau(\varepsilon_t) = 0$ für $\tau \geq 2$. Für die ACF von ε_t bekommen wir also

$$\rho_\tau(\varepsilon_t) = \begin{cases} \frac{\beta}{1+\beta^2} & \text{wenn } \tau = 1 \\ 0 & \text{wenn } \tau \geq 2. \end{cases} \quad (10.50)$$

Für den Prozess

$$X_t = \alpha X_{t-1} + \beta \xi_{t-1} + \xi_t \quad (10.51)$$

wurden Simulationen des ADF Tests durchgeführt, die in Tabelle 10.3 in gekürzter Form zusammengefasst sind.

α	p	β			
		-0.99	-0.9	0	0.9
1	3	0.995	0.722	0.045	0.034
	11	0.365	0.095	0.041	0.039
0.9	3	1.000	0.996	0.227	0.121
	11	0.667	0.377	0.105	0.086

Tabelle 10.3.: ADF-Test: Simulierte Ablehnwahrscheinlichkeiten für den Prozess (10.51) bei einem nominalen Niveau von 5% (nach [Friedmann \(1992\)](#)).

Wie man sieht, wird das nominale Testniveau von 5% unter der Nullhypothese ($\alpha = 1$) besser eingehalten, je größer p ist. Auf der anderen Seite sinkt dann jedoch die Güte des Tests, d.h., der Test vermag nicht mehr gut zwischen dem Prozess mit Einheitswurzel und dem stationären Prozess mit $\alpha = 0.9$ zu unterscheiden. Es besteht also bei der Wahl von p ein Konflikt zwischen Validität und Güte des Tests.

Ist X_t ein trendstationärer Prozess wie in (10.44), so lehnt der ADF Test ebenfalls die (falsche) Nullhypothese einer Einheitswurzel zu selten ab. Asymptotisch geht die Ablehnwahrscheinlichkeit sogar gegen Null. Man kann nun die ADF Regression (10.49) um einen linearen Zeittrend erweitern, d.h. man führt die Regression

$$\Delta X_t = c + \mu t + (\alpha - 1)X_{t-1} + \sum_{i=1}^p \alpha_i \Delta X_{t-i} + \varepsilon_t \quad (10.52)$$

durch und testet wieder auf Signifikanz von α . Die kritischen Werte sind tabelliert. Der ADF Test mit Zeittrend (10.52) hat Macht gegen einen trendstationären Prozess. Andererseits verliert er an Macht gegenüber dem einfachen ADF Test (10.49), wenn der wahre Prozess z.B. ein stationärer AR(1) Prozess ist.

Als empirisches Beispiel betrachten wir die täglichen Aktienkurse 20 großer deutscher Aktiengesellschaften im Zeitraum 2.1.1974 bis 30.12.1996. Tabelle 10.4 zeigt die ADF Test Statistiken für die logarithmierten Aktienkurse mit der Wahl $p = 0$ und $p = 4$. Die Tests wurden mit und ohne linearen Zeittrend durchgeführt. In jeder Regression wurde ein Absolutglied mitgeschätzt.

Nur für den Fall RWE mit linearem Zeittrend lehnt der ADF Test die Nullhypothese einer Einheitswurzel bei einem Testniveau von 10% ab. Da in allen anderen Fällen die Einheitswurzel nicht verworfen wird, scheint Differenzenbildung bei Aktienkursen eine notwendige Operation zu sein, um einen stationären Prozess, d.h. die log Renditen, zu bekommen, der dann weiter untersucht wird. Dieses Ergebnis wird im nächsten Abschnitt noch durch einen weiteren Test untermauert.

10.6.2. Der KPSS Test auf Stationarität

Der KPSS Test von Kwiatkowski, Philipps, Schmidt und Shin testet auf Stationarität, d.h. gegen eine Einheitswurzel. Er vertauscht also im Vergleich zum ADF Test die Hypothesen. Es sind wie beim ADF Test zwei Fälle zu unterscheiden, je nachdem ob man mit oder ohne

p bzw. T	ADF				KPSS			
	ohne Zeittrend		mit Zeittrend		ohne Zeittrend		mit Zeittrend	
	0	4	0	4	8	12	8	12
ALLIANZ	-0.682	-0.616	-1.917	-1.815	24.52**	16.62**	2.36**	1.61**
BASF	0.138	0.341	-1.923	1.700	23.71**	16.09**	1.39**	0.95**
BAYER	-0.112	0.080	-2.257	-2.002	24.04**	16.30**	1.46**	1.00**
BMW	-0.708	-0.657	-2.379	-2.365	23.92**	16.22**	2.01**	1.37**
COMMERZBANK	-0.797	-0.672	-2.533	-2.470	22.04**	14.96**	1.43**	0.98**
DAIMLER	-1.365	-1.291	-1.653	-1.595	22.03**	14.94**	3.34**	2.27**
DEUTSCHE BANK	-1.386	-1.271	-1.923	-1.986	23.62**	16.01**	1.70**	1.16**
DEGUSSA	-0.447	-0.355	-2.622	-2.666	23.11**	15.68**	1.79**	1.22**
DRESDNER	-0.982	-0.938	-2.106	-2.186	22.40**	15.20**	1.79**	1.22**
HOECHST	0.364	0.504	-1.683	-1.548	23.80**	16.15**	1.42**	0.97**
KARSTADT	-1.181	-1.170	-1.848	-1.840	20.40**	13.84**	3.33**	2.26**
LINDE	-1.689	-1.439	-2.481	-2.252	24.40**	16.54**	3.14**	2.15**
MAN	-1.777	-1.575	-3.064	-2.898	21.97**	14.91**	1.59**	1.08**
MANNESMANN	-0.909	-0.795	-2.436	-2.421	21.97**	14.93**	1.89**	1.29**
PREUSSAG	-1.400	-1.380	-1.636	-1.812	23.18**	15.72**	1.53**	1.04**
RWE	-0.094	-0.039	-3.140*	-3.267*	24.37**	16.52**	1.66**	1.14**
SCHERING	0.108	0.041	-2.539	-2.789	24.20**	16.40**	2.35**	1.60**
SIEMENS	-1.352	-1.198	-1.752	-1.939	23.24**	15.76**	1.69**	1.15**
THYSSEN	-1.452	-1.338	-2.284	-2.202	21.97**	14.90**	1.98**	1.35**
VOLKSWAGEN	-0.940	-0.809	-2.313	-2.396	21.95**	14.89**	1.11**	0.76**

Tabelle 10.4.: Einheitswurzeltests: ADF Test (Nullhypothese: Einheitswurzel) und KPSS Test (Nullhypothese: Stationarität). Der augmentierte Teil der ADF Regression hat die Ordnung $p = 0$ und $p = 4$. Die KPSS Statistiken wurden mit dem Stützpunkt $T = 8$ und $T = 12$ berechnet. Die Sterne markieren Signifikanz zum 10% (*) bzw. 1% (**) Niveau.

 SFMAdfKpss.xpl

linearem Zeittrend schätzen will. Das Regressionsmodell mit Zeittrend hat die Form

$$X_t = c + \mu t + k \sum_{i=1}^t \xi_i + \eta_t, \quad (10.53)$$

mit stationärem η_t und u.i.v. ξ_t mit Erwartungswert 0 und Varianz 1. Offensichtlich ist für $k \neq 0$ der Prozess integriert und für $k = 0$ trendstationär. Die Nullhypothese lautet also $H_0 : k = 0$, die Alternativhypothese $H_1 : k \neq 0$.

Man führt nun unter H_0 die Regression (10.53) mit der Methode der kleinsten Quadrate durch und erhält die Residuen $\hat{\eta}_t$. Damit bildet man die Partialsummen

$$S_t = \sum_{i=1}^t \hat{\eta}_i,$$

die bei Gültigkeit von H_0 integriert der Ordnung Eins sind, d.h., die Varianz von S_t steigt linear mit t . Die KPSS Test Statistik ist nun

$$KPSS_T = \frac{\sum_{t=1}^n S_t^2}{n^2 \hat{\omega}_T^2}, \quad (10.54)$$

wobei

$$\hat{\omega}_T^2 = \hat{\sigma}_\eta^2 + 2 \sum_{\tau=1}^T \left(1 - \frac{\tau}{T-1}\right) \hat{\gamma}_\tau$$

ein Schätzer der Spektraldichte an der Frequenz Null ist, mit $\hat{\sigma}_\eta^2$ dem Varianzschätzer von η_t und $\hat{\gamma}_\tau = 1/n \sum_{t=\tau+1}^n \hat{\eta}_t \hat{\eta}_{t-\tau}$ dem Kovarianzschätzer. Eine Schwierigkeit ist wiederum die Bestimmung des Stützpunktes T : für zu kleine T ist der Test bei Autokorrelation verzerrt, für zu große T verliert er hingegen an Güte.

Die Ergebnisse des KPSS Tests in Tabelle 10.49 deuten eindeutig darauf hin, dass die untersuchten Aktienkurse nicht stationär oder trendstationär sind, da in jedem Fall die Nullhypothese zum Niveau 1% abgelehnt wurde. Auch im Fall RWE, bei dem der ADF Test eine Signifikanz zum Niveau 10 % auswies, deutet das hier viel niedrigere Signifikanzniveau auf eine Präferenz der Hypothese einer Einheitswurzel hin.

10.6.3. Varianzquotiententests

Will man testen, ob eine Zeitreihe einer Irrfahrt folgt, so kann man die Tatsache ausnutzen, dass die Varianz einer Irrfahrt linear mit der Zeit steigt, siehe (10.4). Betrachten wir die logarithmierten Preise oder Kurse einer Finanzzeitreihe, $\ln S_t$, so lautet die Nullhypothese

$$H_0 : r_t = \mu + \varepsilon_t, \quad \varepsilon_t \sim N(0, \sigma^2)$$

mit den log Renditen $r_t = \ln S_t - \ln S_{t-1}$, konstantem μ und ε_t weißem Rauschen. Eine Alternativhypothese ist z.B., dass r_t stationär und autokorreliert ist. Man bildet die Summen über die Renditen

$$r_t(q) = r_t + r_{t-1} + \dots + r_{t-q+1}$$

und bestimmt die Varianz von $r_t(q)$. Für $q = 2$ gilt z.B.

$$\begin{aligned} \text{Var}\{r_t(2)\} &= \text{Var}(r_t) + \text{Var}(r_{t-1}) + 2\text{Cov}(r_t, r_{t-1}) \\ &= 2\text{Var}(r_t) + 2\gamma_1 \\ &= 2\text{Var}(r_t)(1 + \rho_1), \end{aligned}$$

wobei die Stationarität von r_t ausgenutzt wurde, und allgemein

$$\text{Var}\{r_t(q)\} = q\text{Var}(r_t) \left(1 + 2 \sum_{\tau=1}^{q-1} \left(1 - \frac{\tau}{q}\right) \rho_\tau \right). \quad (10.55)$$

Unter H_0 gilt $\rho_\tau = 0$ für alle $\tau > 0$, so dass unter H_0

$$\frac{\text{Var}\{r_t(q)\}}{q\text{Var}(r_t)} = 1.$$

Eine Teststatistik lässt sich nun konstruieren, indem man die erwartungstreuen Schätzer

$$\hat{\mu} = \frac{1}{n}(\ln S_n - \ln S_0)$$

für μ ,

$$\hat{\gamma}_0 = \frac{1}{n-1} \sum_{t=2}^n (\ln S_t - \ln S_{t-1} - \hat{\mu})^2$$

für $\text{Var}(r_t)$ und

$$\hat{\gamma}_0(q) = \frac{n}{q(n-q)(n-q+1)} \sum_{t=q+1}^n (\ln S_t - \ln S_{t-q} - q\hat{\mu})^2$$

für $\text{Var}\{r_t(q)\}$ in (10.55) einsetzt. Die Teststatistik lautet also

$$VQ(q) = \frac{\hat{\gamma}_0(q)}{\hat{\gamma}_0} - 1.$$

Wie gezeigt werden kann, ist die asymptotische Verteilung

$$\sqrt{n}VQ(q) \xrightarrow{\mathcal{L}} N\left(0, \frac{2(2q-1)(q-1)}{3q}\right).$$

Die asymptotische Varianz lässt sich durch die folgende Approximation begründen: Nehmen wir an, dass $\hat{\mu} = 0$ und $n \gg q$. Dann ist $\ln S_t - \ln S_{t-q} = \sum_{j=0}^{q-1} r_{t-j}$ und

$$\begin{aligned} VQ(q) &\approx \frac{1}{qn} \sum_{t=q+1}^n \left(\left(\sum_{j=0}^{q-1} r_{t-j} \right)^2 - q\hat{\gamma}_0 \right) / \hat{\gamma}_0 \\ &= \frac{1}{qn} \sum_{t=q+1}^n \left(\sum_{j=0}^{q-1} r_{t-j}^2 + 2 \sum_{j=0}^{q-2} r_{t-j} r_{t-j-1} + \dots + 2r_t r_{t-q+1} - q\hat{\gamma}_0 \right) / \hat{\gamma}_0 \\ &\approx \frac{1}{q} (q\hat{\gamma}_0 + 2(q-1)\hat{\gamma}_1 + \dots + 2\hat{\gamma}_{q-1} - q\hat{\gamma}_0) / \hat{\gamma}_0 \\ &= 2 \sum_{j=1}^{q-1} \frac{q-j}{q} \hat{\rho}_j. \end{aligned}$$

Da unter H_0 die geschätzten Autokorrelationen $\hat{\rho}_j$ skaliert mit \sqrt{n} asymptotisch standardnormalverteilt und unabhängig sind, siehe Abschnitt 11.5, ergibt sich folgende asymptotische Varianz:

$$\begin{aligned} \text{Var}_{as}\{\sqrt{n}VQ(q)\} &= \text{Var}_{as}\left(2\sum_{j=1}^{q-1}\frac{q-j}{q}\sqrt{n}\hat{\rho}_j\right) \\ &= 4\sum_{j=1}^{q-1}\frac{(q-j)^2}{q^2}\text{Var}_{as}(\sqrt{n}\hat{\rho}_j) \\ &= 4\sum_{j=1}^{q-1}\frac{(q-j)^2}{q^2} \\ &= 4(q-1) - \frac{8}{q}\sum_{j=1}^{q-1}j + \frac{4}{q^2}\sum_{j=1}^{q-1}j^2. \end{aligned}$$

Mit den Summenformeln $\sum_{j=1}^{q-1}j = (q-1)q/2$ und $\sum_{j=1}^{q-1}j^2 = q(q-1)(2q-1)/6$ erhält man schließlich

$$\text{Var}_{as}\{\sqrt{n}VQ(q)\} = \frac{2(2q-1)(q-1)}{3q}.$$

10.7. Literaturhinweise

Drei aktuelle Lehrbücher im Bereich der empirischen Finanzmarktanalyse sind [Mills \(1993\)](#), [Gouriéroux \(1997\)](#) und [Campbell, Lo and MacKinlay \(1997\)](#). Die Schwerpunkte bei Gouriéroux und Mills sind eher auf dem ökonometrisch/zeitreihenanalytischen Ansatz (dem wir in diesem Buch auch folgen werden), während Campbell, Lo und MacKinlay viele ökonomische Ansätze verfolgen, die nicht immer in statistischen oder ökonometrischen Modellen enden. Als einführendes und doch umfassendes Buch über Zeitreihenanalyse sei [Schlittgen and Streitberg \(1995\)](#) empfohlen. Desgleichen hat man mit [Copeland and Weston \(1992\)](#) ein einfaches und gutes Buch über Finanztheorie.

Das Experiment der Erwartungshypothesen stammt von [Forsythe, Palfrey and Plott \(1982\)](#). Die Definition von rationalen Erwartungen und Markteffizienz basiert auf [Jarrow \(1992\)](#). Das CAPM wurde entwickelt von [Sharpe \(1964\)](#), [Lintner \(1965\)](#) und [Mossin \(1966\)](#). Die Darstellung der Zinsparitätentheorie folgt [Jarchow and Rühmann \(1994\)](#)[pp.236] und die des Zinsstrukturmodells nach Cox-Ingorsoll-Ross folgt [Ingorsoll \(1987\)](#). Das Standard Optionspreismodell stammt von [Black and Scholes \(1973\)](#).

Einen guten Überblick über Einheitswurzeltests gibt [Hassler \(1994\)](#). Der ADF Test stammt von [Dickey and Fuller \(1979\)](#) und der KPSS Test von [Kwiatkowski, Phillips, Schmidt and Shin \(1992\)](#).

11. ARIMA Zeitreihenmodelle

In diesem Kapitel befassen wir uns mit der klassischen, linearen Zeitreihenanalyse. Zunächst definieren wir den allgemeinen linearen Prozess.

DEFINITION 11.1 (Linearer Prozess)

Besitzt der Prozess X_t die Darstellung

$$X_t = \mu + \sum_{i=-\infty}^{\infty} a_i \varepsilon_{t-i}$$

mit weißem Rauschen ε_t und absolut summierbarem Filter $(a_i) : \sum_{i=-\infty}^{\infty} |a_i| < \infty$, so ist er ein linearer Prozess.

Der lineare Prozess X_t ist kovarianzstationär, da $E(X_t) = \mu$ und

$$\text{Cov}(X_t, X_{t+\tau}) = \sigma^2 \sum_{i=-\infty}^{\infty} \sum_{j=-\infty}^{\infty} a_i a_j I(\tau = i - j) = \sigma^2 \sum_{i=-\infty}^{\infty} a_i a_{i-\tau}$$

mit $\text{Var}(\varepsilon_t) = \sigma^2$ und Indikatorfunktion $I(\cdot)$.

Allgemein werden in der Ökonomie und speziell im Finanzmarktbereich Reihen beobachtet, die ein nichtstationäres Verhalten aufweisen. Im vorigen Kapitel haben wir gesehen, dass ökonomische Modelle, die auf der Annahme rationaler Erwartungen beruhen, häufig implizieren, dass die relevanten Größen wie z.B. Preise einem Random Walk folgen. Um diese Prozesse im Rahmen der klassischen Zeitreihenanalyse behandeln zu können, müssen also in diesem Fall erste Differenzen gebildet werden, um einen stationären Prozess zu erhalten. Wir verallgemeinern die Definition eines differenzenstationären Prozesses in der folgenden Definition.

DEFINITION 11.2 (Integrierter Prozess)

Wir sagen, der Prozess X_t ist integriert der Ordnung d , $I(d)$, falls $(1 - L)^{d-1} X_t$ nichtstationär und $(1 - L)^d X_t$ stationär ist.

Ein weißes Rauschen ist z.B. $I(0)$, ein Random Walk $I(1)$. In den seltensten Fällen werden Prozesse beobachtet, die $I(d)$ sind mit $d > 1$, d.h., in der Regel genügt die Bildung erster Differenzen, um einen stationären Prozess zu erhalten. Im folgenden gehen wir davon aus, dass der beobachtete Prozess Y_t $I(d)$ ist und wir betrachten den transformierten Prozess $X_t = (1 - L)^d Y_t$, d.h., wir konzentrieren uns auf stationäre Prozesse.

11.1. Moving Average Prozesse

Der Moving Average Prozess der Ordnung q , $MA(q)$, ist definiert als

$$X_t = \beta_1 \varepsilon_{t-1} + \dots + \beta_q \varepsilon_{t-q} + \varepsilon_t \quad (11.1)$$

mit weißem Rauschen ε_t . Mit dem Lag-Operator L (siehe Definition 10.13) können wir statt (11.1) auch schreiben

$$X_t = \beta(L)\varepsilon_t \quad (11.2)$$

mit $\beta(L) = 1 + \beta_1 L + \dots + \beta_q L^q$. Der $MA(q)$ Prozess ist stationär, da er durch Linearkombination eines stationären Prozesses entsteht. Die Mittelwertfunktion ist einfach $E(X_t) = 0$. Setzen wir $\beta_0 = 1$, so ist die Kovarianzfunktion

$$\begin{aligned} \gamma_\tau &= \text{Cov}(X_t, X_{t+\tau}) \\ &= \text{Cov}\left(\sum_{i=0}^q \beta_i \varepsilon_{t-i}, \sum_{i=0}^q \beta_i \varepsilon_{t+\tau-i}\right) \\ &= \sum_{i=0}^q \sum_{j=0}^q \beta_i \beta_j \text{Cov}(\varepsilon_{t-i}, \varepsilon_{t+\tau-i}) \\ &= \sum_{i=0}^{q-|\tau|} \beta_i \beta_{i+|\tau|} \sigma^2, \quad |\tau| \leq q. \end{aligned}$$

Für die ACF ergibt sich also für $|\tau| \leq q$

$$\rho_\tau = \frac{\sum_{i=0}^{q-|\tau|} \beta_i \beta_{i+|\tau|}}{\sum_{i=0}^q \beta_i^2}, \quad (11.3)$$

und $\rho_\tau = 0$ für $|\tau| > q$, d.h., die ACF bricht nach q Lags ab.

Als Beispiel betrachten wir den $MA(1)$ Prozess

$$X_t = \beta \varepsilon_{t-1} + \varepsilon_t,$$

für den nach (11.3) gilt $\rho_1 = \beta/(1 + \beta^2)$ und $\rho_\tau = 0$ für $\tau > 1$. Abbildung 11.1 zeigt das Korrelogramm eines $MA(1)$ Prozesses.

Offenbar hat der Prozess

$$X_t = 1/\beta \varepsilon_{t-1} + \varepsilon_t$$

dieselbe ACF, da gilt

$$\rho_1 = \frac{1/\beta}{1 + (1/\beta)^2} = \frac{\beta}{1 + \beta^2}.$$

Mit anderen Worten, der Prozess mit dem Parameter β hat dieselben stochastischen Eigenschaften wie der Prozess mit dem Parameter $1/\beta$. Diesem Identifizierbarkeitsproblem kann man begegnen, indem man verlangt, dass die Lösungen der charakteristischen Gleichung

$$1 + \beta_1 z + \dots + \beta_q z^q = 0 \quad (11.4)$$

außerhalb des komplexen Einheitskreises liegen. In diesem Fall ist der lineare Filter $\beta(L)$ invertierbar, d.h., es existiert ein Polynom $\beta^{-1}(L)$ so dass $\beta(L)\beta^{-1}(L) = 1$ und

$\beta^{-1}(L) = b_0 + b_1 L + b_2 L^2 + \dots$. Abbildung 11.2 zeigt das Korrelogramm des $MA(2)$ Prozesses $X_t = \beta_1 \varepsilon_{t-1} + \beta_2 \varepsilon_{t-2} + \varepsilon_t$ für einige Parameterkonstellationen.

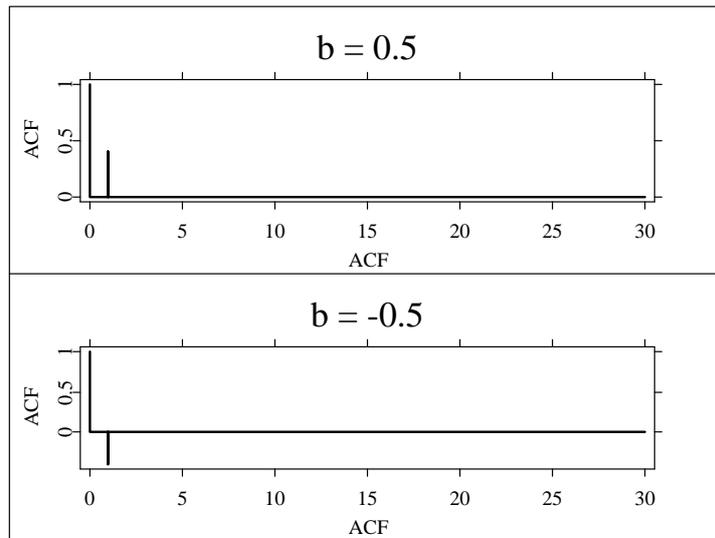


Abbildung 11.1.: ACF eines MA(1) Prozesses mit $\beta = 0.5$ (oben) und $\beta = -0.5$ (unten).

[SFMacfma1.xpl](#)

11.2. Autoregressive Prozesse

Der lineare autoregressive Prozess der Ordnung p ($AR(p)$), wird definiert als

$$X_t = \nu + \alpha_1 X_{t-1} + \dots + \alpha_p X_{t-p} + \varepsilon_t \quad (11.5)$$

Nutzen wir die Definition des Lag-Operators L (siehe Definition 10.13), so können wir für (11.5) auch schreiben

$$\alpha(L)X_t = \nu + \varepsilon_t$$

mit dem Lag-Polynom $\alpha(L) = 1 - \alpha_1 L - \dots - \alpha_p L^p$. Der Prozess X_t ist stationär, wenn alle Wurzeln der charakteristischen Gleichung

$$\alpha(z) = 1 - \alpha_1 z - \dots - \alpha_p z^p = 0. \quad (11.6)$$

außerhalb des komplexen Einheitskreises liegen, d.h., wenn für alle z mit $|z| \leq 1$ gilt

$$\alpha(z) \neq 0. \quad (11.7)$$

In diesem Fall gibt es zu dem linearen Filter $\alpha(L)$ einen inversen Filter $\alpha^{-1}(L)$, so dass gilt $\alpha(L)\alpha^{-1}(L) = 1$ und

$$\alpha^{-1}(L) = a_0 + a_1 L + a_2 L^2 + \dots = \sum_{i=0}^{\infty} a_i L^i.$$

Der Prozess (11.5) lässt sich unter der Bedingung (11.7) auch schreiben als

$$\begin{aligned} X_t &= \alpha^{-1}(1)\nu + \alpha^{-1}(L)\varepsilon_t \\ &= \sum_{i=0}^{\infty} a_i \nu + \sum_{i=0}^{\infty} a_i L^i \varepsilon_t, \end{aligned}$$

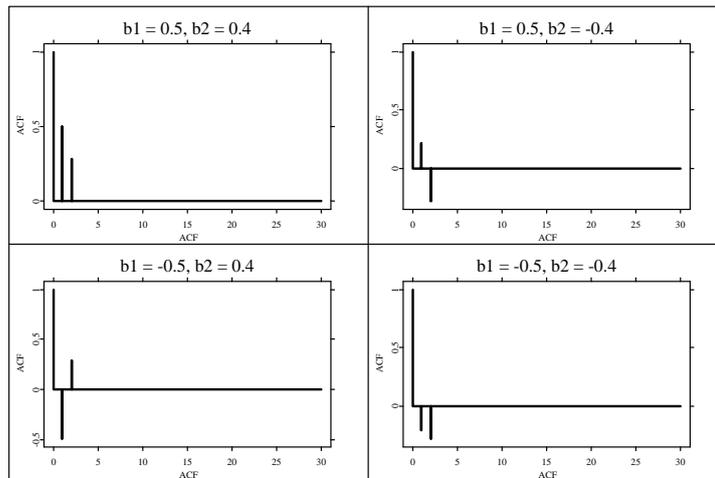


Abbildung 11.2.: ACF eines MA(2) Prozesses mit $(\beta_1 = 0.5, \beta_2 = 0.4)$ (links oben), $(\beta_1 = 0.5, \beta_2 = -0.4)$ (rechts oben), $(\beta_1 = -0.5, \beta_2 = 0.4)$ (links unten) und $(\beta_1 = -0.5, \beta_2 = -0.4)$ (rechts unten). [SFMacfma2.xpl](#)

also als MA(∞) Prozess.

Eine einfache Möglichkeit, für gegebene Parameter die Autokovarianzen eines AR(p) Prozesses zu erhalten und umgekehrt, sind die Yule-Walker Gleichungen. Sie ergeben sich aus der Definition des AR(p) Prozesses in (11.5) durch Multiplikation mit $X_{t-\tau}$ und Erwartungswertbildung.

$$E[X_t X_{t-\tau}] = \alpha_1 E[X_{t-1} X_{t-\tau}] + \dots + \alpha_p E[X_{t-p} X_{t-\tau}]. \quad (11.8)$$

Da für $\nu = 0$, $E[X_t X_{t-\tau}]$ gerade der Definition der Autokovarianzfunktion γ_τ entspricht, kann man auch einfacher schreiben für $\tau = 1, \dots, p$

$$\begin{aligned} \alpha_1 \gamma_0 + \alpha_2 \gamma_1 + \dots + \alpha_p \gamma_{p-1} &= \gamma_1 \\ \alpha_1 \gamma_1 + \alpha_2 \gamma_0 + \dots + \alpha_p \gamma_{p-2} &= \gamma_2 \\ \dots & \\ \alpha_1 \gamma_{p-1} + \alpha_2 \gamma_{p-2} + \dots + \alpha_p \gamma_0 &= \gamma_p \end{aligned} \quad (11.9)$$

oder durch Division mit der Varianz γ_0

$$\rho = R\alpha \quad (11.10)$$

mit $\rho = (\rho_1 \rho_2 \dots \rho_p)^T$, $\alpha = (\alpha_1 \alpha_2 \dots \alpha_p)^T$, und der $p \times p$ -Autokovarianzmatrix

$$R = \begin{pmatrix} 1 & \rho_1 & \dots & \rho_{p-1} \\ \rho_1 & 1 & \dots & \rho_{p-2} \\ \vdots & \vdots & \ddots & \vdots \\ \rho_{p-1} & \rho_{p-2} & \dots & 1 \end{pmatrix}.$$

Die Yule-Walker Gleichungen sind nützlich bei der Bestimmung der ACF für gegebene Parameter oder umgekehrt für die Schätzung der Parameter bei gegebenen (empirischen) Autokorrelationen.

BEISPIEL 11.1 (AR(1))

Der AR(1) Prozess aus Beispiel 10.1 mit $\nu = 0$ hat die charakteristische Gleichung $1 - \alpha z = 0$. Die eindeutige Lösung ist $z = 1/\alpha$ und $|z| > 1$ genau dann, wenn $|\alpha| < 1$. Der inverse Filter zu $\alpha(L) = 1 - \alpha L$ ist dann $\alpha^{-1}(L) = \sum_{i=0}^{\infty} \alpha^i L^i$ und die MA(∞) Darstellung des AR(1) Prozesses ist

$$X_t = \sum_{i=0}^{\infty} \alpha^i \varepsilon_{t-i}.$$

Die ACF des AR(1) Prozesses ist $\rho_\tau = \alpha^\tau$. Für $\alpha > 0$ sind alle Autokorrelationen positiv, für $\alpha < 0$ alternieren die Vorzeichen, siehe Abbildung 11.3.

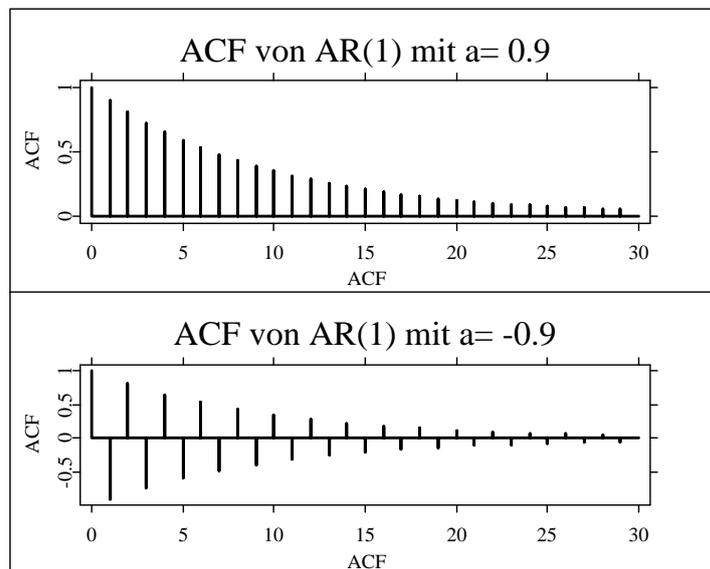


Abbildung 11.3.: ACF eines AR(1) Prozesses mit $\alpha = 0.9$ (oben) und $\alpha = -0.9$ (unten).

[SFMacfar1.xpl](#)

BEISPIEL 11.2 (AR(2))

Der AR(2) Prozess mit $\nu = 0$,

$$X_t = \alpha_1 X_{t-1} + \alpha_2 X_{t-2} + \varepsilon_t$$

ist stationär, wenn für die Wurzeln z_1 und z_2 der quadratischen Gleichung

$$1 - \alpha_1 z - \alpha_2 z^2 = 0$$

$|z_1| > 1$ und $|z_2| > 1$ gilt. Wir bekommen als Lösungen

$$z_{1,2} = -\frac{\alpha_1}{2\alpha_2} \pm \sqrt{\frac{\alpha_1^2}{4\alpha_2^2} + \frac{1}{\alpha_2}}$$

TO DO: und $z_1 z_2 = -1/\alpha_2$. Wegen $|z_1| > 1$ und $|z_2| > 1$ gilt damit $|z_1 z_2| = 1/|\alpha_2| > 1$ und

Ist der Rest des Kap. in den Folien?

$$|\alpha_2| < 1. \tag{11.11}$$

Aus den Yule-Walker Gleichungen für den AR(2) Fall

$$\rho_1 = \alpha_1 + \alpha_2 \rho_1 \tag{11.12}$$

$$\rho_2 = \alpha_1 \rho_1 + \alpha_2 \tag{11.13}$$

$$\tag{11.14}$$

folgt außerdem, dass $\rho_1 = \alpha_1/(1 - \alpha_2)$. Der Fall $\rho_1 = \pm 1$ ist ausgeschlossen, da dann eine Wurzel auf dem Einheitskreis (bei 1 oder -1) liegen würde. Also muss für einen stationären AR(2) Prozess gelten

$$|\rho_1| = |\alpha_1/(1 - \alpha_2)| < 1,$$

woraus wir zusammen mit (11.11) das ‘Stationaritätsdreieck’

$$\alpha_1 + \alpha_2 < 1 \tag{11.15}$$

$$\alpha_2 - \alpha_1 < 1 \tag{11.16}$$

erhalten, d.h. die Region, für die der AR(2) Prozess stationär ist.

Die ACF des AR(2) Prozesses ist rekursiv gegeben mit (11.12), (11.13) und $\rho_\tau = \alpha_1 \rho_{\tau-1} + \alpha_2 \rho_{\tau-2}$ für $\tau > 2$. Abbildung (11.4) stellt einige typische Verläufe dar.

11.3. ARMA Modelle

Das ARMA(p, q) Modell wird definiert als

$$X_t = \nu + \alpha_1 X_{t-1} + \dots + \alpha_p X_{t-p} + \beta_1 \varepsilon_{t-1} + \dots + \beta_q \varepsilon_{t-q} + \varepsilon_t, \tag{11.17}$$

oder als

$$\alpha(L)X_t = \nu + \beta(L)\varepsilon_t$$

mit dem Moving-Average Lag-Polynom $\beta(L) = 1 + \beta_1 L + \dots + \beta_q L^q$ und dem autoregressiven Lag-Polynom $\alpha(L) = 1 - \alpha_1 L - \dots - \alpha_p L^p$. Damit der Prozess (11.17) eine eindeutige Parametrisierung hat, wird gefordert, dass die charakteristischen Polynome $\alpha(z)$ und $\beta(z)$ keine gemeinsamen Nullstellen haben. Der Prozess (11.17) ist stationär, wenn alle Wurzeln der charakteristischen Gleichung (11.6) außerhalb des Einheitskreises liegen. In diesem Fall besitzt (11.17) die MA(∞) Darstellung

$$X_t = \alpha^{-1}(L)\beta(L)\varepsilon_t.$$

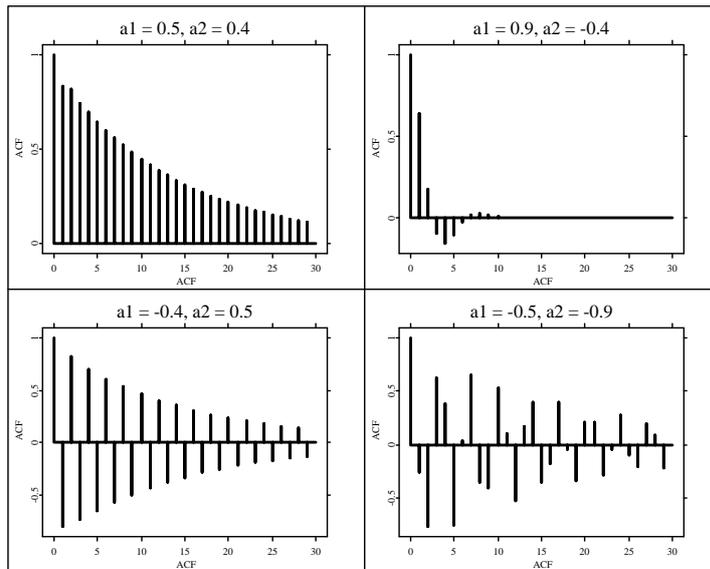


Abbildung 11.4.: ACF eines AR(2) Prozesses mit $(\alpha_1 = 0.5, \alpha_2 = 0.4)$ (links oben), $(\alpha_1 = 0.9, \alpha_2 = -0.4)$ (rechts oben), $(\alpha_1 = -0.4, \alpha_2 = 0.5)$ (links unten) und $(\alpha_1 = -0.5, \alpha_2 = -0.9)$ (rechts unten). [SFMacfar2.xpl](#)

Der Prozess X_t in (11.17) ist invertierbar, wenn alle Wurzeln der charakteristischen Gleichung (11.4) außerhalb des Einheitskreises liegen. In diesem Fall können wir (11.17) schreiben als

$$\beta^{-1}(L)\alpha(L)X_t = \varepsilon_t,$$

also als AR(∞) Prozess. Wir können somit jeden stationären, invertierbaren ARMA(p, q) Prozess durch einen reinen AR bzw. MA Prozess hinreichend großer Ordnung approximieren. Umgekehrt bietet der ARMA(p, q) Prozess eine Möglichkeit sparsamer Parametrisierung.

11.4. Partielle Autokorrelationen

Für einen gegebenen stochastischen Prozess interessiert man sich oft für den Zusammenhang zwischen zwei Zufallsvariablen des Prozesses zu verschiedenen Zeitpunkten. Eine Möglichkeit, den linearen Zusammenhang zu messen, ist die ACF, d.h., die Korrelation zwischen diesen Zufallsvariablen. Eine andere Möglichkeit, den Zusammenhang zwischen X_t und $X_{t+\tau}$ zu messen, ist zunächst X_t und $X_{t+\tau}$ um den linearen Einfluss der dazwischen liegenden Zufallsvariablen $X_{t+1}, \dots, X_{t+\tau-1}$ zu bereinigen und erst dann die Korrelation der so transformierten Zufallsvariablen zu berechnen. Man nennt dies die *partielle* Autokorrelation.

DEFINITION 11.3 (Partielle Autokorrelation)

Die Partielle Autokorrelation k -ter Ordnung ist definiert als

$$\phi_{kk} = \text{Corr}(X_t - \mathcal{P}(X_t | X_{t+1}, \dots, X_{t+k-1}), X_{t+k} - \mathcal{P}(X_{t+k} | X_{t+1}, \dots, X_{t+k-1})) \quad (11.18)$$

wobei $\mathcal{P}(W | Z)$ die beste lineare Projektion von W auf Z sei, d.h. $\mathcal{P}(W | Z) = \Sigma_{WZ} \Sigma_{ZZ}^{-1} Z$ mit $\Sigma_{ZZ} = \text{Var}(Z)$ der Kovarianzmatrix der Regressoren und $\Sigma_{WZ} = \text{Cov}(W, Z)$ der Matrix der Kovarianzen zwischen W und Z .

TO DO:

beste lin. Proj. erklären (Schlittgen S. 517) Die 'beste lineare Projektion' wird also im Sinne des mittleren quadratischen Fehlers verstanden.

Eine äquivalente Definition ist die Lösung des Gleichungssystems

$$P_k \phi_k = \rho_{(k)}$$

mit

$$P_k = \begin{pmatrix} 1 & \rho_1 & \cdots & \rho_{k-1} \\ \rho_1 & 1 & \cdots & \rho_{k-2} \\ \vdots & \vdots & \ddots & \vdots \\ \rho_{k-1} & \rho_{k-2} & \cdots & 1 \end{pmatrix}$$

und $\phi_k = (\phi_{k1}, \dots, \phi_{kk})^T$ und $\rho_{(k)} = (\rho_1, \dots, \rho_k)^T$. Dies sind gerade die Yule-Walker Gleichungen eines $\text{AR}(k)$ Prozesses. Der letzte Koeffizient, ϕ_{kk} ist gerade die partielle Autokorrelation der Ordnung k wie oben definiert. Da nur dieser Koeffizient in diesem Zusammenhang von Interesse ist, kann man das Gleichungssystem nach ϕ_{kk} mit der Cramerschen Regel lösen. Man erhält

$$\phi_{kk} = \frac{|P_k^*|}{|P_k|}$$

wobei P_k^* gleich der Matrix P_k ist, in der die k -te Spalte durch $\rho_{(k)}$ ersetzt wird. $|\cdot|$ bezeichnet hier die Determinante. Da man dies für verschiedene Ordnungen k durchführen kann, erhält man letztendlich eine *Partielle Autokorrelationsfunktion* (PACF). Die PACF lässt sich also für einen gegebenen stochastischen Prozess wie die ACF als Funktion der Ordnung k graphisch darstellen. Wir nennen dies das partielle Autokorrelogramm.

Aus der Definition der PACF folgt sofort, dass zwischen PACF und ACF der Ordnung Eins kein Unterschied besteht:

$$\phi_{11} = \rho_1.$$

Für die zweite Ordnung bekommen wir

$$\phi_{22} = \frac{\begin{vmatrix} 1 & \rho_1 \\ \rho_1 & \rho_2 \end{vmatrix}}{\begin{vmatrix} 1 & \rho_1 \\ \rho_1 & 1 \end{vmatrix}} = \frac{\rho_2 - \rho_1^2}{1 - \rho_1^2} \quad (11.19)$$

BEISPIEL 11.3 (AR(1))

Der $\text{AR}(1)$ Prozess $X_t = \alpha X_{t-1} + \varepsilon_t$ hat die ACF $\rho_\tau = \alpha^\tau$. Für die PACF ergibt sich $\phi_{11} = \rho_1 = \alpha$ und

$$\phi_{22} = \frac{\rho_2 - \rho_1^2}{1 - \rho_1^2} = \frac{\alpha^2 - \alpha^2}{1 - \alpha^2} = 0$$

und, $\phi_{kk} = 0$ für alle $k > 1$. Dies ist plausibel, da der letzte Koeffizient eines $\text{AR}(k)$ Modells für diesen Prozess gerade Null ist für alle $k > 1$. Speziell für $k = 2$ illustrieren wir die

Übereinstimmung mit Definition 11.3: Aus $X_t = \alpha X_{t-1} + \varepsilon_t$ ergibt sich direkt $\mathcal{P}(X_{t+2}|X_{t+1}) = \alpha X_{t+1}$ mit

$$\alpha = \frac{\text{Cov}(X_{t+2}, X_{t+1})}{\text{Var}(X_{t+1})}.$$

Aus der 'Rückwärtsregression' $X_t = \alpha' X_{t+1} + \eta_t$ mit weißem Rauschen η_t folgt ferner $\mathcal{P}(X_t|X_{t+1}) = \alpha' X_{t+1}$ mit

$$\alpha' = \frac{\text{Cov}(X_t, X_{t+1})}{\text{Var}(X_{t+1})}.$$

Für $|\alpha| < 1$ ist der Prozess kovarianzstationär und es gilt $\text{Cov}(X_{t+2}, X_{t+1}) = \text{Cov}(X_t, X_{t+1}) = \gamma_1$ und $\alpha = \alpha' = \rho_1$. Wir erhalten

$$\begin{aligned} \text{Cov}(X_t - \mathcal{P}(X_t|X_{t+1}), X_{t+2} - \mathcal{P}(X_{t+2}|X_{t+1})) &= \text{Cov}(X_t - \rho_1 X_{t+1}, X_{t+2} - \rho_1 X_{t+1}) \\ &= E[(X_t - \rho_1 X_{t+1})(X_{t+2} - \rho_1 X_{t+1})] \\ &= \gamma_2 - 2\rho_1\gamma_1 + \rho_1^2\gamma_0 \end{aligned}$$

und

$$\begin{aligned} \text{Var}(X_{t+2} - \mathcal{P}(X_{t+2}|X_{t+1})) &= E[(X_{t+2} - \rho_1 X_{t+1})^2] \\ &= \gamma_0(1 + \rho_1^2) - 2\rho_1\gamma_1 \\ &= E[(X_t - \rho_1 X_{t+1})^2] \\ &= \text{Var}(X_t - \mathcal{P}(X_t|X_{t+1})). \end{aligned}$$

Somit bekommen wir für die partielle Autokorrelation zweiter Ordnung

$$\begin{aligned} \phi_{22} &= \frac{\text{Corr}(X_t - \mathcal{P}(X_t|X_{t+1}), X_{t+2} - \mathcal{P}(X_{t+2}|X_{t+1}))}{\sqrt{\text{Var}(X_{t+2} - \mathcal{P}(X_{t+2}|X_{t+1}))} \sqrt{\text{Var}(X_t - \mathcal{P}(X_t|X_{t+1}))}} \\ &= \frac{\text{Cov}(X_t - \mathcal{P}(X_t|X_{t+1}), X_{t+2} - \mathcal{P}(X_{t+2}|X_{t+1}))}{\sqrt{\text{Var}(X_{t+2} - \mathcal{P}(X_{t+2}|X_{t+1}))} \sqrt{\text{Var}(X_t - \mathcal{P}(X_t|X_{t+1}))}} \\ &= \frac{\gamma_2 - 2\rho_1\gamma_1 + \rho_1^2\gamma_0}{\gamma_0(1 + \rho_1^2) - 2\gamma_1\rho_1} \\ &= \frac{\rho_2 - \rho_1^2}{1 - \rho_1^2} \end{aligned}$$

welches gerade unserem Ergebnis in (11.19) entspricht. Für den AR(1) Prozess gilt $\rho_2 = \rho_1^2$ und daher $\phi_{22} = 0$.

Es gilt allgemein für AR(p) Prozesse, dass $\phi_{kk} = 0$ für alle $k > p$. In Abbildung 11.5 ist die PACF eines AR(2) Prozesses mit Parametern wie in Abbildung 11.4 abgebildet.

BEISPIEL 11.4 (MA(1))

Für den MA(1) Prozess $X_t = \beta\varepsilon_{t-1} + \varepsilon_t$ mit $\text{Var}(\varepsilon_t) = \sigma^2$ gilt $\gamma_0 = \sigma^2(1 + \beta^2)$, $\rho_1 = \beta/(1 + \beta^2)$ und $\rho_k = 0$ für alle $k > 1$. Für die partiellen Autokorrelationen erhalten wir $\phi_{11} = \rho_1$ und

$$\phi_{22} = \frac{\begin{vmatrix} 1 & \rho_1 \\ \rho_1 & 0 \end{vmatrix}}{\begin{vmatrix} 1 & \rho_1 \\ \rho_1 & 1 \end{vmatrix}} = -\frac{\rho_1^2}{1 - \rho_1^2} \quad (11.20)$$

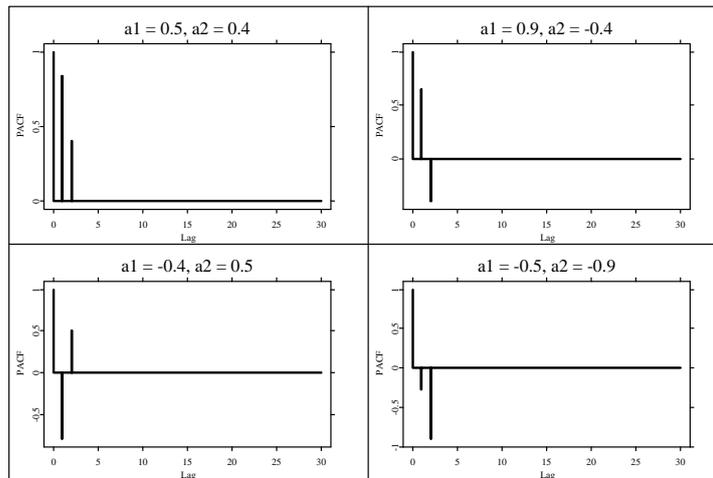


Abbildung 11.5.: PACF eines AR(2) Prozesses mit $(\alpha_1 = 0.5, \alpha_2 = 0.4)$ (links oben), $(\alpha_1 = 0.9, \alpha_2 = -0.4)$ (rechts oben), $(\alpha_1 = -0.4, \alpha_2 = 0.5)$ (links unten) und $(\alpha_1 = -0.5, \alpha_2 = -0.9)$ (rechts unten). [SFMpacfar2.xpl](#)

Für den MA(1) Prozess gilt also stets $\phi_{22} < 0$. Setzt man die Berechnungen mit $k > 2$ fort, so stellt man fest, dass die partiellen Autokorrelationen nicht Null werden.

Allgemein gilt für MA(q) Prozesse, dass die PACF nicht abbricht, im Gegensatz zu autoregressiven Prozessen. Dies ist also ein mögliches Kriterium bei der Spezifikation linearer Modelle. Abbildung 11.6 zeigt die PACF eines MA(2) Prozesses. Man vergleiche die PACF mit der ACF in Abbildung 11.2.

11.5. Schätzung der Momentfunktionen

Im folgenden gehen wir von einem stationären stochastischen Prozess X_t aus, d.h. $E[X_t] = \mu$ und $\text{Cov}(X_t, X_{t+\tau}) = \gamma_\tau$. Bislang sind wir davon ausgegangen, dass wir den Prozess kennen und somit die Momentfunktionen auch bekannt sind. In der Praxis beobachtet man jedoch nur eine Realisation des Prozesses, X_1, \dots, X_n , und damit stellt sich das Problem der Schätzung der Momentfunktionen.

11.5.1. Schätzung der Mittelwertfunktion

Der Parameter $\mu = E[X_t]$ kann durch das einfache arithmetische Stichprobenmittel geschätzt werden:

$$\bar{X}_n = 1/n \sum_{i=1}^n X_i. \quad (11.21)$$

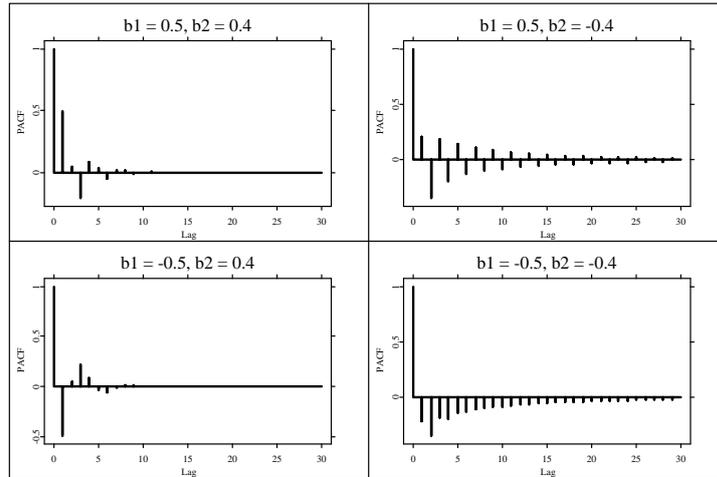


Abbildung 11.6.: PACF eines MA(2) Prozesses mit $(\beta_1 = 0.5, \beta_2 = 0.4)$ (links oben), $(\beta_1 = 0.5, \beta_2 = -0.4)$ (rechts oben), $(\beta_1 = -0.5, \beta_2 = 0.4)$ (links unten) und $(\beta_1 = -0.5, \beta_2 = -0.4)$ (rechts unten). [SFMpacfma2.xpl](#)

Der Schätzer \bar{X}_n ist erwartungstreu, da gilt $E[\bar{X}_n] = \mu$, und seine Varianz ist

$$\begin{aligned}
 \text{Var}(\bar{X}_n) &= \text{Var}\left(\frac{1}{n} \sum_{i=1}^n X_t\right) \\
 &= \frac{1}{n^2} \sum_{t=1}^n \sum_{s=1}^n \text{Cov}(X_t, X_s) \\
 &= \frac{1}{n^2} \sum_{t=1}^n \sum_{s=1}^n \gamma_{t-s} \\
 &= \frac{1}{n} \sum_{\tau=-(n-1)}^{n-1} \frac{n - |\tau|}{n} \gamma_\tau
 \end{aligned}$$

Falls die Autokovarianzfunktion γ_τ absolut summierbar ist, gilt $\text{Var}(\bar{X}_n) < \infty$ und $\lim_{n \rightarrow \infty} \text{Var}(\bar{X}_n) = 0$. Der Schätzer \bar{X} ist dann also ein konsistenter Schätzer für μ . In vielen Fällen gibt es jedoch effizientere Schätzer, die die Korrelationsstruktur des Prozesses ausnutzen.

Die asymptotische Varianz

$$\lim_{n \rightarrow \infty} n \text{Var}(\bar{X}_n) = \gamma_0 + 2 \sum_{\tau=1}^{\infty} \gamma_\tau$$

bezeichnen wir mit $f(0)$, da sie gleich der Spektraldichte an der Frequenz 0 ist. Unter der absoluten Summierbarkeit von γ_τ gilt folgende asymptotische Verteilung des Schätzers:

$$\sqrt{n}(\bar{X}_n - \mu) \xrightarrow{\mathcal{L}} N(0, f(0)). \quad (11.22)$$

11.5.2. Schätzung der Kovarianzfunktion

Ein möglicher Schätzer der Kovarianzfunktion γ_τ ist

$$\hat{\gamma}_{\tau,n} = \frac{1}{n} \sum_{t=1}^{n-\tau} (X_t - \bar{X}_n)(X_{t+\tau} - \bar{X}_n) \quad (11.23)$$

mit dem Mittelwertschätzer \bar{X}_n aus (11.21). Anstelle von n könnte man in (11.23) auch durch $n - \tau$ teilen, jedoch hätte der Schätzer ungünstigere Eigenschaften. Der Schätzer $\hat{\gamma}_{\tau,n}$ ist nicht mehr erwartungstreu, da man folgendes zeigen kann.

$$E[\hat{\gamma}_{\tau,n}] = \left(1 - \frac{\tau}{n}\right) \gamma_\tau - \left(1 - \frac{\tau}{n}\right) \text{Var}(\bar{X}_n) + O(n^{-2}).$$

Positive Autokovarianzen werden also in der Regel durch $\hat{\gamma}_{\tau,n}$ unterschätzt. Asymptotisch ist $\hat{\gamma}_{\tau,n}$ jedoch erwartungstreu: $\lim_{n \rightarrow \infty} E[\hat{\gamma}_{\tau,n}] = \gamma_\tau$. Für die Varianz gilt unter Vernachlässigung von Termen höherer Ordnung

$$\text{Var}(\hat{\gamma}_{\tau,n}) \approx \frac{1}{n} \sum_{j=-\infty}^{\infty} (\gamma_j^2 + \gamma_{j-\tau} \gamma_{j+\tau}) = \frac{1}{n} \text{Var}(\hat{\gamma}_{\tau,\infty})$$

und da $\lim_{n \rightarrow \infty} \text{Var}(\hat{\gamma}_{\tau,n}) = 0$ gilt, ist $\hat{\gamma}_{\tau,n}$ ein konsistenter Schätzer für γ_τ . Außerdem kann asymptotisch gezeigt werden, dass der Kovarianzschätzer sich wie eine normalverteilte Zufallsgröße verhält:

$$\sqrt{n}(\hat{\gamma}_{\tau,n} - \gamma_\tau) \xrightarrow{\mathcal{L}} N(0, \text{Var}(\hat{\gamma}_{\tau,\infty})).$$

11.5.3. Schätzung der ACF

Ein naheliegender Schätzer für die ACF ρ_τ ist

$$\hat{\rho}_{\tau,n} = \frac{\hat{\gamma}_{\tau,n}}{\hat{\gamma}_{0,n}}. \quad (11.24)$$

Wir haben wieder einen Bias der Ordnung $1/n$, d.h.,

$$E(\hat{\rho}_{\tau,n}) = \rho_\tau - O(n^{-1})$$

und $\hat{\rho}_{\tau,n}$ ist asymptotisch erwartungstreu. Für die Varianz gilt

$$\text{Var}(\hat{\rho}_{\tau,n}) = \frac{1}{n} \Sigma_{\rho,\tau\tau} + O(n^{-2}).$$

Der Schätzer $\hat{\rho}_{\tau,n}$ ist konsistent, da $\lim_{n \rightarrow \infty} \text{Var}(\hat{\rho}_{\tau,n}) = 0$. Für die asymptotische Verteilung des Vektors $\hat{\rho}_{(k),n} = (\hat{\rho}_{1,n}, \dots, \hat{\rho}_{k,n})^T$ kann gezeigt werden

$$\sqrt{n}(\hat{\rho}_{(k),n} - \rho_{(k)}) \xrightarrow{\mathcal{L}} N(0, \Sigma_\rho)$$

mit der Kovarianzmatrix Σ_ρ mit typischem Element

$$\begin{aligned}\Sigma_{\rho,kl} &= \sum_{j=-\infty}^{\infty} \rho_j \rho_{j+k+l} + \sum_{j=-\infty}^{\infty} \rho_j \rho_{j+k-l} \\ &+ 2\rho_k \rho_l \sum_{j=-\infty}^{\infty} \rho_j^2 - 2\rho_l \sum_{j=-\infty}^{\infty} \rho_j \rho_{j+k} - 2\rho_k \sum_{j=-\infty}^{\infty} \rho_j \rho_{j+l}.\end{aligned}$$

Speziell gilt für die asymptotische Varianz von $\sqrt{n}(\hat{\rho}_{\tau,n} - \rho_\tau)$,

$$\begin{aligned}\Sigma_{\rho,\tau\tau} &= \sum_{j=-\infty}^{\infty} \rho_j \rho_{j+2\tau} + \sum_{j=-\infty}^{\infty} \rho_j^2 \\ &+ 2\rho_\tau^2 \sum_{j=-\infty}^{\infty} \rho_j^2 - 4\rho_\tau \sum_{j=-\infty}^{\infty} \rho_j \rho_{j+\tau}.\end{aligned}$$

BEISPIEL 11.5 (MA(q))

Für den MA(q) Prozess in (11.1) wissen wir, dass $\rho_\tau = 0$ für alle $\tau > q$. Damit vereinfacht sich die asymptotische Varianz von $\sqrt{n}(\hat{\rho}_{\tau,n} - \rho_\tau)$ für $\tau > q$ zu

$$\Sigma_{\rho,\tau\tau} = 1 + 2 \sum_{i=1}^q \rho_i^2.$$

BEISPIEL 11.6 (weißes Rauschen)

Ist X_t weißes Rauschen, so gilt $E(\hat{\rho}_{\tau,n}) = -\frac{1}{n} + O(n^{-2})$ und $\text{Var}(\hat{\rho}_{\tau,n}) = \frac{1}{n} + O(n^{-2})$ für $\tau \neq 0$. Die asymptotische Kovarianzmatrix von $\sqrt{n}(\hat{\rho}_{(k),n} - \rho_{(k)})$ ist die Einheitsmatrix. Damit lassen sich approximative 95% Konfidenzintervalle für die ACF bilden: $[-\frac{1}{n} \pm \frac{2}{\sqrt{n}}]$.

11.6. Portmanteau Statistiken

Mit Hilfe der Kenntnis über die asymptotische Verteilung der Autokorrelationen können wir nun eine Statistik zur Überprüfung der Hypothese eines weißen Rauschens ableiten. Man kann entweder die Originalreihe X_t oder die Residuen eines ARMA(p, q) Prozesses überprüfen. Die Anzahl der geschätzten Parameter ist so im ersten Fall $k = 0$ und im zweiten Fall $k = p + q$.

Unter der Nullhypothese gilt für jedes m

$$\rho_1 = 0, \dots, \rho_m = 0.$$

Die Alternativhypothese ist entsprechend, dass mindestens ein ρ_i , $1 \leq i \leq m$ ungleich Null ist. Unter der Nullhypothese ist $\sqrt{n}\hat{\rho}_{\tau,n}$ asymptotisch standardnormalverteilt. Die Statistik

$$Q_m = n \sum_{j=1}^m \hat{\rho}_{j,n}^2$$

ist asymptotisch χ^2 verteilt ist mit $m - k$ Freiheitsgraden. Man lehnt die Nullhypothese zu einem Testniveau von α ab, sofern $Q_m > \chi_{m-k;\alpha}^2$, das $(1 - \alpha)$ -Quantil der Chi-Quadrat-Verteilung mit $m - k$ Freiheitsgraden.

Es gibt Untersuchungen, die zeigen, dass Q_m in kleinen Stichproben die asymptotische Verteilung nur unzureichend approximiert. Das kommt daher, dass $\hat{\rho}_{\tau,n}$ ein verzerrter Schätzer für ρ_τ ist. Die Verzerrung ist stärker bei kleineren τ , und so kann man eine asymptotisch äquivalente Statistik Q_m^* definieren,

$$Q_m^* = n(n+2) \sum_{j=1}^m \frac{1}{n-j} \hat{\rho}_{j,n}^2$$

die die empirischen Autokorrelationen kleiner Ordnung weniger stark gewichtet als die großer Ordnung. Die modifizierte Portmanteau Statistik Q_m^* ist somit in kleinen Stichproben häufig näher an der asymptotischen χ^2 Verteilung. Für große n , wie sie im Finanzmarktbereich üblich sind, sind beide Statistiken annähernd gleich gut zu gebrauchen.

TO DO:
Bsp. in
letztes
Kap. einf-
gen (Q)

11.7. Schätzung von AR(p) Modellen

Eine einfache Möglichkeit, die Parameter des autoregressiven Modells

$$X_t = \alpha_1 X_{t-1} + \dots + \alpha_p X_{t-p} + \varepsilon_t$$

mit $\text{Var}(\varepsilon_t) = \sigma^2$ zu schätzen, sind die Yule-Walker Gleichungen aus (11.10), indem man die theoretischen Autokorrelationen durch empirische ersetzt:

$$\begin{pmatrix} 1 & \hat{\rho}_1 & \cdots & \hat{\rho}_{p-1} \\ \hat{\rho}_1 & 1 & \cdots & \hat{\rho}_{p-2} \\ \vdots & \vdots & \ddots & \vdots \\ \hat{\rho}_{p-1} & \hat{\rho}_{p-2} & \cdots & 1 \end{pmatrix} \begin{pmatrix} \hat{\alpha}_1 \\ \hat{\alpha}_2 \\ \vdots \\ \hat{\alpha}_p \end{pmatrix} = \begin{pmatrix} \hat{\rho}_1 \\ \hat{\rho}_2 \\ \vdots \\ \hat{\rho}_p \end{pmatrix}.$$

Auflösen nach $\hat{\alpha}$ liefert die Yule-Walker Schätzer. Diese sind konsistent, asymptotisch normalverteilt mit Kovarianzmatrix $\sigma^2 \Gamma^{-1}$,

$$\Gamma = \begin{pmatrix} \gamma_0 & \gamma_1 & \cdots & \gamma_{p-1} \\ \gamma_1 & \gamma_0 & \cdots & \gamma_{p-2} \\ \vdots & \vdots & \ddots & \vdots \\ \gamma_{p-1} & \gamma_{p-2} & \cdots & \gamma_0 \end{pmatrix}, \quad (11.25)$$

Die Yule-Walker Schätzer sind asymptotisch mit anderen Schätzern wie Kleinste-Quadrate und - im Spezialfall normalverteilter ε_t und dann auch normalverteilter X_t - Maximum Likelihood Schätzern äquivalent. Im letztgenannten Fall sind sie daher auch asymptotisch effizient.

11.8. Schätzung von MA(q) und ARMA(p, q) Modellen

Sobald *moving average* Koeffizienten im Modell mitgeschätzt werden, gestaltet sich die Schätzung schwieriger. Betrachten wir zum Beispiel das einfache MA(1) Modell

$$X_t = \beta \varepsilon_{t-1} + \varepsilon_t \quad (11.26)$$

mit $|\beta| < 1$ und $\text{Var}(\varepsilon_t) = \sigma^2$. Einen einfachen Schätzer für den Parameter β erhält man wieder durch die Yule-Walker Gleichungen $\gamma_0 = \sigma^2(1 + \beta^2)$ und $\gamma_1 = \beta\sigma^2$. Man erhält durch Division der beiden Gleichungen $\rho_1 = \gamma_1/\gamma_0 = \beta/(1 + \beta^2)$ und als Lösungen der quadratischen Gleichung

$$\beta = \frac{1}{2\rho_1} \pm \sqrt{\frac{1}{4\rho_1^2} - 1}. \quad (11.27)$$

Der Yule-Walker Schätzer ersetzt in (11.27) die theoretische Autokorrelation erster Ordnung ρ_1 durch die empirische $\hat{\rho}_1$. Der Schätzer ist sehr einfach, hat aber den Nachteil, dass er asymptotisch ineffizient ist.

Kleinste-Quadrate Schätzer führen zu nichtlinearen Gleichungssystemen, die man nur mit iterativen numerischen Algorithmen lösen kann. Am Beispiel des MA(1) Prozesses (11.26) sei dies verdeutlicht: Der KQ Schätzer ist definiert durch

$$\hat{\beta} = \arg \min_{\beta} \sum_{t=2}^n \varepsilon_t^2 = \arg \min_{\beta} \sum_{t=2}^n (X_t - \beta \varepsilon_{t-1})^2 \quad (11.28)$$

Da die ε_t nicht beobachtet sind, muss man zur Lösung auf die AR(∞)-Darstellung des MA(1) Prozesses zurückgreifen, d.h.

$$\varepsilon_t = X_t + \sum_{k=1}^{\infty} (-\beta)^k X_{t-k}. \quad (11.29)$$

Liegen Daten X_1, \dots, X_n vor, so kann (11.29) approximiert werden durch

$$\varepsilon_t = X_t + \sum_{k=1}^{t-1} (-\beta)^k X_{t-k}.$$

Löst man nun die Bedingung erster Ordnung

$$\frac{\partial}{\partial \beta} \sum_{t=2}^n \varepsilon_t^2 = 0,$$

so erhält man nichtlineare Gleichungen für β , die man nicht explizit lösen kann. Für das Minimierungsproblem (11.28) verwendet man daher im allgemeinen numerische Optimierungsverfahren. Kleinste-Quadrate Schätzer sind asymptotisch effizient und haben asymptotisch dieselben Eigenschaften wie Maximum-Likelihood (ML) Schätzer.

Wir gehen im folgenden von einem stationären und invertierbaren ARMA(p, q) Prozess aus mit der AR(∞)-Darstellung

$$X_t = \sum_{j=1}^{\infty} \pi_j X_{t-j} + \varepsilon_t.$$

Maximum Likelihood Schätzer beruhen auf der Verteilungsannahme

$$\varepsilon_t \sim N(0, \sigma^2),$$

unter der $x = (X_1, \dots, X_n)$ multivariat normalverteilt ist mit Dichte

$$p(x | \theta) = (2\pi\sigma^2)^{-n/2} |\Gamma|^{-1/2} \exp\left(-\frac{1}{2\sigma^2} x^T \Gamma^{-1} x\right)$$

mit Kovarianzmatrix Γ , die in (11.25) gegeben ist, und Parametervektor

$$\theta = (\alpha_1, \dots, \alpha_p, \beta_1, \dots, \beta_q; \sigma^2).$$

Die Likelihoodfunktion L ist nun diese Dichtefunktion aufgefasst als Funktion in dem Parametervektor θ für gegebene Beobachtungen, d.h. $L(\theta | x) = p(x | \theta)$. Man wählt denjenigen Parametervektor, der die Likelihoodfunktion für gegebene Beobachtungen maximiert, d.h., der ML Schätzer ist definiert durch

$$\hat{\theta} = \arg \max_{\theta \in \Theta} L(\theta | x).$$

Unter der Normalverteilungsannahme nimmt der Logarithmus der Likelihoodfunktion

$$\log L(\theta | x) = -\frac{n}{2} \log(2\pi\sigma^2) - \frac{1}{2} \log |\Gamma| - \frac{1}{2\sigma^2} x^T \Gamma^{-1} x \quad (11.30)$$

eine einfachere Form an, ohne dass die monotone Transformation etwas an dem maximierenden $\hat{\theta}$ ändert. Die log-Likelihoodfunktion (11.30) wird auch *exakte* log-Likelihoodfunktion genannt. Man beachte, dass insbesondere die Berechnung der Inversen und der Determinanten der $(n \times n)$ -Matrix Γ für lange Zeitreihen sehr aufwendig ist. Man bildet daher häufig Approximationen an die exakte Likelihood, die gut sind für lange Zeitreihen. Eine Möglichkeit besteht darin, die *bedingten* Verteilungen $p(X_t | X_{t-1}, \dots, X_1; \theta)$ zu verwenden:

$$L(\theta | x) = \prod_{t=1}^n p(X_t | X_{t-1}, \dots, X_1; \theta)$$

Die bedingten Verteilungen sind unter der Normalverteilungsannahme normal mit Erwartungswert $E[X_t | X_{t-1}, \dots, X_1]$ und Varianz $\text{Var}(X_t | X_{t-1}, \dots, X_1)$. Je größer t , desto besser wird die Approximation von $E[X_t | X_{t-1}, \dots, X_1, \dots] = \sum_{j=1}^{\infty} \pi_j X_{t-j}$ durch $\sum_{j=1}^{t-1} \pi_j X_{t-j}$. Die bedingte log-Likelihoodfunktion

$$\log L^b(\theta | x) = -\frac{n}{2} \log(2\pi\sigma^2) - \frac{1}{2} \log \sigma^2 - \frac{1}{2\sigma^2} \sum_{t=1}^n \left(X_t - \sum_{j=1}^{t-1} \pi_j X_{t-j}\right)^2 \quad (11.31)$$

kann mit den Daten X_1, \dots, X_n berechnet und bezüglich der Parameter θ optimiert werden. Als Startwerte für die numerischen Optimierungsalgorithmen können zum Beispiel die (außer in Spezialfällen asymptotisch ineffizienten) Yule-Walker Schätzer verwendet werden.

Zum Vergleich der exakten und bedingten Likelihoodschätzer betrachten wir den MA(1) Prozess (11.26) mit $\beta = 0.5$ und $\varepsilon_t \sim N(0, 1)$. Die Matrix Γ ist banddiagonal mit Elementen $1 + \beta^2$ auf der Hauptdiagonalen und β auf den beiden Nebendiagonalen. Zwei Realisierungen des Prozesses mit $n = 10$ und $n = 20$ sind in Abbildung 11.7 gezeigt. Da der Prozess nur einen Parameter enthält, kann man einfach eine Suche auf einem Gitter im Bereich $(-1, 1)$ durchführen. Dies ist für beide Schätzer in den Abbildungen 11.8 ($n = 10$) und 11.9 ($n = 20$) gezeigt. Bei dem Prozess mit $n = 10$ sieht man noch eine deutliche Diskrepanz zwischen beiden Likelihoodfunktionen, die jedoch schon bei $n = 20$ zu vernachlässigen ist. Beide Schätzer treffen in diesem Fall schon recht gut den wahren Parameter 0.5.

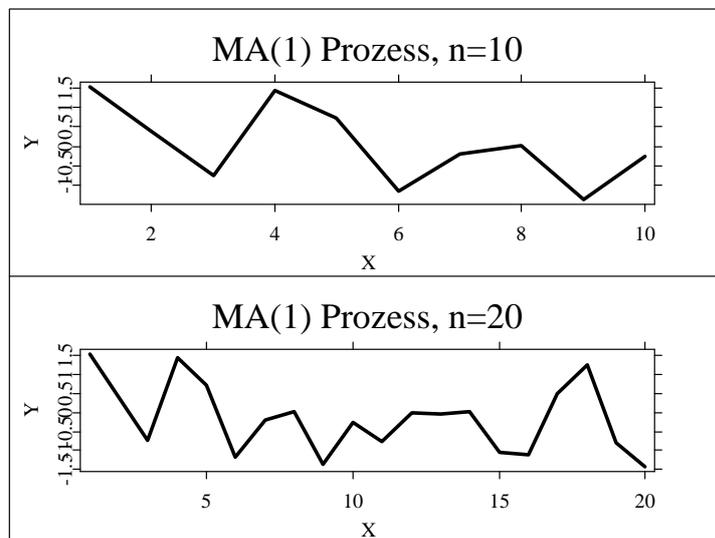


Abbildung 11.7.: Zwei Realisierungen des MA(1) Prozesses mit $\beta = 0.5$, $\varepsilon_t \sim N(0, 1)$, $n = 10$ (oben) und $n = 20$ (unten). [SFMplotma1.xpl](#)

Unter einigen technischen Voraussetzungen sind die ML Schätzer konsistent, asymptotisch effizient und asymptotisch normalverteilt:

$$\sqrt{n}(\hat{\theta} - \theta) \xrightarrow{\mathcal{L}} N(0, J^{-1})$$

mit der Fisherschen Informationsmatrix

$$J = E \left[-\frac{\partial^2 \log L(\theta, x)}{\partial \theta \partial \theta^T} \right]. \quad (11.32)$$

Für die Optimierung der Likelihoodfunktion verwendet man häufig numerische Verfahren. Die notwendige Bedingung für ein Maximum ist

$$\text{grad } l^b(\theta) = 0$$

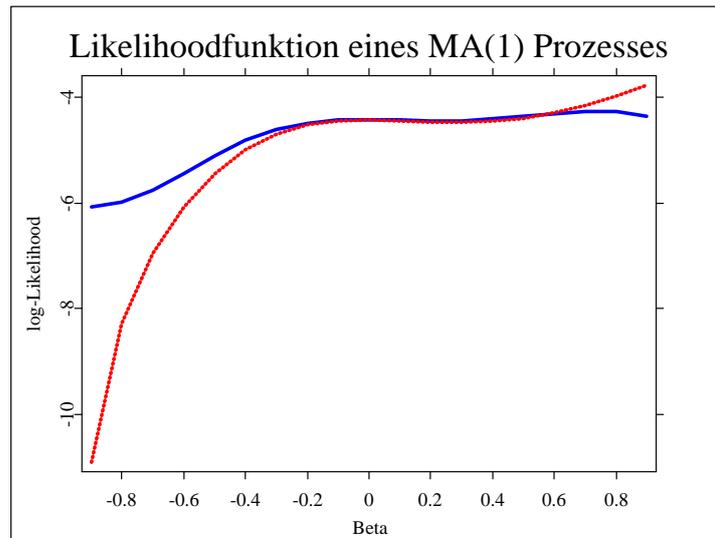


Abbildung 11.8.: Exakte (solid) und bedingte (gestrichelt) Likelihoodfunktion für den MA(1) Prozess aus Abbildung 11.7 mit $n = 10$. Der wahre Parameter ist $\beta = 0.5$. [SFMLikma1.xpl](#)

mit $l^b = \log L(\theta | x)$. Wählt man einen Startwert θ_0 (z.B. den Yule-Walker Schätzer), so erhält man aufgrund der Taylor-Entwicklung

$$\text{grad } l^b(\theta) \approx \text{grad } l^b(\theta_0) + \text{Hess } l^b(\theta_0)(\theta - \theta_0)$$

folgende Relation:

$$\theta = \theta_0 - \text{Hess } l^b(\theta_0)^{-1} \text{grad } l^b(\theta_0).$$

Da man im allgemeinen nicht sofort den maximierenden Parameter trifft, bildet man Iterationen

$$\theta_{j+1} = \theta_j - \text{Hess } l^b(\theta_j)^{-1} \text{grad } l^b(\theta_j)$$

mit $j = 1, 2, \dots$ solange, bis eine Konvergenz erreicht ist, d.h., $\theta_{j+1} \approx \theta_j$. Oft ist es einfacher, den Erwartungswert der Hesseschen Matrix zu verwenden, also die Informationsmatrix aus (11.32):

$$\theta_{j+1} = \theta_j + J(\theta_j)^{-1} \text{grad } l^b(\theta_j). \quad (11.33)$$

Die Schreibweise $J(\theta_j)$ bedeutet hier, dass (11.32) an der Stelle θ_j ausgewertet wird. Die Iteration (11.33) nennt man den *Score-Algorithmus* oder *Fisher Scoring*.

11.9. Literaturhinweise

Umfassende Lehrbücher zur klassischen Zeitreihenanalyse sind z.B. Schlittgen and Streitberg (1995), Brockwell and Davis (1991) und Gouriéroux and Monfort (1996), neben dem klassischen Buch von Box and Jenkins (1976).

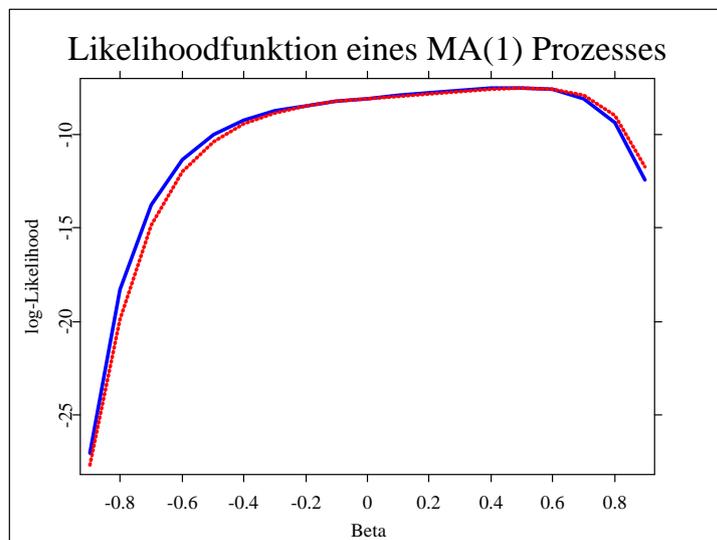


Abbildung 11.9.: Exakte (solid) und bedingte (gestrichelt) Likelihoodfunktion für den MA(1) Prozess aus Abbildung 11.7 mit $n = 20$. Der wahre Parameter ist $\beta = 0.5$. [SFMLikma1.xpl](#)

12. Zeitreihen mit stochastischer Volatilität

Bereits in den vorhergehenden Kapiteln haben wir darauf hingewiesen, dass der Volatilität bei der Modellierung finanzieller Systeme und Zeitreihen eine besondere Rolle zukommt. Die Volatilität ist nicht, wie z.B. die Zinsstruktur, beobachtbar, sondern muss aus den Daten geschätzt werden.

Eine möglichst zuverlässige Schätzung und Prognose der Volatilität ist wichtig für große Kreditinstitute, da Volatilität direkt in Beziehung steht zu Risikomaßen. Zum Beispiel gibt es in der Literatur viele Ansätze, die Risikoprämien als Funktion der Volatilität spezifizieren. Die Problemstellung ist, eine geeignete Modellform für die Volatilität zu finden. Es wurden z.B. makroökonomische Faktoren auf ihren Erklärungsgehalt für die Volatilität untersucht. Eine Schlussfolgerung dieser globalen Ansätze war jedoch, dass makroökonomische Faktoren zwar ein gewisses Erklärungspotential haben, der weitaus wichtigste Faktor jedoch die verzögerten endogenen Renditen waren, so dass man sich in letzter Zeit hauptsächlich auf Zeitreihenmodelle konzentrierte. Wir wollen im folgenden die Familie der ARCH-Modelle, eine der wichtigsten Modellklassen in diesem Bereich, besprechen.

12.1. ARCH- und GARCH Modelle

Seit der Einführung der ARCH Modelle (*autoregressive conditional heteroscedasticity* = *autoregressive bedingte Heteroskedastizität*) in den achtziger Jahren hat es eine enorme Entwicklung in der Finanzökonometrie sowohl in Theorie als auch Praxis gegeben. Es wurde deutlich, dass eine Modellklasse gefunden wurde, die den typischsten empirischen Befund finanzieller Zeitreihen, bedingte Heteroskedastizität, sparsam und einfach modellieren kann. Besonders nach dem Fall des Währungssystems von Bretton Woods und der nachfolgenden Zeit flexibler Wechselkurse in den siebziger Jahren wurde ein solches Modell zunehmend von Forschern und Anwendern benötigt. Die empirische Beobachtung war, dass Volatilitäten dazu tendieren, Cluster zu bilden: nach großen (kleinen) Preisänderungen (positiv oder negativ) folgen tendenziell wieder große (kleine) Preisänderungen. Zum Vergleich zeigen wir in Abbildung 12.1 ein normalverteiltes weißes Rauschen, in Abbildung 12.2 einen GARCH(1,1) Prozess und in Abbildung 12.3 den DAFOX Index (1993-96), siehe <http://finance.wiwi.uni-karlsruhe.de/Forschung/dafox.html> . Offensichtlich eignet sich schon rein visuell ein GARCH Prozess besser zur Modellierung von Aktienrenditen als ein weißes Rauschen.

Seitdem besteht weitgehend Einigkeit darüber, dass Renditen in den meisten Fällen nicht mehr als statistisch unabhängig gelten können, sondern höchstens noch als unkorreliert. Diese

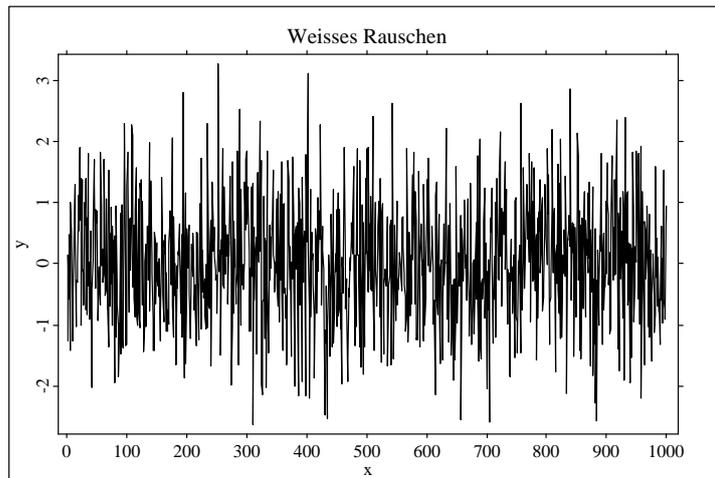


Abbildung 12.1.: Normalverteiltes weißes Rauschen. [SFMtimewr.xpl](#)

Feststellung bezieht sich auf Zeitreihen von relativ hoher Frequenz, d.h. zum Beispiel auf Tagesdaten.

In letzter Zeit wurden einige Besonderheiten der ARCH Modelle aufgedeckt, die von praktischer Relevanz sind. Dazu gehören z.B. die Definition und Modellierung der Persistenz von Schocks, die sich von der Definition in ARMA Modellen unterscheidet, und die Problematik, Asymmetrien zu modellieren. Somit kam es zu zahlreichen Erweiterungen der Standard ARCH Modelle. Heute steht fest, dass das ARCH Modell nur Ausgangspunkt einer empirischen Studie sein kann und man auf umfangreiche Spezifikationstests angewiesen ist.

Wir bezeichnen mit X_t einen stochastischen Prozess in diskreter Zeit und nennen wie in Definition 10.15 $r_t = \log X_t / X_{t-1}$ den relativen Zuwachs oder die *Rendite* des Prozesses X_t . Sind die Renditen unabhängig, identisch verteilt, so folgt X_t einer geometrischen Irrfahrt. ARCH Modelle unterstellen nun, dass die Renditen in einer bestimmten Form von der Information der Vergangenheit abhängen.

Wie bisher bezeichnen wir mit \mathcal{F}_t die ‘Informationsmenge’ zum Zeitpunkt t , die X_t und alle vergangenen Realisierungen des Prozesses X_t umfasst. In dem allgemeinen Modell

$$r_t = \mu_t + \varepsilon_t \tag{12.1}$$

mit $E[\varepsilon_t | \mathcal{F}_{t-1}] = 0$ stellt μ_t z.B. eine Risikoprämie dar, die sich aus ökonomischen Modellen ergeben und zeitabhängig sein kann. Der stochastische Fehler ε_t wird nicht mehr als unabhängig angenommen, sondern nur noch als zentriert und unkorreliert. Bei ARCH Modellen ist die bedingte Varianz von ε_t eine lineare Funktion von verzögerten quadrierten Fehlern.

12.1.1. ARCH(1): Definition und Eigenschaften

Das ARCH Modell der Ordnung 1, ARCH(1), wird folgendermaßen definiert:

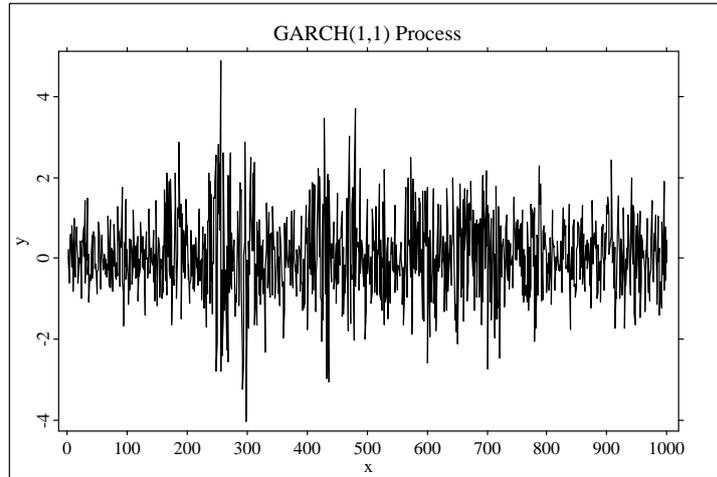


Abbildung 12.2.: Ein GARCH(1,1) Prozess ($\alpha = 0.15$, $\beta = 0.8$). [📄 SFMtimegarc.xpl](#)

DEFINITION 12.1 (ARCH(1)) Der Prozess (ε_t) , $t \in \mathbb{Z}$, ist ARCH(1), wenn $E[\varepsilon_t | \mathcal{F}_{t-1}] = 0$,

$$\sigma_t^2 = \omega + \alpha \varepsilon_{t-1}^2 \quad (12.2)$$

mit $\omega > 0$, $\alpha \geq 0$ und

- $\text{Var}(\varepsilon_t | \mathcal{F}_{t-1}) = \sigma_t^2$ und $Z_t = \varepsilon_t / \sigma_t$ ist u.i.v. (starkes ARCH)
- $\text{Var}(\varepsilon_t | \mathcal{F}_{t-1}) = \sigma_t^2$ (semi-starkes ARCH),
- $\mathcal{P}(\varepsilon_t^2 | 1, \varepsilon_{t-1}, \varepsilon_{t-2}, \dots, \varepsilon_{t-1}^2, \varepsilon_{t-2}^2, \dots) = \sigma_t^2$ (schwaches ARCH),

wobei \mathcal{P} wie in Abschnitt 11.4 die beste lineare Projektion bezeichnet. Offensichtlich ist ein starker ARCH(1) Prozess auch semi-stark und ein semi-starker auch schwach. Andererseits kann z.B. die bedingte Varianz eines schwachen ARCH(1) Prozesses nichtlinear sein (ungleich σ_t^2), womit dieser kein semi-starker ARCH-Prozess wäre.

Setzen wir $Z_t = \varepsilon_t / \sigma_t$, so gilt im semi-starken und im starken ARCH-Modell $E[Z_t] = 0$ und $\text{Var}(Z_t) = 1$. Im starken ARCH-Modell gilt zusätzlich, dass Z_t u.i.v. ist, so dass keine Abhängigkeiten in höheren als den zweiten Momenten modelliert werden können. Eine häufig unterstellte zusätzliche Annahme ist, dass Z_t normalverteilt ist, was bedeutet, dass ε_t bedingt normalverteilt ist:

$$\varepsilon_t | \mathcal{F}_{t-1} \sim N(0, \sigma_t^2). \quad (12.3)$$

Unter (12.3) verschwindet die Unterscheidung zwischen starken und semi-starken ARCH Modellen.

Ursprünglich wurden nur starke und semi-starke ARCH Modelle in der Literatur behandelt. Schwache ARCH-Modelle sind wichtig, weil sie geschlossen sind unter temporaler

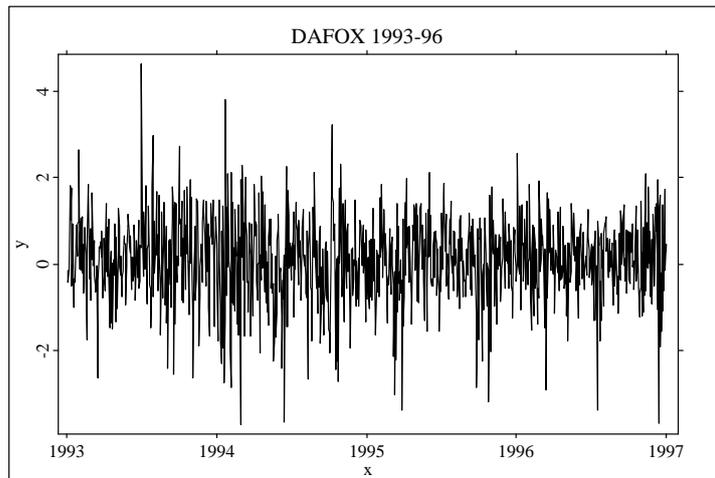


Abbildung 12.3.: Renditen des DAFOX von 1993-96. [SFMtimedax.xpl](#)

Aggregation. Folgen z.B. die Tagesrenditen einem schwachen ARCH Prozess, so tun dies auch die Wochen- und Monatsrenditen bei entsprechenden Parameteranpassungen. Dies gilt in allgemeinen nicht für starke oder semi-starke ARCH Modelle.

Der Prozess ε_t ist nach Definition 12.1 eine Martingaldifferenz und daher auch weißes Rauschen.

SATZ 12.1 *Der Prozess ε_t sei ein schwacher ARCH(1) Prozess mit $\text{Var}(\varepsilon_t) = \sigma^2 < \infty$. Dann folgt, dass ε_t weißes Rauschen ist.*

BEWEIS:

Aus $E[\varepsilon_t | \mathcal{F}_{t-1}] = 0$ folgt $E[\varepsilon_t] = 0$ und $\text{Cov}(\varepsilon_t, \varepsilon_{t-k}) = E[\varepsilon_t \varepsilon_{t-k}] = E[E(\varepsilon_t \varepsilon_{t-k} | \mathcal{F}_{t-1})] = E[\varepsilon_{t-k} E(\varepsilon_t | \mathcal{F}_{t-1})] = 0$. \square

Jedoch ist ε_t kein unabhängiges weißes Rauschen.

SATZ 12.2 (Unbedingte Varianz des ARCH(1)) *Der Prozess ε_t sei ein semi-starker ARCH(1) Prozess mit $\text{Var}(\varepsilon_t) = \sigma^2 < \infty$. Dann gilt*

$$\sigma^2 = \frac{\omega}{1 - \alpha}.$$

BEWEIS:

$\sigma^2 = E[\varepsilon_t^2] = E[E(\varepsilon_t^2 | \mathcal{F}_{t-1})] = E[\sigma_t^2] = \omega + \alpha E[\varepsilon_{t-1}^2] = \omega + \alpha \sigma^2$. Also gilt $\sigma^2 = \omega / (1 - \alpha)$ wenn $\alpha < 1$. \square

Notwendige und hinreichende Bedingung für schwache Stationarität eines semi-starken ARCH(1) Prozesses ist also $\alpha < 1$.

Wenn die Innovation $Z_t = \varepsilon_t / \sigma_t$ symmetrisch um Null verteilt ist, dann sind alle ungeraden Momente von ε_t gleich Null. Unter der Normalverteilungsannahme (12.3) lassen sich Bedingungen für die Existenz der höheren geraden Momente herleiten.

SATZ 12.3 (Viertes Moment) Sei ε_t ein starker ARCH(1) Prozess mit u.i.v. $Z_t \sim N(0, 1)$ und $E[\varepsilon_t^4] = c < \infty$. Dann ist

1.

$$E[\varepsilon_t^4] = \frac{3\omega^2}{(1-\alpha)^2} \frac{1-\alpha^2}{1-3\alpha^2}$$

mit $3\alpha^2 < 1$.

2. die unbedingte Verteilung von ε_t leptokurtisch.

BEWEIS:

1. $c = E[\varepsilon_t^4] = E[E(\varepsilon_t^4 | \mathcal{F}_{t-1})] = E[\sigma_t^4 E(Z_t^4 | \mathcal{F}_{t-1})] = E[Z_t^4] E[(\omega + \alpha_1 \varepsilon_{t-1}^2)^2] = 3(\omega^2 + 2\omega\alpha E[\varepsilon_{t-1}^2] + \alpha^2 E[\varepsilon_{t-1}^4])$. Da $E[\varepsilon_{t-1}^2] = \omega/(1-\alpha)$ und $E[\varepsilon_{t-1}^4] = c$, folgt nach Umformen die Behauptung.

2.

$$\text{Kurt}(\varepsilon_t) = \frac{E[\varepsilon_t^4]}{E[\varepsilon_t^2]^2} = 3 \frac{1-\alpha^2}{1-3\alpha^2} \geq 3.$$

□

Für den Grenzfall $\alpha = 0$ und normalverteilten Innovationen ist $\text{Kurt}(\varepsilon_t) = 3$, während für $\alpha > 0$ gilt $\text{Kurt}(\varepsilon_t) > 3$. Die unbedingte Verteilung ist also bei bedingter Heteroskedastizität leptokurtisch, d.h., die Wölbung im Zentrum der Verteilung ist stärker und die Flanken sind dicker als bei der Normalverteilung, was häufig bei Finanzmarktdaten zu beobachten ist.

Die Dicke der Flanken und damit die Existenz von Momenten hängt von den Parametern des ARCH-Modells ab. Die Varianz des ARCH(1) Prozesses ist endlich, falls $\alpha < 1$ (Satz 12.2) und das vierte Moment im Falle von normalverteilten Innovationen existiert, falls $3\alpha^2 < 1$ (Satz 12.3). Schon in den sechziger Jahren stellte Mandelbrot die Existenz der Varianz für viele Finanzzeitreihen in Frage. Häufig sind empirische Verteilungen so stark leptokurtisch, dass man nicht auf eine endliche Varianz schließen kann. Um empirisch Aussagen über den Grad der Flankenstärke der unbedingten Verteilung machen zu können, nimmt man z.B. an, dass man eine Verteilung vom Pareto-Typ hat, d.h. für große x :

$$P(x) = \Pr(X_t > x) \sim kx^{-a}$$

für ein $a > 0$. Wenn $a > c$, dann gilt $E[|X_t|^c] < \infty$. Die Frage ist, wie man den Parameter a schätzen kann. Ein einfacher Ansatz folgt aus der Feststellung, dass für große x die logarithmierte Funktion $P(x)$ linear ist, d.h.,

$$\log P(x) \approx \log k - a \log x. \tag{12.4}$$

Wir bilden die Ordnungsstatistiken $X_{(1)} > X_{(2)} > \dots > X_{(n)}$ und schätzen die Wahrscheinlichkeit $P(x)$ für $x = X_{(j)}$ durch die relative Häufigkeit

$$\frac{\#\{t; X_t \geq X_{(j)}\}}{n} = \frac{j}{n}.$$

In (12.4) ersetzen wir $p(X_{(j)})$ durch den Schätzer j/n :

$$\log \frac{j}{n} \approx \log k - a \log X_{(j)}, \quad (12.5)$$

woraus man a aus der Regression von j/n auf $X_{(j)}, j = 1, \dots, n$, durch die Methode der kleinsten Quadrate schätzen kann. In der Regel wird man nur einen kleinen Teil der Daten für die Regression verwenden, da die lineare Approximation der von $\log P(x)$ nur in den Flanken gut ist. Man schätzt also die Regression (12.5) nur mit den m größten Ordnungsstatistiken. Abbildung 12.4 zeigt die Regression (12.5) für den DAFOX von 1974 bis 1996 mit $m = 20$. Der Steigungsparameter der KQ-Geraden ist -3.25. Für diese Indexzeitreihe ist demnach nicht nur die Varianz, sondern auch das dritte Moment $E(|X|^3)$ endlich, während das vierte Moment und damit die Kurtosis nicht endlich sind.

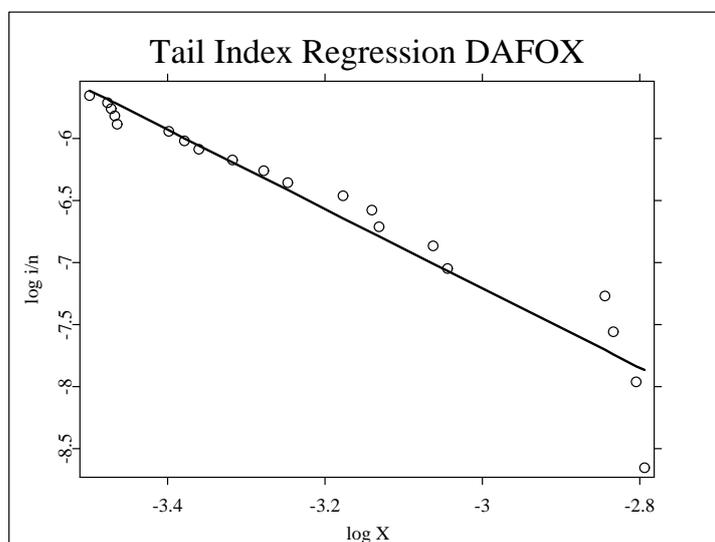


Abbildung 12.4.: Rechte Flanke der logarithmierten empirischen Verteilung der DAFOX Renditen 1974 bis 1996. [SFMtaildax.xpl](#)

Hill (1975) hat einen Schätzer vorgeschlagen, der auf einem Maximum Likelihood Ansatz beruht:

$$\hat{a} = \left(\frac{1}{m-1} \sum_{i=1}^m \log X_{(i)} - \log X_{(m)} \right)^{-1}, \quad (12.6)$$

wobei m die Anzahl der für den Schätzer verwendeten Daten in den Flanken angibt. Die Wahl von m stellt offensichtlich ein Problem dar. Wenn m zu groß gewählt wird, ist die Approximation der Verteilung nicht mehr gut, wenn m zu klein ist, können Bias und Varianz des Schätzers groß sein. Eine einfache Daumenregel sagt z.B., dass m/n etwa 0.5% oder 1% sein sollte. Es wird deutlich, daß man sehr viele Daten benötigt, um a vernünftig schätzen zu können. Als empirisches Beispiel betrachten wir wieder die täglichen Renditen

m	KQ		Hill	
	20	50	20	50
DAFOX	3.25	2.94	3.17	2.88
ALLIANZ	2.29	2.44	2.28	2.26
BASF	4.19	4.29	4.58	4.01
BAYER	3.32	3.20	3.90	3.23
BMW	3.42	3.05	3.05	2.89
COMMERZBANK	6.58	4.67	7.14	5.19
DAIMLER	2.85	2.85	2.43	2.56
DEUTSCHE BANK	3.40	3.26	3.41	3.29
DEGUSSA	3.03	4.16	2.93	3.30
DRESDNER	5.76	4.08	4.20	4.32
HOECHST	4.77	3.68	5.66	4.05
KARSTADT	3.56	3.42	3.11	3.16
LINDE	3.30	3.35	3.87	3.37
MAN	3.83	3.66	3.17	3.45
MANNESMANN	3.19	3.85	2.84	3.22
PREUSSAG	3.52	4.11	3.57	3.68
RWE	3.87	3.78	3.51	3.54
SCHERING	3.34	4.82	3.22	3.64
SIEMENS	6.06	4.50	5.96	5.23
THYSSEN	5.31	5.36	4.67	4.97
VOLKSWAGEN	4.59	3.31	4.86	4.00

Tabelle 12.1.: Naive Kleinste-Quadrate (KQ) und Hill Schätzer des Flanken-Exponenten a mit Stützpunkten m .

deutscher Aktienkurse von 1974 bis 1996, jeweils 5747 Beobachtungen. Die Ergebnisse des naiven Kleinsten-Quadrate (KQ) Schätzers und des Hill Schätzers mit den Stützpunkten $m = 20$ und $m = 50$ sind in Tabelle 12.1 aufgeführt. In allen Fällen sind die Schätzer größer als 2, was die Existenz der Varianzen nahelegt. In einigen Fällen bestehen Zweifel über die Existenz dritter Momente, z.B. bei Allianz und Daimler.

SATZ 12.4 (Repräsentation eines ARCH(1) Prozesses) Sei ε_t ein starker ARCH(1) Prozess mit $\text{Var}(\varepsilon_t) = \sigma^2 < \infty$. Dann ist

$$\varepsilon_t^2 = \omega \sum_{k=0}^{\infty} \alpha^k \prod_{j=0}^k Z_{t-j}^2$$

und die Summe konvergiert im L_1 -Sinne.

BEWEIS:

durch sukzessives Einsetzen von $\varepsilon_s^2 = \sigma_s^2 Z_s^2$ und $\sigma_s^2 = \omega + \alpha \varepsilon_{s-1}^2$. Die Konvergenz folgt aus $\text{E}[\varepsilon_t^2 - \omega \sum_{k=0}^m \alpha^k \prod_{j=0}^k Z_{t-j}^2] = \alpha^{m+1} \text{E}[\varepsilon_{t-m-1}^2 \prod_{j=0}^m Z_{t-j}^2] = \alpha^{m+1} \text{E}[\varepsilon_{t-m-1}^2] \rightarrow 0$ für $m \rightarrow \infty$, da die Z_t unabhängig mit $\text{E}(Z_t^2) = 1$ sind. \square

SATZ 12.5 Sei ε_t ein stationärer starker ARCH(1) Prozess mit $\text{E}(\varepsilon_t^4) = c < \infty$ und $Z_t \sim N(0, 1)$. Dann ist

1.

$$\varepsilon_t^2 = \omega \sum_{k=0}^{\infty} \alpha^k \prod_{j=0}^k Z_{t-j}^2$$

und die Summe konvergiert im L_2 -Sinne.

2. $\eta_t = \sigma_t^2(Z_t^2 - 1)$ ist weißes Rauschen.

3. ε_t^2 ist ein AR(1) Prozess mit $\varepsilon_t^2 = \omega + \alpha\varepsilon_{t-1}^2 + \eta_t$.

BEWEIS:

1. wie in Satz 12.4. Die Konvergenz in L_2 folgt aus $E[(\varepsilon_t^2 - \omega \sum_{k=0}^m \alpha^k \prod_{j=0}^k Z_{t-j}^2)^2] = E[(\alpha^{m+1} \varepsilon_{t-m-1}^2 \prod_{j=0}^m Z_{t-j}^2)^2] = \alpha^{2(m+1)} 3^{m+1} E[\varepsilon_{t-m-1}^4] = \alpha^{2(m+1)} 3^{m+1} c \rightarrow 0$ für $m \rightarrow \infty$, da $3\alpha^2 < 1$ wegen der vorausgesetzten Endlichkeit von $E(\varepsilon_t^4)$ und da die Z_t unabhängig mit $\text{Kurt}(Z_t) = 3$ sind.

2. a) $E[\eta_t] = E[\sigma_t^2]E[Z_t^2 - 1] = 0$

b) $\text{Var}(\eta_t) = E[\sigma_t^4]E[(Z_t^2 - 1)^2] = 2E[(\omega + \alpha\varepsilon_{t-1}^2)^2] = 2(\omega^2 + 2\alpha\omega E[\varepsilon_{t-1}^2] + \alpha^2 E[\varepsilon_{t-1}^4]) = \text{const}$ unabhängig von t .

c) $\text{Cov}(\eta_t, \eta_{t+s}) = E[\sigma_t^2(Z_t^2 - 1)\sigma_{t+s}^2(Z_{t+s}^2 - 1)] = E[\sigma_t^2(Z_t^2 - 1)\sigma_{t+s}^2]E[(Z_{t+s}^2 - 1)] = 0$ für $s \neq 0$.

3. Folgt durch Umformen: $\varepsilon_t^2 = \sigma_t^2 Z_t^2 = \sigma_t^2 + \sigma_t^2(Z_t^2 - 1) = \omega + \alpha\varepsilon_{t-1}^2 + \eta_t$.

□

BEMERKUNG 12.1 Von [Nelson \(1990\)](#) wurde gezeigt, dass der starke ARCH(1) Prozess ε_t streng stationär ist, wenn $E[\log(\alpha Z_t^2)] < 0$. Ist z.B. $Z_t \sim N(0, 1)$, dann ist die Bedingung für starke Stationarität $\alpha < 3.5622$, also schwächer als die Bedingung für Kovarianzstationarität, $\alpha < 1$, die wegen der hier geforderten Endlichkeit der Varianz gebraucht wird.

Die Dynamik des Volatilitätsprozesses im ARCH(1) Fall ist im wesentlichen durch den Parameter α bestimmt. In Satz 12.5 wurde gezeigt, dass die Quadrate eines ARCH(1) Prozesses einem AR(1) Prozess folgen. Die Korrelationsstruktur von empirisch beobachteten quadrierten Renditen ist jedoch häufig komplexer als die eines einfachen AR(1) Prozesses. In Abschnitt 12.1.3 betrachten wir ARCH Modelle der Ordnung q mit $q > 1$, die eine flexiblere Modellierung der Korrelationsstruktur ermöglichen.

12.1.2. Schätzung von ARCH(1) Modellen

Satz 12.5 sagte aus, dass sich ein ARCH(1)-Prozess als AR(1)-Prozess in X_t^2 auffassen lässt. Ein einfacher Yule-Walker Schätzer nutzt diese Eigenschaft aus:

$$\hat{\alpha}^{(0)} = \frac{\sum_{t=2}^n (\varepsilon_t^2 - \hat{\omega}^{(0)})(\varepsilon_{t-1}^2 - \hat{\omega}^{(0)})}{\sum_{t=2}^n (\varepsilon_t^2 - \hat{\omega}^{(0)})}$$

mit $\hat{\omega}^{(0)} = n^{-1} \sum_{t=1}^n \varepsilon_t^2$. Da natürlich die Verteilung von ε_t^2 nicht normal ist, ist der Yule-Walker Schätzer nicht effizient. Er kann aber als Startwert für iterative Schätzverfahren verwendet werden.

Die Schätzung von ARCH Modellen erfolgt üblicherweise mit der Maximum Likelihood (ML) Methode. Nimmt man hierbei eine bedingte Normalverteilung für die Renditen ε_t an, d.h.

$$p(\varepsilon_t | \mathcal{F}_{t-1}) = \frac{1}{\sqrt{2\pi}\sigma_t} \exp\left\{-\frac{1}{2} \frac{\varepsilon_t^2}{\sigma_t^2}\right\}, \quad (12.7)$$

so lässt sich die log-Likelihoodfunktion $l(\omega, \alpha)$ als Funktion der Parameter ω und α aufstellen:

$$\begin{aligned} l(\omega, \alpha) &= \sum_{t=2}^n l_t(\omega, \alpha) + \log p_\varepsilon(\varepsilon_1) \\ &= \sum_{t=2}^n \log p(\varepsilon_t | \mathcal{F}_{t-1}) + \log p_\varepsilon(\varepsilon_1) \\ &= -\frac{n-1}{2} \log(2\pi) - \frac{1}{2} \sum_{t=2}^n \log(\omega + \alpha\varepsilon_{t-1}^2) - \frac{1}{2} \sum_{t=2}^n \frac{\varepsilon_t^2}{\omega + \alpha\varepsilon_{t-1}^2} + \log p_\varepsilon(\varepsilon_1), \end{aligned} \quad (12.8)$$

wobei p_ε die stationäre marginale Dichte von ε_t bezeichnet. Ein Problem ist, dass ein analytischer Ausdruck für p_ε im Fall von ARCH Modellen unbekannt ist, so dass auch (12.9) nicht berechenbar ist. In der bedingten Likelihoodfunktion $l^b = \log p(\varepsilon_n, \dots, \varepsilon_2 | \varepsilon_1)$ fällt der Ausdruck $\log p_\varepsilon(\varepsilon_1)$ weg:

$$\begin{aligned} l^b(\omega, \alpha) &= \sum_{t=2}^n l_t(\omega, \alpha) \\ &= \sum_{t=2}^n \log p(\varepsilon_t | \mathcal{F}_{t-1}) \\ &= -\frac{n-1}{2} \log(2\pi) - 1/2 \sum_{t=2}^n \log(\omega + \alpha\varepsilon_{t-1}^2) - 1/2 \sum_{t=2}^n \frac{\varepsilon_t^2}{\omega + \alpha\varepsilon_{t-1}^2}. \end{aligned} \quad (12.9)$$

Für große n ist die Differenz $l - l^b$ vernachlässigbar.

Abbildung 12.5 zeigt die bedingte Likelihood eines generierten ARCH(1) Prozesses mit $n = 100$. Der Parameter ω wurde dabei so gewählt, dass die unbedingte Varianz stets gleich ist, d.h. also bei einer Varianz von σ^2 , $\omega = (1 - \alpha)\sigma^2$. Die Optimierung der Likelihood eines ARCH(1) Modells kann gegebenenfalls noch mit einer Gittersuche erfolgen. Allerdings möchte man auch die Präzision des Schätzers kennen. Im wesentlichen wird diese durch die zweite Ableitung der Likelihood an der Stelle des Optimums bestimmt, wie sich aus den asymptotischen Eigenschaften der ML Schätzer ergibt (siehe Abschnitt 12.1.6). Außerdem ist man bei Modellen höherer Ordnung auf numerische Verfahren wie dem Score-Algorithmus aus Abschnitt 11.8 angewiesen, so dass die ersten und zweiten partiellen Ableitungen der Likelihood berechnet werden müssen.

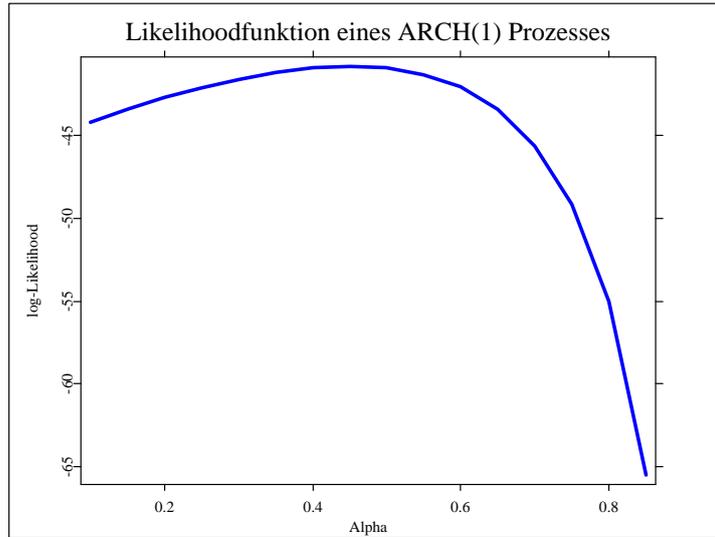


Abbildung 12.5.: Bedingte Likelihoodfunktion eines generierten ARCH(1) Prozesses mit $n = 100$. Der wahre Parameter ist $\alpha = 0.5$. [SFMLikarch1.xpl](#)

Bei dem ARCH(1) Modell lauten diese

$$\frac{\partial l_t^b}{\partial \omega} = \frac{1}{2\sigma_t^2} \left(\frac{\varepsilon_t^2}{\sigma_t^2} - 1 \right) \quad (12.11)$$

$$\frac{\partial l_t^b}{\partial \alpha} = \frac{1}{2\sigma_t^2} \varepsilon_{t-1}^2 \left(\frac{\varepsilon_t^2}{\sigma_t^2} - 1 \right) \quad (12.12)$$

$$\frac{\partial^2 l_t^b}{\partial \omega^2} = -\frac{1}{2\sigma_t^4} \left(2\frac{\varepsilon_t^2}{\sigma_t^2} - 1 \right) \quad (12.13)$$

$$\frac{\partial^2 l_t^b}{\partial \alpha^2} = -\frac{1}{2\sigma_t^4} \varepsilon_{t-1}^4 \left(2\frac{\varepsilon_t^2}{\sigma_t^2} - 1 \right) \quad (12.14)$$

$$\frac{\partial^2 l_t^b}{\partial \omega \partial \alpha} = -\frac{1}{2\sigma_t^4} \varepsilon_{t-1}^2 \left(2\frac{\varepsilon_t^2}{\sigma_t^2} - 1 \right). \quad (12.15)$$

Die Bedingungen erster Ordnung sind $\sum_{t=2}^n \partial l_t^b / \partial \omega = 0$ und $\sum_{t=2}^n \partial l_t^b / \partial \alpha = 0$. Für den Score-Algorithmus müssen die Erwartungswerte der zweiten Ableitungen berechnet werden. Nach Voraussetzung gilt $E[Z_t^2] = E[(\varepsilon_t / \sigma_t)^2] = 1$, so dass der Klammerausdruck $(2\varepsilon_t^2 / \sigma_t^2 - 1)$ den Erwartungswert eins hat. Daraus folgt

$$E \left[\frac{\partial^2 l_t^b}{\partial \omega^2} \right] = -\frac{1}{2} E \left[\frac{1}{\sigma_t^4} \right].$$

Der Erwartungswert von σ_t^{-4} wird konsistent durch $(n-1)^{-1} \sum_{t=2}^n (\omega + \alpha \varepsilon_{t-1}^2)^{-2}$ geschätzt, so dass sich als Schätzer für den Erwartungswert der zweiten Ableitung ergibt:

$$\hat{E} \frac{\partial^2 l_t^b}{\partial \omega^2} = -\frac{1}{2(n-1)} \sum_{t=2}^n \frac{1}{\sigma_t^4}.$$

Ebenso erhält man für den Erwartungswert der zweiten Ableitung bezüglich α

$$E \left[\frac{\partial^2 l_t^b}{\partial \alpha^2} \right] = -\frac{1}{2} E \left[\frac{\varepsilon_{t-1}^4}{\sigma_t^4} \right]$$

und als Schätzer

$$\hat{E} \frac{\partial^2 l_t^b}{\partial \alpha^2} = -\frac{1}{2(n-1)} \sum_{t=2}^n \frac{\varepsilon_{t-1}^4}{\sigma_t^4}.$$

SATZ 12.6 Ist $Z_t \sim N(0, 1)$, so gilt

$$E \left[\left(\frac{\partial l_t^b}{\partial \omega} \right)^2 \right] = -E \left[\frac{\partial^2 l_t^b}{\partial \omega^2} \right]$$

BEWEIS:

Folgt sofort aus $E \left[\left(\frac{\partial l_t^b}{\partial \omega} \right)^2 \right] = E \left[\frac{1}{4\sigma_t^4} (Z_t^4 - 2Z_t^2 + 1) \right] = E \left[\frac{1}{4\sigma_t^4} \right] (3 - 2 + 1)$. \square

Natürlich gilt Satz 12.6 auch für den Parameter α anstelle von ω . In der Tat gilt der Satz auch für wesentlich allgemeinere Modelle, siehe z.B. Abschnitt 12.1.6 über die Schätzung von GARCH Modellen. Der Satz ist nützlich, da man bei komplizierteren Modellen häufig die zweiten Ableitungen durch das Quadrat der ersten Ableitungen, die sich einfacher berechnen lassen, ersetzen kann. Voraussetzung ist allerdings, dass man die Likelihoodfunktion richtig spezifiziert hat, dass also die wahre Verteilung der Fehler normal ist.

Unter den beiden Bedingungen

1. $E[Z_t | \mathcal{F}_{t-1}] = 0$ und $E[Z_t^2 | \mathcal{F}_{t-1}] = 1$
2. $E[\log(\alpha Z_t^2) | \mathcal{F}_{t-1}] < 0$ (strenge Stationarität)

und einigen technischen Bedingungen sind die ML-Schätzer konsistent. Gilt zusätzlich $E[Z_t^4 | \mathcal{F}_{t-1}] < \infty$ und $\omega > 0$, $\alpha > 0$, dann ist $\hat{\theta} = (\hat{\omega}, \hat{\alpha})^T$ asymptotisch normalverteilt:

$$\sqrt{n}(\hat{\theta} - \theta) \xrightarrow{\mathcal{L}} N(0, J^{-1} I J^{-1}) \quad (12.16)$$

mit

$$I = E \left(\frac{\partial l_t(\theta)}{\partial \theta} \frac{\partial l_t(\theta)}{\partial \theta^T} \right)$$

und

$$J = -E \left(\frac{\partial^2 l_t(\theta)}{\partial \theta \partial \theta^T} \right).$$

Ist die wahre Verteilung von Z_t normal, so ist $I = J$ und die asymptotische Kovarianzmatrix vereinfacht sich zu J^{-1} , d.h., der Inversen der Fisherschen Informationsmatrix. Ist sie es nicht, sondern z.B. leptokurtisch, so kann die Maximierung von (12.10) zwar noch konsistente, aber nicht mehr effiziente Schätzungen liefern. In diesem Fall wird die ML Methode als ‘Quasi Maximum Likelihood’ (QML) Methode interpretiert.

n	$k^{-1} \sum_{j=1}^k \hat{\alpha}_j$	$\sqrt{k^{-1} \sum_{j=1}^k (\hat{\alpha}_j - \alpha)^2}$	$\#(\alpha_j \geq 1)$
100	0.852	0.257	27%
250	0.884	0.164	24%
500	0.893	0.107	15%
1000	0.898	0.081	10%

Tabelle 12.2.: Monte Carlo Simulationsergebnisse der QML Schätzung des Parameters $\alpha = 0.9$ eines ARCH(1) Modells bei $k = 1000$ Replikationen. Die letzte Spalte gibt den Anteil der Schätzer an, die größer als 1 waren (nach Shephard (1996)).

In einer Monte Carlo Simulationsstudie von Shephard (1996) wurden 1000 ARCH(1) Prozesse mit $\omega = 0.2$ und $\alpha = 0.9$ generiert und die Parameter mit QML geschätzt. Die Ergebnisse sind in Tabelle 12.2 dargestellt. Offenbar ist schon für moderate Stichprobengrößen ($n = 500$) der Bias vernachlässigbar. Die Varianz allerdings ist noch bei $n = 500$ so groß, dass ein relativ großer Teil (10%) der Schätzer größer als eins ist, was einen kovarianzstationären Prozess impliziert und erhebliche Auswirkungen auf Volatilitätsprognosen

TO DO: hat.
Abdruckerlaubnis
für Tabelle

12.1.3. ARCH(q): Definition und Eigenschaften

Die Definition des ARCH(1) Modells wird erweitert auf den Fall mit $q > 1$ Verzögerungen, von denen die bedingte Varianz abhängt.

DEFINITION 12.2 (ARCH(q)) Der Prozess (ε_t) , $t \in \mathbb{Z}$, ist ARCH(q), wenn $E[\varepsilon_t | \mathcal{F}_{t-1}] = 0$,

$$\sigma_t^2 = \omega + \alpha_1 \varepsilon_{t-1}^2 + \dots + \alpha_q \varepsilon_{t-q}^2 \quad (12.17)$$

mit $\omega > 0$, $\alpha_1 \geq 0, \dots, \alpha_q \geq 0$ und

- $\text{Var}(\varepsilon_t | \mathcal{F}_{t-1}) = \sigma_t^2$ und $Z_t = \varepsilon_t / \sigma_t$ ist u.i.v. (starkes ARCH)
- $\text{Var}(\varepsilon_t | \mathcal{F}_{t-1}) = \sigma_t^2$ (semi-starkes ARCH), oder
- $\mathcal{P}(\varepsilon_t^2 | 1, \varepsilon_{t-1}, \varepsilon_{t-2}, \dots, \varepsilon_{t-1}^2, \varepsilon_{t-2}^2, \dots) = \sigma_t^2$ (schwaches ARCH)

Die bedingte Varianz σ_t^2 im ARCH(q) Modell ist also eine lineare Funktion der q quadrierten Verzögerungen.

SATZ 12.7 Sei ε_t ein semi-starker ARCH(q) Prozess mit $\text{Var}(\varepsilon_t) = \sigma^2 < \infty$. Dann ist

$$\sigma^2 = \frac{\omega}{1 - \alpha_1 - \dots - \alpha_q}$$

mit $\alpha_1 + \dots + \alpha_q < 1$.

BEWEIS:wie in Satz 12.2. □

Falls hingegen $\alpha_1 + \dots + \alpha_q \geq 1$, dann existiert die unbedingte Varianz nicht und der Prozess ist nicht kovarianzstationär.

SATZ 12.8 (Repräsentation eines ARCH(q) Prozesses) Sei ε_t ein (semi-)starker ARCH(q) Prozess mit $E[\varepsilon_t^4] = c < \infty$. Dann ist

1. $\eta_t = \sigma_t^2(Z_t^2 - 1)$ weißes Rauschen.
2. ε_t^2 ein AR(q) Prozess mit $\varepsilon_t^2 = \omega + \sum_{i=1}^q \alpha_i \varepsilon_{t-i}^2 + \eta_t$.

BEWEIS:wie in Satz 12.5. □

Ein Problem der ARCH(q) Modelle ist, dass für einige Anwendungen hohe Ordnungen q verwendet werden müssen, da der Einfluss weit zurückliegender Werte auf die Volatilität nur langsam an Stärke verliert. Einige Anwender empfehlen als empirische Daumenregel, mindestens die Ordnung $q = 14$ zu verwenden. Der Nachteil einer hohen Modellordnung ist, dass viele Parameter unter Restriktionen geschätzt werden müssen. Die Restriktionen teilen sich auf in Nichtnegativitäts- und Stationaritätsbedingungen. Will man effiziente Schätzmethoden verwenden, z.B. die Maximum Likelihood Methode, so kann sich die Schätzung bei hochdimensionalen Parameterräumen als numerisch sehr aufwendig herausstellen.

Eine Möglichkeit, die Zahl der Parameter zu reduzieren und trotzdem einen langen Zeitraum der Vergangenheit mit einzubeziehen ist, eine linear abfallende Gewichtsverteilung über die Verzögerungen (*Lags*) anzunehmen, d.h.

$$\sigma_t^2 = \omega + \alpha \sum_{i=1}^q w_i \varepsilon_{t-i}^2,$$

mit

$$w_i = \frac{2(q+1-i)}{q(q+1)},$$

so dass nur zwei Parameter geschätzt werden müssen. In Abschnitt 12.1.5 beschreiben wir eine Generalisierung der ARCH Modelle, die einerseits eine sparsame Parametrisierung, andererseits eine flexiblere Lagstruktur ermöglicht.

12.1.4. Schätzung von ARCH(q) Modellen

Für das allgemeine ARCH(q) Modell aus (12.17) lautet die bedingte Likelihood

$$l^b(\theta) = \sum_{t=q+1}^n l_t(\theta) \tag{12.18}$$

$$= -\frac{n-1}{2} \log(2\pi) - 1/2 \sum_{t=2}^n \log \sigma_t^2 - 1/2 \sum_{t=q+1}^n \frac{\varepsilon_t^q + 1}{\sigma_t^2} \tag{12.19}$$

mit dem Parametervektor $\theta = (\omega, \alpha_1, \dots, \alpha_q)'$. Hat man im ARCH(1) Modell noch die Möglichkeit, über eine Gittersuche wie in Abbildung 12.5 das Optimum zu finden, ist dies für höherdimensionale Parameterräume sehr aufwendig. Die Maximierung von (12.19) bezüglich θ ist ein nichtlineares Optimierungsproblem, für das verschiedene numerische Verfahren verwendet werden können. Bereits bei ARMA Modellen in Abschnitt 11.8 wurde ein Optimierungsalgorithmus, der *Score*-Algorithmus, vorgestellt, der auch im Falle von ARCH-Modellen praktikabel ist. Hierbei sind die ersten und zweiten partiellen Ableitungen der (bedingten) Likelihood nach den Parametern zu bilden. Bei dem ARCH(q) Modell lauten die ersten Ableitungen

$$\frac{\partial l_t^b}{\partial \theta} = \frac{1}{2\sigma_t^2} \frac{\partial \sigma_t^2}{\partial \theta} \left(\frac{\varepsilon_t^2}{\sigma_t^2} - 1 \right) \quad (12.20)$$

mit

$$\frac{\partial \sigma_t^2}{\partial \theta} = (1, \varepsilon_{t-1}^2, \dots, \varepsilon_{t-q}^2)^T.$$

Die Bedingungen erster Ordnung sind $\sum_{t=q+1}^n \partial l_t / \partial \theta = 0$. Bezüglich der zweiten Ableitungen und der asymptotischen Eigenschaften der QML Schätzer siehe Abschnitt 12.1.6.

12.1.5. Generalisiertes ARCH (GARCH)

Das ARCH(q) Modell kann durch Erweiterung um autoregressive Terme der Volatilität verallgemeinert werden.

DEFINITION 12.3 (GARCH(p, q)) *Der Prozess (ε_t) , $t \in \mathbb{Z}$, ist GARCH(p, q), wenn $E[\varepsilon_t | \mathcal{F}_{t-1}] = 0$,*

$$\sigma_t^2 = \omega + \sum_{i=1}^q \alpha_i \varepsilon_{t-i}^2 + \sum_{j=1}^p \beta_j \sigma_{t-j}^2, \quad (12.21)$$

und

- $\text{Var}(\varepsilon_t | \mathcal{F}_{t-1}) = \sigma_t^2$ und $Z_t = \varepsilon_t / \sigma_t$ ist u.i.v. (*starkes GARCH*)
- $\text{Var}(\varepsilon_t | \mathcal{F}_{t-1}) = \sigma_t^2$ (*semi-starkes GARCH*), oder
- $\mathcal{P}(\varepsilon_t^2 | 1, \varepsilon_{t-1}, \varepsilon_{t-2}, \dots, \varepsilon_{t-1}^2, \varepsilon_{t-2}^2, \dots) = \sigma_t^2$ (*schwaches GARCH*).

Hinreichende, aber nicht notwendige Bedingungen für

$$\sigma_t^2 > 0 \quad f.s., \quad (P[\sigma_t^2 > 0] = 1) \quad (12.22)$$

sind $\omega > 0$, $\alpha_i \geq 0$, $i = 1, \dots, q$ und $\beta_j \geq 0$, $j = 1, \dots, p$. Im Falle des GARCH(1,2) Modells

$$\begin{aligned} \sigma_t^2 &= \omega + \alpha_1 \varepsilon_{t-1}^2 + \alpha_2 \varepsilon_{t-2}^2 + \beta_1 \sigma_{t-1}^2 \\ &= \frac{\omega}{1-\beta} + \alpha_1 \sum_{j=0}^{\infty} \beta_1^j \varepsilon_{t-j-1}^2 + \alpha_2 \sum_{j=0}^{\infty} \beta_1^j \varepsilon_{t-j-2}^2 \\ &= \frac{\omega}{1-\beta} + \alpha_1 \varepsilon_{t-1}^2 + (\alpha_1 \beta_1 + \alpha_2) \sum_{j=0}^{\infty} \beta_1^j \varepsilon_{t-j-2}^2 \end{aligned}$$

mit $0 \leq \beta_1 < 1$ sind notwendige und hinreichende Bedingungen für (12.22) $\omega > 0$, $\alpha_1 \geq 0$ und $\alpha_1\beta_1 + \alpha_2 \geq 0$, vorausgesetzt, dass die Summe $\sum_{j=0}^{\infty} \beta_1^j \varepsilon_{t-j-2}^2$ konvergiert.

SATZ 12.9 (Repräsentation eines GARCH(p, q) Prozesses) Sei ε_t ein (semi-)starker GARCH(p, q) Prozess mit $E[\varepsilon_t^4] = c < \infty$. Dann ist

1. $\eta_t = \sigma_t^2(Z_t^2 - 1)$ weißes Rauschen.
2. ε_t^2 ein ARMA(m, p) Prozess mit

$$\varepsilon_t^2 = \omega + \sum_{i=1}^m \gamma_i \varepsilon_{t-i}^2 - \sum_{j=1}^p \beta_j \eta_{t-j} + \eta_t, \quad (12.23)$$

mit $m = \max(p, q)$, $\gamma_i = \alpha_i + \beta_i$, $\alpha_i = 0$, wenn $i > q$, und $\beta_i = 0$, wenn $i > p$.

BEWEIS:

wie in Satz 12.5. □

Folgt ε_t einem GARCH Prozess, so sagt Satz 12.9, dass ε_t^2 einem ARMA Modell folgt mit bedingt heteroskedastischem Fehler η_t . Liegen z.B. alle Wurzeln des Polynoms $(1 - \beta_1 z - \dots - \beta_p z^p)$ außerhalb des Einheitskreises, so ist der ARMA Prozess (12.23) invertierbar und als AR(∞) Prozess darstellbar. Aus Satz 12.8 folgt dann, dass das GARCH(p, q) Modell als ARCH(∞) Modell repräsentiert werden kann. Man kann also in einigen Fällen analoge Schlussfolgerungen zu ARMA Modellen ziehen, z.B. bei der Bestimmung der Modellordnung (p, q). Allerdings gibt es auch wesentliche Unterschiede, z.B. die Definition der Persistenz von Schocks.

SATZ 12.10 (Unbedingte Varianz eines GARCH(p, q) Prozesses) Sei ε_t ein semi-starker GARCH(p, q) Prozess mit $\text{Var}(\varepsilon_t) = \sigma^2 < \infty$. Dann ist

$$\sigma^2 = \frac{\omega}{1 - \sum_{i=1}^q \alpha_i - \sum_{j=1}^p \beta_j},$$

mit $\sum_{i=1}^q \alpha_i + \sum_{j=1}^p \beta_j < 1$.

BEWEIS:

wie in Satz 12.2. □

Allgemeine Bedingungen für die Existenz höherer Momente von GARCH(p, q) Modellen werden in He and Teräsvirta (1998) angegeben. Für Modelle kleiner Ordnungen und unter Verteilungsannahmen lassen sich einfache Ergebnisse herleiten:

SATZ 12.11 (Viertes Moment eines GARCH(1,1) Prozesses) Sei ε_t ein (semi-)starker GARCH(1,1) Prozess mit $\text{Var}(\varepsilon_t) = \sigma^2 < \infty$ und $Z_t \sim N(0, 1)$. Dann ist $E[\varepsilon_t^4] < \infty$ genau dann, wenn $3\alpha_1^2 + 2\alpha_1\beta_1 + \beta_1^2 < 1$. Die Kurtosis $\text{Kurt}(\varepsilon_t)$ ist dann gegeben durch

$$\text{Kurt}[\varepsilon_t] = \frac{E[\varepsilon_t^4]}{(E[\varepsilon_t^2])^2} = 3 + \frac{6\alpha_1^2}{1 - \beta_1^2 - 2\alpha_1\beta_1 - 3\alpha_1^2}. \quad (12.24)$$

BEWEIS:

Folgt nach einigen Rechnungen aus $E[\varepsilon_t^4] = 3E[(\omega + \alpha_1\varepsilon_{t-1}^2 + \beta_1\sigma_{t-1}^2)^2]$ und der Stationarität von ε_t . \square

Die Funktion (12.24) ist in Abbildung 12.6 dargestellt. Für alle Werte $\alpha_1 > 0$ ist $\text{Kurt}[\varepsilon_t] > 3$, d.h., die Verteilung von ε_t ist leptokurtisch. Nur im Grenzfall $\alpha_1 = 0$ ist die Kurtosis 3, da die bedingte Heteroskedastizität verschwindet und ein Gaußsches weißes Rauschen entsteht. Man beachte, dass die Kurtosis für gegebenes α_1 nur langsam in β_1 steigt. Andererseits steigt sie sehr viel schneller für gegebenes β_1 in α_1 .

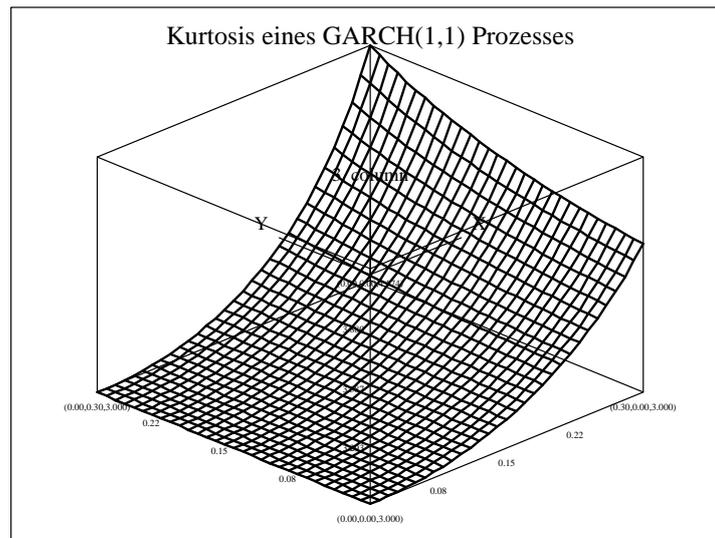


Abbildung 12.6.: Kurtosis eines GARCH(1,1) Prozesses nach (12.24). Die linke Achse zeigt den Parameter β_1 , die rechte α_1 . [SFMkurgarch.xpl](#)

BEMERKUNG 12.2 Von *Nelson (1990)* wurde gezeigt, dass der starke GARCH(1,1) Prozess X_t streng stationär ist, wenn $E[\log(\alpha_1 Z_t^2 + \beta_1)] < 0$. Ist z.B. $Z_t \sim N(0, 1)$, dann ist die Bedingung für starke Stationarität schwächer als die Bedingung für Kovarianzstationarität, $\alpha_1 + \beta_1 < 1$.

In der praktischen Anwendung hat sich häufig herausgestellt, dass Modelle kleiner Ordnung die Daten ausreichend beschrieben haben. In den meisten Fällen wurden GARCH(1,1) Modelle angepasst.

Ein wesentlicher Nachteil der Standard ARCH und GARCH Modelle ist, dass sie eine Asymmetrie der Volatilität bezüglich des Vorzeichens vergangener Schocks nicht abzubilden vermögen. Das folgt einfach daraus, dass in (12.17) und (12.21) die verzögerten Schocks nur in quadratischer Form eingehen, sie haben also einen Größeneffekt, aber keinen Vorzeicheneffekt. Schlechte Nachrichten (ausgedrückt durch negative Vorzeichen) haben denselben Einfluss auf die Volatilität wie gute Nachrichten (positive Vorzeichen), vorausgesetzt der Betrag ist derselbe. Man hat jedoch empirisch häufig beobachtet, dass schlechte Nachrichten

einen größeren Effekt auf die Volatilität haben als gute Nachrichten. In Kapitel 12.2 wollen wir auf Erweiterungen der Standardmodelle eingehen, die dieser Beobachtung Rechnung tragen.

12.1.6. Schätzung von GARCH(p, q) Modellen

Aufgrund der ARMA-Darstellung von GARCH-Prozessen (siehe Satz 12.9) kommen zunächst wieder Yule-Walker Schätzer $\tilde{\theta}$ in Betracht. Diese Schätzer sind, wie man zeigen kann, konsistent und asymptotisch normalverteilt, $\sqrt{n}(\tilde{\theta} - \theta) \xrightarrow{\mathcal{L}} N(0, \tilde{\Sigma})$. Jedoch sind sie im Falle von GARCH Modellen asymptotisch nicht effizient in dem Sinne, dass die Matrix $\tilde{\Sigma} - J^{-1}IJ^{-1}$ positiv definit ist, wobei $J^{-1}IJ^{-1}$ die asymptotische Kovarianzmatrix der QML-Schätzer ist, siehe (12.27). Über die Effizienz der Yule-Walker und QML-Schätzer in endlichen Stichproben gibt es in der Literatur einige Untersuchungen, siehe Abschnitt 12.3. Man wird aus Effizienzgründen in den meisten Fällen Maximum Likelihood-Methoden wählen.

Die Likelihoodfunktion für das allgemeine GARCH(p, q) Modell aus (12.21) ist identisch mit (12.19) mit dem erweiterten Parametervektor $\theta = (\omega, \alpha_1, \dots, \alpha_q, \beta_1, \dots, \beta_p)^T$. Abbildung 12.7 zeigt die Likelihood eines generierten GARCH(1,1) Prozesses mit $\omega = 0.1$, $\alpha = 0.1$, $\beta = 0.8$ und $n = 500$. Der Parameter ω wurde dabei so gewählt, dass die unbedingte Varianz stets gleich ist, d.h. also bei einer Varianz von σ^2 , $\omega = (1 - \alpha - \beta)\sigma^2$. Wie man sieht ist die Funktion recht flach in der Nähe des Optimums, so dass die Schätzung relativ unpräzise sein wird, d.h., mit einer großen Varianz. Zusätzlich haben wir in Abbildung 12.8 den Contourplot der Likelihoodfunktion dargestellt.

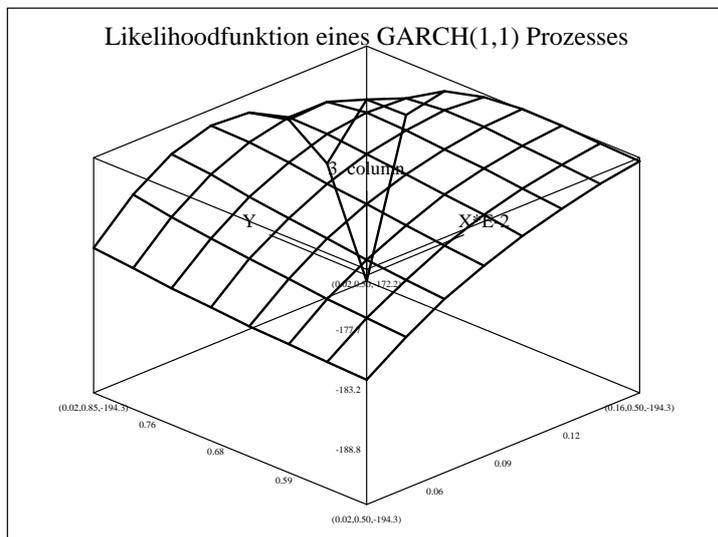


Abbildung 12.7.: Likelihoodfunktion eines generierten GARCH(1,1) Prozesses mit $n = 500$. Die linke Achse zeigt den Parameter β , die rechte α . Die wahren Parameter sind $\alpha = 0.1$ und $\beta = 0.8$. [SFMLikgarch.xpl](#)

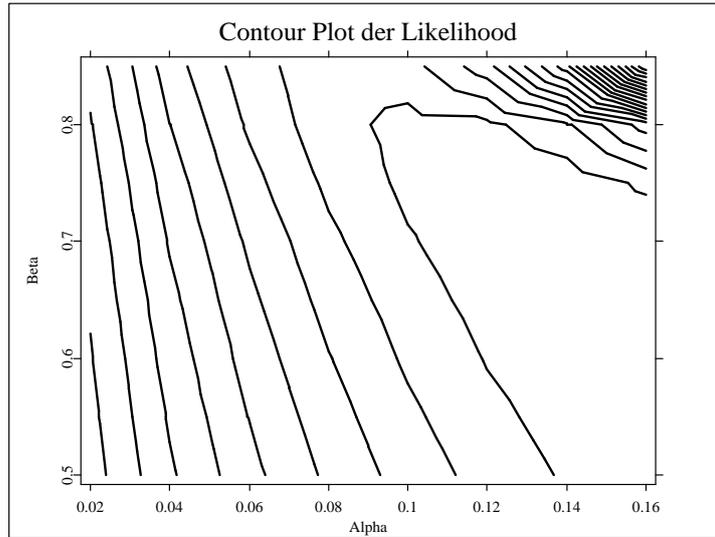


Abbildung 12.8.: Contour Plot der Likelihoodfunktion eines generierten GARCH(1,1) Prozesses mit $n = 500$. Die senkrechte Achse zeigt den Parameter β , die horizontale α . Die wahren Parameter sind $\alpha = 0.1$ und $\beta = 0.8$.

[SFMLikgarch.xpl](#)

Die ersten partiellen Ableitungen von (12.19) sind

$$\frac{\partial l_t}{\partial \theta} = \frac{1}{2\sigma_t^2} \frac{\partial \sigma_t^2}{\partial \theta} \left(\frac{\varepsilon_t^2}{\sigma_t^2} - 1 \right) \quad (12.25)$$

mit

$$\frac{\partial \sigma_t^2}{\partial \theta} = \vartheta_t + \sum_{j=1}^p \frac{\partial \sigma_{t-j}^2}{\partial \theta}.$$

und $\vartheta_t = (1, \varepsilon_{t-1}^2, \dots, \varepsilon_{t-q}^2, \sigma_{t-1}^2, \dots, \sigma_{t-p}^2)^T$. Die Bedingungen erster Ordnung sind $\sum_{t=q+1}^n \partial l_t / \partial \theta = 0$. Die Matrix der zweiten Ableitungen nimmt folgende Form an:

$$\begin{aligned} \frac{\partial^2 l_t(\theta)}{\partial \theta \partial \theta^T} &= \frac{1}{2\sigma_t^4} \frac{\partial \sigma_t^2}{\partial \theta} \frac{\partial \sigma_t^2}{\partial \theta^T} - \frac{1}{2\sigma_t^2} \frac{\partial^2 \sigma_t^2(\theta)}{\partial \theta \partial \theta^T} \\ &- \frac{\varepsilon_t^2}{\sigma_t^6} \frac{\partial \sigma_t^2}{\partial \theta} \frac{\partial \sigma_t^2}{\partial \theta^T} + \frac{\varepsilon_t^2}{2\sigma_t^4} \frac{\partial^2 \sigma_t^2(\theta)}{\partial \theta \partial \theta^T} \end{aligned} \quad (12.26)$$

Unter den Bedingungen

1. $E[Z_t | \mathcal{F}_{t-1}] = 0$ und $E[Z_t^2 | \mathcal{F}_{t-1}] = 1$,
2. strenge Stationarität von ε_t

und einigen technischen Bedingungen sind die ML-Schätzer konsistent. Gilt zusätzlich $E[Z_t^4 | \mathcal{F}_{t-1}] < \infty$, dann ist $\hat{\theta}$ asymptotisch normalverteilt:

$$\sqrt{n}(\hat{\theta} - \theta) \xrightarrow{\mathcal{L}} N_{p+q+1}(0, J^{-1} I J^{-1}) \quad (12.27)$$

mit

$$I = E \left(\frac{\partial l_t(\theta)}{\partial \theta} \frac{\partial l_t(\theta)}{\partial \theta^T} \right)$$

und

$$J = -E \left(\frac{\partial^2 l_t(\theta)}{\partial \theta \partial \theta^T} \right).$$

SATZ 12.12 (Äquivalenz von I und J) Sei $Z_t \sim N(0, 1)$. Dann gilt $I = J$.

BEWEIS:

Nach Erwartungswertbildung von (12.26) erhält man

$$E \left[\frac{\partial^2 l_t(\theta)}{\partial \theta \partial \theta^T} \right] = -E \frac{1}{2\sigma_t^4} \frac{\partial \sigma_t^2}{\partial \theta} \frac{\partial \sigma_t^2}{\partial \theta^T}.$$

Für I ergibt sich

$$E \left[\frac{\partial l_t(\theta)}{\partial \theta} \frac{\partial l_t(\theta)}{\partial \theta^T} \right] = E \left[\frac{1}{4\sigma_t^4} \frac{\partial \sigma_t^2}{\partial \theta} \frac{\partial \sigma_t^2}{\partial \theta^T} (Z_t^4 - 2Z_t^2 + 1) \right] \quad (12.28)$$

$$= E \left[\frac{1}{4\sigma_t^4} \frac{\partial \sigma_t^2}{\partial \theta} \frac{\partial \sigma_t^2}{\partial \theta^T} \right] (\text{Kurt}(Z_t | \mathcal{F}_{t-1}) - 1). \quad (12.29)$$

Aus der Annahme $Z_t \sim N(0, 1)$ folgt $\text{Kurt}(Z_t | \mathcal{F}_{t-1}) = 3$ und damit die Behauptung. \square

Hat man die Verteilung von Z_t richtig spezifiziert, so ist $I = J$ und die asymptotische Varianz vereinfacht sich zu J^{-1} , d.h., der Inversen der Fisherschen Informationsmatrix. Ist sie es nicht, sondern z.B. leptokurtisch, so kann die Maximierung von (12.10) zwar noch konsistente, aber nicht mehr effiziente Schätzungen liefern. In diesem Fall wird die ML Methode als ‘Quasi Maximum Likelihood’ (QML) Methode interpretiert.

Konsistente Schätzer für die Matrizen I und J erhält man durch Ersetzen der Erwartungswerte durch einfache Mittelwerte.

12.2. Erweiterungen der GARCH-Modelle

Standard GARCH Modelle unterstellen einen symmetrischen Effekt von positiven und negativen Fehlern auf die Volatilität. Mit anderen Worten, gute und schlechte Nachrichten haben in diesem Modell die gleiche Auswirkung auf die Volatilität. Man hat nun empirisch festgestellt, dass diese Annahme insbesondere bei Aktienrenditen häufig verletzt ist indem nach negativen Nachrichten die Volatilität stärker steigt als nach positiven. Dieser als *Leverage Effekt* in die Literatur eingegangene Effekt geht zurück auf [Black \(1976\)](#), der bemerkte, dass:

”a drop in the value of the firm will cause a negative return on its stock, and will usually increase the leverage of the stock. [...] That rise in the debt-equity ratio will surely mean a rise in the volatility of the stock”.

Negative Nachrichten implizieren also durch einen reduzierten Marktwert der Unternehmung einen höheren Verschuldungsgrad, der zu einer höheren Volatilität führt. Dies ist eine sehr einfache, aber plausible Erklärung für den Leverage Effekt, allerdings sollte das Risiko bzw. die Volatilität erst bei großen Veränderungen des Marktwertes reagieren. Man stellt jedoch empirisch meistens schon bei kleinen Veränderungen eine Erhöhung der Volatilität fest. Weiterhin sagt Black nichts über die Auswirkung positiver Renditen auf die Volatilität, die empirisch eine weniger starke Erhöhung, aber eben doch eine Erhöhung der Volatilität zur Folge haben. Auf jeden Fall besteht aus empirischer Sicht kein Zweifel daran, dass Volatilität asymmetrisch auf das Vorzeichen von Schocks reagiert und man daher im Rahmen von ARCH-Modellen nach Erweiterungen der Standardmodelle suchen muss. In der Tat sind in den letzten Jahren eine Reihe von parametrischen Erweiterungen der Standardmodelle vorgeschlagen worden, von denen wir hier zwei der wichtigsten herausstellen wollen: das exponential GARCH (EGARCH) und das threshold GARCH (TGARCH) Modell.

12.2.1. Exponential GARCH

Bezeichne Z_t weiterhin eine Folge von *u.i.v.* auf Erwartungswert 0 und Varianz 1 standardisierten Zufallsvariablen. Das allgemeine exponentielle GARCH (EGARCH) Modell von Nelson (1991) ist gegeben durch

$$\log \sigma_t^2 = \omega_t + \sum_{k=1}^{\infty} \beta_k g(Z_{t-k}), \quad (12.30)$$

wobei ω_t, β_k deterministische Koeffizienten sind und

$$g(Z_t) = \theta Z_t + \gamma(|Z_t| - E|Z_t|). \quad (12.31)$$

Es ist einfach zu sehen, dass $E[g(Z_t)] = 0$.

Das EGARCH Modell in (12.30) zeigt einige Unterschiede zum Standard GARCH Modell:

- Volatilität, gemessen durch die bedingte Varianz σ_t^2 , ist eine explizite multiplikative Funktion verzögerter Innovationen. Bei Standard GARCH Modellen ist andererseits die Volatilität eine additive Funktion in den verzögerten Fehlern ε_t , was eine komplizierte funktionale Abhängigkeit von den Innovationen bewirkt, siehe schon für den einfachen ARCH(1) Fall die Repräsentation in Satz 12.4.
- Volatilität kann asymmetrisch auf gute und schlechte Nachrichten reagieren.
- Für allgemeine Verteilungen von Z_t fallen die Parameterrestriktionen für strenge und Kovarianzstationarität zusammen.
- Die Parameter in (12.30) and (12.31) sind nicht auf positive Werte restringiert.

Die Funktion $g(\cdot)$ in (12.31) ist stückweise linear. Sie enthält zwei Parameter die den ‘Größeneffekt’ und ‘Vorzeicheneffekt’ von Schocks auf Volatilität bestimmen. Der erstere ist der typische ARCH Effekt, der zweite der Asymmetrieffekt, z.B. der *Leverage Effekt*. Der

Term $\gamma(|Z_t| - E|Z_t|)$ bestimmt den Größeneffekt, der Term θZ_t hingegen den Vorzeicheneffekt. Der Parameter γ ist somit typischerweise positiv, θ negativ.

Zur Schätzung von EGARCH Modellen wird anstelle der allgemeinen MA(∞)-Darstellung ein (12.30) ARMA(p, q) Modell angenommen, d.h.

$$\Delta(L) \log \sigma_t^2 = \omega + \Psi(L)g(Z_t), \quad (12.32)$$

mit Lag-Polynomen $\Delta(L)$ und $\Psi(L)$ der Ordnung p beziehungsweise q .

Ein allgemeiner Vorteil von EGARCH Modellen ist, dass keine Parameterrestriktionen existieren und somit mögliche Instabilitäten von Optimierungsroutinen vermieden werden. Auf der anderen Seite sind die theoretischen Eigenschaften von QML Schätzern von EGARCH Modellen weitgehend ungeklärt.

Sei $\omega_t = \omega = 0$ und $\sum_{k=1}^{\infty} \beta_k^2 < \infty$. Dann ist σ_t^2 streng stationär und ergodisch (Nelson (1991), Theorem 2.1). Weiterhin existiert unter diesen Bedingungen die unbedingte Varianz falls Z_t eine *generalisierte Fehlerverteilung* (GED) hat mit Parameter $\zeta > 1$, der die Dicke der Flanken bestimmt, besitzt. (Nelson (1991), Theorem 2.2); Die GED ist leptokurtisch, falls $\zeta < 2$. Die Normalverteilung ist ein Spezialfall der GED ($\zeta = 2$). Nelson gibt ferner recht komplizierte Formeln für die unbedingten Momente an. Ein Problem ist, dass unter anderen leptokurtischen Verteilungen wie der Student-t die unbedingte Varianz nicht existiert. Der Grund ist das exponentielle Wachstum der bedingten Varianz mit der Größe der Schocks, was dazu führt, dass die unbedingte Varianz explodiert wenn die Wahrscheinlichkeit für extreme Schocks hinreichend groß ist. Die Existenz der unbedingten Momente hängt somit entscheidend von der Verteilung der Innovationen ab, was eine eher ungewünschte Eigenschaft der EGARCH Modelle ist. In empirischen Studien hat man zudem häufig festgestellt, dass EGARCH den Effekt großer Schocks auf die Volatilität übergewichtet und daher schlechtere Anpassungen resultieren als bei Standard GARCH Modellen, siehe z.B. die empirische Studie von Engle and Ng (1993).

12.2.2. Threshold ARCH Modelle

Die Idee von Threshold ARCH (TARCH) Modellen ist, den Träger der Verteilung der Innovationen in disjunkte Intervalle zu unterteilen und eine stückweise lineare Funktion für die bedingte Standardabweichung (Zakoian (1991)) beziehungsweise bedingte Varianz (Glosten, Jagannathan and Runkle (1993)) anzupassen. Falls es nur zwei Intervalle gibt wird die Trennung normalerweise bei Null sein, d.h., man unterscheidet den Einfluss positiver und negativer Innovationen auf die Volatilität. In diesem Fall kann das TARCH Modell der Ordnung q geschrieben werden als

$$\sigma_t^\delta = \omega + \sum_{i=1}^q \alpha_i \varepsilon_{t-i}^\delta + \sum_{i=1}^q \alpha_i^- \varepsilon_{t-i}^\delta I(\varepsilon_{t-i} < 0), \quad (12.33)$$

mit der Indikatorfunktion $I(\cdot)$ und $\delta = 1$ wie bei Zakoian (1991) oder $\delta = 2$ wie bei Glosten et al. (1993).

Rabemananjara and Zakoian (1993) erweiterten dieses Modell durch Hinzufügen der verzögerten bedingten Standardabweichung (beziehungsweise Varianz) als Regressor, was

als TGARCH Modell bezeichnet werden kann. Sie geben auch Bedingungen für die Kovarianzstationarität.

Anstelle einer stückweise linearen Funktion nehmen [Gouriéroux and Monfort \(1992\)](#) eine Treppenfunktion (stückweise konstant) als Modell für die Volatilität. Sei $A_j, j = 1, \dots, J$ eine Partition des Trägers der Verteilung der Innovationen. Dann ist das *qualitative threshold ARCH* Modell (QTARCH) der Ordnung 1

$$y_t = \sum_{j=1}^J m_j I(y_{t-1} \in A_j) + \sum_{j=1}^J s_j I(y_{t-1} \in A_j) Z_t, \quad (12.34)$$

wobei m_j und s_j Skalare sind. In (??) sind bedingter Erwartungswert und bedingte Standardabweichung als Treppenfunktion modelliert. Man bemerke, dass (??) eine homogene Markoff-Kette der Ordnung Eins ist. Modelle höherer Ordnung lassen sich einfach herleiten. [Gouriéroux and Monfort \(1992\)](#) geben eine ausführliche Diskussion der statistischen Eigenschaften von (??) und zeigen die Konsistenz und asymptotische Normalverteilung von QML Schätzern.

Die Threshold Modelle sind gekennzeichnet durch einen abrupten Übergang zwischen zwei Regimes, falls die treibende Innovation einen Schwellenwert überschreitet. Zieht man bei der Modellierung einen glatteren Übergang vor, so kann man die Indikatorfunktion in (12.33) durch eine beliebige stetige Funktion ersetzen, die für kleine Werte von ε_{t-i} gegen Null geht und für große Werte gegen Eins. Eine häufige Wahl ist z.B. die logistische Funktion. Das Modell ist also

$$\sigma_t^\delta = \omega + \sum_{i=1}^q \alpha_i \varepsilon_{t-i}^\delta + \sum_{i=1}^q \alpha_i^- \varepsilon_{t-i}^\delta F_\gamma(\varepsilon_{t-i}) \quad (12.35)$$

$$F_\gamma(u) = \frac{1}{1 + \exp(-\gamma u)}, \quad (12.36)$$

mit dem zu schätzenden Parameter $\gamma > 0$. Die Indikatorfunktion ist ein Grenzfall der logistischen Funktion für $\gamma \rightarrow \infty$.

Schließlich sei auf eine weitere Modellklasse verwiesen, die recht allgemein ist und ebenfalls Asymmetrie erzeugen kann: das asymmetrische Power ARCH (APARCH) Modell von [Ding, Granger and Engle \(1993\)](#),

$$\sigma_t^\delta = \omega + \sum_{i=1}^q \alpha_i (|\varepsilon_{t-i}| - \gamma_i \varepsilon_{t-i})^\delta + \sum_{j=1}^p \beta_j \sigma_{t-j}^\delta, \quad (12.37)$$

wobei $\delta > 0$ ein zu schätzender Parameter ist. Ein Nachteil ist allerdings, dass das EGARCH Modell nicht eingeschlossen ist, so dass ein direkter Test zwischen GARCH und EGARCH nicht möglich ist. Ein sehr allgemeines ARCH Modell, das *augmented GARCH* Modell von [Duan \(1997\)](#), umfasst auch das EGARCH Modell.

12.2.3. Risiko und Rendite

In der Finanztheorie spielt die Beziehung zwischen Risiko und Rendite eine große Rolle. Viele theoretische Modelle wie z.B. das CAPM implizieren eine lineare Beziehung zwischen

der erwarteten Rendite des Marktportefeuilles und der Varianz. Ist das Risiko (die Varianz) nicht über die Zeit konstant, so sagt die Theorie, dass der bedingte Erwartungswert der Marktrendite eine lineare Funktion der bedingten Varianz ist. Die Idee von [Engle, Lilien and Robins \(1987\)](#) war folglich, die bedingte Varianz mit GARCH zu schätzen und sie in den bedingten Erwartungswert einzubeziehen. Dies ist das sogenannte *ARCH in mean* (ARCH-M) Modell.

Sei y_t ein kovarianzstationärer Renditeprozess eines breiten Marktindex und σ_t^2 seine bedingte Varianz, spezifiziert als ein ARCH(1) Modell. Das ARCH-M Modell ist

$$\begin{aligned} y_t &= \lambda g(\sigma_t^2) + \varepsilon_t \\ \sigma_t^2 &= \omega + \alpha \varepsilon_{t-1}^2, \end{aligned} \tag{12.38}$$

wobei $g(\cdot)$ eine bekannte parametrische Funktion ist. Das CAPM impliziert $g(x) = x$, in den meisten empirischen Anwendungen wurde jedoch die Wurzel oder der Logarithmus vorgezogen. Der Ausdruck $\lambda g(\sigma_t^2)$ kann als Risikoprämie interpretiert werden.

12.2.4. Schätzergebnisse für DAX-Renditen

Wir haben die verschiedenen besprochenen ARCH-Modelle auf die Renditen des Deutschen Aktien Index (DAX) angewendet. Genauer haben wir nicht den börslich notierten DAX verwendet, sondern den DAFOX der Karlsruher Kapitalmarktdatenbank, der zu Forschungszwecken erstellt wird und z.B. auszahlungs- und nominalwertbereinigt ist. Wir betrachten tägliche Renditen von 1974 bis 1996 (5748 Beobachtungen).

Die Renditen weisen eine deutliche Autokorrelation erster Ordnung auf. Es gibt mehrere Möglichkeiten, diese Autokorrelation zu modellieren. Die beiden wichtigsten Modelle, die wir hier betrachten, sind autoregressive Modelle und ARCH-M Modelle. Letztere sind ökonomisch leicht interpretierbar, d.h., eine zeitveränderliche Risikoprämie impliziert eine Autokorrelation der Renditen. Diese ökonomische Motivation fehlt bei den autoregressiven Modellen: Der AR-Term lässt sich nicht als Risikoprämie interpretieren, da er negativ werden kann, was unter der üblichen Annahme einer Risikoaversion der Marktteilnehmer nicht vorkommen kann. Jedoch bieten autoregressive Modelle häufig eine bessere Anpassung an die Daten als ARCH-M Modelle. Das Grundmodell ist also

$$y_t = \mu_t + \sigma_t Z_t$$

mit $\mu_t = \phi y_{t-1}$ (AR(1)) beziehungsweise $\mu_t = \lambda \sigma_t$ (ARCH-M), und σ_t GARCH, TGARCH oder EGARCH.

Die Schätzergebnisse sind in Tabelle [12.3](#) zusammengestellt. Sie lassen sich wie folgt zusammenfassen:

- Der Parameter ν ist bei allen AR-Modellen signifikant positiv, was den langfristigen Aufwärtstrend der Aktienkurse widerspiegelt. Bei den ARCH-M Modellen ist das Vorzeichen des Trends nicht allein durch ν , sondern auch durch λ gegeben. Der Effekt eines negativen ν kann durch ein positives λ überkompensiert werden, wie hier der Fall bei dem GARCH-M und dem EGARCH-M Modell.
- Die ARCH-Effekte sind sehr deutlich, d.h., der Parameter α zum Beispiel im GARCH Modell ist signifikant.

	AR-G	AR-TG	AR-EG	GARCH-M	TGARCH-M	EGARCH-M
λ				0.107 (2.14)	0.051 (1.09)	0.338 (3.31)
ν	3.75E-04 (4.414)	3.03E-04 (3.290)	2.57E-04 (2.608)	-2.84E-04 (-0.79)	4.71E-05 (0.14)	-0.002 (-2.72)
ϕ	0.131 (6.808)	0.137 (8.079)	0.136 (5.940)			
ω	8.42E-07 (5.312)	8.22E-07 (5.444)	-5.259 (-3.692)	8.81E-07 (5.28)	8.78E-07 (5.45)	-3.086 (-4.79)
α	0.079 (3.004)	0.045 (2.073)		0.076 (2.74)	0.049 (1.94)	
α^-		0.058 (1.752)			0.044 (1.77)	
β	0.914 (44.07)	0.918 (38.18)	0.490 (3.41)	0.917 (41.43)	0.919 (36.57)	0.711 (10.87)
γ			0.431 (13.95)			0.388 (11.25)
θ			-0.171 (-3.65)			-0.085 (-1.97)
logL	20030.86	20049.18	19713.81	19996.19	20008.22	19827.41

Tabelle 12.3.: *Schätzergebnisse verschiedener ARCH Modelle, angewendet auf DAFOX-Renditen 1974–1996. In Klammern stehen die t-Statistiken basierend auf QML asymptotischen Standardfehlern.*

- Es besteht eine hohe Persistenz von Schocks in der Volatilität. Diese Persistenz wird im GARCH Fall durch die Summe von α und β gemessen und ist jeweils nahe bei 1.
- Außer für die EGARCH Spezifikation der Volatilität beschreiben die AR(1) Modelle besser die Daten als die ARCH-M Modelle.
- Es ist ein Leverage Effekt vorhanden: Die entsprechenden Parameter in den TGARCH und EGARCH Modellen haben das entsprechende Vorzeichen. Im TGARCH Fall ist die t-Statistik für α^- 1.75, bzw. 1.77, im EGARCH Fall ist die t-Statistik für θ -3.65 bzw. -1.97. Negative Schocks steigern die Volatilität mehr als positive Schocks.
- Unter den asymmetrischen Modellen ist das TGARCH Modell dem EGARCH Modell vorzuziehen, da eine bessere Anpassung an die Daten vorliegt bei gleicher Parameteranzahl.

12.3. Literaturhinweise

Die empirischen Eigenschaften von Finanzmarktdaten, insbesondere die Leptokurtosis und das Clustering von Volatilitäten, wurden bereits systematisch untersucht von [Mandelbrot](#)

(1963) und Fama (1965). ARCH Modelle wurden von Engle (1982) eingeführt und von Bollerslev (1986) als GARCH verallgemeinert. Für einen exzellenten Überblick über ARCH Modelle konsultiere man die Monographie von Gouriéroux (1997). Eine extensive Diskussion der Schätzung des Flanken-Exponenten und des Schätzers von Hill (1975) ist zu finden in der Monographie von Embrechts, Klüppelberg and Mikosch (1997). Ein Effizienzvergleich des Yule-Walker Schätzers mit ML Schätzern von ARCH- und GARCH Modellen ist z.B. in der Dissertation von Maercker (1997) gegeben. Die Theorie der QML-Schätzung wird in White (1982) und Gouriéroux, Monfort and Trognon (1984) entwickelt. Weiss (1986) wandte die Theorie auf ARCH-Modelle, Lee and Hansen (1994) und Lumsdaine (1996) auf GARCH Modelle an. Bollerslev and Wooldridge (1992) betrachten QML Schätzung für allgemein bedingt heteroskedastische Modelle. Die Idee der *Smooth Transition* Spezifikation stammt von Teräsvirta (1994), der sie für AR Modelle verwendete. Auf GARCH Modelle wurden sie angewendet von Lubrano (1998). Hafner and Herwartz (1999) diskutieren verschiedene Verfahren zum Testen auf Signifikanz des AR Parameters unter bedingter Heteroskedastizität. Weiterhin vergleichen sie die empirischen Ergebnisse von AR Modellen versus ARCH-M Modellen, angewendet auf verschiedene deutsche Aktienkurse.

13. Nichtparametrische Konzepte für Finanzzeitreihen

Bei der Analyse von (Finanz-)Zeitreihen besteht eines der wichtigsten Ziele in der Erstellung von Prognosen. Anhand von in der Vergangenheit beobachteten Daten möchte man daher Aussagen über den zukünftigen Mittelwert, die zukünftige Volatilität etc. gewinnen, d.h. man möchte die unter der Vergangenheit bedingten Erwartungswerte und Varianzen des zugrundeliegenden Prozesses schätzen. Ein Verfahren zur Bestimmung solcher Schätzwerte werden wir in diesem Kapitel vorstellen.

Es sei (Y_t) , $t = 0, 1, 2, \dots$, eine Zeitreihe. Wir betrachten ein nichtlineares autoregressives heteroskedastisches Modell der Form

$$Y_t = f(Y_{t-1}) + s(Y_{t-1})\xi_t, \quad t = 1, 2, \dots \quad (13.1)$$

Dabei sind die Innovationen ξ_t u.i.v. Zufallsvariablen mit $E[\xi_t] = 0$ und $E[\xi_t^2] = 1$, $f : \mathbb{R} \rightarrow \mathbb{R}$ und $s : \mathbb{R} \rightarrow (0, \infty)$ sind unbekannte Funktionen, und Y_0 ist unabhängig von (ξ_t) . Unter diesen Voraussetzungen gilt wegen Satz 3.1

$$E[Y_t | Y_{t-1} = x] = f(x) + E[s(Y_{t-1})\xi_t | Y_{t-1} = x] = f(x) + s(x)E[\xi_t] = f(x),$$

wobei bei der vorletzten Gleichheit die Unabhängigkeit von ξ_t und Y_{t-1} verwendet wurde. Eine ähnliche Rechnung ergibt $s^2(x) = \text{Var}[Y_t | Y_{t-1} = x]$. Die unbekannt Funktionen f und s beschreiben also den bedingten Mittelwert und die bedingte Volatilität des Prozesses, die wir schätzen wollen.

Mit der speziellen Wahl $f(x) = \alpha x$ und $s \equiv \sigma > 0$ ist der Prozess Y_t ein AR(1)-Prozess. Auch jeder starke ARCH(1)-Prozess (Y_i) genügt dem Modell (13.1). In diesem Fall ist $f \equiv 0$, und es gilt $s(x) = \sqrt{\omega + \alpha x^2}$ mit den Parametern $\omega > 0$ und $\alpha \geq 0$, vgl. Definition 12.1. Bezüglich der an die Gestalt des bedingten Mittelwertes und der bedingten Varianz gestellten Bedingungen stellt also das obige Modell eine weitreichende Verallgemeinerung zu den (starken) ARCH-Modellen dar. Der Vorteil dieses *nichtparametrischen* Ansatzes besteht darin, dass das Modell keine strukturellen Annahmen über die Funktionen f und s beinhaltet, da sich solcherlei Annahmen durch beobachtete Daten oft nicht bestätigen lassen.

In einem engen Zusammenhang zu unserem Modell stehen die von [Gouriéroux and Monfort \(1992\)](#) untersuchten Qualitative Threshold ARCH-Modelle (QTARCH), die für den Fall genau eines Lags (QTARCH(1)) ein Spezialfall von (13.1) ist, wo die unbekannt Funktionen $f : \mathbb{R} \rightarrow \mathbb{R}$ und $s : \mathbb{R} \rightarrow (0, \infty)$ die Gestalt von Treppenfunktionen besitzen - s. (12.34). Andererseits kann man (13.1) unter gewissen Regularitätsforderungen an f und s auch als Grenzmodell des QTARCH(1)-Modells für $J \rightarrow \infty$ auffassen, indem man f und s durch Elementarfunktionen approximiert.

Die Arbeit von Gouriéroux und Monfort ist die erste, in welcher der bedingte Mittelwert und die bedingte Varianz *gemeinsam* in einem nichtparametrischen Modell betrachtet werden. Diese Idee wird mit dem hier vorgestellten Ansatz von Härdle and Tsybakov (1997) aufgegriffen, der unabhängig auch von ? betrachtet worden ist. Im folgenden werden wir unter der Modellannahme (13.1) eine Klasse von Schätzern für die bedingte Volatilität $v(x) = s^2(x)$ und den bedingten Mittelwert $f(x)$ der Zeitreihe (Y_i) konstruieren, die auf lokal polynomialer Regression basieren. Zusätzlich zur Modellannahme (13.1) werden dazu gewisse Regularitätsforderungen, jedoch keine Strukturforderungen, an f und s gestellt. Als Hauptergebnis dieses Kapitels werden wir beweisen, dass diese gemeinsame Schätzung des bedingten Erwartungswertes und der bedingten Varianz asymptotisch normalverteilt ist.

13.1. Nichtparametrische Regression

In diesem Abschnitt führen wir einige grundlegende Begriffe und Ideen aus der Theorie der nichtparametrischen Regression ein und erklären insbesondere die Methode der lokal polynomialen Regression. Anschließend erläutern wir, wie sich diese auf (Finanz-)Zeitreihen übertragen lässt. Eine ausführliche Darstellung findet sich in [Härdle, Müller, Sperlich and Werwatz \(2001\)](#).

Bei der nichtparametrischen Regression interessiert man sich für den (funktionalen) Zusammenhang zwischen einer erklärenden Variable X und einer abhängigen Variable Y , d.h. man möchte Schätzungen für die unbekannte Funktion $m(x) = E[Y | X = x]$ erhalten. Dabei werden – im Gegensatz zur parametrischen Statistik – keine speziellen Annahmen über die Gestalt der Funktion m gemacht. Es werden lediglich gewisse Regularitäts- bzw. Glattheitsforderungen an m gestellt.

Ein Verfahren zur Schätzung von m ist die Methode der lokal polynomialen Regression (LP-Methode). Die Idee beruht darauf, dass sich die Funktion m unter geeigneten Voraussetzungen lokal durch ihr Taylorpolynom approximieren lässt, d.h. in einer Umgebung eines vorgegebenen Punktes x_0 gilt

$$m(x) \approx \sum_{k=0}^p \frac{m^{(k)}(x_0)}{k!} (x - x_0)^k. \quad (13.2)$$

Um eine Schätzung für m im Punkt x_0 zu gewinnen, versucht man daher, auf der Basis von Beobachtungen $(X_1, Y_1), \dots, (X_n, Y_n)$ ein Polynom zu bestimmen, das m in einer Umgebung von x_0 möglichst gut approximiert. Als Maß für die Qualität der Approximation wählt man üblicherweise den quadratischen Abstand an den Beobachtungspunkten, d.h. man möchte den Ausdruck

$$\sum_{i=1}^n \left(Y_i - \sum_{j=0}^p \beta_j (X_i - x_0)^j \right)^2 \quad (13.3)$$

bezüglich $\beta = (\beta_0, \dots, \beta_p)$ minimieren. Da jedoch die Darstellung (13.2) nur lokal gültig ist, hat man noch zu berücksichtigen, dass einige der beobachteten X_i möglicherweise nicht nahe genug bei x_0 liegen und (13.2) deswegen auf diese nicht mehr zutrifft. Man muss also die Beobachtungen geeignet *lokalisieren*, d.h. nur solche Beobachtungen berücksichtigen, die nahe genug bei x_0 liegen.

Ein in der nichtparametrischen Regression gängiges Verfahren zur Lokalisierung besteht darin, die Daten mit Hilfe eines Kerns zu gewichten. Unter einem Kern versteht man eine Funktion $K : \mathbb{R} \rightarrow [0, \infty)$ mit $\int K(u) du = 1$. Die meisten gebräuchlichen Kerne sind zusätzlich symmetrisch und verschwinden außerhalb eines um den Nullpunkt gelegenen Intervalls.

Ist K ein Kern und $h > 0$, so ist durch

$$K_h(u) = \frac{1}{h} K\left(\frac{u}{h}\right)$$

der mit der *Bandbreite* h reskalierte Kern K_h definiert, der sich wieder zu 1 integriert. Wenn z.B. der Ausgangskern K außerhalb des Intervalls $[-1, 1]$ verschwindet, so ist K_h außerhalb

des Intervalls $[-h, h]$ Null. Indem man in (13.3) die i -te Abweichung mit $K_h(x - X_i)$ gewichtet, gelangt man so zu einem Minimierungsproblem, das sich aufgrund der nun vorgenommenen Lokalisierung unabhängig vom Punkt x_0 formulieren lässt. Der Koeffizientenvektor $\hat{\beta} = \hat{\beta}(x) = (\hat{\beta}_0(x), \dots, \hat{\beta}_p(x))$ des zum Punkt x zu bestimmenden Polynoms ist also gegeben durch

$$\hat{\beta} = \arg \min_{\beta} \sum_{i=1}^n \left(Y_i - \sum_{j=0}^p \beta_j (x - X_i)^j \right)^2 K_h(x - X_i). \quad (13.4)$$

Es ist offensichtlich, dass $\hat{\beta}$ wesentlich von der Wahl des Kerns und der Bandbreite abhängt. Verschiedene Verfahren zur Bestimmung von K und h werden in Härdle et al. (2001) vorgestellt.

Mit den Bezeichnungen

$$\mathbf{X} = \begin{pmatrix} 1 & X_1 - x & (X_1 - x)^2 & \dots & (X_1 - x)^p \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & X_n - x & (X_n - x)^2 & \dots & (X_n - x)^p \end{pmatrix}, \mathbf{Y} = \begin{pmatrix} Y_1 \\ \vdots \\ Y_n \end{pmatrix},$$

$$\mathbf{W} = \begin{pmatrix} K_h(x - X_1) & & & & 0 \\ & & & \ddots & \\ & & 0 & & K_h(x - X_n) \end{pmatrix}$$

lässt sich die Lösung $\hat{\beta}$ des gewichteten Kleinst-Quadrate-Problems (13.4) explizit schreiben als

$$\hat{\beta}(x) = (\mathbf{X}^T \mathbf{W} \mathbf{X})^{-1} \mathbf{X}^T \mathbf{W} \mathbf{Y} \quad (13.5)$$

Die Schätzung $\hat{m}(x)$ für $m(x)$ erhält man nun, indem man das approximierende Polynom im Punkt x auswertet:

$$\hat{m}(x) = \hat{\beta}_0(x). \quad (13.6)$$

Die restlichen Komponenten von $\hat{\beta}(x)$ liefern wegen der Gleichungen (13.2) und (13.3) Schätzer für die Ableitungen von m : $\hat{m}^{(j)}(x) = j! \hat{\beta}_j(x)$, $j = 1, \dots, p$, mit denen wir uns aber hier nicht weiter beschäftigen. Im Spezialfall $p = 0$ ist $\hat{m}(x)$ ein gewöhnlicher Kernschätzer vom Nadaraya-Watson-Typ - s. Härdle (1990).

Die soeben hergeleitete Methode der lokal-polynomialen Approximation oder kurz LP-Methode wollen wir nun auf eine Zeitreihe (Y_i) anwenden. Wie zuvor schon erwähnt, interessiert man sich hier vor allem für die Erstellung von Prognosen.

Für den einfachsten Fall einer Ein-Schritt-Prognose bedeutet dies, dass wir den funktionalen Zusammenhang zwischen Y_{i-1} und einer Funktion $\lambda(Y_i)$ von Y_i untersuchen wollen, d.h. wir möchten eine Schätzung für die unbekannte Funktion

$$m(x) = E[\lambda(Y_i) | Y_{i-1} = x]$$

erhalten. Um die oben erklärte LP-Methode anwenden zu können, fassen wir eine vorgegebene Stichprobe Y_0, \dots, Y_n als Beobachtungen der Form $(Y_0, Y_1), \dots, (Y_{n-1}, Y_n)$ auf. Damit diese Beobachtungen identisch verteilt sind und insbesondere die Funktion m unabhängig vom Zeitindex i ist, muss der Prozess (Y_i) geeignete Bedingungen erfüllen. Dies ist z.B. stets der Fall, wenn (Y_i) stationär ist. Indem wir also in (13.4) $X_i = Y_{i-1}$ setzen und Y_i durch $\lambda(Y_i)$ ersetzen, erhalten wir in dieser Situation

$$\hat{\beta} = \arg \min_{\beta} \sum_{i=1}^n \left(\lambda(Y_i) - \sum_{j=0}^p \beta_j (x - Y_{i-1})^j \right)^2 K_h(x - Y_{i-1}), \quad (13.7)$$

und die Schätzung für $m(x)$ ist wieder durch $\hat{\beta}_0(x)$ gegeben.

13.2. Konstruktion des Schätzers

Die im letzten Abschnitt vorgestellte LP-Methode werden wir nun unter der Annahme eines nichtparametrischen, autoregressiven Modells gemäß (13.1) zur Schätzung der Volatilitätsfunktion $s(x)$ des Prozesses (Y_i) auf der Basis von Beobachtungen Y_0, \dots, Y_n verwenden.

Die bedingte Volatilität $s_i(x)$ bzw. die bedingte Varianz $v_i(x)$ zum Zeitpunkt i sind definiert durch

$$v_i(x) = s_i^2(x) = \mathbb{E}[Y_i^2 | Y_{i-1} = x] - \mathbb{E}^2[Y_i | Y_{i-1} = x]. \quad (13.8)$$

Unter der Annahme des Modells (13.1) sind diese unabhängig vom Zeitindex i . Eine Schätzung für $v(x) = s^2(x)$ durch die LP-Methode beruht jedoch darauf, dass die zweidimensionalen Randverteilungen (Y_{i-1}, Y_i) unabhängig von i sind. Im folgenden werden wir sehen, dass sich (Y_i) einem stationären Prozess annähert, wodurch der folgende Ansatz gerechtfertigt ist: In Anlehnung an die Darstellung (13.8) der bedingten Varianz $v(x)$ suchen wir einen Schätzer \hat{v}_n für v der Gestalt

$$\hat{v}_n(x) = \hat{g}_n(x) - \hat{f}_n^2(x), \quad (13.9)$$

d.h. wir suchen einen Schätzer $\hat{g}_n(x)$ für $g(x) = f^2(x) + s^2(x)$ und einen Schätzer $\hat{f}_n(x)$ für $f(x)$.

Um diese beiden Schätzer durch die LP-Methode zu definieren, haben wir nach den Ausführungen des letzten Abschnitts die beiden folgenden Minimierungsprobleme zu lösen:

$$\begin{aligned} \bar{c}_n(x) &= \arg \min_{c \in \mathbb{R}^{p+1}} \sum_{i=1}^n (Y_i^2 - c^T U_{in})^2 K \left(\frac{Y_{i-1} - x}{h_n} \right), \\ c_n(x) &= \arg \min_{c \in \mathbb{R}^{p+1}} \sum_{i=1}^n (Y_i - c^T U_{in})^2 K \left(\frac{Y_{i-1} - x}{h_n} \right). \end{aligned} \quad (13.10)$$

Hierbei ist $K : \mathbb{R} \rightarrow \mathbb{R}$ ein Kern und $\{h_n\}$ eine Folge positiver Zahlen (Bandbreiten) mit $\lim_{n \rightarrow \infty} h_n = 0$. Die in (13.10) auftretenden Vektoren U_{in} sind definiert durch

$$U_{in} = F(u_{in}), \quad u_{in} = \frac{Y_{i-1} - x}{h_n}. \quad (13.11)$$

mit der durch

$$F_k(u) = \frac{u^k}{k!}$$

gegebenen \mathbb{R}^{p+1} -wertigen Funktion $F(u) = (F_0(u), \dots, F_p(u))^T$. Nach der LP-Methode haben wir \hat{g}_n und \hat{f}_n zu definieren durch

$$\hat{g}_n(x) = \bar{c}_n(x)^T F(0) \quad \text{und} \quad \hat{f}_n(x) = c_n(x)^T F(0),$$

was unserem obigen Ansatz entsprechend

$$\hat{v}_n(x) = \bar{c}_n(x)^T F(0) - (c_n(x)^T F(0))^2 \tag{13.12}$$

zur Folge hat.

Dieser Schätzer ist eine unmittelbare Modifikation des durch lokalpolynomiale, nichtparametrische Regression erklärten Schätzers in [Tsybakov \(1986\)](#).

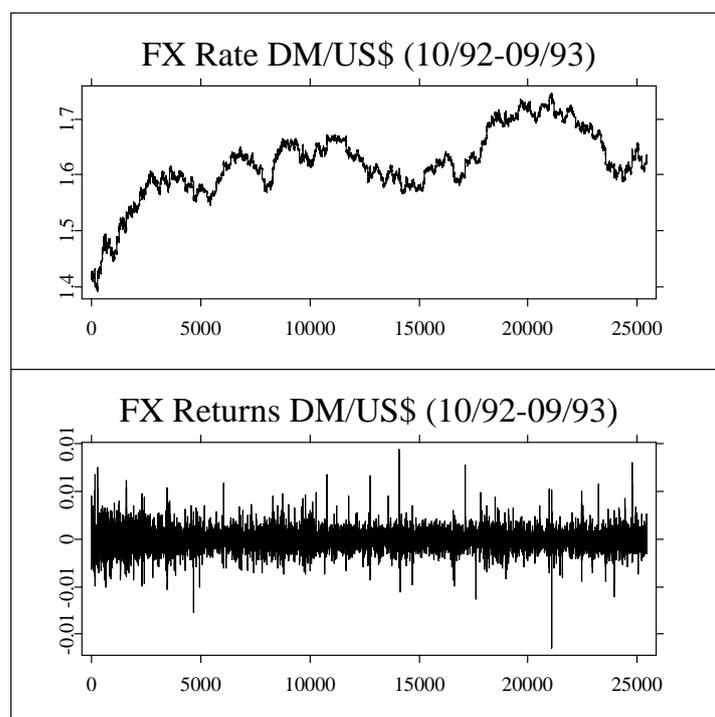


Abbildung 13.1.: DM/US\$-Wechselkurs und dessen Renditen [SFmdmusrate.xpl](#)

Zur Illustration des Schätzers betrachten wir zunächst ein Beispiel. Abbildung 13.1 oben zeigt den DM/US\$-Wechselkurs vom 1. Oktober 1992 bis zum 30. September 1993 in 20-Minuten-Intervallen. Es gibt $n = 25144$ Beobachtungen. Wir haben die Renditen dieser Reihe berechnet (siehe Abbildung 13.1 unten) und den Schätzer (13.12) auf die Zeitreihe der Renditen angewendet. Unter dem Modell der geometrischen Brownschen Bewegung für den Kurs, das dem Black-Scholes Ansatz zugrundeliegt (s. Abschnitt 6.1), müssten die Renditen

eine gewöhnliche Brownsche Bewegung sein. Ihre Volatilität und damit auch $v(x)$ wären somit konstant und unabhängig von x . Die in Abbildung 13.2 gezeigte geschätzte bedingte Varianzfunktion weist jedoch eine U-förmige Struktur auf, auch "smiling face"-Struktur oder *Smile* genannt. Die geschätzte Volatilitätsfunktion $\hat{s}(x) = \sqrt{\hat{v}(x)}$ sieht qualitativ analog aus. Dies bedeutet, dass die erwarteten Risiken der Renditen deutlich höher sind, wenn am Vortag extreme Werte registriert wurden.

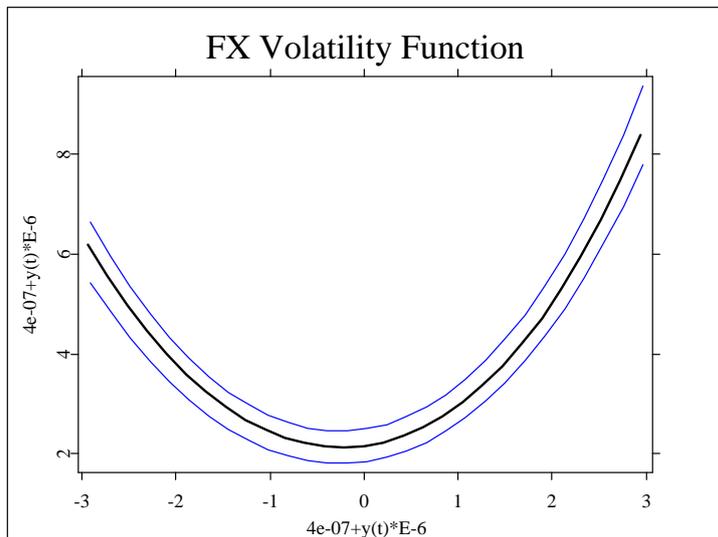


Abbildung 13.2.: Die geschätzte bedingte Varianzfunktion $\hat{v}(x)$ der DM/US\$-Renditen [SFMdmusvol.xpl](#)

Als Alternative zu Gleichung 13.9 ist es auch möglich, erst die Stichprobenresiduen

$$\hat{Z}_i = Y_i - \hat{f}_n(Y_{i-1}), \quad i = 1, \dots, n,$$

zu bestimmen. Sie approximieren die wahren Residuen $Z_i = Y_i - f(Y_{i-1})$, die unter der Annahme von Modell (13.1)

$$E[Z_i | Y_{i-1} = x] = 0, \quad E[Z_i^2 | Y_{i-1} = x] = v(x)$$

erfüllen. Die Volatilität lässt sich dann wie im vorherigen Abschnitt direkt durch die nichtparametrische Regression von \hat{Z}_i^2 auf Y_{i-1} schätzen. ? haben gezeigt, dass diese Vorgehensweise bei heteroskedastischen Regressionsmodellen Vorteile hat. Im Zeitreihenfall erlaubt sie zumindestens, für das Schätzen von $f(x)$ und $v(x) = s^2(x)$ unterschiedliche Bandbreiten zu benutzen, ohne Gefahr zu laufen, dass der Varianzschätzer negative Werte annimmt. Dies kann sinnvoll sein, wenn zu erwarten ist, dass die lokalen Fluktuationen von f und s von sehr unterschiedlichem Ausmaß sind.

13.3. Asymptotische Normalität

Wir werden nun zeigen, dass der durch (13.12) definierte Schätzer $\hat{v}_n(x)$ asymptotisch normalverteilt ist. Dazu setzen wir einige technische Bedingungen an das Modell voraus. Diese

stellen u.a. die Ergodizität des Prozesses (Y_i) sicher. Es gelte:

(A1) $E[\xi_1^2] = 1$, $E[\xi_1] = E[\xi_1^3] = 0$, und

$$m_4 = E[(\xi_1^2 - 1)^2] < \infty.$$

(A2) ξ_1 besitzt eine Wahrscheinlichkeitsdichte p , die

$$\inf_{x \in \mathcal{K}} p(x) > 0$$

für jede kompakte Teilmenge $\mathcal{K} \subset \mathbb{R}$ erfüllt.

(A3) Es gebe Konstanten $C_1, C_2 > 0$, so dass

$$|f(y)| \leq C_1(1 + |y|), \quad (13.13)$$

$$|s(y)| \leq C_2(1 + |y|), \quad y \in \mathbb{R}. \quad (13.14)$$

(A4) Für die Funktion s gilt

$$\inf_{y \in \mathcal{K}} s(y) > 0,$$

für jede kompakte Teilmenge $\mathcal{K} \subset \mathbb{R}$.

(A5) $C_1 + C_2 E|\xi_1| < 1$.

Durch (A2) und (A4) ist sichergestellt, dass der Prozess (Y_i) nicht ausstirbt, wohingegen (A3) und (A5) Bedingungen dafür sind, dass (Y_i) nicht explodiert. Diese einfach zu formulierenden Bedingungen können mit größerem technischen Aufwand wie in ? abgeschwächt werden. Insbesondere muss die lineare Wachstumsbedingung (A3) im wesentlichen nur asymptotisch für $|y| \rightarrow \infty$ erfüllt sein.

Das Modell (13.1) impliziert, dass (Y_i) insbesondere eine Markoff-Kette ist. Aus dem folgenden Lemma von [Ango Nze \(1992\)](#) folgt die Ergodizität dieser Kette. Es basiert auf Anwendungen der Resultate von [Nummelin and Tuominen \(1982\)](#) und Tweedie (1983).

LEMMA 13.1 *Unter den Bedingungen (A1) - (A5) ist die Markoff-Kette (Y_i) geometrisch ergodisch, d.h. (Y_i) ist ergodisch mit stationärem Wahrscheinlichkeitsmaß π , und es gibt ein $\rho \in [0, 1)$, so dass für fast alle y gilt*

$$\|P^n(\cdot | y) - \pi\|_{TV} = O(\rho^n).$$

Dabei bezeichnet

$$P^n(B | y) = P(Y_n \in B | Y_0 = y), \quad B \in \mathcal{B},$$

die bedingte Verteilung von Y_n gegeben $Y_0 = y$, und

$$\|\nu\|_{TV} = \sup \left\{ \sum_{i=1}^k |\nu(B_i)|; k \in \mathbb{N}, B_1, \dots, B_k \in \mathcal{B} \text{ paarweise disjunkt} \right\}$$

ist die Totalvariation eines signierten Maßes ν auf der Borelschen σ -Algebra \mathcal{B} über \mathbb{R} .

Zur Herleitung der asymptotischen Normalität von $\hat{v}_n(x)$ in einem festen Punkt $x \in \mathbb{R}$ benötigen wir weitere Bedingungen. Dabei setzen wir zur Notationsvereinfachung $l = p + 1$.

- (A6) Die Funktionen f und s sind im Punkt x $(l - 1)$ -mal stetig differenzierbar, und es existieren die einseitigen Ableitungen $f_{\pm}^l(x)$, $s_{\pm}^l(x)$ l -ter Ordnung.
- (A7) Die stationäre Verteilung π besitzt eine beschränkte, stetige Wahrscheinlichkeitsdichte γ , die in einer Umgebung von x strikt positiv ist.
- (A8) Der Kern $K : \mathbb{R} \rightarrow \mathbb{R}^+$ ist beschränkt mit kompaktem Träger, und es gilt $K > 0$ auf einer Menge positiven Lebesguemaßes.
- (A9) Die Bandbreiten h_n sind von der Form $h_n = \beta n^{-1/(2l+1)}$, wobei $\beta > 0$ ist.
- (A10) Der Startwert Y_0 ist eine reelle Konstante.

Gemäß Lemma 1 in [Tsybakov \(1986\)](#) folgt aus (A8), dass die Matrizen

$$A = \int F(u) F(u)^T K(u) du \quad \text{und}$$

$$Q = \int F(u) F(u)^T K^2(u) du$$

positiv definit sind. Wir setzen

$$\mathcal{D} = A^{-1}QA^{-1} \quad \text{und}$$

$$f^{(l)}(x; u) = \begin{cases} f_+^{(l)}(x), & u \geq 0, \\ f_-^{(l)}(x), & u < 0, \end{cases}$$

Damit definieren wir die asymptotischen Fehler

$$b_f(x) = A^{-1} \frac{\beta^l}{l!} \int F(u) u^l K(u) f^{(l)}(x; u) du \quad \text{und}$$

$$b_g(x) = A^{-1} \frac{\beta^l}{l!} \int F(u) u^l K(u) g^{(l)}(x; u) du.$$

Ferner seien

$$c(x) = \begin{pmatrix} f(x) \\ f'(x)h_n \\ \vdots \\ f^{(l-1)}(x) \frac{h_n^{l-1}}{(l-1)!} \end{pmatrix} \quad \text{und} \quad \bar{c}(x) = \begin{pmatrix} g(x) \\ g'(x)h_n \\ \vdots \\ g^{(l-1)}(x) \frac{h_n^{l-1}}{(l-1)!} \end{pmatrix}.$$

Die Aussage des folgenden Satzes ist das zentrale Resultat dieses Kapitels.

SATZ 13.1 *Unter den Voraussetzungen (A1) - (A10) gilt*

$$(\bar{c}_n(x) - \bar{c}(x))^T F(0) \xrightarrow{p} 0, \quad (c_n(x) - c(x))^T F(0) \xrightarrow{p} 0 \quad (13.15)$$

und

$$n^{l/(2l+1)} \begin{pmatrix} \bar{c}_n(x) - \bar{c}(x) \\ c_n(x) - c(x) \end{pmatrix} \xrightarrow{\mathcal{D}} N(b(x), \Sigma(x)) \quad (13.16)$$

für $n \rightarrow \infty$, wobei

$$b(x) = \begin{pmatrix} b_g(x) \\ b_f(x) \end{pmatrix} \quad \text{und} \quad \Sigma(x) = \frac{s^2(x)}{\beta\gamma(x)} \begin{pmatrix} 4f^2(x) + s^2(x)m_4 & 2f(x) \\ 2f(x) & 1 \end{pmatrix} \otimes \mathcal{D}.$$

Hierbei bezeichnet $\mathcal{D}' \otimes \mathcal{D}$ das Kronecker-Produkt von Matrizen \mathcal{D}' und \mathcal{D} .

BEWEIS:

Die Normalgleichungen für das erste Kleinste-Quadrate-Problem in (13.10) sind gegeben durch

$$n^{\frac{l}{2l+1}} B_n \bar{c}_n(x) = n^{-\frac{l}{2l+1}} \sum_{i=1}^n Y_i^2 U_{in} K(u_{in}) \quad (13.17)$$

mit der Matrix

$$B_n = n^{-\frac{2l}{2l+1}} \sum_{i=1}^n U_{in} U_{in}^T K(u_{in}).$$

Andererseits gilt nach Definition von B_n

$$n^{\frac{l}{2l+1}} B_n \bar{c}(x) = n^{-\frac{l}{2l+1}} \sum_{i=1}^n U_{in} U_{in}^T \bar{c}(x) K(u_{in}), \quad (13.18)$$

woraus sich zusammen mit (13.17)

$$n^{\frac{l}{2l+1}} B_n (\bar{c}_n(x) - \bar{c}(x)) = n^{-\frac{l}{2l+1}} \sum_{i=1}^n (Y_i^2 - U_{in}^T \bar{c}(x)) U_{in} K(u_{in}) \quad (13.19)$$

ergibt. Aus der Modellannahme (13.1) folgt

$$\begin{aligned} Y_i^2 &= \left(f(Y_{i-1}) + s(Y_{i-1})\xi_i \right)^2 \\ &= f^2(Y_{i-1}) + 2f(Y_{i-1})s(Y_{i-1})\xi_i + (\xi_i^2 - 1) s^2(Y_{i-1}) + s^2(Y_{i-1}) \\ &= g(Y_{i-1}) + \alpha_i \end{aligned} \quad (13.20)$$

mit

$$\alpha_i = 2f(Y_{i-1}) s(Y_{i-1}) \xi_i + s^2(Y_{i-1})(\xi_i^2 - 1).$$

Nach Definition von U_{in} und $\bar{c}(x)$ gilt $U_{in}^T \bar{c}(x) = \sum_{j=0}^{l-1} \frac{1}{j!} g^{(j)}(x) (Y_{i-1} - x)^j$. Durch Taylorentwicklung von $g = f^2 + s^2$ erhalten wir daher unter Verwendung der Integraldarstellung des Restgliedes

$$\begin{aligned} g(Y_{i-1}) - U_{in}^T \bar{c}(x) &= \frac{(Y_{i-1} - x)^l}{(l-1)!} \int_0^1 g^{(l)}(x + t(Y_{i-1} - x))(1-t)^{l-1} dt \\ &= r_g(Y_{i-1}, x). \end{aligned} \quad (13.21)$$

Aus (13.19), (13.20) und (13.21) erhalten wir

$$\begin{aligned}
& n^{\frac{l}{2l+1}} B_n (\bar{c}_n(x) - \bar{c}(x)) \\
&= n^{-\frac{l}{2l+1}} \sum_{i=1}^n (g(Y_{i-1}) - U_{in}^T \bar{c}(x)) U_{in} K(u_{in}) \\
&\quad + n^{-\frac{l}{2l+1}} \sum_{i=1}^n (2f(Y_{i-1}) s(Y_{i-1}) \xi_i + (\xi_i^2 - 1) s^2(Y_{i-1})) U_{in} K(u_{in}) \\
&= \bar{b}_n(x) + \bar{q}_n(x)
\end{aligned} \tag{13.22}$$

mit

$$\bar{b}_n(x) = n^{-\frac{l}{2l+1}} \sum_{i=1}^n r_g(Y_{i-1}, x) U_{in} K(u_{in}) \quad \text{und} \quad \bar{q}_n(x) = n^{-\frac{l}{2l+1}} \sum_{i=1}^n \alpha_i U_{in} K(u_{in}),$$

In analoger Weise erhält man

$$n^{\frac{l}{2l+1}} B_n (c_n(x) - c(x)) = b_n(x) + q_n(x) \tag{13.23}$$

mit

$$b_n(x) = n^{-\frac{l}{2l+1}} \sum_{i=1}^n r_f(Y_{i-1}, x) U_{in} K(u_{in}) \quad \text{und} \quad q_n(x) = n^{-\frac{l}{2l+1}} \sum_{i=1}^n \beta_i U_{in} K(u_{in}),$$

wobei hier $\beta_i = s(Y_{i-1}) \xi_i$ gesetzt wurde.

In Anlehnung an die Darstellungen (13.22) und (13.23) verläuft die weitere Beweisführung zu Satz 13.1 nun in folgenden Schritten:

a) Zuerst zeigen wir, dass

$$B_n \xrightarrow{p} B \quad \text{für } n \rightarrow \infty \tag{13.24}$$

elementweise erfüllt ist. Dabei ist die Matrix $B = \beta \gamma(x) A$ positiv definit.

b) Als nächstes beweisen wir die Beziehungen

$$\bar{b}_n(x) \xrightarrow{p} B b_g(x) \quad \text{für } n \rightarrow \infty \tag{13.25}$$

und

$$b_n(x) \xrightarrow{p} B b_f(x) \quad \text{für } n \rightarrow \infty. \tag{13.26}$$

c) Der gemeinsame Zufallsvektor $(\bar{q}_n(x), q_n(x))^T$ ist asymptotisch normalverteilt:

$$\begin{pmatrix} \bar{q}_n(x) \\ q_n(x) \end{pmatrix} \xrightarrow{\mathcal{D}} N(0, \Sigma_0) \quad \text{für } n \rightarrow \infty \quad (13.27)$$

mit der Kovarianzmatrix

$$\Sigma_0 = s^2(x)\beta\gamma(x) \begin{pmatrix} 4f^2(x) + s^2(x)m_4 & 2f(x) \\ 2f(x) & 1 \end{pmatrix} \otimes Q.$$

d) Es gilt

$$\begin{aligned} n^{-l/(2l+1)} q_n^T(x) F(0) &\xrightarrow{p} 0 \quad \text{und} \\ n^{-l/(2l+1)} \bar{q}_n^T(x) F(0) &\xrightarrow{p} 0 \end{aligned} \quad (13.28)$$

für $n \rightarrow \infty$.

Sind die Aussagen a) bis d) gezeigt, so ergibt sich daraus in folgender Weise die Aussage des Satzes:

Aus b) und d) folgt

$$B_n (\bar{c}_n(x) - \bar{c}(x))^T F(0) = n^{-l/(2l+1)} \bar{b}_n(x) F(0) + n^{-l/(2l+1)} \bar{q}_n(x) F(0) \xrightarrow{p} 0$$

für $n \rightarrow \infty$. Wegen a) und der Definitheit der Grenzmatrix impliziert dies $(\bar{c}_n(x) - \bar{c}(x))^T F(0) \xrightarrow{p} 0$. Ebenso schließt man, um $(c_n(x) - c(x))^T F(0) \xrightarrow{p} 0$ zu zeigen.

Die asymptotische Normalität (13.16) ersieht man auf ähnliche Weise:

Wegen b) und c) gilt

$$n^{\frac{l}{2l+1}} B_n \begin{pmatrix} \bar{c}_n(x) - \bar{c}(x) \\ c_n(x) - c(x) \end{pmatrix} = \begin{pmatrix} \bar{b}_n(x) \\ b_n(x) \end{pmatrix} + \begin{pmatrix} \bar{q}_n(x) \\ q_n(x) \end{pmatrix} \xrightarrow{\mathcal{D}} N \left(\begin{pmatrix} B b_g(x) \\ B b_f(x) \end{pmatrix}, \Sigma_0 \right),$$

woraus gemäß a) die Gültigkeit von (13.16) folgt. □

Es bleiben noch a) bis d) zu beweisen. Dazu benötigen wir einige Hilfsaussagen.

LEMMA 13.2 (Davydov, 1973)

Es sei (Y_i) eine geometrisch ergodische Markoff-Kette, so dass Y_0 gemäß dem stationären Maß π der Kette verteilt ist. Dann ist die Kette geometrisch stark mischend, d.h. sie ist stark mischend (α -mischend) mit Mischungskoeffizienten $\alpha(n)$, die $\alpha(n) \leq c_0 \rho_0^n$ für geeignete $0 < \rho_0 < 1$ und $c_0 > 0$ erfüllen.

Es sei (\mathcal{F}_k) die kanonische Filtration des Prozesses (Y_k) , d.h. $\mathcal{F}_k = \sigma(Y_k, Y_{k-1}, \dots, Y_0)$ bezeichnet die von Y_0, \dots, Y_k erzeugte σ -Algebra.

LEMMA 13.3 (Liptser and Shirjaev (1980), Corollary 6)

Für jedes $n > 0$ sei die Folge $\eta^n = (\eta_{nk}, \mathcal{F}_k)$ eine quadratisch integrierbare Martingaldifferenz, d.h.

$$E[\eta_{nk} | \mathcal{F}_{k-1}] = 0, \quad E[\eta_{nk}^2] < \infty, \quad 1 \leq k \leq n, \quad (13.29)$$

und es gelte

$$\sum_{k=1}^n E[\eta_{nk}^2] = 1, \quad \forall n \geq n_0 > 0. \quad (13.30)$$

Dann sind die Bedingungen

$$\sum_{k=1}^n E[\eta_{nk}^2 | \mathcal{F}_{k-1}] \xrightarrow{p} 1 \quad \text{für } n \rightarrow \infty, \quad (13.31)$$

$$\sum_{k=1}^n E[\eta_{nk}^2 I(|\eta_{nk}| > \varepsilon) | \mathcal{F}_{k-1}] \xrightarrow{p} 0 \quad \text{für } n \rightarrow \infty, \quad \text{für jedes } \varepsilon > 0, \quad (13.32)$$

hinreichend für die Verteilungskonvergenz

$$\mathcal{L}\left(\sum_{k=1}^n \eta_{nk}\right) \longrightarrow N(0, 1) \quad \text{für } n \rightarrow \infty.$$

LEMMA 13.4 Es seien ϕ_1 eine stetige, beschränkte Funktion und ϕ_2 eine beschränkte Funktion. Unter den Bedingungen (A1) bis (A10) gilt für jeden (13.1) erfüllenden Prozess $Y_i, i \geq 0$,

$$n^{-\frac{2l}{2l+1}} \sum_{i=1}^n \phi_1(Y_{i-1}) \phi_2(u_{in}) K(u_{in}) \xrightarrow{p} \beta\gamma(x) \phi_1(x) \int \phi_2(u) K(u) du \quad (13.33)$$

$$n^{-\frac{2l}{2l+1}} \sum_{i=1}^n E[\phi_1(Y_{i-1}) \phi_2(u_{in}) K(u_{in})] \longrightarrow \beta\gamma(x) \phi_1(x) \int \phi_2(u) K(u) du$$

für $n \rightarrow \infty$.

BEWEIS:

Wir beweisen die Aussage zunächst für den Fall, dass die Markoff-Kette bereits im Gleichgewicht startet, und führen den allgemeinen Fall darauf zurück.

Dazu sei (Y_i^*) eine Markoff-Kette, die (13.1) erfüllt und für die Y_0^* der in Lemma 13.1 eingeführten stationären Verteilung π von (Y_i) genügt. Diese Kette ist per Konstruktion stationär, und durch Anwenden von Lemma 13.2 erhalten wir, dass (Y_i^*) geometrisch stark mischend ist. Daraus ergibt sich

$$n^{-\frac{2l}{2l+1}} \sum_{i=1}^n \phi_1(Y_{i-1}^*) \phi_2(u_{in}^*) K(u_{in}^*) - n^{-\frac{1}{2l+1}} E[\phi_1(Y_1^*) \phi_2(u_{1n}^*) K(u_{1n}^*)] \xrightarrow{p} 0 \quad (13.34)$$

für $n \rightarrow \infty$, wobei $u_{in}^* = (Y_{i-1}^* - x)/h_n$ gesetzt wurde. Für den zweiten Summanden in (13.34) gilt

$$\begin{aligned} n^{\frac{1}{2l+1}} \mathbb{E}[\phi_1(Y_1^*) \phi_2(u_{1n}^*) K(u_{1n}^*)] &= \beta \frac{1}{h_n} \int \phi_1(y) \phi_2\left(\frac{y-x}{h_n}\right) K\left(\frac{y-x}{h_n}\right) \gamma(y) dy \\ &= \beta \gamma(x) \phi_1(x) \int \phi_2(u) K(u) du (1 + o(1)) \end{aligned} \quad (13.35)$$

für $n \rightarrow \infty$. Zusammen mit (13.34) folgt hieraus, dass (Y_i^*) (13.33) erfüllt. Setzen wir nun

$$\zeta_i = \phi_1(Y_{i-1}) \phi_2(u_{in}) K(u_{in}), \quad \zeta_i^* = \phi_1(Y_{i-1}^*) \phi_2(u_{in}^*) K(u_{in}^*),$$

und wählen eine Folge $\{\delta_n\}$ mit $\delta_n = o(n^{\frac{2l}{2l+1}})$ und $\lim_{n \rightarrow \infty} \delta_n = \infty$. Es folgt

$$\begin{aligned} n^{-\frac{2l}{2l+1}} \sum_{i=1}^n |\mathbb{E}[\zeta_i - \zeta_i^*]| &\leq n^{-\frac{2l}{2l+1}} \left[\sum_{i=1}^{\delta_n-1} |\mathbb{E}[\zeta_i - \zeta_i^*]| + \sum_{i=\delta_n}^n |\mathbb{E}[\zeta_i - \zeta_i^*]| \right] \\ &\leq 2n^{-\frac{2l}{2l+1}} \delta_n \|\phi_1 \phi_2 K\|_\infty + n^{-\frac{2l}{2l+1}} \sum_{i=\delta_n}^n |\mathbb{E}[\zeta_i - \zeta_i^*]| \\ &= n^{-\frac{2l}{2l+1}} \sum_{i=\delta_n}^n |\mathbb{E}[\zeta_i - \zeta_i^*]| + o(1) \quad \text{für } n \rightarrow \infty. \end{aligned} \quad (13.36)$$

Aus der geometrischen Ergodizität von (Y_i) gemäß Lemma 13.1 ergibt sich für den linken Summanden des letzten Ausdrucks

$$\begin{aligned} n^{-\frac{2l}{2l+1}} \sum_{i=\delta_n}^n |\mathbb{E}[\zeta_i - \zeta_i^*]| &= n^{-\frac{2l}{2l+1}} \sum_{i=\delta_n}^n |\mathbb{E}[\phi_1(Y_{i-1}) \phi_2(u_{in}) K(u_{in}) \\ &\quad - \phi_1(Y_{i-1}^*) \phi_2(u_{in}^*) K(u_{in}^*)]| \\ &\leq n^{-\frac{2l}{2l+1}} \sum_{i=\delta_n}^n \|\phi_1 \phi_2 K\|_\infty \int |\gamma_i(y) - \gamma(y)| dy \\ &= O\left(n^{-\frac{2l}{2l+1}} \sum_{i=\delta_n}^n \rho^i\right) = o(1) \quad \text{für } n \rightarrow \infty, \end{aligned} \quad (13.37)$$

wobei γ_i die Dichte von Y_{i-1} bezeichnet. Es gilt also

$$\lim_{n \rightarrow \infty} n^{-\frac{2l}{2l+1}} \sum_{i=1}^n |\mathbb{E}[\zeta_i - \zeta_i^*]| = 0.$$

Daher folgt mit Hilfe der Markoff-Ungleichung, dass sich (13.33) von (Y_i^*) auf (Y_i) überträgt. \square

BEWEIS:

(zu Satz 13.1, Fortsetzung)

Es sind noch die auf Seite 207 formulierten Behauptungen a) bis d) zu beweisen.

a) Nach Definition von B_n gilt für die Elemente dieser Matrix

$$(B_n)_{j,k} = n^{-\frac{2l}{2l+1}} \sum_{i=1}^n \frac{u_{in}^{k+j-2}}{(k-1)!(j-1)!} K(u_{in}).$$

Diese besitzen demnach die in Lemma 13.4 geforderte Gestalt, und es folgt

$$(B_n)_{j,k} \xrightarrow{p} \frac{\beta \gamma(x)}{(k-1)!(j-1)!} \int u^{k+j-2} K(u) du,$$

was nach Definition der Matrix A gleichbedeutend mit $B_n \xrightarrow{p} \beta \gamma(x) A = B$ ist. Die Definitheit überträgt sich von A auf B .

b) Mit f und s erfüllt auch $g = f^2 + s^2$ die Bedingung (A6). Für das Restglied aus der Taylorentwicklung von g gilt

$$\begin{aligned} r_g(Y_{i-1}, x) &= u_{in}^l h_n^l \frac{1}{(l-1)!} \int_0^1 g^{(l)}(x + t(Y_{i-1} - x))(1-t)^{l-1} dt \\ &= u_{in}^l n^{-\frac{l}{2l+1}} \phi_3(Y_{i-1}) \end{aligned}$$

mit

$$\phi_3(Y_{i-1}) = \frac{\beta^l}{(l-1)!} \int_0^1 g^{(l)}(x + t(Y_{i-1} - x))(1-t)^{l-1} dt.$$

Hiermit lässt sich $\bar{b}_n(x)$ umschreiben zu

$$\bar{b}_n(x) = n^{-\frac{2l}{2l+1}} \sum_{i=1}^n \phi_3(Y_{i-1}) u_{in}^l U_{in} K(u_{in}),$$

d.h. die Elemente von $\bar{b}_n(x)$ erfüllen die Voraussetzungen von Lemma 13.4.

Wir wählen wieder (Y_i^*) wie im Beweis zu Lemma 13.4 und setzen $U_{in}^* = F(u_{in}^*)$. Aus (13.36) und (13.37) ergibt sich

$$\bar{b}_n(x) - n^{-\frac{2l}{2l+1}} \sum_{i=1}^n \phi_3(Y_{i-1}^*) (u_{in}^*)^l U_{in}^* K(u_{in}^*) \xrightarrow{p} 0 \quad (13.38)$$

für $n \rightarrow \infty$. Da (Y_i^*) α -mischend ist, erhalten wir wie in (13.34)

$$n^{-\frac{2l}{2l+1}} \sum_{i=1}^n \phi_3(Y_{i-1}^*) (u_{in}^*)^l U_{in}^* K(u_{in}^*) - n^{\frac{1}{2l+1}} \mathbb{E} \left[\phi_3(Y_1^*) (u_{1n}^*)^l U_{1n}^* K(u_{1n}^*) \right] \xrightarrow{p} 0$$

für $n \rightarrow \infty$. Der rechte Term dieses Ausdrucks lässt sich umschreiben zu

$$n^{\frac{1}{2l+1}} \mathbb{E} \left[\phi_3(Y_1^*) (u_{1n}^*)^l U_{1n}^* K(u_{1n}^*) \right] = \beta \int \phi_3(x + uh_n) u^l F(u) K(u) \gamma(x + uh_n) du.$$

Ferner gilt

$$\lim_{n \rightarrow \infty} \phi_3(x + uh_n) = \beta^l g^{(l)}(x; u)/l \quad (13.39)$$

für jedes $u \in \mathbb{R}$. Zusammen mit (13.39) und (A7), folgt daraus

$$\begin{aligned} \lim_{n \rightarrow \infty} \beta \int \phi_3(x + uh_n) u^l F(u) K(u) \gamma(x + uh_n) du \\ = \frac{\beta^{l+1}}{l!} \left(\int F(u) u^l K(u) g^{(l)}(x; u) du \right) \gamma(x) \\ = A \gamma(x) \beta b_g(x) = B b_g(x). \end{aligned}$$

Hiermit ist (13.25) gezeigt. Der Beweis zu (13.26) verläuft analog.

c) Wir definieren die Matrizen

$$\begin{aligned} \Sigma_n^{11} &= n^{-\frac{2l}{2l+1}} \sum_{i=1}^n E[\alpha_i^2 | \mathcal{F}_{i-1}] U_{in} U_{in}^T K^2(u_{in}), \\ \Sigma_n^{12} &= n^{-\frac{2l}{2l+1}} \sum_{i=1}^n E[\alpha_i \beta_i | \mathcal{F}_{i-1}] U_{in} U_{in}^T K^2(u_{in}), \\ \Sigma_n^{22} &= n^{-\frac{2l}{2l+1}} \sum_{i=1}^n E[\beta_i^2 | \mathcal{F}_{i-1}] U_{in} U_{in}^T K^2(u_{in}) \end{aligned}$$

und konstruieren daraus die Blockmatrix

$$\Sigma_n = \begin{pmatrix} \Sigma_n^{11} & \Sigma_n^{12} \\ \Sigma_n^{12} & \Sigma_n^{22} \end{pmatrix}.$$

Die Elemente von Σ_n^{11} , Σ_n^{12} und Σ_n^{22} erfüllen die Voraussetzungen von Lemma 13.4. Insbesondere sind die dort auftretenden zusammengesetzten Funktionen $\phi_1(Y_{i-1})$ in diesem Fall gegeben durch

$$\begin{aligned} E[\alpha_i^2 | \mathcal{F}_{i-1}] &= 4f^2(Y_{i-1}) s^2(Y_{i-1}) + s^4(Y_{i-1}) m_4, \\ E[\alpha_i \beta_i | \mathcal{F}_{i-1}] &= 2f(Y_{i-1}) s^2(Y_{i-1}) \quad \text{bzw.} \\ E[\beta_i^2 | \mathcal{F}_{i-1}] &= s^2(Y_{i-1}), \end{aligned}$$

wobei (A1) verwendet wurde. Man beachte, dass die entsprechenden Funktionen ϕ_1 wegen (A6) in einer Umgebung von x stetig und beschränkt sind. Da K außerhalb eines Kompaktums verschwindet, ist dies ausreichend für Lemma 13.4. Somit erhalten wir

$$\Sigma_n \xrightarrow{p} \Sigma_0 \quad \text{und} \quad E[\Sigma_n] \longrightarrow \Sigma_0 \quad (13.40)$$

für $n \rightarrow \infty$.

Zum Beweis von (13.27) genügt es nach dem Satz von Cramér-Wold zu zeigen, dass

$$a^T \begin{pmatrix} \bar{q}_n(x) \\ q_n(x) \end{pmatrix} \xrightarrow{\mathcal{D}} N(0, a^T \Sigma_0 a) \quad \text{für } n \rightarrow \infty \quad (13.41)$$

für jeden Vektor $a \in \mathbb{R}^{2l}$ mit euklidischer Norm $\|a\| = 1$ erfüllt ist. Dazu wählen wir gemäß (13.40) ein $n_0 \in \mathbb{N}$, so dass $E[\Sigma_n] > \frac{1}{2}\Sigma_0$ für alle $n \geq n_0$ gilt, und setzen für $n \geq n_0$

$$\eta_{ni} = \frac{n^{-\frac{l}{2l+1}}}{\sqrt{a^T E[\Sigma_n] a}} a^T \begin{pmatrix} \alpha_i U_{in} \\ \beta_i U_{in} \end{pmatrix} K(u_{in}).$$

Dann ist

$$\sum_{i=1}^n \eta_{ni} = \frac{1}{\sqrt{a^T E[\Sigma_n] a}} a^T \begin{pmatrix} \bar{q}_n(x) \\ q_n(x) \end{pmatrix},$$

und (13.41) ist äquivalent zu

$$\sum_{k=1}^n \eta_{nk} \xrightarrow{\mathcal{D}} N(0, 1) \quad \text{für } n \rightarrow \infty. \quad (13.42)$$

Wir zeigen nun, dass die (η_{nk}) die Voraussetzungen (13.29) bis (13.32) von Lemma 13.3 erfüllen, woraus (13.42) folgt.

Zunächst merken wir an, dass $E[\alpha_i | \mathcal{F}_{i-1}] = 0$ f.s. und $E[\beta_i | \mathcal{F}_{i-1}] = 0$ f.s. gilt, woraus (13.29) folgt. Außerdem prüft man leicht nach, dass

$$\sum_{k=1}^n E[\eta_{nk}^2 | \mathcal{F}_{k-1}] = \frac{a^T \Sigma_n a}{a^T E[\Sigma_n] a}.$$

Daher ist (13.30) erfüllt, und aus (13.40) erhalten wir (13.31).

Es ist noch (13.32) nachzuweisen. Für $n \geq n_0$ ist

$$\eta_{nk}^2 \leq \frac{n^{-\frac{2l}{2l+1}}}{a^T E[\Sigma_n] a} (a^T Z_{nk})^2 \leq \frac{2n^{-\frac{2l}{2l+1}}}{a^T \Sigma_0 a} (a^T Z_{nk})^2 \leq \kappa_1 n^{-\frac{2l}{2l+1}} |Z_{nk}|^2,$$

mit einer geeigneten Konstanten $\kappa_1 > 0$ und

$$Z_{nk} = \begin{pmatrix} \alpha_k U_{kn} \\ \beta_k U_{kn} \end{pmatrix} K(u_{kn}).$$

Da K beschränkt ist und kompakten Träger besitzt, und da f und s lokal beschränkt sind, gibt es eine Konstante $\kappa_2 > 0$, so dass

$$\begin{aligned} \eta_{nk}^2 &\leq \kappa_1 n^{-\frac{2l}{2l+1}} (\alpha_k^2 + \beta_k^2) |U_{kn}|^2 K^2(u_{kn}) \\ &\leq \kappa_2 n^{-\frac{2l}{2l+1}} (1 + |\xi_k|^4) K(u_{kn}). \end{aligned}$$

Hieraus folgt

$$\begin{aligned} &E[\eta_{nk}^2 I(|\eta_{nk}| \geq \varepsilon) | \mathcal{F}_{k-1}] \\ &\leq \kappa_2 n^{-\frac{2l}{2l+1}} K(u_{kn}) E \left[(1 + |\xi_1|^4) I(\sqrt{1 + |\xi_1|^4} \geq \varepsilon n^{\frac{l}{2l+1}} \kappa_2^{-1} \|K\|_\infty^{-1}) \right] \\ &= \kappa_2 n^{-\frac{2l}{2l+1}} K(u_{kn}) \cdot o(1) \end{aligned}$$

für $n \rightarrow \infty$, wobei $o(1)$ von k unabhängig ist. Somit erhalten wir

$$\sum_{k=1}^n \mathbb{E}[\eta_{nk}^2 I(|\eta_{nk}| \geq \varepsilon) | \mathcal{F}_{k-1}] \leq o(1) \sum_{k=1}^n n^{-\frac{2l}{2l+1}} K(u_{kn}) \quad \text{für } n \rightarrow \infty. \quad (13.43)$$

Gemäß Lemma 13.4 gilt für die letzte Summe

$$n^{-\frac{2l}{2l+1}} \sum_{k=1}^n K(u_{kn}) \xrightarrow{p} \beta\gamma(x) \int K(u) du \quad \text{für } n \rightarrow \infty. \quad (13.44)$$

Aus (13.43) und (13.44) folgt (13.32), d.h. die Voraussetzungen von Lemma 13.3 sind in der Tat erfüllt, und damit ist auch (13.41) gezeigt.

d) Es ist

$$\begin{aligned} n^{-l/(2l+1)} q_n^T(x) F(0) &= n^{-2l/(2l+1)} \sum_{i=1}^n \beta_i U_{in}^T F(0) K(u_{in}) \\ &= n^{-2l/(2l+1)} \sum_{i=1}^n \beta_i u_{in} K(u_{in}) = n^{-2l/(2l+1)} \sum_{i=1}^n (\beta_i - \mathbb{E}[\beta_i | \mathcal{F}_{i-1}]) u_{in} K(u_{in}). \end{aligned}$$

Nach (A8) ist der Kern K beschränkt, und es gilt $d^* = \max\{|u| : u \in \text{supp } K\} < \infty$. Daher gibt es eine Konstante $\kappa_0 > 0$, so dass

$$\begin{aligned} \mathbb{E}[(n^{-l/(2l+1)} q_n^T(x) F(0))^2] &= n^{-\frac{4l}{2l+1}} \mathbb{E}\left[\left(\sum_{i=1}^n (\beta_i - \mathbb{E}[\beta_i | \mathcal{F}_{i-1}]) u_{in} K(u_{in})\right)^2\right] \\ &\leq \kappa_0 n^{-\frac{4l}{2l+1}} \sum_{i=1}^n \mathbb{E}\left[(\beta_i - \mathbb{E}[\beta_i | \mathcal{F}_{i-1}])^2 I(|u_{in}| \leq d^*)\right]. \end{aligned}$$

Ist n hinreichend groß, so gilt für die Summanden der letzten Summe

$$\begin{aligned} \mathbb{E}\left[(\beta_i - \mathbb{E}[\beta_i | \mathcal{F}_{i-1}])^2 I(|u_{in}| \leq d^*)\right] &= \mathbb{E}\left[s^2(Y_{i-1}) \xi_i^2 I\left(\frac{|Y_{i-1} - x|}{h_n} \leq d^*\right)\right] \\ &= \mathbb{E}\left[s^2(Y_{i-1}) I\left(\frac{|Y_{i-1} - x|}{h_n} \leq d^*\right)\right] \leq \sup_{|y-x| \leq h_n d^*} s^2(y) < \infty. \end{aligned}$$

Damit ist $n^{-l/(2l+1)} q_n^T(x) F(0) \xrightarrow{p} 0$ gezeigt. Ebenso beweist man

$$n^{-l/(2l+1)} \bar{q}_n^T(x) F(0) \xrightarrow{p} 0.$$

□

Als unmittelbare Konsequenz aus Satz 13.1 ergibt sich:

SATZ 13.2 *Unter den Bedingungen (A1) bis (A10) gilt*

$$n^{l/(2l+1)}(\hat{v}_n(x) - v(x)) \xrightarrow{\mathcal{D}} N(b_v(x), \sigma_v^2(x)) \quad \text{für } n \rightarrow \infty,$$

wobei

$$\begin{aligned} b_v(x) &= F^T(0) (b_g(x) - 2f(x) b_f(x)) \quad \text{und} \\ \sigma_v^2(x) &= \frac{s^4(x)m_4}{\beta\gamma(x)} F^T(0) \mathcal{D} F(0) \end{aligned}$$

ist.

BEWEIS:

Aus $g(x) = \bar{c}(x)^T F(0)$, $f(x) = c(x)^T F(0)$, $v(x) = g(x) - f^2(x)$ und der Konstruktion von \hat{v}_n erhalten wir

$$\begin{aligned} \hat{v}_n(x) - v(x) &= (\bar{c}_n(x) - \bar{c}(x))^T F(0) \\ &\quad - (2c(x)^T F(0) + (c_n(x) - c(x))^T F(0)) ((c_n(x) - c(x))^T F(0)). \end{aligned}$$

Also gilt

$$\begin{aligned} n^{l/(2l+1)}(\hat{v}_n(x) - v(x)) &= n^{l/(2l+1)} \Psi(x) \begin{pmatrix} \bar{c}_n(x) - \bar{c}(x) \\ c_n(x) - c(x) \end{pmatrix} \\ &\quad + n^{l/(2l+1)} ((c_n(x) - c(x))^T F(0))^2 \end{aligned} \quad (13.45)$$

mit der Transformationsmatrix

$$\Psi(x) = \begin{pmatrix} F(0) \\ -2f(x) F(0) \end{pmatrix}^T.$$

Gemäß (13.15) gilt $(c_n(x) - c(x))^T F(0) \xrightarrow{p} 0$ für $n \rightarrow \infty$, woraus zusammen mit (13.16) $n^{l/(2l+1)} ((c_n(x) - c(x))^T F(0))^2 \xrightarrow{p} 0$ folgt. Die Grenzverteilung von $n^{l/(2l+1)}(\hat{v}_n(x) - v(x))$ ist also allein durch den ersten Summanden auf der rechten Seite von (13.45) gegeben. Für diese erhalten wir unter Beachtung von (13.16)

$$n^{l/(2l+1)}(\hat{v}_n(x) - v(x)) \xrightarrow{\mathcal{D}} N(\Psi(x)b(x), \Psi(x)\Sigma(x)\Psi(x)^T)$$

für $n \rightarrow \infty$. Eine einfache Rechnung ergibt $\Psi(x)b(x) = b_v(x)$ sowie $\Psi(x)\Sigma(x)\Psi(x)^T = \sigma_v^2(x)$, womit die Behauptung gezeigt ist. \square

Über die in Satz 13.2 gezeigte asymptotische Normalität hinaus haben ? gezeigt, dass auch Bootstrap-Verfahren für nichtparametrische Volatilitätsschätzer anwendbar sind. Dabei betrachten sie gewöhnliche Kernschätzer, d.h. den Spezialfall $p = 0$ der LP-Schätzer in (13.4), aber die Resultate lassen sich direkt auch auf allgemeine LP-Schätzer übertragen - s. auch ?.

Betrachten wir zur Illustration den Fall $l = 2$. Wir setzen voraus, dass f und s zweimal stetig differenzierbar sind, und dass der Kern K den Bedingungen

$$\int K(u) du = 1 \quad \text{und} \quad K(u) = K(-u)$$

genügt. Dann gilt

$$\begin{aligned} A &= \begin{pmatrix} 1 & 0 \\ 0 & \sigma_K^2 \end{pmatrix} \quad \text{mit} \quad \sigma_K^2 = \int u^2 K(u) du, \\ Q &= \begin{pmatrix} \int K^2(u) du & 0 \\ 0 & \int u^2 K^2(u) du \end{pmatrix}, \\ b_f(x) &= A^{-1} \frac{\beta^2 f''(x)}{2} \begin{pmatrix} \sigma_K^2 \\ 0 \end{pmatrix} = \begin{pmatrix} \sigma_K^2 \beta^2 f''(x)/2 \\ 0 \end{pmatrix}, \\ b_g(x) &= A^{-1} \frac{\beta^2 g''(x)}{2} \begin{pmatrix} \sigma_K^2 \\ 0 \end{pmatrix} = \begin{pmatrix} \sigma_K^2 \beta^2 g''(x)/2 \\ 0 \end{pmatrix}, \\ \mathcal{D} &= \begin{pmatrix} \int K^2(u) du & 0 \\ 0 & \frac{1}{\sigma_K^4} \int u^2 K^2(u) du \end{pmatrix}, \end{aligned}$$

und daher

$$b_v(x) = \frac{\sigma_K^2 \beta^2}{2} \left((f^2(x) + s^2(x))'' - 2f(x) f''(x) \right) = \frac{\sigma_K^2 \beta^2}{2} \left(v''(x) + 2(f'(x))^2 \right)$$

und

$$\sigma_v^2(x) = \frac{s^4(x) m_4}{\beta \gamma(x)} \int K^2(u) du = \frac{v^2(x) m_4}{\beta \gamma(x)} \int K^2(u) du.$$

Insbesondere ergibt sich für den aus der asymptotischen Verteilung berechneten normierten quadratischen Fehler von \hat{v}_n :

$$\begin{aligned} \mathbb{E} \left[n^{2l/2l+1} (\hat{v}_n(x) - v(x))^2 \right] &\approx b_v^2(x) + \sigma_v^2(x) \\ &= \frac{v^2(x) m_4}{\beta \gamma(x)} \int K^2(u) du + \frac{\sigma_K^4 \beta^4}{4} \left(v''(x) + 2(f'(x))^2 \right)^2. \end{aligned}$$

Minimierung dieses Ausdrucks bezüglich K und β führt zu dem Epanechnikov-Kern

$$K(u) = K^*(u) = \frac{3}{4} (1 - u^2)_+$$

und zu dem folgenden Wert für β :

$$\beta(K) = \left(\frac{v^2(x) m_4 \int K^2(u) du}{\gamma(x) \sigma_K^4 (v''(x) + 2(f'(x))^2)^2} \right)^{1/5}.$$

Damit erhalten wir

$$\beta^* = \beta(K^*) = \left(\frac{125 v^2(x) m_4}{4 \gamma(x) (v''(x) + 2(f'(x))^2)^2} \right)^{1/5}.$$

13.4. Literaturhinweise

Das Modell (13.1) wurde im Zusammenhang mit Finanzzeitreihen ausgiebig studiert, insbesondere unter Annahme der ARCH-Struktur (Engle, 1982). Bis zu jener Zeit war die wissenschaftliche Stoßrichtung jedoch überwiegend dem (linearen) bedingten Mittelwert gewidmet, oder es wurde angenommen, dass die bedingte Varianz fest ist oder eine – wie z.B. in den ARCH-Modellen – spezielle Form hat. Zu Beginn der Achtziger Jahre wurde dieser Nachteil in der Finanzliteratur von Engle (1982) und Robinson (1983; 1984) und in der Statistikk-literatur von Collomb (1984) und Vieu (1995) aufgezeigt. Es wurden jedoch auch nichtparametrische und semiparametrische Approximationen vorgeschlagen (Gregory, 1989; Engle and Gonzalez-Rivera, 1991). Seitdem ist das Interesse für die hier betrachtete nichtparametrische Situation, in der die Gestalt der Funktionen f und s nicht vordefiniert ist, in der Wirtschafts- und Statistikk-literatur deutlich gewachsen.

Die QTARCH-Modelle (??) von Gouriéroux and Monfort (1992) bilden eine Verallgemeinerung der Schwellenmodelle (Threshold Models) für den bedingten Mittelwert von Tong (1983). Die Methode von Gouriéroux and Monfort (1992) und der Artikel von McKeague and Zhang (1994) basieren auf Histogramm-Schätzungen der Volatilität. Die Arbeiten von Chen and Tsay (1993a; 1993b) konzentrieren sich auf additive Modellierung für die Mittelwertfunktion f . Additive oder multiplikative Strukturen der Volatilität werden in Härdle, Lütkepohl and Chen (1997), Yang, Härdle and Nielsen (1999) und Hafner (1997) betrachtet. Das allgemeine nichtparametrische ARCH-Modell wird von Härdle, Tsybakov and Yang (1996) behandelt. In ? wird der Zusammenhang zwischen dem nichtparametrischen AR-ARCH-Modell und der zeitdiskreten Version der dem Black-Scholes Ansatzes zugrundeliegenden geometrischen Brownschen Bewegung diskutiert. ? untersuchen im Zusammenhang mit einem speziellen stochastischen Volatilitätsmodell nichtparametrische Dekonvolutions-schätzer für die Volatilitätsfunktion als ersten Schritt in Richtung auf die nichtparametrische Behandlung allgemeiner GARCH-Modelle.

Die Idee der lokal polynomialen Schätzung stammt von Stone (1977), Cleveland (1979) und Katkovnik (1979; 1985), die diese bei nichtparametrischen Regressionsmodellen angewendet haben. Statistische Eigenschaften von LP-Schätzungen bei nichtparametrischen Regressionsmodellen (Konvergenz, Konvergenzrate und punktweise asymptotische Normalität) wurden von Tsybakov (1986) hergeleitet. Referenzen neuerer Werke auf diesem Gebiet findet man in Fan and Gijbels (1996).

Neben den statistischen Studien des Modells (13.1) sind auch die verwendeten wahrscheinlichkeitstheoretischen Eigenschaften des dort konstruierten Prozesses (Y_i) von Bedeutung. Diese werden in den Werken von Doukhan and Ghindès (1981), Chan and Tong (1985), Mokkadem (1987), Diebolt and Guégan (1990) und Ango Nze (1992) untersucht. In diesen Artikeln werden die Ergodizität, geometrische Ergodizität und Mischungseigenschaften des Prozesses (Y_i) hergeleitet.

Teil III.

Spezifische Finanzanwendungen

14. Optionsbewertung mit flexiblen Volatilitätsschätzern

Seit ihrer Einführung durch Engle und Bollerslev werden Modelle mit autoregressiver, bedingter Heteroskedastizität (*autoregressive conditional heteroskedasticity models* oder ARCH) erfolgreich auf Finanzmarktdaten angewandt. Daher liegt es nahe, Optionspreismodelle zu betrachten, bei denen die zugrundeliegenden Instrumente einem ARCH-Prozess folgen. Aus empirischer Sicht hat sich herausgestellt, dass die Form der *News impact Kurve*, die als Funktion der heutigen Volatilität in Abhängigkeit gestriger Renditen definiert ist, der dominante Faktor bei der Preisbestimmung ist. So ist es zum Beispiel wichtig zu wissen, ob die News impact Kurve symmetrisch oder asymmetrisch ist. Um Fehlbewertungen durch diese Asymmetrien auszuschließen, werden flexible Volatilitätsmodelle benötigt. So können EGARCH-Modelle (s. Abschnitt 12.2) herangezogen werden, wenn Aktienpreise und Volatilität korreliert sind. Jedoch leidet dieses Modell an der Schwäche, dass das Problem der Stationaritätsbedingungen und der Asymptotik der Quasi-Maximum-Likelihood-Schätzung (QMLE) noch nicht vollständig gelöst sind. Ein anderer Ansatz ist es, wie in den Threshold GARCH-Modellen Schwellen in die News impact Kurve einzuführen und so Raum für Asymmetrie zu schaffen.

In diesem Abschnitt möchten wir uns auf die Spezifikation der Volatilität konzentrieren. Wir betrachten TGARCH-Prozesse und liefern umfangreiche Monte-Carlo-Simulationen für drei typische Parameterkonstellationen. Speziell vergleichen wir simulierte GARCH-Optionspreise mit Optionspreisen auf der Grundlage von TGARCH- und Black-Scholes-Modellen. Im empirischen Teil des Kapitels zeigen wir, dass Marktpreise von Call-Optionen in der Tat die Asymmetrie widerspiegeln, die in der News impact Kurve von DAX-Zeitreihen festgestellt wurde.

14.1. Optionsbewertung mit ARCH-Modellen

Wir betrachten eine Ökonomie in diskreter Zeit, in der Zinssätze und Erträge nach jedem Zeitintervall von konstanter, äquidistanter Länge ausgezahlt werden. Sei $S_t, t = 0, 1, 2, \dots$ der Preis der Aktie zum Zeitpunkt t und $Y_t = (S_t - S_{t-1})/S_{t-1}$ die zugehörige einperiodige Rendite ohne Dividendenzahlung. Man nehme an, es gebe einen Preis für das Risiko, gemessen in Form einer Risikoprämie, die auf den risikolosen Zins r addiert wird, um die erwartete Rendite der Folgeperiode zu erhalten. Es scheint vernünftig, die Risikoprämie in Abhängigkeit von der bedingten Varianz zu modellieren. Wir legen das ARCH-M-Modell (s. Abschnitt 12.2.3) mit einer Risikoprämie zugrunde, die eine lineare Funktion der bedingten

Standardabweichung ist:

$$Y_t = r + \lambda\sigma_t + \varepsilon_t \quad (14.1)$$

$$\mathcal{L}(\varepsilon_t | \mathcal{F}_{t-1}) = N(0, \sigma_t^2) \quad (14.2)$$

$$\sigma_t^2 = \omega + \alpha\varepsilon_{t-1}^2 + \beta\sigma_{t-1}^2. \quad (14.3)$$

In (14.3) sind ω , α und β konstante Parameter, die den Stationaritäts- und Positivitätsbedingungen genügen. Der konstante Parameter λ kann als der Preis einer Einheit Risiko aufgefaßt werden. \mathcal{F}_t bezeichnet wie üblich die Menge der bis einschließlich zum Zeitpunkt t zur Verfügung stehenden Information. Um die Notation zu erleichtern beschränken wir unsere Diskussion auf den GARCH(1,1)-Fall.

Das obige Modell wird unter dem empirischen Maß P geschätzt. Um ähnlich wie bei Black und Scholes in kontinuierlicher Zeit (s. Abschnitt 6.1) eine Bewertung mit dem Prinzip der Arbitragefreiheit vorzunehmen, müssen Annahmen über die Bewertung des Risikos getroffen werden. Viele Arbeiten untersuchen Optionspreise bei stochastischer Volatilität unter der Annahme, dass die Volatilität ein systematisches Risiko von Null besitzt, d.h. dass die Risikoprämie für die Volatilität Null ist. Duan (1995) identifiziert ein zu P äquivalentes Martingalmaß Q , indem er annimmt, dass die bedingte Verteilung der Renditen normalverteilt ist und außerdem gilt

$$\text{Var}^P(Y_t | \mathcal{F}_{t-1}) = \text{Var}^Q(Y_t | \mathcal{F}_{t-1}) \quad (14.4)$$

P -fast sicher. Er zeigt unter dieser Annahme, dass ein repräsentativer Agent mit z.B. konstanter relativer Risikoaversion bei normalverteilten relativen Änderungen des aggregierten Konsums seinen Erwartungsnutzen maximiert. Die Annahme (14.4) beinhaltet eine konstante Risikoprämie für die Volatilität, die direkt in deren Mittelwert eingeht.

Um ein Martingal unter dem neuen Maß zu erhalten, muß ein neuer Störterm, η_t , eingeführt werden, der den Effekt der in der Zeit variierenden Risikoprämie abbildet. Wenn wir $\eta_t = \varepsilon_t + \lambda\sigma_t$ definieren, führt (14.4) zu folgendem Modell unter dem neuen Maß Q :

$$Y_t = r + \eta_t \quad (14.5)$$

$$\mathcal{L}(\eta_t | \mathcal{F}_{t-1}) = N(0, \sigma_t^2) \quad (14.6)$$

$$\sigma_t^2 = \omega + \alpha(\eta_{t-1} - \lambda\sigma_{t-1})^2 + \beta\sigma_{t-1}^2. \quad (14.7)$$

Im Falle des GARCH(1,1)-Modells ist nach Satz ?? die Varianz der stationären Verteilung, unter dem empirischen Maß P gerade $\text{Var}^P(\varepsilon_t) = \omega/(1 - \alpha - \beta)$. Für das Duan-Maß Q steigt die Varianz der stationären Verteilung auf $\text{Var}^Q(\eta_t) = \omega/(1 - \alpha(1 + \lambda^2) - \beta)$, weil der Volatilitätsprozeß unter dem neuen Maß von Innovationen gesteuert wird, die nicht aus einer zentrierten, sondern einer nichtzentrierten Chi-Quadrat-Verteilung stammen. Weiter unten werden wir sehen, dass die Veränderung der unbedingten Varianz in entscheidendem Maße von der Spezifikation der News impact Kurve abhängt.

Die Beschränkung auf eine quadratische oder symmetrische News impact Kurve ist nicht immer wünschenswert, wie zahlreiche empirische Studien von Aktienrenditen aufgezeigt haben. Im Rahmen des obigen Modells kann diese Annahme zu einer nichtlinearen News impact

Funktion $g(\cdot)$ abgeschwächt werden. Das folgende Modell ist ein semiparametrisches Analogon zu dem GARCH-Modell. Unter dem empirischen Maß P erhalten wir

$$\begin{aligned} Y_t &= r + \lambda\sigma_t + \varepsilon_t \\ \mathcal{L}(\varepsilon_t | \mathcal{F}_{t-1}) &=_{\mathcal{P}} N(0, \sigma_t^2) \\ \sigma_t^2 &= g(\varepsilon_{t-1}) + \beta\sigma_{t-1}^2. \end{aligned}$$

Unter dem Duan-Martingalmaß Q wandelt sich das Modell zu

$$\begin{aligned} Y_t &= r + \eta_t \\ \mathcal{L}(\eta_t | \mathcal{F}_{t-1}) &=_{\mathcal{Q}} N(0, \sigma_t^2) \\ \sigma_t^2 &= g(\eta_{t-1} - \lambda\sigma_{t-1}) + \beta\sigma_{t-1}^2. \end{aligned}$$

Man beachte, dass - sobald ein Schätzer von $g(\cdot)$ unter P bekannt ist - er umgehend unter dem Maß Q eingesetzt werden kann.

In dieser allgemeinen Spezifikation ist das Schätzen ohne zusätzliche Information über $g(\cdot)$ eine heikle Angelegenheit, da iterative Schätzverfahren notwendig werden, um zugleich die Parameter λ, β und nichtparametrisch die Funktion g zu schätzen. Wir betrachten daher ein spezifisches flexibles parametrisches Modell: das Threshold GARCH-Modell, s. Abschnitt 12.2. Bei diesem Modell kann die News impact Funktion folgendermaßen geschrieben werden:

$$g(x) = \omega + \alpha_1 x^2 I(x < 0) + \alpha_2 x^2 I(x \geq 0)$$

Um dieses Modell zu motivieren, passen wir ein sehr einfaches nichtparametrisches Modell $Y_t = \sigma(Y_{t-1})\xi_t$ an Renditen des deutschen Aktienindex DAX an, wobei ξ_t unabhängig, identisch verteilt mit Erwartungswert 0 und Varianz 1 sind. Der Schätzer der News impact Kurve $\sigma^2(\cdot)$ ist in Abbildung 14.2 wiedergegeben. Um eine Vorstellung über die zugrundeliegende Verteilung der Renditen zu geben, tragen wir in Abbildung 14.1 einen nichtparametrischen Schätzer der Renditeverteilung gegen eine geglättete Normalverteilung ab. Ganz offensichtlich ist $g(\cdot)$ nicht symmetrisch um Null. Das TGARCH-Modell bildet dieses Phänomen ab, wenn $\alpha_1 > \alpha_2$ ist. Auch andere parametrische Modelle können diese Eigenschaft gut beschreiben, aber das TGARCH-Modell hat sich im Falle von Aktienrenditen als hinreichend flexibel und technisch beherrschbar erwiesen - vgl. z.B. Rabemananjara and Zakoian (1993).

Man erinnere sich daran, dass die Innovationen normalverteilt sind. Daraus folgt für das TGARCH-Modell, dass die unbedingte Varianz ähnlich wie in Satz ?? unter dem Maß P $\text{Var}^P(\varepsilon_t) = \omega/(1 - \bar{\alpha} - \beta)$ ist, wobei $\bar{\alpha} = (\alpha_1 + \alpha_2)/2$. Der folgende Satz liefert die unbedingte Varianz für $\eta_t = \varepsilon_t + \lambda\sigma_t$ unter Q .

SATZ 14.1 Die unbedingte Varianz des TGARCH(1,1)-Modells unter dem äquivalenten Martingalmaß Q von Duan ist

$$\text{Var}^Q(\eta_t) = \frac{\omega}{1 - \psi(\lambda)(\alpha_1 - \alpha_2) - \alpha_2(1 + \lambda^2) - \beta} \quad (14.8)$$

wobei

$$\psi(u) = u\varphi(u) + (1 + u^2)\Phi(u)$$

und $\varphi(u), \Phi(u)$ die Dichte und Verteilungsfunktion der Standardnormalverteilung sind.

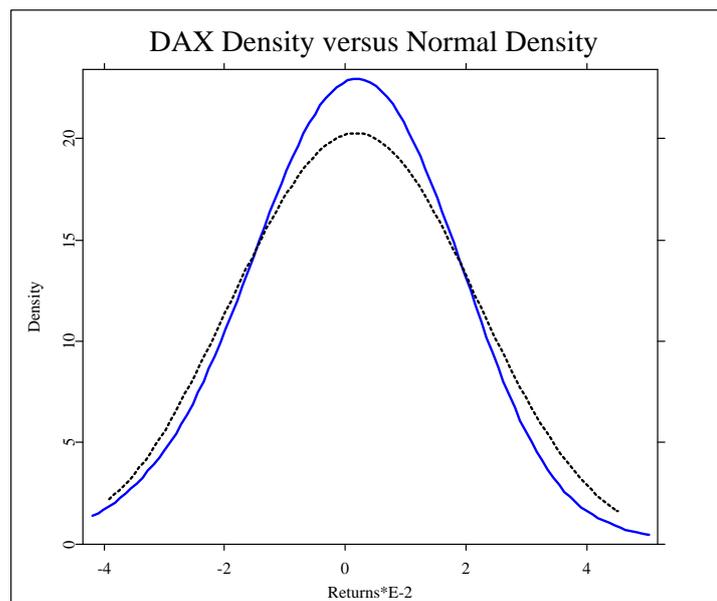


Abbildung 14.1.: Kernschätzung der Dichte von DAX-Renditen (durchgezogen) gegen eine Kernschätzung einer Normalverteilung (gestrichelt) mit gleichem Mittelwert und Varianz. Wir benutzen eine Bandbreite von 0.03 und eine quadratische Kernfunktion $K(u) = 15/16(1 - u^2)^2 I(|u| < 1)$. Die Randregionen wurden aus der Abbildung entfernt.

[SFMDaxReturnDistribution.xpl](#)

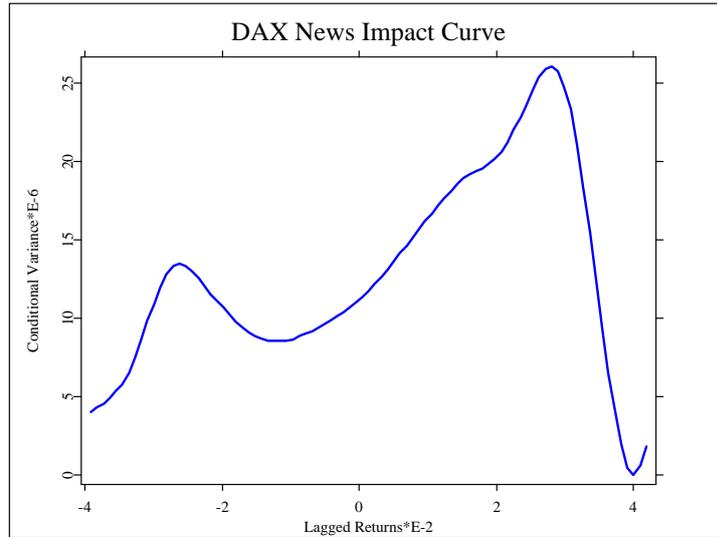


Abbildung 14.2.: Lokal lineare Schätzung der News impact Kurve für den DAX. Das Modell ist $Y_t = \sigma(Y_{t-1})\xi_t$. Gezeigt wird der Schätzer der Funktion $\sigma^2(y)$ mit einer Bandbreite 0.03. Die Randregionen wurden aus der Abbildung entfernt.

 [SFMNewsImpactCurve.xpl](#)

BEWEIS:

Sei $Z_t = \eta_t/\sigma_t - \lambda$. Unter Q gilt, $\mathcal{L}(Z_t | \mathcal{F}_{t-1}) = N(-\lambda, 1)$. Die bedingte Varianz σ_t^2 kann geschrieben werden als

$$\sigma_t^2 = \omega + \alpha_1 \sigma_{t-1}^2 Z_{t-1}^2 I(Z_{t-1} < 0) + \alpha_2 \sigma_{t-1}^2 Z_{t-1}^2 I(Z_{t-1} \geq 0) + \beta \sigma_{t-1}^2.$$

Man bilde Erwartungswerte; dann zeigt sich, dass für den Integralausdruck über den negativen Bereich folgt:

$$\begin{aligned} E^Q[Z_t^2 I(Z_t < 0) | \mathcal{F}_{t-1}] &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^0 z^2 e^{-\frac{1}{2}(z+\lambda)^2} dz \\ &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\lambda} (u - \lambda)^2 e^{-\frac{1}{2}u^2} du \\ &= \frac{\lambda}{\sqrt{2\pi}} e^{-\frac{1}{2}\lambda^2} + (1 + \lambda^2)\Phi(\lambda) \end{aligned} \tag{14.9}$$

$$=: \psi(\lambda). \tag{14.10}$$

Wegen

$$E^Q[Z_t^2 | \mathcal{F}_{t-1}] = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} z^2 e^{-\frac{1}{2}(z+\lambda)^2} dz = 1 + \lambda^2$$

folgern wir für den positiven Bereich

$$E^Q[Z_t^2 I(Z_t \geq 0) | \mathcal{F}_{t-1}] = 1 + \lambda^2 - \psi(\lambda). \tag{14.11}$$

Daher erhalten wir

$$E^Q[\sigma_t^2] = \omega + \alpha_1 \psi(\lambda) E^Q[\sigma_{t-1}^2] + \alpha_2 [1 + \lambda^2 - \psi(\lambda)] E^Q[\sigma_{t-1}^2] + \beta E^Q[\sigma_{t-1}^2]. \quad (14.12)$$

Da die unbedingte Varianz unabhängig von t ist, folgt der Satz. \square

Die Funktion ψ ist positiv und $\psi(\lambda) > 1/2$ für den realistischen Fall $\lambda > 0$. Wir können folgenden Aussagen über die Veränderung der unbedingten Varianz machen: Für $\alpha_1 = \alpha_2$, (14.8) erhält man das GARCH(1,1)-Resultat. Für $\alpha_1 > \alpha_2$ (den Fall des Leverage-Effekts ist der Anstieg der unbedingten Varianz sogar noch stärker als im symmetrischen GARCH-Fall. Für $\alpha_1 < \alpha_2$, ist die unbedingte Varianz kleiner als im Leverage-Fall, und wir können zwei Fälle unterscheiden: Wenn die Ungleichung

$$\alpha_1 < \alpha_2 \frac{2\psi(\lambda) - 1 - 2\lambda^2}{2\psi(\lambda) - 1} \quad (14.13)$$

erfüllt ist, dann wird die unbedingte Varianz unter Q sogar kleiner sein als unter P . Falls (14.13) nicht erfüllt ist, erhalten wir wie oben $\text{Var}^P(\varepsilon_t) \leq \text{Var}^Q(\eta_t)$. Jedoch nimmt der Quotient auf der rechten Seite von (14.13) negative Werte für realistische Werte einer Einheit der Risikoprämie an, (z.B. für kleine positive Werte), so dass in den meisten empirischen Studien (14.13) nicht erfüllt sein wird.

Natürlich hat die stationäre Varianz Auswirkungen auf den Optionspreis: je größer (kleiner) die Varianz, desto höher (niedriger) der Optionspreis. Dies gilt insbesondere für Optionen mit langer Laufzeit, bei denen das langfristige Mittel der Volatilität die wichtigste Determinante des Optionspreises darstellt. Daher kann eine Option zu niedrig bewertet sein, wenn ein GARCH-Modell herangezogen wird und zugleich ein Leverage-Effekt vorliegt.

Eine zweite Besonderheit des Duan-Ansatzes ist, dass unter Q und bei positiven Risikoprämien die heutige Innovation negativ mit der morgigen bedingten Varianz der GARCH-Risikoprämie korreliert ist, während die Korrelation unter P Null ist. Genauer gesagt, erhalten wir $\text{Cov}^Q(\eta_t/\sigma_t, \sigma_{t+1}^2) = -2\lambda\alpha \text{Var}^Q(\eta_t)$ mit dem GARCH-Parameter α . Dies legt nahe, dass kleine Prognosen der Volatilität unter Q (die den Optionspreis beeinflussen) nicht nur von quadrierten vergangenen Innovationen abhängen, sondern auch von deren Vorzeichen. Insbesondere führt für $\lambda > 0$ eine negative (positive) vergangene Innovation dazu, dass die Volatilität steigt (fällt) und mit ihr der Optionspreis. Der folgende Satz besagt, dass die Kovarianz von der Asymmetrie der News impact Funktion abhängt, wenn wir ein TGARCH anstelle eines GARCH-Modells zugrundelegen.

SATZ 14.2 *Für das TGARCH(1,1)-Modell ist die Kovarianz zwischen der Innovation in t und der bedingten Varianz in $t + 1$ unter dem äquivalenten Martingalmaß Q von Duan gegeben durch*

$$\text{Cov}^Q\left(\frac{\eta_t}{\sigma_t}, \sigma_{t+1}^2\right) = -2 \text{Var}^Q(\eta_t) [\lambda\alpha_2 + \{\varphi(\lambda) + \lambda\Phi(\lambda)\}(\alpha_1 - \alpha_2)], \quad (14.14)$$

wobei $\text{Var}^Q(\eta_t)$ aus dem Satz zuvor folgt.

BEWEIS:

Zunächst wird die bedingte Kovarianz ermittelt:

$$\begin{aligned}
\text{Cov}_{t-1}^Q\left(\frac{\eta_t}{\sigma_t}, \sigma_{t+1}^2\right) &= \mathbb{E}_{t-1}^Q\left[\frac{\eta_t}{\sigma_t}\sigma_{t+1}^2\right] = \omega\mathbb{E}_{t-1}^Q\left[\frac{\eta_t}{\sigma_t}\right] \\
&+ \alpha_1\mathbb{E}_{t-1}^Q\left[\frac{\eta_t}{\sigma_t}(\eta_t - \lambda\sigma_t)^2 I(\eta_t - \lambda\sigma_t < 0)\right] \\
&+ \alpha_2\mathbb{E}_{t-1}^Q\left[\frac{\eta_t}{\sigma_t}(\eta_t - \lambda\sigma_t)^2 I(\eta_t - \lambda\sigma_t \geq 0)\right] \\
&+ \beta\sigma_t\mathbb{E}_{t-1}^Q[\eta_t], \tag{14.15}
\end{aligned}$$

wobei $\mathbb{E}_t(\cdot)$ und $\text{Cov}_t(\cdot)$ abkürzend für $\mathbb{E}(\cdot | \mathcal{F}_t)$ bzw. $\text{Cov}(\cdot | \mathcal{F}_t)$ steht. Wegen (14.6) sind der erste und der vierte Erwartungswert auf der rechten Seite von (14.15) Null. Der zweite bedingte Erwartungswert ist

$$\mathbb{E}_{t-1}^Q\left[\frac{\eta_t}{\sigma_t}(\eta_t - \lambda\sigma_t)^2 I(\eta_t - \lambda\sigma_t < 0)\right] = -2\sigma_t^2\left[\frac{1}{\sqrt{2\pi}}\exp\left(-\frac{1}{2}\lambda^2\right) + \lambda\Phi(\lambda)\right]. \tag{14.16}$$

Da $\mathbb{E}_{t-1}^Q\left[\frac{\eta_t}{\sigma_t}(\eta_t - \lambda\sigma_t)^2\right] = -2\lambda\sigma_t^2$, können wir für den dritten bedingten Erwartungswert in (14.15) schreiben:

$$\mathbb{E}_{t-1}^Q\left[\frac{\eta_t}{\sigma_t}(\eta_t - \lambda\sigma_t)^2 I(\eta_t - \lambda\sigma_t \geq 0)\right] = -2\sigma_t^2\left[\lambda - \frac{1}{\sqrt{2\pi}}\exp\left(-\frac{1}{2}\lambda^2\right) - \lambda\Phi(\lambda)\right]. \tag{14.17}$$

Fügen wir (14.16) und (14.17) in (14.15) ein, so folgt

$$\text{Cov}_{t-1}^Q\left(\frac{\eta_t}{\sigma_t}, \sigma_{t+1}^2\right) = -2\sigma_t^2[\lambda\alpha_2 + \{\varphi(\lambda) + \lambda\Phi(\lambda)\}(\alpha_1 - \alpha_2)]. \tag{14.18}$$

Beachtet man, dass $\text{Cov}^Q(\eta_t/\sigma_t, \sigma_{t+1}^2) = \mathbb{E}^Q[\text{Cov}_{t-1}^Q(\eta_t/\sigma_t, \sigma_{t+1}^2)]$, so folgt die Behauptung unmittelbar. \square

Wir nehmen im folgenden an, dass eine positive Risikoprämie λ je Einheit vorliege. Wir können wiederum drei Fälle unterscheiden: Für $\alpha_1 = \alpha_2$ (im symmetrischen Fall) erhalten wir $\text{Cov}^Q(\eta_t/\sigma_t, \sigma_{t+1}^2) = -2\lambda\alpha_2\text{Var}^Q(\eta_t)$, d.h. das GARCH(1,1)-Ergebnis. Für $\alpha_1 < \alpha_2$ (den Fall des umgekehrten Leverage-Effektes) steigt die Kovarianz, und wenn

$$\lambda\alpha_2 + \left[\frac{1}{\sqrt{2\pi}}\exp\left(-\frac{1}{2}\lambda^2\right) + \lambda\Phi(\lambda)\right](\alpha_1 - \alpha_2) < 0, \tag{14.19}$$

ist die Korrelation positiv. Im letzten Fall, $\alpha_1 > \alpha_2$ (der Leverage-Fall), ist die Kovarianz negativ und steigt dem Betrage nach an.

Dies zeigt außerdem, dass das Rückkehrverhalten der Volatilität zur stationären Varianz unter Q anders als im symmetrischen GARCH-Fall ausfällt. Die negative Kovarianz ist im Leverage-Fall sogar noch größer. Dies könnte darauf hinweisen, dass Optionen zu hoch (zu niedrig) bewertet sind, wenn die vergangene Innovation positiv (negativ) ist, der Preisprozeß einem TGARCH-Prozeß mit $\alpha_1 > \alpha_2$ folgt und fälschlicherweise ein GARCH-Modell ($\alpha_1 = \alpha_2$) für Volatilitätsprognosen benutzt wird.

14.2. Eine Monte-Carlo-Studie

Da der diskontierte Preisprozeß ein Martingal unter dem äquivalenten Martingalmaß Q ist, können wir die Methode der risiko-neutralen Bewertung nach [Cox and Ross \(1976\)](#) heranziehen. Der Q -Preis C_t eines Calls zum Zeitpunkt t ist gegeben durch den abgezinste bedingten Erwartungswert der Auszahlung bei Fälligkeit - s. (6.23) -

$$C_t = (1 + r)^{-\tau} E^Q[\max(S_T - K, 0) | \mathcal{F}_t] \quad (14.20)$$

wobei T das Fälligkeitsdatum, $\tau = T - t$ die Restlaufzeit und K den Ausübungspreis bezeichnet. Für europäische Optionen folgt der arbitragefreie Preis P_t eines Puts aus der Put-Call-Parität (Satz 2.3), d.h. $P_t = C_t - S_t + (1 + r)^{-\tau} K$. Da es unter einem GARCH- oder TGARCH-Modell keinen analytischen Ausdruck für den Erwartungswert in (14.20) gibt, müssen wir den Optionspreis numerisch berechnen. Die Verteilung der Auszahlungsfunktion $\max(S_T - K, 0)$ zum Fälligkeitstermin wird simuliert, indem m Aktienpreisprozesse

$$S_{T,i} = S_t \prod_{s=t+1}^T (1 + Y_{s,i}), \quad i = 1, \dots, m, \quad (14.21)$$

generiert werden, wobei $Y_{s,i}$ die Rendite der i^{ten} Replikation zum Zeitpunkt s ist. Anschließend wird der Mittelwert der Auszahlungsfunktion mit dem risikolosen Zins abgezinzt

$$C_t = (1 + r)^{-\tau} \frac{1}{m} \sum_{i=1}^m \max(S_{T,i} - K, 0). \quad (14.22)$$

In der Simulationsstudie benutzen wir folgende Parameter: $r = 0$, $S_0 = 100$, $\tau = 30$ Tage, $m = 400000$, $\lambda = 0.01$. Die *Moneyness* S_0/K variiert zwischen 0.85 und 1.15, was der üblichen Bandbreite der gehandelten Optionen entspricht. Wir vergleichen hier nicht die Effekte verschiedener Maturitäten T , da viele Eigenschaften, die wie etwa der Smile von der stochastischen Volatilität hervorgerufen werden, mit wachsender Laufzeit verschwinden. Im allgemeinen bleiben die Effekte qualitativ zwar gleich, werden aber in quantitativer Hinsicht immer unbedeutender. Dies wurde in zahlreichen Experimenten festgestellt; daher konzentrieren wir uns auf kurze Laufzeiten.

Die Auswirkungen einer asymmetrischen News impact Funktion auf die Optionspreise wird in drei Situationen untersucht, die durch den Grad der kurzfristigen Autokorrelation der quadrierten Renditen und die Persistenz, d.h. den Wert von $\alpha + \beta$, charakterisiert werden. Für den GARCH(1,1)-Prozeß kann gezeigt werden, dass die Autokorrelation ρ_1 erster Ordnung der quadrierten Residuen gegeben ist durch

$$\rho_1 = \alpha(1 - \alpha\beta - \beta^2)/(1 - 2\alpha\beta - \beta^2), \quad (14.23)$$

und $\rho_j = (\alpha + \beta)\rho_{j-1}$, $j \geq 2$. Dies sind die Autokorrelationen eines ARMA(1,1)-Prozesses, da der quadrierte GARCH(1,1)-Prozess einem ARMA(1,1)-Modell genügt (s. Satz 12.9). Tabelle 1 listet die Parameterkonstellationen und Charakteristika der drei Typen auf:

Typ 1 ist gekennzeichnet durch eine hohe Persistenz und geringe Korrelation erster Ordnung, Typ 2 durch hohe Persistenz und große Autokorrelation erster Ordnung und Typ

Typ	α	β	$\alpha + \beta$	ρ_1
1	0.1	0.85	0.95	0.1791
2	0.5	0.45	0.95	0.8237
3	0.1	0.5	0.6	0.1077

Tabelle 14.1.: Charakterisierung der Typen der GARCH(1,1)-Modelle

3 durch niedrige Persistenz und kleine Autokorrelation erster Ordnung. Typ 1 ist typisch für Finanzmarktzeitreihen (sowohl tägliche als auch binnentägliche), da man üblicherweise beobachtet, dass die Autokorrelationsfunktion der quadrierten Renditen schnell nach den ersten Lags abfällt, dann aber nur noch langsam abklingt. Typ 2 beschreibt eine Situation mit sehr starken ARCH Effekten, und Typ 3 ähnelt dem Verhalten stark aggregierter Daten wie Monats- oder Quartalsdaten. In allen Fällen ist der Parameter ω so festgelegt, dass $\sigma^2 = 0.0002$, d.h. dass die unbedingte Varianz konstant bleibt.

Mit Blick auf die nichtlineare News impact Funktion $g(\cdot)$ wählen wir das Threshold ARCH-Modell mit zwei asymmetrischen Fällen. Im ersten Fall, den wir als Leverage-Fall bezeichnen, ist

$$g_1(x) = \omega + 1.2\alpha x^2 I(x < 0) + 0.8\alpha x^2 I(x \geq 0)$$

und im zweiten, dem Fall des ‘umgekehrten Leverage-Effektes’, ist

$$g_2(x) = \omega + 0.8\alpha x^2 I(x < 0) + 1.2\alpha x^2 I(x \geq 0).$$

Für Typ 1 und den Fall des Leverage-Effektes sind die Simulationsergebnisse in Abbildung 14.3 festgehalten. Wir tragen die absoluten und die relativen Differenzen der GARCH- und TGARCH-Preise zum entsprechenden Black-Scholes-Preis ab. Die relative Differenz ist definiert als die absolute Differenz dividiert durch den Black-Scholes-Preis. Wegen der kleinen Schrittweite (wir nehmen Schritte von 0.01 für die Moneyness) wirken die Funktionen sehr glatt. Für den GARCH-Fall erhalten wir das wohlbekanntes Resultat, dass die Preisdifferenz zu Black-Scholes eine U-Form bezüglich der Moneyness aufweist. Infolge der monoton in der Moneyness ansteigenden Call-Preise ist die relative Differenz absolut für Optionen aus dem Geld am größten, während die relative Differenz vernachlässigbar klein wird, je höher die Moneyness. Dies könnte die häufig beobachtbare Schiefe des Smile-Effektes erklären. Für die TGARCH-Optionspreise beobachten wir im Prinzip eine ähnliche Abweichung von Black-Scholes, allerdings mit einem entscheidenden Unterschied: Im Fall des Leverage-Effektes sind die Preise von Optionen aus dem Gelde niedriger und von Optionen im Gelde höher als im GARCH-Modell. Dies ist auch plausibel: Wenn eine Option weit aus dem Geld und der Fälligkeitstermin nah ist, besteht der einzige Weg zu einer positiven Auszahlung bei Fälligkeit darin, dass der Kurs des zugrundeliegenden Instruments mehrfach hintereinander in großen Schritten im Wert steigt. Dies ist jedoch weniger wahrscheinlich im Leverage-Fall, da positive Renditen einen kleineren Effekt auf die Volatilität als im symmetrischen Fall haben, vorausgesetzt die oben genannten Parameterkonstellationen treffen zu.

Tabelle 14.2 zeigt nur die Ergebnisse für die Typen 2 und 3 und den Fall des umgekehrten Leverage-Effektes und für ausgewählte Werte der Moneyness. Für den Leverage-Effekt-Fall ist die beschriebene Abweichung der TGARCH-Preise von jenen der GARCH-Preise auch

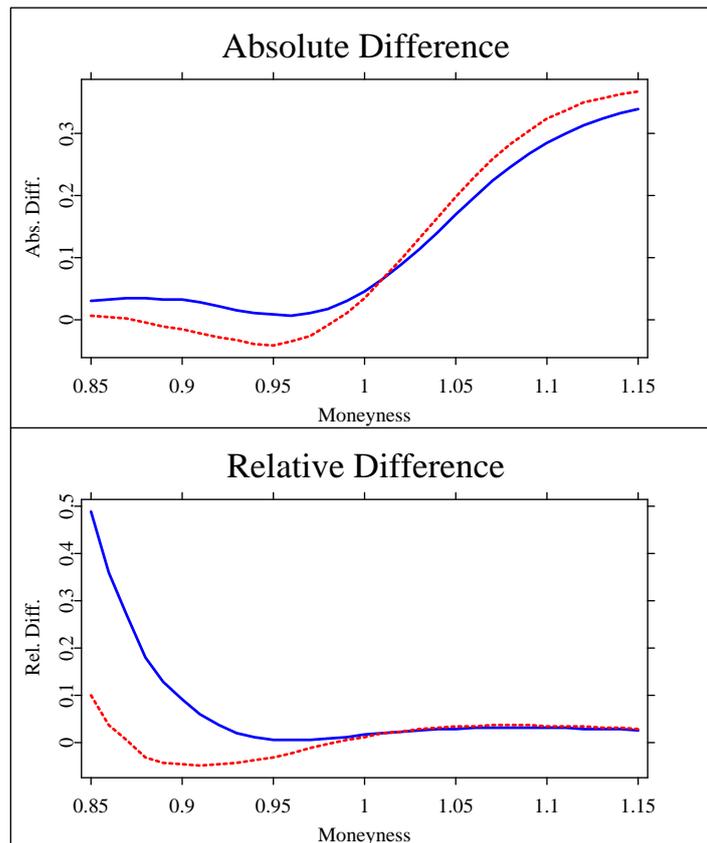


Abbildung 14.3.: Die Differenz zwischen simulierten GARCH- (durchgezogen) und TGARCH- (gestrichelt) Optionspreisen zu BS-Preisen als Funktion der Moneyness für Typ 1 und den Leverage-Fall. Die obere Abbildung zeigt die absoluten Differenzen, die untere die absoluten Differenzen dividiert durch den BS-Preis.

[SFMOptpricesim.xpl](#)

Typ	Moneyness	GARCH		TGARCH			
		% diff	SE	Leverage-Effekt % diff	SE	umgek. Lev.-Eff. % diff	SE
Typ 1	0.85	35.947	1.697	0.746	1.359	75.769	2.069
	0.90	-0.550	0.563	-12.779	0.498	11.606	0.631
	0.95	-6.302	0.261	-9.786	0.245	-3.153	0.278
	1.00	-3.850	0.132	-4.061	0.125	-3.806	0.139
	1.05	-1.138	0.057	-0.651	0.052	-1.692	0.061
	1.10	-0.020	0.025	0.347	0.022	-0.400	0.028
	1.15	0.162	0.012	0.347	0.010	-0.013	0.014
Typ 2	0.85	199.068	5.847	104.619	4.433	293.704	7.884
	0.90	0.489	1.136	-23.964	0.891	22.140	1.469
	0.95	-30.759	0.370	-39.316	0.305	-24.518	0.454
	1.00	-20.975	0.167	-22.362	0.141	-20.804	0.198
	1.05	-6.038	0.077	-5.427	0.063	-7.148	0.095
	1.10	-0.302	0.042	0.202	0.033	-0.966	0.054
	1.15	0.695	0.027	0.991	0.021	0.351	0.037
Typ 3	0.85	-2.899	1.209	-11.898	1.125	6.687	1.297
	0.90	-5.439	0.496	-8.886	0.479	-1.982	0.513
	0.95	-4.027	0.249	-4.970	0.245	-3.114	0.254
	1.00	-2.042	0.128	-2.077	0.126	-2.025	0.130
	1.05	-0.710	0.055	-0.559	0.053	-0.867	0.056
	1.10	-0.157	0.023	-0.047	0.022	-0.267	0.023
	1.15	-0.009	0.010	0.042	0.010	-0.059	0.011

Tabelle 14.2.: Simulationsergebnisse für ausgewählte Werte der Moneyness. Die prozentualen Differenzen der GARCH- und TGARCH-Optionspreise zu den Black-Scholes-Preisen und die zugehörigen Standardfehler (SE) der Simulation.

	ω	α	α_1	α_2	β	λ	$-2\log L$
GARCH	1.66E-05 (1.04E-06)	0.1438 (0.0061)			0.7756 (0.012)	0.0691 (0.0178)	-7,697.66
TGARCH	1.91E-05 (1.359E-06)		0.2005 (0.0084)	0.0454 (0.0113)	0.7736 (0.0157)	0.0385 (0.0175)	-7,719.24

Tabelle 14.3.: Die GARCH- und TGARCH-Schätzergebnisse für DAX-Renditen, 01.01.1988-30.12.91 (QMLE-Standardfehler in Klammern)

für die Typen 2 und 3 sichtbar. Im Fall des umgekehrten Leverage-Effektes kehren sich die Argumente um: Nunmehr ist es wahrscheinlicher, dass eine Option aus dem Geld doch noch im Geld enden wird, so dass TGARCH-Preise von Optionen, die weit aus dem Gelde sind, höher als die GARCH-Preise sind. Wie man erwarten würde, sind die Abweichungen der simulierten Preise zu den Black-Scholes-Preisen und zwischen den GARCH- und TGARCH-Preisen für Typ 2 am größten, d.h. für starke kurzfristige ARCH-Effekte, und am kleinsten für den Typ mit niedriger Persistenz, den Typ 3. Das letztere war zu erwarten, da die Unterschiede verschwinden sollten, je eher der homoskedastische Fall erreicht wird.

14.3. Anwendung zur Bewertung von DAX-Calls

Die Bewertungsmethode mit GARCH wird auf den deutschen Aktienindex und Optionsdaten angewendet. Als Aktienindex benutzen wir die tägliche Schlußnotierung des DAX vom 1. Januar 1988 bis zum 31. März 1992. Die Schlussnotierung wird gewöhnlich etwa um 13:30 Uhr Ortszeit (Frankfurt) fixiert. Für die Optionsdaten auf diesen Index ziehen wir Aufzeichnungen der Transaktionspreise an der Deutschen Terminbörse (DTB) von Januar bis März 1992 heran. Um die Beobachtungszeitpunkte von Index und Optionen zu synchronisieren, interpolieren wir zwischen dem letzten Optionspreis vor 13:30 Uhr und dem ersten danach, sofern der Unterschied nicht mehr als zwei Stunden beträgt.

Es kann keine Evidenz für autokorrelierte DAX-Renditen gefunden werden, aber quadrierte und absolute Renditen sind in hohem Maße autokorreliert. Wir schätzen ein GARCH(1,1)-M-Modell

$$Y_t = \lambda \sigma_t + \varepsilon_t \quad (14.24)$$

$$\mathcal{L}(\varepsilon_t | \mathcal{F}_{t-1}) = N(0, \sigma_t^2) \quad (14.25)$$

$$\sigma_t^2 = \omega + \alpha \varepsilon_{t-1}^2 + \beta \sigma_{t-1}^2 \quad (14.26)$$

für den DAX bis zum 30.12.1991 mit der Quasi-Maximum-Likelihood-Methode - s. Abschnitt 12.1.6. Eine mögliche Konstante in (14.24) ist nicht signifikant und wird daher von vorneherein weggelassen. Tabelle 14.3 zeigt die Resultate der Schätzung. Alle Parameter sind signifikant von Null verschieden. Der Grad der Persistenz $\alpha + \beta = 0.9194$ ist signifikant kleiner als 1, und daher ist die unbedingte Varianz endlich - vgl. Satz ???. Der Parameter der Risikoprämie λ ist positiv, wie die ökonomische Theorie erwarten läßt.

Die Quasi-Maximum-Likelihood-Schätzer des TGARCH-Modells

$$\sigma_t^2 = \omega + \alpha_1 \varepsilon_{t-1}^2 I(\varepsilon_{t-1} < 0) + \alpha_2 \varepsilon_{t-1}^2 I(\varepsilon_{t-1} \geq 0) + \beta \sigma_{t-1}^2 \quad (14.27)$$

werden ebenfalls in Tabelle 14.3 berichtet. Berücksichtigt man den Wert der Log-Likelihood, so ist die Güte des TGARCH-Modells besser als jene des GARCH-Modells. Ein Likelihood-Quotiententest lehnt das GARCH-Modell auf jedem vernünftigen Konfidenzniveau ab. α_1 und α_2 sind signifikant verschieden; damit ist die Asymmetrie der News impact Funktion signifikant. Da $\alpha_1 > \alpha_2$ beobachten wir den gewöhnlichen Leverage-Effekt für Finanzmarktzeitreihen.

Nachdem das Modell für die Zeitspanne von 1988 bis 1991 an die Daten angepasst worden ist, ist der nächste Schritt, die Optionspreise für den Betrachtungszeitraum von Januar bis März 1992 mit der oben beschriebenen Simulationsmethode zu berechnen und mit den beobachteten Marktpreisen zu vergleichen. In dieser Arbeit konzentrieren wir uns auf Kaufoptionen. Da die DAX-Optionen, die an der DTB gehandelt werden, europäischen Typs sind, können die Ergebnisse für Putoptionen aus der Put-Call-Parität gewonnen werden. Wir betrachten neun Call-Optionen mit den Fälligkeitsterminen 17. Januar, 20. März und 19. Juni 1992. Um ungefähr die Fälle von Optionen im, am und aus dem Geld untersuchen zu können, wählen wir die Ausübungspreise 1550, 1600 und 1650 für die Januaroption und 1600, 1650 und 1700 für die März- und Junioptionen. Wir simulieren die Preise der Januaroption vom 3. bis 16. Januar (zehn Tage), der Märzoptionen vom 3. Januar bis 19. März (57 Tage) und die Preise der Junioptionen vom 3. Januar bis 31. März (64 Tage). Die Junioption mit einem Ausübungspreis von 1700 wurde am 16. Januar eingeführt, so dass wir für die ersten zehn Handelstage im Januar über keine Beobachtungen dieser Option verfügen. Wegen nur geringer Handelsaktivität sind gelegentlich Marktpreise nicht verfügbar, was die Zahl der Beobachtungen, als k in Tabelle 14.4 aufgeführt, weiter reduziert.

Eine offene Frage ist, wie der Startwert des Volatilitätsprozesses zu wählen ist. Wir setzen den Startwert gleich dem laufenden Schätzer für die Volatilität (GARCH oder TGARCH), indem der Volatilitätsprozess extrapoliert und die Parameter konstant gehalten werden. Eine Alternative wäre das Heranziehen der implizierten Volatilität aus Abschnitt 6.3.4.

Zur Berechnung der Black-Scholes-Preise zum Zeitpunkt t wird die implizierte Volatilität zum Zeitpunkt $t - 1$ benutzt. Um ein Maß der Schätzgüte zu erhalten, definieren wir die relativen Residuen als

$$u_t \equiv \frac{C_t - C_{\text{Markt},t}}{C_{\text{Markt},t}}$$

wobei C_t entweder der Black-Scholes-, der GARCH oder der TGARCH-Preis und $C_{\text{Markt},t}$ der am Markt beobachtete Preis ist. Residuen sollten als relative Größen betrachtet werden, da ein Händler stets die billige Option, die in gleicher Höhe ‘unterbewertet’ ist wie eine teure Option, vorziehen wird, einfach, weil er seine Position in der billigen Option vervielfacht. Ein ähnliches Argument gilt für den Verkauf von ‘überbewerteten’ Optionen. Aus Symmetriegründen können wir ein quadratisches Verlustkriterium, d.h.

$$U = \sum_{t=1}^k u_t^2$$

benutzen. Die Resultate für die drei Modelle sind in Tabelle 14.4 zu sehen.

T	K	k	BS	GARCH	TGARCH
Jan	1550	10	0.017	0.014	0.014
	1600	10	0.099	0.029	0.028
	1650	10	4.231	1.626	1.314
Mär	1600	47	1.112	0.961	0.954
	1650	53	1.347	1.283	1.173
	1700	56	1.827	1.701	1.649
Jun	1600	53	1.385	1.381	1.373
	1650	56	2.023	1.678	1.562
	1700	51	2.460	2.053	1.913
Summe		346	14.500	10.725	9.980

Tabelle 14.4.: Das Verlustkriterium U für DAX-Calls mit der Fälligkeit in T und dem Ausübungspreis K unter Verwendung von BS-, GARCH- und TGARCH-Optionspreisen. Die Zahl der Beobachtungen ist durch k gegeben.

Alles in allem schneiden sowohl das GARCH- als auch das TGARCH-Optionsbewertungsmodell substanziell besser ab als das Black-Scholes-Modell. Für Optionen im und am Geld ist die Verbesserung der TGARCH-Prognose über das GARCH-Modell klein. Wenn die Option jedoch aus dem Geld ist, ergibt sich eine große Reduktion des Verlustkriteriums. In der Simulationstudie haben Optionen aus dem Geld am empfindlichsten auf stochastische Volatilität und Leverage-Effekte reagiert. In der Situation mit realen Daten wird dies für die Januar-1650-Option am deutlichsten, wo Black-Scholes sehr schlecht und TGARCH besser als GARCH abschneidet. Für die März- und Junioptionen sind die Unterschiede weniger ausgeprägt. Dies erklärt sich aus der Tatsache, dass der Index zu einem Niveau von 1736 Punkten am 20. März und von 1717 Punkten am 31. März anstieg, so dass die Optionen mit einem Ausübungspreis von 1700 ins Geld kamen. Dies ist ebenfalls die Erklärung für die Tatsache, dass U am höchsten für die Januar-1650-Optionen ist. Hier gibt es nur zehn Handelstage, aber diese Option war während einiger Tage weit aus dem Geld. So schloss beispielsweise am 8. Januar der DAX bei 1578 Punkten.

Da in allen Fällen TGARCH besser als GARCH abschneidet, schließen wir, dass der Markt diese Asymmetrie der Volatilität beachtet. Daher ist eine korrekte Spezifikation des Volatilitätsmodells von größter Bedeutung bei der Optionspreisbestimmung.

14.4. Literaturhinweise

Engle (1982)

Nelson (1991) von Zakoian (1994) Glosten et al. (1993) Härdle and Hafner (2000)

vgl. z.B. Hull and White (1987) sowie Renault and Touzi (1996). Melino and Turnbull (1990) betrachteten eine von Null verschiedene, konstante und exogene Risikoprämie für die Volatilität. Wie die empirische Evidenz von Wiggins (1987) nahelegt, dürfte die Vernachlässigung von Änderungen der Volatilität nicht gerechtfertigt sein.

Das alternative Konzept zur Minimierung des quadratischen Verlustes eines Hedgeportfo-

lios, wie eingeführt von Föllmer and Sondermann (1991) und Föllmer and Schweizer (1991), wird im allgemeinen zu einem anderen Maß zur Preisfindung führen.

Wenn β hinreichend klein ist, kann man an einem gewissen Lag abschneiden und ein gewöhnliches additives semiparametrisches Modell schätzen, siehe Carroll, Härdle and Mammen (1999).

Ähnlich waren Bossaerts and Hillion (1993) vorgegangen, die “15 Minuten alte” implizierte Volatilitäten in die Black/Scholes-Formel eingehen ließen, was sich als erfolgreich herausstellte.

15. Value at Risk und Backtesting

15.1. Vorhersage und VaR-Modelle

Value at Risk (VaR) Modelle werden in einer Vielzahl von Finanzanwendungen eingesetzt. Ihr Ziel ist es, die Gewinne oder Verluste eines Portfolios, die in einer kommenden Periode auflaufen können, in geeigneter Weise zu quantifizieren. Die Unsicherheit über die Entwicklung eines Portfolios drücken wir durch eine ‘‘Vorhersageverteilung’’ P_{t+1} für Periode $t + 1$ aus.

$$P_{t+1} = \mathcal{L}(L_{t+1} | \mathcal{F}_t)$$

ist die bedingte Verteilung einer Zufallsvariable L_{t+1} , die die möglichen Gewinne und Verluste eines Portfolios in der folgenden Periode bis zur Zeit $t + 1$ repräsentiert, und \mathcal{F}_t bezeichnet die Information in den bis zur Zeit t vorliegenden historischen Daten. Ein Schätzer für diese Verteilung ergibt sich aus einem statistischen Vorhersagemodell, demzufolge die möglichen bedingten Verteilungen von L_{t+1} aus einer parametrischen Klasse $\mathcal{P}_{t+1} = \{P_{t+1}^{\theta(t)} | \theta(t) \in \Theta\}$ stammen. Der endlich-dimensionale Parameter $\theta(t)$ wird aus typischerweise $n = 250$ historischen Renditebeobachtungen vor der Zeit t geschätzt, d.h. in etwa aus den Kursnotierungen der Handelstage eines Jahres. Bezeichnet $\hat{\theta}(t)$ diesen Schätzer, so wird $\mathcal{L}(L_{t+1} | \mathcal{F}_t)$ durch $P_{t+1}^{\hat{\theta}(t)}$ approximiert.

Ein wichtiges Beispiel für \mathcal{P}_{t+1} ist das Delta-Normal Modell, RiskMetrics (1996), das wir im folgenden beschreiben. In diesem Modell nehmen wir an, dass das Portfolio aus d linearen (oder linearisierten) Instrumenten mit Marktwerten $X_{k,t}$, $k = 1, \dots, d$, besteht und dass die gemeinsame bedingte Verteilung der zugrundeliegenden log-Renditen

$$Y_{t+1} \in \mathbb{R}^d, Y_{k,t+1} = \ln X_{k,t+1} - \ln X_{k,t}, k = 1, \dots, d,$$

gegeben die Information bis zur Zeit t , multivariat normalverteilt ist, d.h.

$$\mathcal{L}(Y_{t+1} | \mathcal{F}_t) = N_d(0, \Sigma_t) \tag{15.1}$$

wobei Σ_t die (bedingte) Kovarianzmatrix des Zufallsvektors Y_{t+1} bezeichnet. Wir betrachten zuerst eine einzige Position ($d = 1$), die aus λ_t Anteilen an einem einzelnen Wert mit aktuellem Marktpreis $X_t = x$ besteht. Mit $w_t = \lambda_t x$ bezeichnen wir die *Exposure* dieser Position zur Zeit t , d.h. ihren Wert gegeben $X_t = x$. Die bedingte Verteilung der Wertänderung $L_{t+1} = \lambda_t(X_{t+1} - X_t)$ ist dann approximativ:

$$\mathcal{L}(L_{t+1} | \mathcal{F}_t) = \mathcal{L}(\lambda_t(X_{t+1} - x) | \mathcal{F}_t) \tag{15.2}$$

$$= \mathcal{L}\left(w_t \frac{X_{t+1} - x}{x} \mid \mathcal{F}_t\right) \tag{15.3}$$

$$\approx \mathcal{L}(w_t Y_{t+1} \mid \mathcal{F}_t) = N(0, w_t^2 \sigma_t^2) \tag{15.4}$$

TO DO:
Referenz
RiskMe-
trics

mit $\sigma_t^2 = \text{Var}(Y_{t+1} \mid \mathcal{F}_t)$. Hierbei haben wir die Taylorapproximation

$$\ln X_{t+1} - \ln x = \frac{X_{t+1} - x}{x} + o(X_{t+1} - x) \quad (15.5)$$

verwendet. Die Verallgemeinerung auf ein ganzes Portfolio aus $\lambda_t^1, \dots, \lambda_t^d$ Anteilen von d (linearen) Instrumenten liegt auf der Hand. Sei hierzu w_t der d -dimensionale Exposure-Vektor zur Zeit t

$$w_t = (w_t^1, \dots, w_t^d)^T = (\lambda_t^1 x^1, \dots, \lambda_t^d x^d)^T. \quad (15.6)$$

$$L_{t+1} = \sum_{k=1}^d \lambda_t^k (X_{k,t+1} - X_{k,t})$$

sei die Wertänderung des Portfolios. Wie für eine einzelne Position ist die bedingte Verteilung von L_{t+1} gegeben die Information \mathcal{F}_t approximativ gleich der bedingten Verteilung von

$$w_t^T Y_{t+1} = \sum_{k=1}^d w_t^k Y_{k,t+1}.$$

Diese Verteilung gehört nach dem Delta-Normal Modell zur Familie

$$\mathcal{P}_{t+1} = \{N(0, \sigma_t^2) : \sigma_t^2 \in [0, \infty)\}, \quad (15.7)$$

mit $\sigma_t^2 = w_t^T \Sigma_t w_t$. Das Ziel der VaR-Analyse besteht darin, den Parameter $\theta(t) = \sigma_t$ zu schätzen und hierdurch die Vorhersageverteilung P_{t+1} zu approximieren.

Wir betrachten daher das Problem des Schätzens der Vorhersageverteilung ausgehend von folgenden Modellannahmen. Die Wertänderung des Portfolios sei von der Form

$$L_{t+1} = \sigma_t Z_{t+1} \quad (15.8)$$

$$\sigma_t^2 = w_t^T \Sigma_t w_t, \quad (15.9)$$

wobei die Z_t u.i.v. $N(0,1)$ -verteilte Zufallsgrößen sind, w_t der Exposure-Vektor zur Zeit t und Σ_t die (bedingte) Kovarianzmatrix des Vektors Y_{t+1} der log-Renditen ist. Wir fassen die letzten n Realisierungen $Y_t = y_t, \dots, Y_{t-n+1} = y_{t-n+1}$ des zugrundeliegenden log-Renditevektors zu einer $(n \times d)$ -Matrix $\mathcal{Y}_t = (y_i^T)_{i=t-n+1, \dots, t}$ zusammen. Aus diesen Beobachtungen berechnen wir zwei Schätzer von Σ_t , und zwar zunächst den naiven RMA, d.h. *rectangular moving average*:

$$\hat{\Sigma}_t = \frac{1}{n} \mathcal{Y}_t^T \mathcal{Y}_t. \quad (15.10)$$

Da der Erwartungswert des Renditevektors Y_t nach dem Delta-Normal Modell Null ist, ist dies gerade die empirische Kovarianzmatrix. Der zweite sogenannte EMA-Schätzer, d.h. *exponentially moving average*, basiert auf einer Idee von Taylor (1986) und benutzt ein exponentielles Gewichtsschema. Seien für ein γ , $0 < \gamma < 1$

$$\tilde{y}_{t-k} = \gamma^k y_{t-k}, k = 0, \dots, n-1, \quad \tilde{\mathcal{Y}}_t = (\tilde{y}_i^T)_{i=t-n+1, \dots, t}$$

die in der Zeit exponentiell gewichteten log-Renditevektoren und die aus ihnen gebildete $(n \times d)$ -Matrix, so schätzen wir Σ_t durch

$$\hat{\Sigma}_t = (1 - \gamma)^{-1} \tilde{\mathcal{Y}}_t^T \tilde{\mathcal{Y}}_t. \quad (15.11)$$

Die Normierung ist sinnvoll, da die Summe $\sum_{i=1}^n \gamma^{i-1} = \frac{1-\gamma^n}{1-\gamma}$ für $\gamma \rightarrow 1$ gegen n konvergiert, der RMA-Schätzer also Grenzfall des EMA-Schätzers ist. Beide Schätzer können nun in (15.8) und (15.9) eingesetzt werden, und wir erhalten mit

$$\hat{P}_{t+1} = N(0, \hat{\sigma}_t^2), \quad \hat{\sigma}_t^2 = w_t^T \hat{\Sigma}_t w_t$$

eine Approximation der Vorhersageverteilung, d.h. der bedingten Verteilung von L_{t+1} . Es sollte jedoch bemerkt werden, dass das Bundesaufsichtsamt für das Kreditwesen derzeit das RMA Verfahren vorschreibt.

Der *Value at Risk* VaR wird für gegebenes Niveau α durch

$$VaR_t = F_{t+1}^{-1}(\alpha) := \inf\{x; F_{t+1}(x) \geq \alpha\} \quad (15.12)$$

bestimmt und durch

$$\widehat{VaR}_t = \hat{F}_{t+1}^{-1}(\alpha) := \inf\{x; \hat{F}_{t+1}(x) \geq \alpha\} \quad (15.13)$$

geschätzt. Hierbei bezeichnen F_{t+1}, \hat{F}_{t+1} die Verteilungsfunktionen von P_{t+1}, \hat{P}_{t+1} . Hauptaugenmerk bei der Beurteilung eines VaR-Verfahrens ist die Güte der Vorhersage. Sie kann empirisch durch realisierte Werte $(\hat{P}_t, L_t), t = 1, \dots, N$, überprüft werden. Falls die Modellannahme, z.B. (15.8) und (15.9), für die Form der Vorhersageverteilung korrekt ist, so sollten die Stichproben $U_t = F_t(L_t), t = 1, \dots, N$, exakt und $\hat{U}_t = \hat{F}_t(L_t), t = 1, \dots, N$, näherungsweise unabhängig identisch uniform auf dem Intervall $[0, 1]$ verteilte Zufallsgrößen sein. Dann ist die Güte der an die Daten angepassten Vorhersageverteilung zufriedenstellend.

15.2. Backtesting mit erwartetem Shortfall

Im folgenden betrachten wir den erwarteten Shortfall von L_{t+1} als Alternative zum VaR und entwickeln Backtesting-Methoden für dieses Risikomaß. Der *erwartete Shortfall*, auch *Tail-VaR* genannt, ist im Delta-Normal Modell, d.h. unter der Annahme von (15.8) und (15.9), durch

$$E(L_{t+1} \mid L_{t+1} > VaR_t) = E(L_{t+1} \mid L_{t+1} > z_\alpha \sigma_t) \quad (15.14)$$

$$= \sigma_t E(L_{t+1}/\sigma_t \mid L_{t+1}/\sigma_t > z_\alpha) \quad (15.15)$$

definiert. Hierbei bezeichnet $z_\alpha = \Phi^{-1}(\alpha)$ das α -Quantil der Standardnormalverteilung, wobei Φ die Verteilungsfunktion der Standardnormalverteilung ist.

Unter dem Modell (15.8) und (15.9) ist $Z_{t+1} = L_{t+1}/\sigma_t$ standardnormalverteilt. Für einen festen Schwellenwert u erhalten wir daher

$$\vartheta = E(Z_{t+1} | Z_{t+1} > u) = \frac{\varphi(u)}{1 - \Phi(u)} \quad (15.16)$$

$$\zeta^2 = \text{Var}(Z_{t+1} | Z_{t+1} > u) = 1 + u \cdot \vartheta - \vartheta^2, \quad (15.17)$$

wobei φ die Dichte der Standardnormalverteilung ist. Für eine gegebene Beobachtungsreihe von Vorhersageverteilungen und Realisationen $(\hat{F}_{t+1}(\cdot/\hat{\sigma}_t), L_{t+1}/\hat{\sigma}_t)$ betrachten wir (15.16) als den uns interessierenden Parameter. Ersetzen wir den Erwartungswert durch ein Stichprobenmittel und die nicht beobachtbaren Z_{t+1} durch

$$\hat{Z}_{t+1} = \frac{L_{t+1}}{\hat{\sigma}_t}, \quad (15.18)$$

wobei σ_t in (15.9) durch (15.10) oder (15.11) geschätzt werden kann, erhalten wir als Schätzer für ϑ

$$\hat{\vartheta} = \frac{1}{N(u)} \sum_{t=0}^n \hat{Z}_{t+1} I(\hat{Z}_{t+1} > u). \quad (15.19)$$

$I(Z > u)$ bezeichnet dabei die Indikatorvariable, die 1 ist, wenn Z die Schwelle u überschreitet, und 0 sonst. $N(u)$ ist die zufällige Anzahl der Überschreitungen der Schwelle u :

$$N(u) = \sum_{t=1}^n I(\hat{Z}_{t+1} > u).$$

Inferenz über den erwarteten Shortfall, d.h. über die Differenz $\hat{\vartheta} - \vartheta$, erhalten wir aus folgendem asymptotischen Resultat:

$$\mathcal{L}\left(\sqrt{N(u)} \frac{\hat{\vartheta} - \vartheta}{\hat{\zeta}}\right) \longrightarrow N(0, 1) \quad (15.20)$$

(15.20) kann benutzt werden, um die Adäquatheit des Delta-Normal Modells zu überprüfen.

15.3. Backtesting in Aktion

Die Daten, die wir für diesen Abschnitt verwenden, stammen aus einem Bond-Portfolio einer deutschen Bank der Jahre 1994 und 1995. Das Portfolio wurde nicht verändert, so dass die Exposurevektoren $w_t = w$ zeitunabhängig sind. Wir gehen von der Gültigkeit der Annahmen (15.8) und (15.9) aus. Die VaR-Vorhersagen basieren auf den beiden in Abschnitt 15.1 vorgestellten Prädiktionsregeln, die den RMA bzw. den EMA mit $\gamma = 0.94$ zum Schätzen des Parameters σ_t der Vorhersageverteilung einsetzen. Angesichts der Bondkrise 1994 ist es interessant wie beide Verfahren unter diesem Stressfaktor reagieren.

Als Signifikanzniveau betrachten wir $\alpha = 1\%$ für hohe Verluste und $\alpha = 99\%$ für große Gewinne. Als explorative Verfahren setzen wir Zeitreihenplots der realisierten P/L- (d.h.

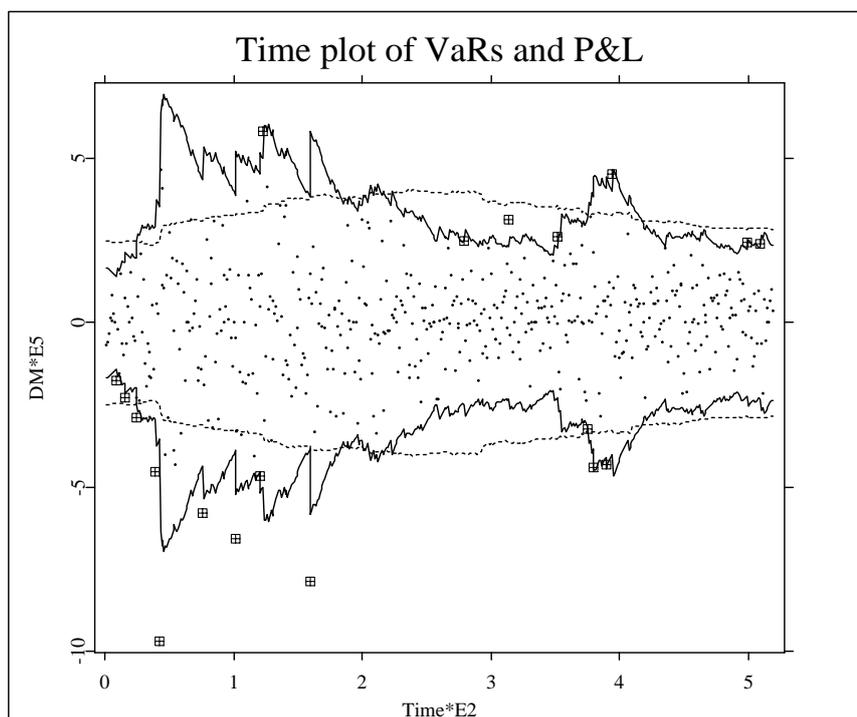


Abbildung 15.1.: Die Punkte zeigen die beobachteten Veränderungen L_t des Portfolio-Wertes. Die gestrichelte Linie zeigt die vorhergesagten VaRs basierend auf der RMA (99% und 1%). Die durchgehende Linie zeigt das gleiche für EMA. [SFMVaRtimeplot.xpl](#)

profit-loss-) Daten L_t im Vergleich zu den jeweiligen mit (15.13) berechneten VaR-Schätzern \widehat{VaR}_t ein. Wenn das Modell und das Schätzverfahren für den Parameter σ_t der Vorhersageverteilung adäquat sind, so sollten je etwa 1% der Daten unterhalb des 1%- bzw. oberhalb des 99%-VaR-Schätzers liegen. Zusätzlich sind in Abbildung 15.1 die Überschreitungen für den Fall des mit EMA geschätzten VaR markiert. Wir erkennen, dass in 1994 (1995) insgesamt 9 (4) Überschreitungen für die EMA Methode festgestellt wurden. Dies ist in starkem Kontrast zu den beobachteten 13 (3) Werten für die RMA Methode zu sehen. Es wird deutlich, dass das RMA-Verfahren vor allem während der Bondkrise in 1994 zu oft zu Überschreitungen des 1%-VaR-Schätzers führt, d.h. dass die Wahrscheinlichkeit hoher Verluste unterschätzt wird. Diese Tendenz zur Risikounterschätzung wird offensichtlich durch die Fensterbreite von 250 Tagen erzeugt, wenn der Markt von einer ruhigen in eine volatile Phase übergeht. Im umgekehrten Fall überschätzt der RMA das Risiko. Der EMA adaptiert sich schneller an die Marktphasen, da durch die exponentiell abfallenden Gewichte weiter zurückliegende Daten den Schätzer in geringerem Maße beeinflussen.

Die nicht so gute Vorhersagequalität des RMA, insbesondere für die linke Flanke der Verteilung, wird auch dadurch deutlich, dass für einen bestimmten Tag der VaR um 400% überschritten wird. Falls das Modell (15.8) - (15.9) korrekt wäre, müsste die Variable (15.21)

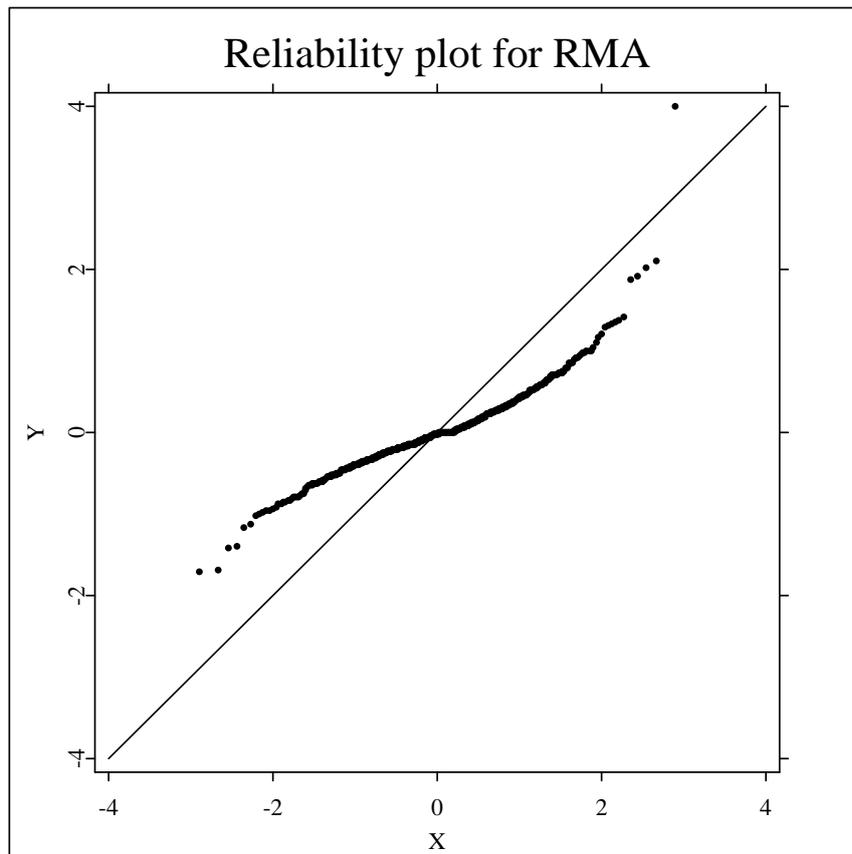


Abbildung 15.2.: Q-Q-Diagramm der L_{t+1}/\widehat{VaR}_t für RMA in 1994. [SFMVaRqqplot.xpl](#)

eine Standardabweichung von ca. 0.41 haben. Die aus den Daten berechnete empirische Standardabweichung dieser Größe liegt jedoch bei 0.62. In der Volatilitätsskala der RMA wird das Risiko daher im Mittel um $\frac{0.62-0.41}{0.41} \approx 50\%$ unterschätzt. Der EMA-Plot von Abbildung 15.1 zeigt, dass hier eine bessere Kalibrierung vorliegt. Die empirische Standardabweichung von (15.21) liegt in diesem Fall bei 0.5, was einer Risikounterschätzung von nur etwa 25% entspricht.

Als weiteres diagnostisches Maß wird der Q-Q-Plot der Variablen

$$\frac{L_{t+1}}{\widehat{VaR}_t} = \frac{L_{t+1}}{2.33\hat{\sigma}_t} \quad (15.21)$$

eingesetzt, siehe Abbildung 15.2 und Abbildung 15.3. Falls die VaR-Vorhersagen \widehat{VaR}_t perfekt wären, sollte der Q-Q Plot eine gerade Linie bilden und den Bereich von $[-1, 1]$ ausfüllen.

Ein Vergleich der Graphen in Abbildung 15.2 und Abbildung 15.3 zeigt, dass die EMA-Methode besser kalibriert ist als die RMA-Methode. Die RMA-Methode zeigt deutlich Ausreißer an beiden Enden. Die Intervallgrenzen von $[-1, 1]$ werden in beiden Fällen deutlich

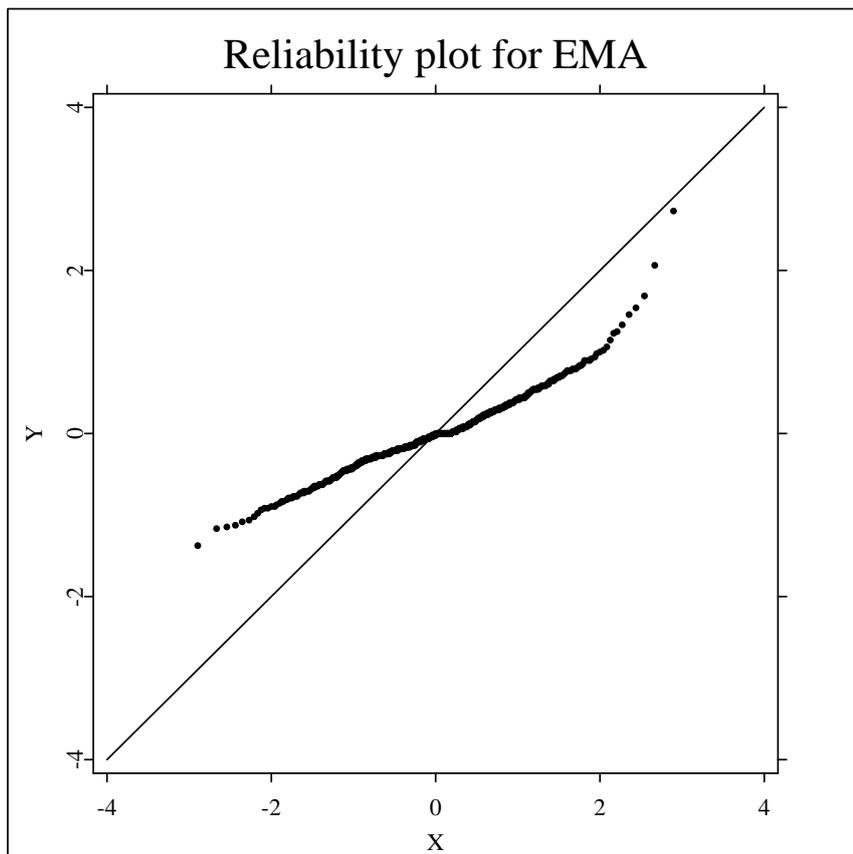


Abbildung 15.3.: Q-Q-Diagramm der L_{t+1}/\widehat{VaR}_t für EMA in 1994.  [SFMVaRqqplot.xpl](#)

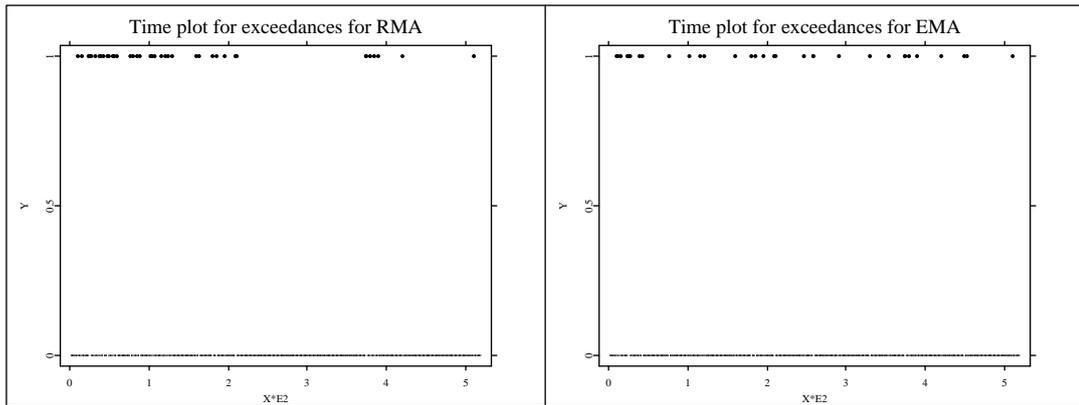


Abbildung 15.4.: Zeitdiagramm der Überschreitungen über das 80% Niveau von VaR für RMA (links) und EMA. Die bessere Lösung mit EMA ist einleuchtend.

[SFMVaRtimeplot.xpl](#)

überschritten. Dies weist auf eine mögliche Inadäquatheit der Annahme der Normalverteilung hin. Q-Q-Plots für das Jahr 1995 sind nicht aufgeführt. Auch sie zeigen jedoch deutlich die Überlegenheit von EMA gegenüber RMA.

Eine weitere wesentliche Annahme unseres Modells ist die Unabhängigkeit der reskalierten Zufallsvariablen Z_t . Abbildung 15.4 zeigt die Überschreitungen

$$\{t, I(L_{t+1} > \widehat{VaR}_t)\}, \quad t = 1, \dots, 260, \quad (15.22)$$

als Funktion von t aufgetragen. Die der Unabhängigkeit der Z_t widersprechende Häufung von Überschreitungen ist bei der RMA-Methode weitaus deutlicher sichtbar als bei der EMA-Methode.

Bereits die explorative Analyse zeigt deutlich Unterschiede zwischen RMA und EMA. Als Ergänzung vergleichen wir die beiden Schätzverfahren jetzt mit geeigneten Tests im Rahmen des Modells (15.8) - (15.9). Wir betrachten wieder die Stichprobenresiduen \hat{Z}_{t+1} aus (15.18) und setzen den Schwellenwert in (15.16) auf den Wert $u = 0.8416$, d.h. auf das 80%-Quantil der Verteilung von $Z_{t+1} = \frac{L_{t+1}}{\hat{\sigma}_t}$. Daraus ergibt sich $\vartheta = 1.4$ nach (15.16). Auf Grund der asymptotischen Verteilung (15.20) können wir die Signifikanz der Hypothese

$$H_0 : \vartheta \stackrel{(<)}{=} 1.4 \quad (15.23)$$

überprüfen. Eine für den vorliegenden Stichprobenumfang bessere Approximation als die Standardnormalverteilung erhalten wir mit der Student $t(20)$ -Verteilung, wenn wir generalisierte Freiheitsgrade heranziehen.

$$\mathcal{L}(\hat{Z}_{t+1}) = \mathcal{L}\left(\frac{L_{t+1}}{\hat{\sigma}_t}\right) \approx t(20). \quad (15.24)$$

Der sich daraus ergebende Wert von ϑ unterscheidet sich von dem obigen Wert um 5%, die entsprechenden Varianzen ζ^2 jedoch um 18%. Wir betrachten daher auch die Hypothese

$$H_0 : \vartheta \stackrel{(\leq)}{=} 1.47. \quad (15.25)$$

Die folgenden Tabellen (15.1) bis (15.4) fassen unsere Ergebnisse zusammen.

Method	$\vartheta = 1.4$	$\zeta = 0.46$	$\sqrt{N(u)} \frac{\hat{\vartheta} - \vartheta}{\hat{\zeta}}$	significance	$N(u)$
EMA	$\hat{\vartheta} = 1.72$	$\hat{\zeta} = 1.01$	2.44	0.75%	61
RMA	$\hat{\vartheta} = 1.94$	$\hat{\zeta} = 1.3$	3.42	0.03%	68

Tabelle 15.1.: $H_0 : \vartheta \stackrel{(\leq)}{=} 1.4$

Method	$\vartheta = 1.47$	$\zeta = 0.546$	$\sqrt{N(u)} \frac{\hat{\vartheta} - \vartheta}{\hat{\zeta}}$	significance	$N(u)$
EMA	$\hat{\vartheta} = 1.72$	$\hat{\zeta} = 1.01$	2.01	2.3%	61
RMA	$\hat{\vartheta} = 1.94$	$\hat{\zeta} = 1.3$	3.04	0.14%	68

Tabelle 15.2.: $H_0 : \vartheta \stackrel{(\leq)}{=} 1.47$

Aus den Tabellen (15.1) und (15.2) wird ersichtlich, dass die beobachteten Überschreitungen für EMA weitaus besser kalibriert sind als für die RMA-Methode. Für eine Stichprobe von 260 Werten erwarten wir 52 Überschreitungen (Standardabweichung 6.45). Für den EMA beobachten wir 61 ($61 - 52 \approx 1.5 \cdot$ Standardabweichung) Überschreitungen und für den RMA 68 ($68 - 52 \approx 2.5 \cdot$ Standardabweichung). Natürlich beeinflussen die Ausreißer den Test erheblich. Wir wiederholen daher die Analyse ohne die extremen Ausreißer und erhalten (15.3) und (15.4).

Method	$\vartheta = 1.4$	$\zeta = 0.46$	$\sqrt{N(u)} \frac{\hat{\vartheta} - \vartheta}{\hat{\zeta}}$	significance	$N(u)$
EMA	$\hat{\vartheta} = 1.645$	$\hat{\zeta} = 0.82$	2.31	1%	60
RMA	$\hat{\vartheta} = 1.83$	$\hat{\zeta} = 0.93$	3.78	0.00%	67

Tabelle 15.3.: $H_0 : \vartheta \stackrel{(\leq)}{=} 1.4$ —largest outlier excluded

Zusammenfassend können wir feststellen, dass die EMA Methode bessere Kalibrationsergebnisse liefert als die RMA Methode. Beide Methoden sind jedoch stark ausreißersensitiv und sollten parallel betrachtet werden. Auch die EMA-Methode leidet darunter, dass unsere auf dem Delta-Normal-Modell basierenden Annahmen (15.8) - (15.9) nur in grober Näherung erfüllt sind. Die Residuen Z_t sind weder normalverteilt noch unabhängig, wobei allerdings die EMA-Methode dank ihres exponentiell abfallenden Gedächtnisses weniger unter den Auswirkungen der Abhängigkeit leidet.

Method	$\vartheta = 1.47$	$\varsigma = 0.546$	$\sqrt{N(u)} \frac{\hat{\vartheta} - \vartheta}{\hat{\varsigma}}$	significance	$N(u)$
EMA	$\hat{\vartheta} = 1.645$	$\hat{\varsigma} = 0.82$	1.65	5%	60
RMA	$\hat{\vartheta} = 1.83$	$\hat{\varsigma} = 0.93$	3.1	0.15%	67

Tabelle 15.4.: $H_0 : \vartheta \stackrel{(<)}{=} 1.47$ —largest outlier excluded

15.4. Literaturhinweise

McAllister and Mingo (1996) führen die Vorteile von (15.14) in einem RAROC (risk-adjusted return on capital) setup an. Artzner, Dealban, Eber and Heath (1997), Jaschke and Küchler (1999) zeigen, dass (15.14) als eine sinnvolle Approximation für eine worst case Situation darstellt. Leadbetter, Lindgren and Rootzen (1983) zeigen, wie (15.14) im Kontext von Extremwerttheorien verwendet wird.

TO DO:
 Zitat:
 Lead-
 better,
 Lindgren,
 ... fehlt
 TO DO:
 Artzner,
 Delbaen
 et al.: er-
 warteter
 Short-
 fall ist
 kohärentes
 Risikomaß

16. Neuronale Netze

Ein neuronales Netz ist ein nichtlineares System, das eine Reihe von reellen Eingangsvariablen x_1, \dots, x_p über mehrere Zwischenschritte in eine oder mehrere Ausgangsvariable y_1, \dots, y_q überführt. Es repräsentiert eine Funktion $\nu : \mathbb{R}^p \rightarrow \mathbb{R}^q$:

$$(y_1, \dots, y_q)^T = \nu(x_1, \dots, x_p),$$

die eine spezielle, durch die Netzstruktur gegebene Form hat. Graphisch wird dies wie in Abbildung 16.1 in Form eines gerichteten Graphen dargestellt, dessen Knoten zu verschiedenen Schichten gruppiert sind. In der *Eingangsschicht* (*Input layer*) steht jeder Knoten für eine der Eingangsvariablen; in der *Ausgangsschicht* (*Output layer*) repräsentiert jeder Knoten eine der Ausgangsvariablen. Dazwischen liegen eine oder mehrere *verborgene Schichten* (*Hidden layers*), deren Knoten weder Quellen noch Senken des Graphen sind. Das Netz in Abbildung 16.1 besitzt nur eine verborgene Schicht. Darüberhinaus handelt es sich um ein *Feedforward-Netz*, da es keine Kanten enthält, die in einem Knoten beginnen und in demselben oder einem anderen Knoten aus derselben oder einer vorausgehenden Schicht enden.

Abbildung 16.2 zeigt ein *Feedback-Netz*, in dem Rückkopplungen zwischen Knoten der bei-

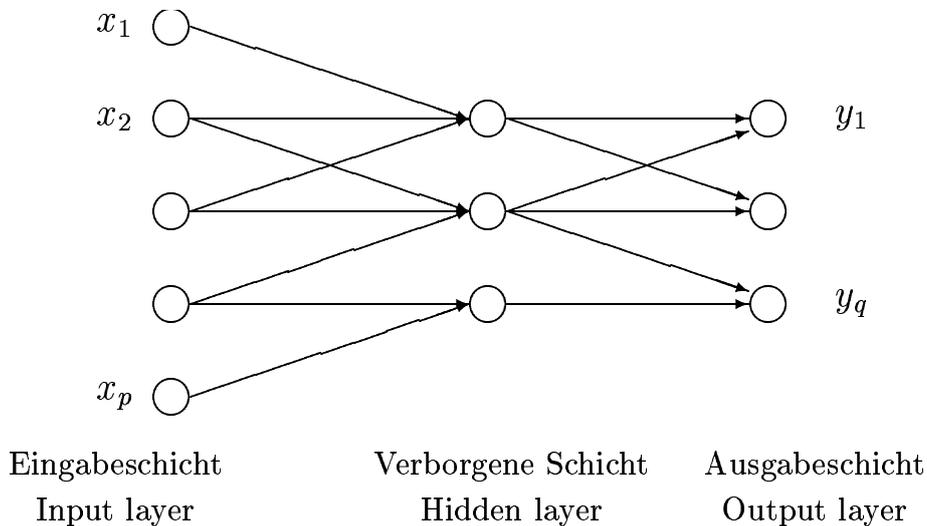


Abbildung 16.1.: Neuronales-Feedforward-Netz mit einer verborgenen Schicht

den verborgenen Schichten vorkommen. Im folgenden beschränken wir uns auf Feedforward-Netze.

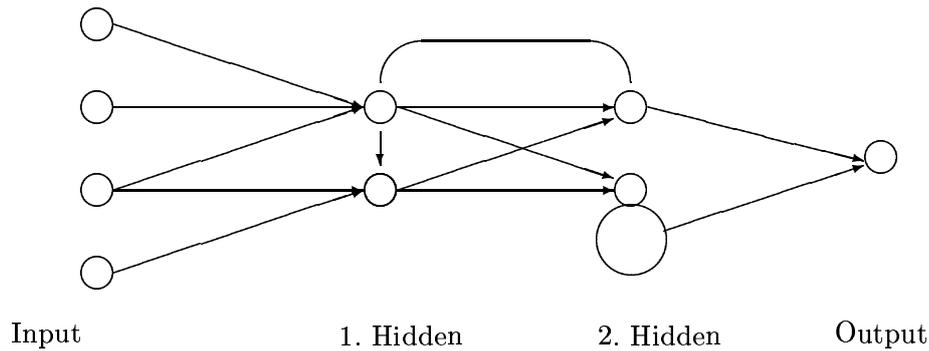


Abbildung 16.2.: Neuronales-Feedback-Netz mit zwei verborgenen Schichten

Neuronale Netze werden in der Finanzstatistik zur Darstellung von Funktionen benutzt, die zum Beispiel die Ausfallwahrscheinlichkeit eines Kredits, die Vorhersage eines Wechselkurses oder die Volatilität einer Aktie sein können. Es handelt sich dabei um einen nichtparametrischen Ansatz, der im Vergleich zu den in Kapitel 17 betrachteten lokalen Glättungsverfahren einen hohen Modellierungs- und Rechenaufwand erfordert, dafür aber auch dann noch praktikabel ist, wenn man sehr viele Variable beim Vorhersagen oder Risikoquantifizieren berücksichtigen möchte, d.h. wenn die Dimension p des Funktionsargumentes groß ist.

TO DO:
Referenzen
auf Kapitel
17

Da neuronale Netze ein in der Statistik noch vergleichsweise unbekanntes Werkzeug sind, geben wir im ersten Abschnitt eine elementare Einführung in die Struktur neuronaler Netze. Sie ermöglichen es, komplexe Funktionen aus einfachen Bausteinen zusammensetzen. Im zweiten Abschnitt beschreiben wir ein populäres numerisches Verfahren zur Anpassung von neuronalen Netzen an Daten, bevor wir anschließend verschiedene Anwendungen auf Finanzprobleme und die dazu benötigten Grundlagen vorstellen.

16.1. Vom Perzeptron zum nichtlinearen Neuron

Das Perzeptron ist ein einfaches mathematisches Modell für das Funktionieren einer Nervenzelle, die Signale von Sinneszellen und anderen Nervenzellen empfängt (die Eingangsvariablen) und davon abhängig „feuert“, d.h. ein Signal an nachgeordnete Nervenzellen sendet, oder inaktiv bleibt. Trotz aller Nachteile war das Perzeptron sehr einflussreich für die Denkweise im Zusammenhang mit neuronalen Netzen, so dass es einen guten Einstieg in die Diskussion der Bausteine bietet, aus denen neuronale Netze gebildet werden. Das Perzeptron arbeitet in zwei Schritten:

- die Eingangsvariablen x_1, \dots, x_p werden mit *Gewichten* w_1, \dots, w_p multipliziert und addiert,

– das Ergebnis wird einer *Schwellenoperation* unterzogen.

$x = (x_1, \dots, x_p)^T$, $w = (w_1, \dots, w_p)^T$ bezeichne die Input- und Gewichtsvektoren, und für ein b sei $\psi(u) = 1_{(b, \infty)}(u)$ die zugehörige Schwellenfunktion. Die Ausgangsvariable $y = \psi(w^T x)$ des Perzeptrons ist also 1 (die Nervenzelle „feuert“), wenn die Summe der gewichteten Eingangssignale oberhalb der Schwelle b liegt, und sonst 0 (die Nervenzelle bleibt inaktiv). Die Wirkungsweise des Perzeptrons hängt von den Gewichten w_1, \dots, w_p und dem

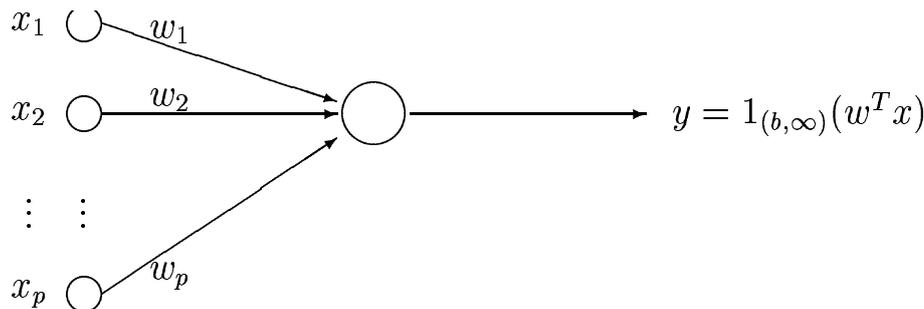


Abbildung 16.3.: Das Perzeptron

Schwellenwert b ab. Eine äquivalente Darstellung erhält man, indem man als zusätzliche Eingangsvariable die Konstante $x_0 \equiv 1$ einführt, diese mit dem Gewicht $w_0 = -b$ versieht und als Schwellenwert 0 wählt, da dann

$$1_{(b, \infty)} \left(\sum_{i=1}^p w_i x_i \right) = 1_{(0, \infty)} \left(\sum_{i=0}^p w_i x_i \right).$$

Diese Darstellung ist oft bequemer, da bei den frei wählbaren Systemparametern nicht zwischen Gewichten und Schwellenwert unterschieden werden muss.

Ein Perzeptron kann trainiert werden, Klassifikationsaufgaben des folgenden Typs zu lösen: Gegeben seien Objekte, die zu einer von zwei Klassen C_0 oder C_1 gehören. Entscheide anhand der an einem Objekt gemachten Beobachtungen x_1, \dots, x_p , ob es zu C_0 oder zu C_1 gehört.

Das durch die Gewichte w_0, \dots, w_p charakterisierte Perzeptron klassifiziert ein Objekt als zu C_0 bzw. C_1 gehörig, wenn die Ausgangsvariable $y = y(x_1, \dots, x_p)$ 0 bzw. 1 ist. Damit die Klassifikationsaufgabe „möglichst gut“ gelöst wird, müssen die Gewichte w_0, \dots, w_p „gelernt“ werden. Dazu steht ein *Trainingsset* $(x^{(1)}, z^{(1)}), \dots, (x^{(T)}, z^{(T)})$ von T Inputvektoren $x^{(t)} = (x_1^{(t)}, \dots, x_p^{(t)})^T$ zur Verfügung, deren korrekte Klassifikationen $z^{(1)}, \dots, z^{(T)} \in \{0, 1\}$ bekannt sind. Mit Hilfe von *Lernregeln* werden aus dem Trainingsset geeignete Gewichte $\hat{w}_0, \dots, \hat{w}_p$ bestimmt.

In statistischer Sprache besteht die Aufgabe darin, aus den Daten $(x^{(t)}, z^{(t)}), t = 1, \dots, T$, die

Parameter des Perzeptrons zu schätzen. Eine *Lernregel* ist ein Schätzverfahren, das Schätzwerte $\hat{w}_0, \dots, \hat{w}_p$ liefert.

Eine Lernregel ist z.B. die Delta- oder *Widrow-Hoff-Lernregel*: Die Inputvektoren $x^{(t)}, t = 1, \dots, T$, werden nacheinander als Eingangsvariablen des Perzeptrons benutzt, und die Ausgangsvariablen $y^{(t)}, t = 1, \dots, T$, werden mit ihren Zielwerten, den korrekten Klassifikationen $z^{(t)}, t = 1, \dots, T$, verglichen. Wenn in einem Schritt $y^{(t)} = z^{(t)}$, so bleiben die Gewichte unverändert. Wenn dagegen $y^{(t)} \neq z^{(t)}$, so wird der Gewichtsvektor $w = (w_0, \dots, w_p)^T$ folgendermaßen angepasst:

$$w_{neu} = w + \eta(z^{(t)} - y^{(t)}) x^{(t)}$$

η ist ein kleiner Relaxationsfaktor, der eventuell langsam gegen 0 gehen muss, um eine Konvergenz des Lernalgorithmus zu ermöglichen. Der Startwert für w wird willkürlich vorgegeben oder zufällig bestimmt, z.B. uniform verteilt aus $[0, 1]^{p+1}$.

Das Lernen endet nicht, wenn alle Inputvektoren dem Netz präsentiert worden sind, sondern nach der Eingabe von $x^{(T)}$ wird als nächster Inputvektor wieder $x^{(1)}$ benutzt. Das Trainingsset wird mehrmals durchlaufen, bis das Netz alle Objekte im Trainingsset korrekt klassifiziert bzw. wenigstens ein gegebenes Qualitätskriterium für das Ausmaß der Fehlklassifikationen klein geworden ist.

Die Gewichte w_0, \dots, w_p sind nur bis auf einen positiven Skalenfaktor identifizierbar, d.h. für $\alpha > 0$ führen $\alpha w_0, \dots, \alpha w_p$ zur selben Klassifikation. Beim Anwenden einer Lernregel wie der von Widrow und Hoff kann es daher vorkommen, dass $\|w\|$ immer größer wird, was zu numerischen Problemen führen kann. Um dies zu vermeiden, verwendet man die Technik des sogenannten *weight decay*, d.h. modifizierte Lernregeln, bei denen $\|w\|$ stabil bleibt.

BEISPIEL 16.1 (Lernen der ODER-Funktion)

Sei $p = 2$ und $x_1, x_2 \in \{0, 1\}$. Die zu lernende Klassifikation sei das logische ODER:

$$\begin{aligned} z &= 1, \text{ wenn } x_1 = 1 \text{ oder } x_2 = 1, \\ z &= 0, \text{ wenn } x_1 = 0 \text{ und } x_2 = 0. \end{aligned}$$

Als Trainingsset dienen die Inputvektoren incl. der ersten Koordinate $x_0 = 1$

$$x^{(1)} = \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix}, x^{(2)} = \begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix}, x^{(3)} = \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}, x^{(4)} = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}$$

mit den korrekten Klassifikationen $z^{(1)} = z^{(2)} = z^{(3)} = 1, z^{(4)} = 0$. Das Perzeptron mit Gewichten w_0, w_1, w_2 klassifiziert ein Objekt als 1 genau dann, wenn

$$w_0 x_0 + w_1 x_1 + w_2 x_2 > 0,$$

und als 0 sonst. Als Startvektor benutzen wir $w = (0, 0, 0)^T$, und wir setzen $\eta = 1$. Die einzelnen Schritte des Widrow-Hoff-Lernens sehen nun folgendermaßen aus:

1. $w^{(1)}$ liefert $y^{(1)} = 0 \neq z^{(1)}$. Die Gewichte werden geändert:
 $w_{neu} = (0, 0, 0)^T + (1 - 0)(1, 1, 0)^T = (1, 1, 0)^T$
2. $x^{(2)}$ wird mit diesem Gewichtsvektor korrekt klassifiziert.
3. $x^{(3)}$ wird mit diesem Gewichtsvektor korrekt klassifiziert.
4. Für $x^{(4)}$ ist $w^T x^{(4)} = 1$, so dass die Gewichte erneut geändert werden:
 $w_{neu} = (1, 1, 0)^T + (0 - 1)(1, 0, 0)^T = (0, 1, 0)^T$
5. $x^{(1)}$ wird erneut als Eingabe benutzt und korrekt klassifiziert.
6. Da $w^T x^{(2)} = 0$:
 $w_{neu} = (0, 1, 0)^T + (1 - 0)(1, 0, 1)^T = (1, 1, 1)^T$
7. Da $w^T x^{(3)} = 3 > 0$, wird $x^{(3)}$ korrekt klassifiziert.
8. $x^{(4)}$ wird falsch klassifiziert, so dass
 $w_{neu} = (1, 1, 1)^T + (0 - 1)(1, 0, 0)^T = (0, 1, 1)^T$

Damit endet das Verfahren, da das Perzeptron mit diesem Gewicht alle Inputvektoren im Trainingsset korrekt klassifiziert. Das Perzeptron hat die ODER-Funktion auf der Menge $\{0, 1\}^2$ gelernt.

Man unterscheidet verschiedene Arten des Lernens für neuronale Netze:

Supervised Learning: Vergleiche Netzausgabe $y = y(x_1, \dots, x_p)$ mit Zielvorgabe $z = z(x_1, \dots, x_p)$. Falls $y \neq z$, verändere Gewichte nach Lernregel.

Reinforcement Learning: Zu jeder Netzausgabe $y = y(x_1, \dots, x_p)$ erfährt man, ob sie „richtig“ oder „falsch“ ist - im zweiten Fall aber nicht, was der richtige Wert ist. Falls y „falsch“, verändere Gewichte nach Lernregel.

Unsupervised Learning: Keine Rückmeldung während des Lernens. Ähnlich wie in der Clusteranalyse werden mit Hilfe redundanter Informationen Zufallsfehler aus den Daten herausgefiltert.

Für $y \in \{0, 1\}$ fallen Supervised und Reinforcement Learning zusammen. Zu diesem Typ zählt die Widrow-Hoff-Lernregel für das Perzeptron.

Das Perzeptron kann nicht alle gewünschten Klassifikationen lernen. Das klassische Gegenbeispiel ist das logische ENTWEDER - ODER (XOR = "exclusive or"):

$$\begin{aligned} z &= 1, \text{ wenn entweder } x_1 = 1 \text{ oder } x_2 = 1, \\ z &= 0, \text{ wenn } x_1 = x_2 = 0 \text{ oder } x_1 = x_2 = 1. \end{aligned}$$

Einem Perzeptron mit Gewichten w_0, w_1, w_2 entspricht eine Hyperebene $w_0 + w_1 x_1 + w_2 x_2 = 0$ im Raum \mathbb{R}^2 der Inputs $(x_1, x_2)^T$, die die Menge der durch das Perzeptron als 0 klassifizierten Objekte von den als 1 klassifizierten Objekten trennt. Man überlegt sich leicht, dass es

keine Hyperebene gibt, die die beim ENTWEDER - ODER als 1 zu klassifizierenden Inputs $\begin{pmatrix} 1 \\ 0 \end{pmatrix}$, $\begin{pmatrix} 0 \\ 1 \end{pmatrix}$ von den als 0 zu klassifizierenden $\begin{pmatrix} 0 \\ 0 \end{pmatrix}$, $\begin{pmatrix} 1 \\ 1 \end{pmatrix}$ trennt.

DEFINITION 16.1 (linear separierbar)

Für $p \geq 1$ heißen zwei Teilmengen $\mathcal{X}_0, \mathcal{X}_1 \subseteq \mathbb{R}^p$ linear separierbar, wenn $w \in \mathbb{R}^p$, $w_0 \in \mathbb{R}$ existieren mit

$$\begin{aligned} w_0 + w^T x &> 0 \quad \text{für } x \in \mathcal{X}_1, \\ w_0 + w^T x &\leq 0 \quad \text{für } x \in \mathcal{X}_0. \end{aligned}$$

Das Perzeptron mit p Eingangsvariablen x_1, \dots, x_p (zzgl. der Konstanten $x_0 \equiv 1$) kann genau die Klassifikationen lernen, die linear separierbaren Mengen von Inputs entsprechen.

Wenn keine perfekte Klassifikation durch ein Perzeptron möglich ist, kann man wenigstens versuchen, eine "gute" Klassifikation zu finden, d.h. die Gewichte w_0, \dots, w_p so zu bestimmen, dass ein Maß für das Ausmaß der Fehlklassifikationen minimiert wird. Ein Beispiel für einen solchen Ansatz liefert die *Kleinste-Quadrate-(KQ-) Klassifikation*:

Gegeben das Trainingsset $(x^{(1)}, z^{(1)}), \dots, (x^{(T)}, z^{(T)})$. Bestimme zu fest vorgegebenem w_0 die Gewichte w_1, \dots, w_p so, dass

$$\begin{aligned} Q(w) &= Q(w_1, \dots, w_p) = \sum_{i=1}^T (z^{(i)} - y^{(i)})^2 = \min! \\ \text{mit } y^{(i)} &= 1_{(0,\infty)}(w_0 + w^T x^{(i)}), \quad w = (w_1, \dots, w_p)^T \end{aligned}$$

w_0 kann beliebig vorgegeben werden, da die Gewichte w_0, \dots, w_p wie oben beschrieben nur bis auf ein Skalenfaktor bestimmt sind. $Q(w)$ ist im Fall des Perzeptrons, das eine binäre Klassifikation vornimmt, einfach die Zahl der Fehlklassifikationen. Die obige Form lässt sich aber direkt auf andere Probleme übertragen. Das erreichbare Minimum von $Q(w)$ ist genau dann 0 (perfekte Klassifikation des Trainingssets), wenn die beiden Mengen

$$\mathcal{X}_0^{(T)} = \{x^{(i)}, i \leq T; z^{(i)} = 0\}, \quad \mathcal{X}_1^{(T)} = \{x^{(i)}, i \leq T; z^{(i)} = 1\}$$

linear separierbar sind.

Die Widrow-Hoff-Lernregel löst das KQ-Klassifikationsproblem; es gibt aber eine Reihe anderer Lernregeln oder Schätzverfahren, die dies ebenfalls leisten. Das Perzeptron erweist sich für viele Anwendungen als zu inflexibel. Daher betrachtet man allgemeinere Typen von *Neuronen* als Bausteine neuronaler Netze:

$x = (x_1, \dots, x_p)^T$, $w = (w_1, \dots, w_p)^T$ seien Input- und Gewichtsvektoren. Zu $\beta, \beta_0 \in \mathbb{R}$ sei

$$\psi_\beta(t) = \frac{1}{1 + \exp(-\frac{t+\beta}{\beta_0})}$$

die logistische Funktion, wegen ihrer Gestalt oft auch "die" *sigmoide Funktion* genannt, obwohl man an ihrer Stelle auch andere Funktionen mit sigmoider Gestalt benutzen kann,

z.B. die Verteilungsfunktion einer Normalverteilung. Die Ausgangsvariable des Neurons ist $y = \psi_\beta(w^T x)$.

Für $\beta_0 \rightarrow 0+$ geht $\psi_\beta(t)$ in eine Schwellenfunktion über:

$$\psi_\beta(t) \longrightarrow 1_{(0,\infty)}(t + \beta) \quad \text{für } \beta_0 \longrightarrow 0+ ,$$

so dass das Perzeptron Grenzfall eines Neurons mit logistischer Aktivierungsfunktion ist.

β_0 wird oft nicht explizit erwähnt, da es als Skalenfaktor in die anderen Parameter w_1, \dots, w_p, β

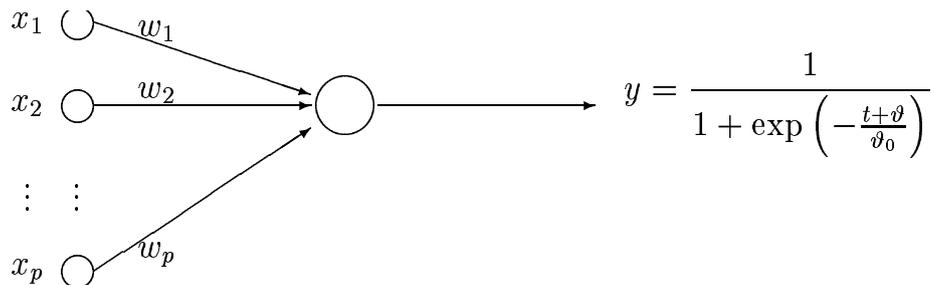


Abbildung 16.4.: Neuron mit sigmoider Übertragungsfunktion

des Neurons integriert werden kann. Setzt man zusätzlich $w_0 = \beta$ und $x_0 \equiv 1$, so lässt sich die Ausgabevariable auch in der Form schreiben:

$$y = \psi(w_0 + w^T x) = \psi\left(\sum_{k=0}^p w_k x_k\right) \quad \text{mit } \psi(t) = \frac{1}{1 + e^{-t}}.$$

Durch Verbinden mehrerer Neuronen mit sigmoider oder - im Grenzfall - Schwellen- Aktivierungsfunktion zu einem Feedforward-Netz erhält man ein *Multilayer-Perzeptron* (MLP, d.h. mehrschichtiges Perzeptron) genanntes neuronales Netz. Abbildung 16.5 zeigt ein solches neuronales Netz mit zwei Eingangsvariablen zzgl. der Konstanten $x_0 \equiv 1$, zwei sigmoiden Neuronen in der verborgenen Schicht, die über ein weiteres sigmoides Neuron zur Ausgangsvariable verbunden werden, wobei $\psi(t) = \{1 + e^{-t}\}^{-1}$ wie oben.

Entsprechend lassen sich auch neuronale Netze mit mehr als einer verborgenen Schicht kon-

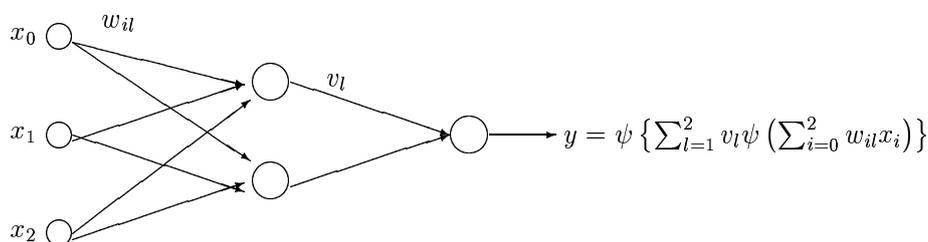


Abbildung 16.5.: Multilayer-Perzeptron mit einer verborgenen Schicht

struieren, die mehrmals eine Ausgangsvariable liefern. Die Verbindungen müssen auch nicht

vollständig sein, d.h. Kanten zwischen Knoten aufeinanderfolgender Schichten können fehlen bzw. äquivalenterweise können einige Gewichte von vorneherein 0 gesetzt werden. Statt der logistischen Funktion oder ähnlicher sigmoider Funktionen können in einzelnen Neuronen auch Schwellenfunktionen wie beim Perzeptron, einfache Linearkombinationen oder auch sogenannte *Radial Basis Functions* (radiale Basisfunktionen oder RBF) auftreten. Zu letzteren gehören die Dichte der Standardnormalverteilung und ähnliche symmetrische Kernfunktionen. Man spricht dann aber nicht mehr von einem MLP, sondern von einem RBF-Netz. Abbildung 16.6 zeigt ein unvollständiges neuronales Netz mit zwei Ausgangsvariablen. Die

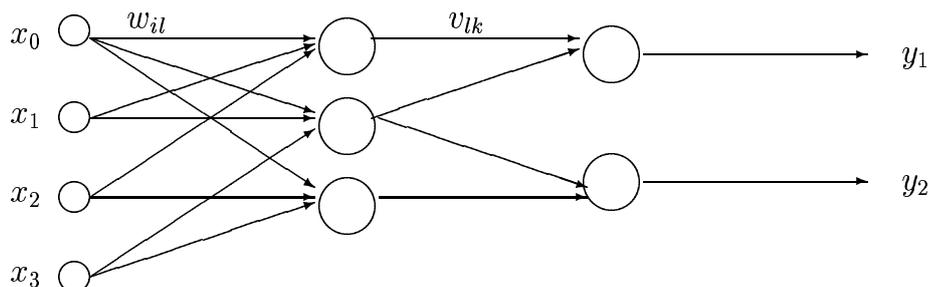


Abbildung 16.6.: Multilayer-Perzeptron mit zwei verborgenen Schichten

Gewichte $w_{13}, w_{22}, w_{31}, v_{12}$ und v_{31} sind auf 0 gesetzt, und die entsprechenden Kanten tauchen im Netzgraphen nicht auf. Die Ausgangsvariable y_1 ist z.B.

$$y_1 = v_{11}\psi(w_{01} + w_{11}x_1 + w_{21}x_2) + v_{21}\psi(w_{02} + w_{12}x_1 + w_{32}x_3),$$

eine Linearkombination der Ergebnisse der oberen beiden Neuronen der verborgenen Schicht.

Bisher haben wir nur den in der Literatur am häufigsten betrachteten Fall erwähnt, dass ein Neuron auf eine Linearkombination von Größen aus der vorhergehenden Schicht wirkt. Gelegentlich wird aber auch der Fall betrachtet, dass der Output eines Neurons von der Form $\psi(\prod_{i=1}^p w_i x_i)$ bzw. $\psi(\max_{i=1, \dots, p} x_i)$ ist.

Neuronale Netze vom MLP-Typ können für Klassifikationsprobleme, aber auch für Regressions- und Vorhersageprobleme benutzt werden. Um ein für die jeweilige Aufgabe geeignetes Netz zu finden, müssen die Gewichte anhand eines Trainingssets gelernt werden, d.h. die Netzparameter anhand von Daten geschätzt werden. Da wir uns auf den Fall des Supervised Learnings beschränken, heißt dies, dass als Trainingsset $(x^{(1)}, z^{(1)}), \dots, (x^{(T)}, z^{(T)})$ gegeben sind. Die $x^{(i)} \in \mathbb{R}^p$ sind Inputvektoren, die $z^{(i)} \in \mathbb{R}^q$ sind die korrespondierenden Zielgrößen, d.h. die als Output des Netzes gewünschten Vektoren. Die Zielvektoren $z^{(i)}$ werden mit den aktuellen Ausgangsvariablen $y^{(i)} \in \mathbb{R}^q$ des Netzes verglichen. Die Gewichte werden so bestimmt, dass die Abweichungen zwischen den $z^{(i)}$ und den $y^{(i)}$ klein sind. Ein Beispiel dafür ist der bereits beim Perzeptron betrachtete *Kleinste-Quadrate-(KQ-)Ansatz*:

Gegeben das Trainingsset $(x^{(1)}, z^{(1)}), \dots, (x^{(T)}, z^{(T)})$. Die Gewichte $w_{0l}, l = 1, \dots, r, x_0 \equiv 1$ seien vorgegeben, wobei r die Zahl der Neuronen in der ersten verborgenen Schicht ist. Bestimme die Gewichte aller anderen im Netz auftretenden Kanten (zwischen Eingangsschicht,

den verborgenen Schichten und der Ausgangsschicht) so, dass

$$\sum_{k=1}^T \|z^{(k)} - y^{(k)}\|^2 = \min!$$

In dem Netz der Abbildung 16.6 wird die Minimierung z.B. bzgl. der Gewichte $w_{11}, w_{12}, w_{21}, w_{23}, w_{32}, w_{33}, v_{11}, v_{21}, v_{22}, v_{32}$ vorgenommen. Die Gewichte w_{01}, w_{02}, w_{03} können frei vorgegeben werden, um wie beim Perzeptron die Willkürlichkeit von Skalenfaktoren zu umgehen.

Statt des KQ-Ansatzes werden auch andere Verlustfunktionen minimiert, z.B. gewichtete quadratische Abstände oder - vor allem bei Klassifikationen - der Kullback-Leibler-Abstand:

$$\sum_{k=1}^T \sum_i \left\{ z_i^{(k)} \log \frac{z_i^{(k)}}{y_i^{(k)}} + (1 - z_i^{(k)}) \log \frac{1 - z_i^{(k)}}{1 - y_i^{(k)}} \right\} = \min!$$

Da nur die $y_i^{(k)}$ von den Gewichten abhängen, ist dies äquivalent zur Minimierung der Kreuzentropie zwischen den z_i und den y_i , die jeweils in $(0, 1)$ liegen:

$$- \sum_{k=1}^T \sum_i \left\{ z_i^{(k)} \log y_i^{(k)} + (1 - z_i^{(k)}) \log(1 - y_i^{(k)}) \right\} = \min!$$

16.2. Backpropagation

Das bekannteste Verfahren, mit dem Feedforward-Netze ihre Gewichte aus dem Trainingsset lernen, ist Backpropagation. Dahinter verbirgt sich nichts anderes als ein numerisches Verfahren zur Lösung des (nichtlinearen) Kleinste-Quadrate-Problems, das sparsam mit dem Speicher umgeht, dafür aber eventuell langsame Konvergenz und numerische Instabilitäten aufweist.

Zur Illustration betrachten wir ein neuronales Netz mit einer Ausgangsvariable y (d.h. $q = 1$) und einer verborgenen Schicht mit nur einem Neuron:

$$y = \psi(w_0 + w^T x).$$

ψ kann die logistische, aber auch eine andere Übertragungsfunktion sein. Das Trainingsset sei $(x^{(1)}, z^{(1)}), \dots, (x^{(T)}, z^{(T)})$. Das Gewicht w_0 der Konstanten halten wir fest, um den willkürlichen Skalenfaktor zu eliminieren. Die zu minimierende Funktion

$$Q(w) = \sum_{k=1}^T (z^{(k)} - y^{(k)})^2$$

hängt somit nur von den Gewichten w_1, \dots, w_p der Eingangsvariablen ab.

Ein elementares numerisches Verfahren zur Minimierung von Q ist das *Gradientenabstiegsverfahren*. Ausgehend von einem Gewicht $w(N)$ ermittelt man die nächste Näherung, indem man ein Stück in Richtung des steilsten Abstiegs von Q geht:

$$\begin{aligned} w(N+1) &= w(N) - \eta \operatorname{grad} Q(w(N)), \\ \operatorname{grad} Q(w) &= - \sum_{k=1}^T 2(z^{(k)} - y^{(k)})\psi'(w^T x^{(k)})x^{(k)}. \end{aligned}$$

Zur Beschleunigung der Konvergenz kann die kleine Konstante $\eta > 0$ zusätzlich im Lauf der Iteration gegen 0 gehen.

Beim Gradientenabstiegsverfahren wird die Qualität der Gewichte $w(N)$, d.h. des aktuellen Netzes, simultan anhand aller Daten im Trainingsset beurteilt. Das Netz wird auf alle $x^{(1)}, \dots, x^{(T)}$ angewandt, und erst dann wird der Gewichtsvektor geändert.

Backpropagation ist ebenfalls eine Art Gradientenabstiegsverfahren mit dem Unterschied, dass das Netz sukzessive auf die einzelnen $x^{(k)}$ angewandt wird und nach jedem Einzelschritt bereits die Gewichte verändert werden, und zwar in Richtung des steilsten Abstiegs der Funktion $Q_k(w) = (z^{(k)} - y^{(k)})^2$:

$$\begin{aligned} w(n+1) &= w(n) - \eta \operatorname{grad} Q_k(w(n)), \quad k = 1, \dots, T, \\ \operatorname{grad} Q_k(w) &= -2(z^{(k)} - y^{(k)})\psi'(w^T x^{(k)})x^{(k)}. \end{aligned}$$

Ist dabei das Trainingsset einmal durchlaufen worden, fängt die Iteration wieder von vorne an. T Schritte Backpropagation entsprechen dabei grob gesagt einem Schritt des Gradientenabstiegsverfahrens. Auch beim Backpropagation-Algorithmus kann es nötig sein, η langsam gegen 0 gehen zu lassen.

Die Widrow-Hoff-Lernregel ist im Prinzip ein Backpropagation-Algorithmus. Die Schwellenfunktion $\psi(t) = 1_{(0,\infty)}(w_0 + t)$ ist zwar nicht differenzierbar, aber nach Präsentation von $x^{(k)}$ werden auch hier die Gewichte in Richtung des steilsten Abstiegs von $Q_k(w)$ verändert, d.h. in Richtung $x^{(k)}$ für $z^{(k)} = 1, y^{(k)} = 0$ und in Richtung $-x^{(k)}$ für $z^{(k)} = 0, y^{(k)} = 1$. Bei richtiger Klassifikation bleiben die Gewichte hier unverändert.

Natürlich kann man zur Bestimmung der Gewichte eines neuronalen Netzes jeden numerischen Algorithmus einsetzen, der das Minimum der nichtlinearen Funktion $Q(w)$ bestimmen kann. Bei einigen Anwendungen haben sich z.B. konjugierte Gradientenverfahren als schnell und zuverlässig erwiesen. All diese Algorithmen laufen allerdings Gefahr, in einem lokalen Minimum von $Q(w)$ zu landen. In der Literatur über neuronale Netze wird zwar gelegentlich behauptet, dass im Zusammenhang mit dem Trainieren von Netzen solche lokalen Minima nicht auftreten. Aus der Erfahrung der Statistik mit Maximum-Likelihood-Schätzern für hochdimensionale Parameter heraus ist damit aber durchaus zu rechnen, da das Trainieren von neuronalen Netzen für z.B. Zwecke der Regressionsanalyse sich unter entsprechenden Normalitätsannahmen als Maximum-Likelihood-Schätzverfahren interpretieren lässt.

16.3. Neuronale Netze in der nichtparametrischen Regressionsanalyse

Neuronale Netze vom MLP-Typ beschreiben eine Abbildung der Inputvariablen $x \in \mathbb{R}^p$ auf die Ausgangsvariable $y \in \mathbb{R}^q$. Wir beschränken uns in diesem Abschnitt auf den Fall, dass das Netz nur eine verborgene Schicht aufweist und dass die Ausgangsvariable univariat ist ($q = 1$). Dann hat $y \in \mathbb{R}$ als Funktion von x die Form

$$y = v_0 + \sum_{h=1}^H v_h \psi(w_{0h} + \sum_{j=1}^p w_{jh} x_j) \equiv \nu_H(x; \vartheta) \quad (16.1)$$

wobei H die Anzahl der Neuronen in der verborgenen Schicht und ψ die vorgegebene Übertragungsfunktion ist. Der Parametervektor

$$\vartheta = (w_{01}, \dots, w_{p1}, w_{02}, \dots, w_{pH}, v_0, \dots, v_H)^T \in \mathbb{R}^{(p+1)H+H+1}$$

enthält alle Gewichte des Netzes. Bereits diese Netze mit einer verborgenen Schicht besitzen eine universelle Approximationseigenschaft: jede messbare Funktion $m : \mathbb{R}^p \rightarrow \mathbb{R}$ lässt sich beliebig genau durch Funktionen $\nu_H(x, \vartheta)$ approximieren, wenn ψ eine monoton wachsende Funktion mit beschränktem Wertebereich ist. Genauer gesagt gilt das folgende Resultat (Hornik et al., 1989):

SATZ 16.1 Sei $\psi : \mathbb{R} \rightarrow [0, 1]$ monoton wachsend mit $\lim_{u \rightarrow -\infty} \psi(u) = 0$, $\lim_{u \rightarrow \infty} \psi(u) = 1$, und sei $J = \{\nu_H(x; \vartheta); H \geq 1, \vartheta \in \mathbb{R}^{(p+1)H+H+1}\}$ die Menge der durch ein MLP mit einer verborgenen Schicht gegebenen Funktionen von \mathbb{R}^p nach \mathbb{R} .

- a) Zu jeder Borel-messbaren Funktion $f : \mathbb{R}^p \rightarrow \mathbb{R}$ existiert eine Folge $\nu_n \in J$, $n \geq 1$, mit $\mu(\{x; |f(x) - \nu_n(x)| > \varepsilon\}) \rightarrow 0$ für $n \rightarrow \infty$ für jedes $\varepsilon > 0$, wobei μ ein beliebiges Wahrscheinlichkeitsmaß auf der Borel- σ -Algebra von \mathbb{R}^p ist.
- b) Zu jeder stetigen Funktion $f : \mathbb{R}^p \rightarrow \mathbb{R}$ existiert eine Folge $\nu_n \in J$, $n \geq 1$, mit $\sup_{x \in C} |f(x) - \nu_n(x)| \rightarrow 0$ für $n \rightarrow \infty$, wobei C eine beliebige kompakte Teilmenge von \mathbb{R}^p ist.

Der Wertebereich von ψ kann anstelle von $[0, 1]$ jedes beschränkte Intervall sein, ohne dass sich an der Gültigkeit der Approximationseigenschaft etwas ändert.

Der Gewichtsvektor ϑ ist nicht eindeutig durch die Netzwerkfunktion ν_H bestimmt. Ist zum Beispiel die Übertragungsfunktion antisymmetrisch um 0, d.h. $\psi(-u) = -\psi(u)$, so ändert sich $\nu_H(x; \vartheta)$ nicht, wenn

a) die Neuronen der verborgenen Schicht vertauscht werden, was einer entsprechenden Vertauschung der Koordinaten von ϑ entspricht, oder wenn

b) alle Eingangsgewichte w_{0h}, \dots, w_{ph} und das Ausgangsgewicht v_h eines Neurons mit -1 multipliziert werden.

Um diese Mehrdeutigkeit zu umgehen, beschränken wir die Parametermenge auf einen Fundamentalbereich im Sinne von Rüger und Ossen (1997), der zu jeder Netzwerkfunktion $\nu_H(x; \vartheta)$ genau einen zugehörigen Parametervektor ϑ enthält. Im Fall der antisymmetrischen Übertragungsfunktion beschränken wir uns zum Beispiel auf Gewichtsvektoren mit $v_1 \geq v_2 \geq \dots \geq v_H \geq 0$. Um die folgenden Überlegungen zu vereinfachen, nehmen wir außerdem an, dass ϑ in einer genügend großen kompakten Teilmenge $\Theta_H \subset \mathbb{R}^{(p+1)H+H+1}$ eines Fundamentalbereichs enthalten ist.

Wegen ihrer universellen Approximationseigenschaft sind neuronale Netze ein geeignetes Werkzeug, um nichtparametrische Schätzer für Regressionsfunktionen zu konstruieren. Wir betrachten hierzu das folgende heteroskedastische Regressionsmodell:

$$Z_t = f(X_t) + \varepsilon_t, \quad t = 1, \dots, N,$$

wobei X_1, \dots, X_N unabhängig und identisch verteilte d -variate Zufallsgrößen mit Dichte $p(x), x \in \mathbb{R}^d$, sind. Die Residuen $\varepsilon_1, \dots, \varepsilon_N$ sind unabhängige, reellwertige Zufallsgrößen mit

$$E(\varepsilon_t | X_t = x) = 0, \quad E(\varepsilon_t^2 | X_t = x) = s_\varepsilon^2(x) < \infty.$$

Wir setzen voraus, dass der bedingte Erwartungswert $f(x)$ und die bedingte Varianz $s_\varepsilon^2(x)$ von Z_t gegeben $X_t = x$ stetige, beschränkte Funktionen auf \mathbb{R}^d sind. Um die Regressionsfunktion f zu schätzen, passen wir ein neuronales Netz mit einer verborgenen Schicht und einer ausreichend großen Zahl H von Neuronen an die Eingangsvariablen X_1, \dots, X_N und die Zielgrößen Z_1, \dots, Z_N an, d.h. für vorgegebenes H bestimmen wir den nichtlinearen Kleinste-Quadrate-Schätzer $\hat{\vartheta}_N = \operatorname{argmin}_{\vartheta \in \Theta_H} D_N(\vartheta)$ mit

$$\hat{D}_N(\vartheta) = \frac{1}{N} \sum_{t=1}^N (Z_t - \nu_H(X_t; \vartheta))^2.$$

Unter geeigneten Bedingungen konvergiert $\hat{\vartheta}_N$ für $N \rightarrow \infty$ und festes H in Wahrscheinlichkeit gegen den Parametervektor $\vartheta_0 \in \Theta_H$, der der besten Approximation von $f(x)$ durch Funktionen des Typs $\nu_H(x; \vartheta), \vartheta \in \Theta_H$, entspricht:

$$\vartheta_0 = \operatorname{argmin}_{\vartheta \in \Theta_H} D(\vartheta) \quad \text{mit} \quad D(\vartheta) = E(f(X_t) - \nu_H(X_t; \vartheta))^2.$$

Unter etwas stärkeren Voraussetzungen folgt auch die asymptotische Normalität von $\hat{\vartheta}_N$ und damit des Schätzers $\hat{f}_H(x) = \nu_H(x; \hat{\vartheta}_N)$ für die Regressionsfunktion $f(x)$.

Der Schätzfehler $\hat{\vartheta}_N - \vartheta_0$ lässt sich in zwei asymptotisch unabhängige Teilkomponenten zerlegen: $\hat{\vartheta}_N - \vartheta_0 = (\hat{\vartheta}_N - \vartheta_N) + (\vartheta_N - \vartheta_0)$, wobei die Hilfsgröße

$$\vartheta_N = \operatorname{argmin}_{\vartheta \in \Theta_H} \frac{1}{N} \sum_{t=1}^N (f(X_t) - \nu_H(X_t; \vartheta))^2$$

die Stichprobenversion von $D(\vartheta)$ minimiert. Damit gilt (Franke und Neumann, 2000):

SATZ 16.2 *Sei ψ beschränkt und zweimal stetig differenzierbar mit beschränkten Ableitungen. $D(\vartheta)$ habe ein eindeutiges globales Minimum ϑ_0 im Innern von Θ_H , und die Hesse-Matrix $\nabla^2 D(\vartheta_0)$ von D an der Stelle ϑ_0 sei positiv definit. Zusätzlich zu den oben geforderten Bedingungen an das Regressionsmodell gelte*

$$\begin{aligned} 0 < \delta \leq s_\varepsilon^2(x) &\leq \Delta < \infty \quad \text{für alle } x, \\ E(|\varepsilon_t|^n | X_t = x) &\leq C_n < \infty \quad \text{für alle } x, n \geq 1 \end{aligned}$$

mit geeigneten Konstanten $\delta, \Delta, C_n, n \geq 1$. Dann gilt für $N \rightarrow \infty$:

$$\sqrt{N} \begin{pmatrix} \hat{\vartheta}_N - \vartheta_0 \\ \vartheta_N - \vartheta_0 \end{pmatrix} \xrightarrow{\mathcal{L}} N \left(0, \begin{pmatrix} \Sigma_1 & 0 \\ 0 & \Sigma_2 \end{pmatrix} \right)$$

mit Kovarianzmatrizen

$$\begin{aligned} \Sigma_i &= (\nabla^2 D(\vartheta_0))^{-1} B_i(\vartheta_0) (\nabla^2 D(\vartheta_0))^{-1}, i = 1, 2, \\ B_1(\vartheta) &= 4 \int s_\varepsilon^2(x) \nabla \nu_H(x; \vartheta) \nabla \nu_H(x; \vartheta)^T p(x) dx \\ B_2(\vartheta) &= 4 \int (f(x) - \nu_H(x; \vartheta))^2 \nabla \nu_H(x; \vartheta) \nabla \nu_H(x; \vartheta)^T p(x) dx \end{aligned}$$

wobei $\nabla \nu_H$ den Gradienten der Netzfunktion bzgl. der Parameter ϑ bezeichnet.

Aus dem Satz folgt sofort, dass $\sqrt{N}(\hat{\vartheta}_N - \vartheta_0)$ asymptotisch $N(0, \Sigma_1 + \Sigma_2)$ -verteilt ist. Σ_1 steht hier für die durch die Beobachtungsfehler ε_t verursachte Variabilität des Schätzers $\hat{\vartheta}_N$. Σ_2 repräsentiert den Anteil an asymptotischer Variabilität, der durch die Misspezifikation der Regressionsfunktion verursacht wird, d.h. durch die Tatsache, dass $f(x)$ für festes H und kein ϑ von der Form $\nu_H(x; \vartheta)$ ist. Im korrekt spezifizierten Fall, wo $f(x) = \nu_H(x; \vartheta_0)$ ist, verschwindet diese Kovarianzkomponente, da dann $B_2(\vartheta_0) = 0$ und damit $\Sigma_2 = 0$.

Σ_1, Σ_2 lassen sich wie üblich durch Stichprobenkovarianzmatrizen schätzen. Um Tests und Konfidenzbänder für $f(x)$ zu konstruieren, bieten sich alternativ zur Ausnutzung der asymptotischen Verteilung Bootstrap- bzw. im heteroskedastischen Fall Wild Bootstrap-Verfahren an (Franke und Neumann, 2000).

16.2 bezieht sich auf den theoretischen Wert des Kleinste-Quadrate-Schätzers $\hat{\vartheta}_N$, der in der Praxis numerisch bestimmt werden muss. $\tilde{\vartheta}_N$ sei eine solche numerische Approximation von $\hat{\vartheta}_N$. Die Qualität des resultierenden Schätzers $\tilde{\vartheta}_N$ kann dabei von dem benutzten numerischen Verfahren abhängen. White (1989) hat insbesondere gezeigt, dass der Backpropagation-Algorithmus unter entsprechenden Voraussetzungen zu einem asymptotisch ineffizienten Schätzer $\tilde{\vartheta}_N$ führt, d.h. die asymptotische Kovarianzmatrix von $\sqrt{N}(\tilde{\vartheta}_N - \vartheta_0)$ ist größer als die von $\sqrt{N}(\hat{\vartheta}_N - \vartheta_0)$ in dem Sinn, dass die Differenz beider Matrizen positiv definit ist. Allerdings hat White auch gezeigt, dass durch Anfügen eines einzigen globalen Minimierungsschrittes

der aus dem Backpropagation berechnete Schätzer so modifiziert werden kann, dass er für $N \rightarrow \infty$ genauso effizient ist wie der theoretische Kleinste-Quadrate-Schätzer $\hat{\vartheta}_N$.

Bisher haben wir die Anzahl der Neuronen H in der verborgenen Schicht des Netzes und damit die Dimension des Parametervektors ϑ festgehalten. Der auf dem Netz basierende Schätzer $\hat{f}_H(x) = \nu_H(x; \hat{\vartheta}_N)$ konvergiert dann gegen $\nu_H(x; \vartheta_0)$, so dass im allgemeinen der Bias $E(\hat{f}_H(x)) - m(x)$ für $N \rightarrow \infty$ nicht verschwindet, sondern gegen $\nu_H(x; \vartheta_0) - f(x)$ konvergiert. Mit Standardargumenten folgt aus 16.2 direkt:

KOROLLAR 16.1 *Unter den Voraussetzungen von 16.2 gilt für $N \rightarrow \infty$*

$$\sqrt{N}(\nu_H(x; \hat{\vartheta}_N) - \nu_H(x; \vartheta_0)) \xrightarrow{\mathcal{L}} N(0, \sigma_\infty^2)$$

mit $\sigma_\infty^2 = \nabla \nu_H(x; \vartheta_0)^T (\Sigma_1 + \Sigma_2) \nabla \nu_H(x; \vartheta_0)$.

TO DO:

Korollar
mit einem
r !!

Um einen konsistenten Schätzer für $f(x)$ zu erhalten, muss die Zahl H der Neuronen, die bei dem nichtparametrischen Schätzer $\hat{f}_H(x)$ die Rolle eines Glattheitsparameters spielt, mit N wachsen. Wegen der universellen Approximationseigenschaft neuronaler Netze konvergiert dann $\nu_H(x; \vartheta_0)$ gegen $f(x)$, so dass der Bias asymptotisch verschwindet. Da mit wachsendem H die Dimension des Parametervektors ϑ wächst, darf H aber nicht zu schnell gegen ∞ gehen, damit die Varianz von $\hat{f}_H(x)$ weiterhin gegen 0 konvergiert. Bei der praktischen Wahl von H befindet man sich daher in dem für die nichtparametrische Statistik typischen Bias-Varianz-Dilemma: ein kleines H resultiert in einer glatten Schätzfunktion \hat{f}_H mit geringer Varianz und großem Bias, während ein großes H zu kleinem Bias, aber hoher Variabilität des dann weniger glatten Schätzers \hat{f}_H führt.

White (1990) hat in einem passenden Rahmen gezeigt, dass der auf neuronalen Netzen beruhende Regressionsschätzer $\hat{f}_H(x)$ in Wahrscheinlichkeit gegen $f(x)$ konvergiert und somit konsistent ist, wenn mit $N \rightarrow \infty$ auch $H \rightarrow \infty$ mit geeigneter Rate.

Daraus folgt, daß neuronale Netze mit frei wählbarer Zahl H an Neuronen in der verborgenen Schicht brauchbare nichtparametrische Funktionsschätzer in der Regressions- und, wie wir im folgenden Abschnitt diskutieren, in der Zeitreihenanalyse zur Verfügung stellen. Sie haben den Vorteil, daß die approximierende Funktion $\nu_H(x; \vartheta)$ der Form (16.1) aus einfachen Bausteinen, den Neuronen, zusammengesetzt ist, die nur aus einer gegebenen nichtlinearen Transformation einer affin-linearen Kombination der Variablen $x = (x_1, \dots, x_p)^T$ bestehen. Das macht u.a. die numerische Berechnung von Kleinste-Quadrate-Schätzern für ϑ auch dann noch möglich, wenn die Dimension p der Eingangsvariablen und die Zahl H der Neuronen groß und die Dimension $(p+1)H + H + 1$ des Parametervektors somit sehr groß ist. Im Gegensatz zu den im Kapitel 17 vorgestellten lokalen Glättungsverfahren sind neuronale Netze daher auch zum Schätzen von Funktionen auf hochdimensionalen Räumen einsetzbar. Ein Grund dafür ist die Nicht-Lokalität des Funktionsschätzers $\hat{f}_H(x)$, der nicht nur von den Beobachtungen (X_t, Z_t) mit kleiner Norm $\|X_t - x\|$ abhängt und damit in der Praxis weniger unter dem Fluch der Dimensionalität, d.h. hier der selbst für großes N geringen

lokalen Dichte der Beobachtungen X_t in hochdimensionalen Räumen, leidet.

Von der Theorie her reicht es aus, neuronale Netze vom MLP-Typ mit einer verborgenen Schicht zu betrachten. In der Praxis erzielt man aber manchmal eine vergleichbar gute Anpassung an die Daten bei wesentlich sparsamerer Parametrisierung, wenn man mehrere verborgene Schichten hintereinander schaltet. Eine Netzfunktion mit zwei verborgenen Schichten aus H bzw. G Neuronen hat z.B. die Gestalt

$$\nu(x; \vartheta) = v_0 + \sum_{g=1}^G v_g \psi \left(w'_{0g} + \sum_{h=1}^H w'_{hg} \psi(w_{0h} + \sum_{j=1}^p w_{jh} x_j) \right),$$

wobei ϑ für den Vektor aller Gewichte v_g, w'_{hg}, w_{jh} steht. Eine solche Funktion mit kleinem H, G kann eine sparsamer parametrisierte Approximation der Regressionsfunktion $f(x)$ liefern als eine Netzfunktion mit nur einer verborgenen Schicht aus einer großen Zahl an Neuronen. In einer Fallstudie zur Entwicklung von Handelsstrategien für Währungsportfolios (Franke und Klein, 1999) hat sich zum Beispiel herausgestellt, daß mit zwei verborgenen Schichten wesentlich bessere Ergebnisse erzielt werden konnten als mit nur einer.

Zusätzlich kann die Zahl der zu schätzenden Parameter weiter reduziert werden, wenn in dem neuronalen Netz einige Verbindungen gekappt, d.h. die entsprechenden Gewichte von vorneherein auf Null gesetzt werden. Die große Flexibilität, die neuronale Netze bei der Approximation von Regressionsfunktionen bieten, bereitet aber auch bei der Modellbildung Probleme, da man sich für eine Netzstruktur entscheiden und dabei die Fragen muß:

1. Wieviele verborgene Schichten hat das Netz?
2. Wieviele Neuronen hat jede einzelne verborgene Schicht?
3. Welchen Knoten (Inputs, verborgene Neuronen, Outputs) des Netzes sollen verbunden sein, d.h. welche Gewichte werden von vorneherein auf Null gesetzt?

Gesucht wird dabei ein Netz, für das die Netzfunktion $\nu(x; \vartheta)$ sparsam parametrisiert ist, gleichzeitig aber für geeignetes ϑ eine ausreichend gute Approximation der Regressionsfunktion $m(x)$ ermöglicht.

Ähnlich wie in der klassischen linearen Regressionsanalyse gibt es mittlerweile ein umfangreiches Instrumentarium, um die Spezifizierung der Netzstruktur datenadaptiv vorzunehmen. Der Einfachheit halber konzentrieren wir uns auf ein Feedforward-Netz mit nur einer verborgenen Schicht aus H Neuronen.

a) *Wiederholte Signifikanz-Tests:* Wie im schrittweisen Aufbau eines linearen Regressionsmodells wird von einem einfachen Netz ausgehend ein zusätzliches Neuron mit der Nummer H und Ausgangsgewicht v_H hinzugefügt. Ob dadurch die Anpassungsqualität des Netzes wesentlich verbessert wird, entscheidet ein Test der Hypothese $H_0 : v_H = 0$ gegen die Alternative $H_1 : v_H \neq 0$. Da unter H_0 die Eingangsgewichte w_{0H}, \dots, w_{pH} des fraglichen Neurons

nicht identifizierbar sind, d.h. keinen Einfluss auf den Wert der Netzfunktion ν_H haben, handelt es sich dabei um kein Standard-Testproblem. White (1989) und Tervirta et al. (1993) haben aber Lagrange-Multiplikatoren-Tests entwickelt, die sich zum Testen der Signifikanz eines zusätzlichen Neurons eignen. Umgekehrt ist es auch möglich, von einem komplexen Netz mit großem H ausgehend Neuronen sukzessive zu entfernen, bis der hierzu verwandte Test die Hypothese $H_0 : v_H = 0$ ablehnt. Zur Reduzierung der Parameterzahl kann es auch sinnvoll sein, einzelne Eingangsverbindungen zu kappen, d.h. das entsprechende Gewicht auf Null zu setzen. Für den Test der Hypothese $H_0 : w_j h = 0$ gegen die Alternative $H_1 : w_j h \neq 0$ lassen sich klassische Wald-Tests auf der Grundlage von asymptotischen Resultaten wie 16.2 einsetzen (vgl. z.B. Anders (1997) für Anwendungen in der Finanzstatistik).

b) *Kreuzvalidierung und Validierung*: Die herkömmliche Kreuzvalidierung scheidet wegen des damit verbundenen Rechenaufwands meist aus, um die Modellordnung, d.h. hier in erster Linie der Zahl H von Neuronen in der verborgenen Schicht, zu bestimmen. Um die Leave-one-out Schätzer für den Modellparameter zu berechnen, müsste man N -mal ein neuronales Netz an die jeweils um eine Beobachtung reduzierte Stichprobe anpassen, und das für jede in Betracht gezogene Netzstruktur. Eine verwandte und beim Einsatz neuronaler Netze in der Regressions- und Zeitreihenanalyse weit verbreitete Vorgehensweise ist das Herausnehmen eines Teils der Daten aus der Stichprobe, um die Qualität von Modellen anhand dieses sogenannten Validierungssets zu messen. Neben den zur Berechnung der Kleinst-Quadrate-Schätzer $\hat{\vartheta}_N$ benutzten Daten (X_t, Z_t) , $t = 1, \dots, N$, steht dann eine zweite unabhängige Teilstichprobe (X_t, Z_t) , $t = N + 1, \dots, N + M$, zur Verfügung. Durch Minimierung von Anpassungsmaßen wie z.B.

$$V(H) = \frac{1}{M} \sum_{t=N+1}^{N+M} (Z_t - \nu_H(X_t; \hat{\vartheta}_N))^2$$

läßt sich die Modellordnung H bestimmen und auch die Qualität von unvollständigen Netzstrukturen bestimmen, bei denen einzelne Eingangsgewichte auf Null gesetzt worden sind.

c) *Netzinformationskriterium*: Zum Vergleich von Netzstrukturen lassen sich auch bekannte Ordnungswahlverfahren wie das Akaike Information Criterion (AIC) benutzen, das von Murata et al. (1994) als *Network Information Criterion* (NIC) auf den Fall neuronaler Netze spezialisiert worden ist. Dabei wird implizit angenommen, dass die Residuen ε_t normalverteilt sind mit gemeinsamer Varianz σ_ε^2 .

16.4. Vorhersage von Finanzzeitreihen mit neuronalen Netzen

Zur Vorhersage der zukünftigen Entwicklung von Finanzzeitreihen eignen sich besonders Modelle mit autoregressiver Struktur. Hier ist der Wert der Zeitreihe zur Zeit $t + 1$ eine Funktion

endlich vieler Beobachtungen aus der Vergangenheit zuzüglich einer von der Vergangenheit unabhängigen Innovation:

$$Z_{t+1} = f(Z_t, \dots, Z_{t-p+1}) + \varepsilon_{t+1}, \quad -\infty < t < \infty, \quad (16.2)$$

wobei ε_t , $-\infty < t < \infty$, unabhängig, identisch verteilt mit $E(\varepsilon_t) = 0$, $Var(\varepsilon_t) = \sigma_\varepsilon^2 < \infty$ sind. Die formale Analogie dieses nichtlinearen autoregressiven Modells der Ordnung p - kurz: NLAR(p)-Modell - zu dem im vorigen Abschnitt betrachteten Regressionsmodell ist offensichtlich, wobei der p -variater Zufallsvektor $(Z_t, \dots, Z_{t-p+1})^T$ an die Stelle der d -variater unabhängigen Variable X_t tritt. Die Autoregressionsfunktion $f : \mathbb{R}^p \rightarrow \mathbb{R}$ liefert in diesem Modell sofort die beste Vorhersage für Z_{t+1} gegeben die Werte der Zeitreihe bis zur Zeit t :

$$\hat{Z}_{t+1|t}^0 = f(Z_t, \dots, Z_{t-p+1}).$$

Da f im allgemeinen nicht bekannt ist, liegt es im Hinblick auf den vorigen Abschnitt nahe, die Autoregressionsfunktion durch ein neuronales Netz zu approximieren, wenn Beobachtungen der Zeitreihe Z_1, \dots, Z_{N+1} als Daten vorliegen. Beim Trainieren des Netzes, d.h. beim Schätzen der Netzgewichte, werden als Eingangsgröße der Vektor $(Z_t, \dots, Z_{t-p+1})^T$ und als Zielgröße Z_{t+1} für $t = p, \dots, N$, benutzt. Wir beschränken uns der Einfachheit halber wieder auf ein MLP mit einer verborgenen Schicht. Bezeichnet $\hat{\vartheta}_N$ wieder den Kleinste-Quadrate-Schätzer für den Gewichtsvektor:

$$\hat{\vartheta}_N = \operatorname{argmin}_{\vartheta \in \Theta_H} \frac{1}{N-p+1} \sum_{t=p}^N (Z_{t+1} - \nu_H(Z_t, \dots, Z_{t-p+1}; \vartheta))^2$$

wobei ν_H wie im vorigen Abschnitt definiert ist, so erhalten wir als nichtparametrische, auf einem neuronalen Netz basierender Vorhersage für Z_{t+1} :

$$\hat{Z}_{t+1|t} = \nu_H(Z_t, \dots, Z_{t-p+1}; \hat{\vartheta}_N).$$

Die Ergebnisse über asymptotische Normalität von Parameter- und Funktionsschätzer und über die Konsistenz von $\nu_H(\cdot; \hat{\vartheta}_N)$ als Schätzer von f für wachsendes H bleiben auch im Zeitreihenfall erhalten, wenn im wesentlichen der stochastische Prozeß $\{Z_t, -\infty < t < \infty\}$ α -mischend mit exponentiell abfallenden Mischungskoeffizienten ist (vgl. White 1989b, 1990). Franke et al. (2000) haben für den Fall $p = 1$ Bedingungen an die Autoregressionsfunktion f und an die Verteilung der Innovationen ε_t formuliert, die für den NLAR(1)-Prozeß sogar die stärkere β -Mischungseigenschaft mit exponentiell abfallenden Koeffizienten garantieren. Neben technischen Details muß im wesentlichen

$$\lim_{|x| \rightarrow \infty} |f(x)/x| < 1$$

erfüllt sein, während es für die Innovationsverteilung genügt, die Existenz einer nirgends verschwindenden Dichte zu fordern. Die letzte Bedingung kann deutlich abgeschwächt werden.

Die Bedingung an die Autoregressionsfunktion ist vergleichsweise schwach und auch einleuchtend, wenn man an die Stationaritätsbedingung $|\alpha| < 1$ für lineare AR(1)-Prozesse

$Z_{t+1} = \alpha Z_t + \varepsilon_{t+1}$ denkt, wo $f(x) = \alpha x$ ist. Entsprechend genügen auch für NLAR(p)-Prozesse höherer Ordnung ($p > 1$) schwache Bedingungen an f , die in erster Linie die Stationarität garantieren, um neuronale Netze als ein brauchbares Werkzeug zum nichtparametrischen Schätzen von f erscheinen zu lassen.

Für die praktische Vorhersage von Finanzzeitreihen möchte man nicht nur die letzten Werte der Zeitreihe selbst benutzen, sondern auch andere, zur Zeit t verfügbare Wirtschaftsdaten wie Wechselkurse, Indexwerte, Rohstoffpreise oder auch die in der technischen Analyse gebräuchlichen nichtlinearen Transformationen von Kursdaten. Für diesen Zweck eignen sich nichtlineare autoregressive Prozesse mit exogenen Komponenten der Ordnung p - kurz: NLARX(p)-Prozesse:

$$Z_{t+1} = f(Z_t, \dots, Z_{t-p+1}, X_t) + \varepsilon_t, \quad -\infty < t < \infty, \quad (16.3)$$

wobei die Innovation ε_t , $-\infty < t < \infty$, wieder unabhängig und identisch verteilt mit $E(\varepsilon_t) = 0$, $Var(\varepsilon_t) = \sigma_\varepsilon^2 < \infty$, sind und X_t der Wert eines d -variaten stochastischen Prozesses ist, in dem alle zur Zeit t gegebene externe Information enthalten ist, die für die Vorhersage benutzt werden soll.

Die praktische Anwendung der Vorhersage von Finanzzeitreihen mit neuronalen Netzen illustrieren wir mit einer Pilotstudie, die in Zusammenarbeit mit der Commerzbank AG durchgeführt wurde (vgl. Franke, 1999). Ziel war die Entwicklung einer Handelsstrategie für ein Portfolio aus den 28 wichtigsten, im niederländischen CBS-Index enthaltenen Aktien. Wir beschränken uns dabei auf Buy-and-Hold-Strategien mit einem Zeithorizont von einem Vierteljahr (60 Handelstage), d.h. zu Beginn eines Quartals wird das Portfolio aufgebaut und dann drei Monate lang unverändert gehalten. Am Ende dieses Zeitraums soll der Portfoliowert möglichst groß sein.

Als Grundlage der Handelsstrategie verwenden wir Drei-Monats-Vorhersagen der Aktienkurse. S_t bezeichne den Kurs einer der zur Auswahl stehenden 28 Aktien. Als Modell für die Zeitreihe S_t verwenden wir einen NLARX-Prozess der Form (16.3); die Systemfunktion f approximieren wir durch eine Netzfunktion $\nu_H(S_t, A_t, X_t; \vartheta)$. Dabei ist A_t ein Vektor aus festen nichtlinearen Transformationen von S_t, \dots, S_{t-p+1} , die der technischen Marktanalyse entnommen werden, z.B. gleitende Mittel, Momentum oder Bollinger-Bänder (Müller und Nietzer, 1993, Welcker, 1994). Der Zufallsvektor X_t repräsentiert wieder ausgewählte Intermarketdaten wie Indexpreise, Wechselkurse, internationale Zinssätze usw. Wie bei einem Vorhersagehorizont von 60 Zeiteinheiten in die Zukunft nicht anders zu erwarten war, sind die eigentlichen Vorhersagen des Aktienkurses in 60 Tagen

$$\hat{S}_{t+60|t} = \nu_H(S_t, A_t, X_t; \hat{\vartheta}_N)$$

nicht sehr zuverlässig. Für die Entscheidung, ob eine Aktie in das Portfolio aufgenommen wird oder nicht, ist aber in erster Linie der allgemeine Trend der Wertentwicklung wichtig und nicht der exakte Kurs der Aktie am Ende des Anlagezeitraums. Um diesen Aspekt zu würdigen, wird bei der Zusammenstellung des Portfolios nur berücksichtigt, ob auf der

Grundlage der netzbasierten Vorhersage $\hat{S}_{t+60|t}$ zu erwarten ist, dass der Kurs deutlich steigt (um mehr als 5 %), deutlich fällt (um mehr als 5 %) oder ob er andernfalls im wesentlichen auf demselben Niveau bleibt. Das netzbasierte Portfolio wird aus denjenigen Aktien zusammengestellt (mit relativen Anteilen, die sich nach ihrem jeweiligen Gewicht im CBS-Index richten), für die $(\hat{S}_{t+60|t} - S_t)/S_t > 0.05$ ist. Dabei wird für alle 28 in Betracht gezogenen Aktien *dieselbe* Netzfunktion $\nu_H(s, a, x; \vartheta)$ benutzt, deren kursabhängigen Argumente s , a allerdings die aktienspezifischen Werte annehmen.

Zur Auswahl eines geeigneten Netzes und zur Schätzung des Netzgewichtevektors ϑ wurden Daten von 1993 bis 1995 benutzt. Bei der Wahl der Netzstruktur wurden statistische Modellselektionstechniken sowie die Erfahrung von Experten genutzt. Das resultierende Netz ist ein Multilayer-Perzeptron mit einer verborgenen Schicht aus nur $H = 3$ Neuronen. Der Eingabevektor (S_t, A_t, X_t) hat die Dimension 25, so dass insgesamt ein Parametervektor $\vartheta \in \mathbb{R}^{82}$ zu schätzen war.

Um die Qualität der netzbasierten Handelsstrategie zu überprüfen, wird sie auf Daten des Jahres 1996 angewandt. Zu Beginn eines jeden Quartals wird auf der Grundlage der netzbasierten Prognosen ein Portfolio aus den 28 CBS-Aktien zusammengestellt. Am Ende des Quartals wird der Wertzuwachs in Prozent betrachtet. Als Vergleichsmaßstab wird der Wertzuwachs eines Portfolios betrachtet, dessen Zusammensetzung genau den CBS-Index widerspiegelt. Da in dem betrachteten Jahr der Markt im großen und ganzen in einer steigenden Phase war, ist es erfahrungsgemäß nicht einfach, den Index zu schlagen. Wie die Abbildung 16.7 zeigt, erzielt das Netzportfolio dennoch in jedem Quartal einen höheren prozentualen Wertzuwachs wie das Indexportfolio, und zwar sowohl in Quartalen wie dem 1. und 4., wo der Index deutlich steigt, als auch in Quartalen wie dem 2., wo der Index geringfügig fällt. Allerdings muss das Ergebnis mit etwas Vorsicht interpretiert werden. Auch in der Trainingsphase (1993-1995) ist der CBS-Index tendenziell nach oben gegangen, so dass das Netz sich auf Trendprognosen in einem generell wachsenden Markt spezialisieren konnte. Vermutlich braucht man andere Netze als Grundlage von Handelsstrategien, wenn der Markt in eine längerfristige Seitwärtsbewegung umschwenkt oder gar der Index deutlich fällt.

16.5. Risikoquantifizierung mit neuronalen Netzen

In den vorhergehenden Kapiteln sind mit der Volatilität und dem Value-at-Risk die beiden gängigen Risikomaße für Anlagen vorgestellt worden. Beide werden meist als bedingte Standardabweichung bzw. als bedingtes Quantil bzgl. der durch die Vergangenheit gegebenen Informationsbasis definiert. Wie andere nichtparametrische Verfahren können auch neuronale Netze dazu benutzt werden, diese Risikomaße zu schätzen. Der Vorteil der auf neuronalen Netzen basierenden Volatilitäts- und VaR-Schätzer besteht darin, daß die für die Risikoabschätzung benutzte Information durch einen hochdimensionalen Datenvektor

repräsentiert werden kann, ohne daß die Praktikabilität des Verfahrens darunter leidet. Es ist so zum Beispiel möglich, das bedingte 5%-Quantil des Returnprozesses einer Aktie aus dem DAX gegeben einzelne aktuelle Returns aller DAX-Werte und zusätzlicher ökonomischer Daten wie Zinssätze, Wechselkurse usw. zu schätzen. In dem folgenden Abschnitt skizzieren wir kurz die Vorgehensweise.

Wie in (13.1) gehen wir zum Schätzen der Volatilität von einem Modell der Form

$$Z_{t+1} = f(Z_t, \dots, Z_{t-p+1}, X_t) + s(Z_t, \dots, Z_{t-p+1}, X_t) \xi_{t+1} \quad (16.4)$$

aus, wobei ξ_t unabhängig, identisch verteilte Zufallsgrößen mit $E(\xi_t) = 0$, $E(\xi_t^2) = 1$ sind. $X_t \in \mathbb{R}^d$ repräsentiert wie im vorigen Abschnitt die exogene zur Zeit t verfügbare Information, die wir bei der Abschätzung des Risikos der Zeitreihe Z_t benutzen wollen. Die durch (16.4) gegebene Zeitreihe ist ein nichtlinearer AR(p)-ARCH(p)-Prozeß mit exogenen Komponenten.

Zur Abkürzung schreiben wir $Z_t(p) = (Z_t, \dots, Z_{t-p+1})^T \in \mathbb{R}^p$. Dann gilt für $z \in \mathbb{R}^p$, $x \in \mathbb{R}^d$

$$\begin{aligned} E[Z_{t+1}|Z_t(p) = z, X_t = x] &= f(z, x) \\ \text{Var}[Z_{t+1}|Z_t(p) = z, X_t = x] &= s^2(z, x) \\ &= E[Z_{t+1}^2|Z_t(p) = z, X_t = x] - f^2(z, x). \end{aligned}$$

Die bedingte Erwartungsfunktion $f(z, x)$ approximieren wir wie im vorigen Abschnitt durch eine neuronale Netzfunktion $\nu_H(z, x; \vartheta)$ der Form (16.1). Mit dem nichtlinearen Kleinste-Quadrate-Schätzer $\hat{\vartheta}_N$ für ϑ erhalten wir als Schätzer für f :

$$\hat{f}_H(z, x) = \nu_H(z, x; \hat{\vartheta}_N).$$

Völlig analog können wir auch den bedingten Erwartungswert

$$E[Z_{t+1}^2|Z_t(p) = z, X_t = x] = g(z, x)$$

schätzen, indem wir die Funktion durch ein neuronales Netz mit Ausgabefunktion $\nu_G(z, x; \delta)$ approximieren und dessen Parameter δ durch den Kleinste-Quadrate-Schätzer $\hat{\delta}$ in einer wie Θ_H gewählten, genügend großen kompakten Teilmenge $\Delta_G \subset \mathbb{R}^{(p+d+1)G+G+1}$ eines Fundamentalbereichs schätzen:

$$\hat{\delta}_N = \underset{\delta \in \Delta_G}{\operatorname{argmin}} \frac{1}{N - p + 1} \sum_{t=p}^N (Z_{t+1}^2 - \nu_G(Z_t(p), X_t; \delta))^2,$$

$$\hat{g}_G(z, x) = \nu_G(z, x; \hat{\delta}_N).$$

Als Schätzer für die bedingte Volatilität erhalten wir sofort:

$$\hat{s}_{H,G}^2(z, x) = \hat{g}_G(z, x) - \hat{f}_H^2(z, x).$$

Die Nichtnegativität dieses Schätzers ist im allgemeinen aber nur für $G = H$ garantiert. Um diese Einschränkung zu vermeiden, kann man der Vorgehensweise von Fan und Yao folgen, die ein ähnliches Problem für Kernschätzer bedingter Varianzen in heteroskedastischen Regressionsmodellen studiert haben. Bei diesem Ansatz werden die Residuen

$$\varepsilon_{t+1} = Z_{t+1} - f(Z_t(p), X_t) = s(Z_t(p), X_t) \xi_{t+1}$$

durch die Stichprobenresiduen

$$\hat{\varepsilon}_{t+1} = Z_{t+1} - \hat{f}_H(Z_t(p), X_t), \quad t = p, \dots, N,$$

approximiert. Da die ξ_{t+1} Erwartungswert 0 und Varianz 1 besitzen, ist

$$E[\varepsilon_{t+1}^2 | Z_t(p) = z, X_t = x] = s^2(z, x).$$

Wir können diese Funktion direkt mit einem neuronalen Netz mit G Neuronen und der Ausgabefunktion $\nu_G(z, x; \delta)$ approximieren, dessen Parameter wir durch

$$\hat{\delta}_N = \operatorname{argmin}_{\delta \in \Delta_G} \frac{1}{N - p + 1} \sum_{t=p}^N (\hat{\varepsilon}_{t+1}^2 - \nu_G(Z_t(p), X_t; \delta))^2$$

schätzen. Der resultierende Schätzer für die bedingte Volatilität, der über die $\hat{\varepsilon}_t$ auch von H abhängt, ist dann

$$\hat{s}_{H,G}(z, x) = \nu_G(z, x; \hat{\delta}_N).$$

Er ist für beliebige G, H automatisch nichtnegativ. Da die Zahl der Neuronen im wesentlichen die Glattheit der Netzfunktion bestimmt, kann es sinnvoll sein, zur Approximation von f und s^2 unterschiedliche Netze mit $H \neq G$ Neuronen zu wählen, wenn zu vermuten ist, daß die Glattheit dieser beiden Funktionen sich deutlich voneinander unterscheiden.

Wenn in dem Modell (16.4) zusätzlich die Verteilung der Innovationen ξ_t spezifiziert wird, so erhalten wir zusammen mit den Schätzern für f und s^2 sofort auch einen Schätzer für den bedingten Value-at-Risk. Ist die Verteilung von ξ_t z.B. $N(0, 1)$, so ist die bedingte Verteilung von Z_{t+1} gegeben die Information $Z_t(p)$ und X_t zur Zeit t ebenfalls eine Normalverteilung mit Mittelwert $f(Z_t(p), X_t)$ und Varianz $s^2(Z_t(p), X_t)$. Ist q_α° das α -Quantil der Standardnormalverteilung, so ist der VaR des Prozesses $\{Z_t\}$, d.h. das bedingte α -Quantil von Z_{t+1} gegeben $Z_t(p), X_t$:

$$\operatorname{VaR}_{t+1} = f(Z_t(p), X_t) + s(Z_t(p), X_t)q_\alpha^\circ.$$

Einen auf neuronalen Netzen basierenden Schätzer für diesen bedingten Value-at-Risk erhalten wir, indem wir f und s durch die entsprechenden Schätzer ersetzen:

$$\widehat{\operatorname{VaR}}_{t+1} = \hat{f}_H(Z_t(p), X_t) + \hat{s}_{H,G}^2(Z_t(p), X_t)q_\alpha^\circ. \quad (16.5)$$

Dabei können wir ohne weiteres die Standardnormalverteilung durch andere Verteilungen ersetzen, z.B. durch eine auf Erwartungswert 0 und Varianz 1 standardisierte t -Verteilung. q_α° ist dann das entsprechende α -Quantil der Innovationsverteilung, d.h. der Verteilung der ξ_t .

Der Schätzer (16.5) für den Value-at-Risk setzt voraus, daß Z_t ein nichtlinearer ARX-ARCHX-Prozeß der Form (16.4) ist. Vor allem aber hat er den Nachteil, von der kritischen Annahme einer spezifischen Verteilung der ξ_t abzuhängen. Vor allem die herkömmliche Vorgehensweise, bei stochastischen Volatilitätsmodellen von einer Standardnormalverteilung der Innovationen auszugehen, ist in letzter Zeit in der Finanzstatistik auf der Grundlage empirischer Befunde kritisiert worden (Pictet...). Die Dicke der Flanken von Finanzzeitreihenverteilungen scheint zumindestens manchmal so ausgeprägt zu sein, daß zu ihrer adäquaten Modellierung auch die Verteilung der Innovationen als leptokurtisch angenommen werden muß. Wegen der Einfachheit der Darstellung werden hierbei oft t -Verteilungen mit wenigen Freiheitsgraden betrachtet. Um die Willkür bei der Wahl der Innovationsverteilung zu umgehen, ist es möglich, bedingte Quantile direkt ohne Rückgriff auf ein Modell der Form (16.4) zu schätzen. Dieser Ansatz geht auf die Regressionsquantile von Koenker und Bassett zurück und wurde im Zusammenhang mit Kernschätzern von Abberger auf Zeitreihen angewandt.

Wir gehen davon aus, daß Z_t eine stationäre Zeitreihe ist. Wie in Kapitel 17 bezeichnet P_{t+1} die Vorhersageverteilung, d.h. die bedingte Verteilung von Z_{t+1} gegeben $Z_t(p), X_t$. Mit F_{t+1} bezeichnen wir die zugehörige bedingte Verteilungsfunktion

$$F_{t+1}(y|z, x) = pr(Z_{t+1} \leq y | Z_t(p) = z, X_t = x)$$

für $y \in \mathbb{R}$, $z \in \mathbb{R}^p$, $x \in \mathbb{R}^d$. $q_\alpha(z, x)$ sei das bedingte α -Quantil, d.h. die Lösung der Gleichung $F_{t+1}(q_\alpha(z, x)|z, x) = \alpha$. Die bedingte Quantilenfunktion $q_\alpha(z, x)$ löst das Minimierungsproblem

$$E\{\alpha(Z_{t+1} - q)^+ + (1 - \alpha)(Z_{t+1} - q)^- | Z_t(p) = z, X_t = x\} = \min_{q \in \mathbb{R}}! \quad (16.6)$$

wobei $y^+ = y - 1_{[0, \infty)}(y)$ und $y^- = |y| \cdot 1_{(-\infty, 0]}(y)$ Positiv- und Negativteil von $y \in \mathbb{R}$ bezeichnen. Um die Quantilenfunktion direkt mit einem neuronalen Netz mit Q Neuronen zu schätzen, approximieren wir $q_\alpha(z, x)$ durch eine Netzfunktion $\nu_Q(z, x; \gamma)$ der Form (16.1), deren Gewichtsparameter γ in einem Fundamentalbereich $\Gamma_Q \subset \mathbb{R}^{(p+d+1)Q+Q+1}$ liegt. γ schätzen wir jedoch nicht über einen Kleinste-Quadrate-Ansatz, sondern durch Minimieren des Stichprobenanalogons von (16.6):

$$\hat{\gamma}_N = \operatorname{argmin}_{\gamma \in \Gamma_Q} \frac{1}{N - p + 1} \sum_{t=p}^N \{\alpha[Z_{t+1} - \nu_Q(Z_t(q), X_t)]^+ + (1 - \alpha)[Z_{t+1} - \nu_Q(Z_t(q), X_t)]^-\}.$$

Als Schätzer für die Quantilenfunktion erhalten wir

$$\hat{q}_{Q\alpha}(z, x) = \nu_Q(z, x; \hat{\gamma}_N)$$

und damit als Schätzer für den bedingten Value-at-Risk gegeben $Z_t, \dots, Z_{t-p+1}, X_t$

$$\widehat{VaR}_{t+1} = \hat{q}_{Q\alpha}(Z_t, \dots, Z_{t-p+1}, X_t).$$

White hat unter geeigneten Annahmen gezeigt, daß der Funktionsschätzer $\hat{q}_{Q\alpha}(z, x)$ in Wahrscheinlichkeit gegen $q_\alpha(z, x)$ konvergiert, wenn der Stichprobenumfang $N \rightarrow \infty$ und gleichzeitig mit geeigneter Rate die Zahl der Neuronen $Q \rightarrow \infty$.

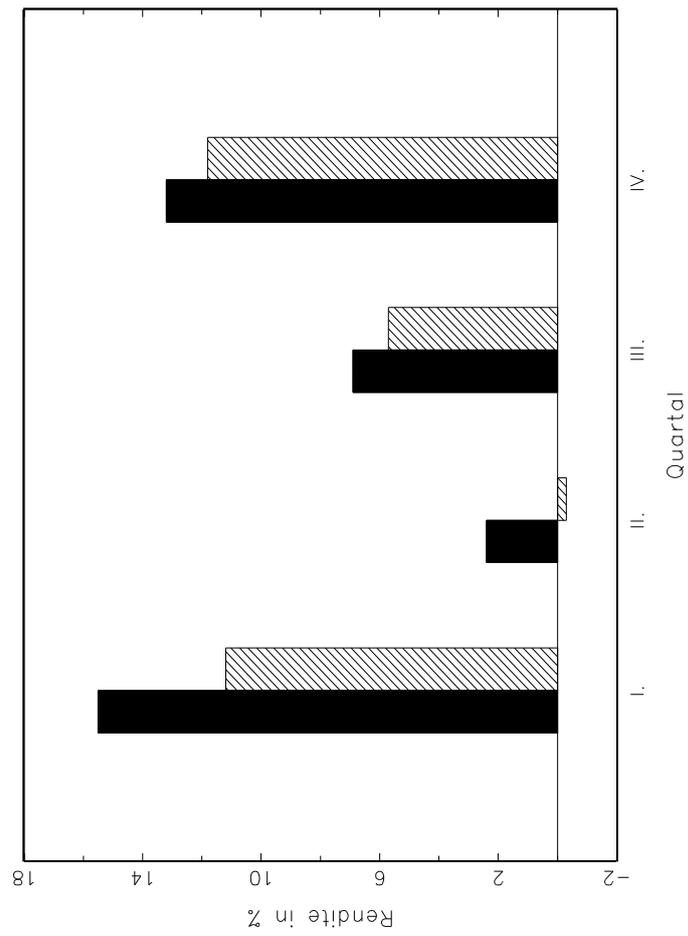


Abbildung 16.7.: Quartalsrenditen des Netzportfolios (schwarz) und des Indexportfolios (schraffiert) in 1996

17. Volatilitätsrisiko von Optionsportfolios

17.1. Einführung

In diesem Kapitel analysieren wir die maßgeblichen Faktoren der Struktur­dynamik implizierter Volatilitäten "am Geld" (*at the money, ATM*). Wir orientieren uns dabei eng an ?. Die Datengrundlage sind tägliche VDAX-Werte. Unter Verwendung der Hauptkomponentenanalyse betrachten wir einen Ansatz zur Modellierung des Risikos von Optionsportfolios auf der Basis des "Maximum Loss".

Zwischen der Bewertung von Optionen und der Prozessvarianz des zugrundeliegenden Finanztitels besteht ein enger Zusammenhang. Unter der Annahme einer geometrischen Brown'schen Bewegung des Kurses haben wir in Kapitel 6.3 die Black/Scholes-Formel (B/S) für die Bewertung europäischer Optionen hergeleitet. Mit dieser Formel ist der Optionspreis zu einem gegebenen Zeitpunkt eine Funktion des Volatilitätsparameters, wenn folgende Größen gegeben sind: τ (Restlaufzeit in Jahren), K (Ausübungskurs), r (sicherer, laufzeitkongruenter Zins) und S (Preis des zugrundeliegenden Finanzobjekts zum aktuellen Zeitpunkt). Alternativ kann man die zu einem bestimmten Zeitpunkt am Markt beobachtbaren Optionspreise mit Hilfe der B/S-Formel auf die sogenannten "implizierten" Volatilitäten abbilden (siehe Kapitel 6.3.4). Bei dieser Vorgehensweise findet man regelmäßig eine U-förmige Gestalt der resultierenden Volatilitätsfläche über verschiedene Restlaufzeiten und Ausübungskurse vor. Dieses Phänomen wird auch als "Volatilitäts-Smile" bezeichnet. Abbildung 17.1 verdeutlicht die typische Form der Volatilitätsfläche am Beispiel der DAX-Option. Gezeigt wird die implizierte Volatilität als Funktion der Moneyness und der Restlaufzeit τ . Dabei bezeichnet der Begriff *Moneyness* $\frac{S}{K}$ den Quotienten aus dem aktuellen Kurs S des zugrundeliegenden Finanztitels und dem Ausübungskurs K der jeweiligen Option. Hierbei ist zu beachten, dass am Markt nur Optionen für diskrete Basispreise und Restlaufzeiten gehandelt werden. Bei der Bestimmung der Volatilitätsfläche müssen daher Glättungsmethoden zum Einsatz kommen.

Bei Beobachtung der Volatilitätsfläche im Zeitablauf werden starke Veränderungen von Lage und Struktur deutlich. Für eine Reihe finanzmarktorientierter Anwendungen ist die Identifikation dieser zeitlichen Dynamik von zentraler Bedeutung. Dies trifft insbesondere auf das Risikomanagement von Optionsportfolios zu. Für die Ermittlung der Volatilitätsdynamik bietet sich die Hauptkomponentenanalyse als ein geeignetes Instrument an, [Skiadopoulos, Hodges and Clewlow \(1998\)](#). Die gesamte Laufzeitenstruktur kann durch eine geringe Anzahl von Hauptkomponenten hinreichend gut repräsentiert bzw. die Dimension des Fak-

TO DO:
Verweis
HSchmidt:2
- Haerd-
le und
Schmidt
- in Lite-
raturver-
zeichnis
aufnehmen

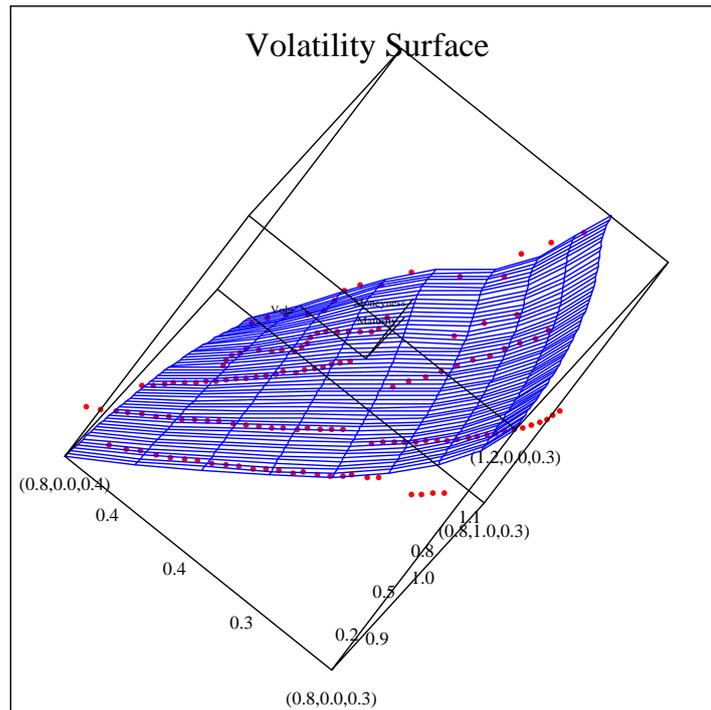


Abbildung 17.1.: Implizierte Volatilitätsfläche der DAX-Option am 18. Juli 1998

[SFMVolSurfPlot.xpl](#)

torraumes zum Zwecke der Risikoanalyse deutlich reduziert werden.

DAX-Optionen gehören zu den meistgehandelten Kontrakten der Deutsch-Schweizerischen Terminbörse "EUREX". An jedem Handelstag findet man eine beachtliche Anzahl liquider Serien mit unterschiedlichen Basispreisen und Laufzeiten (K, τ) am Markt vor, die grundsätzlich zur Berechnung implizierter Volatilitäten herangezogen werden können. In Anbetracht der in der Praxis oft begrenzten Datenverarbeitungskapazitäten ist die für eine zeitnahe Risikosteuerung notwendige Kalkulation zahlreicher Volatilitäten sowie der zugehörigen partiellen Ableitungen eines umfangreichen Optionsportfolios jedoch nicht praktikabel. Selbst bei entsprechend vorhandenen Ressourcen ist die isolierte Behandlung jeder implizierten Volatilität als eine separate Risikoquelle problematisch, denn sie resultiert in unstrukturierten bzw. "rauen" Volatilitätsflächen. Nutzt man entsprechend generierte Volatilitäten zur Kalibrierung von Optionspreis- bzw. Risikomodellen, so kann dies zu schwerwiegenden Spezifikationsfehlern führen und die Resultate entsprechender Handels- und Absicherungsstrategien entscheidend verschlechtern. Als Resultat einer Hauptkomponentenanalyse können im Gegensatz dazu mit einem überschaubaren Ressourcenaufwand "glatte" Volatilitätsflächen generiert werden. Diese erlauben eine bessere Modell-Kalibrierung und eine genauere Schätzung von Portfolio-Sensitivitäten, [Fengler, Härdle and Villa \(2001\)](#).

17.2. Datenbeschreibung

Für unsere Untersuchung der Dynamik implizierter Volatilitäten nutzen wir den von der Deutsche Börse AG zur Verfügung gestellten Volatilitätsindex (VDAX) bzw. die Schlußstände der zugehörigen VDAX-Subindizes. Diese Indizes spiegeln die implizierten Volatilitäten der DAX-Option "am Geld" für Laufzeiten von einem bis 24 Monaten wider. Die entsprechenden Werte werden aus am Markt beobachteten Preisen unter Verwendung der Black-Scholes-Formel (6.24) ermittelt:

$$C(S, \tau) = e^{(b-r)\tau} S \Phi(y + \sigma\sqrt{\tau}) - e^{-r\tau} K \Phi(y),$$

wobei Φ die Verteilungsfunktion der Standardnormalverteilung ist und

$$y = \frac{\ln \frac{S}{K} + (b - \frac{1}{2}\sigma^2)\tau}{\sigma\sqrt{\tau}},$$

Der einzige nicht unmittelbar am Markt beobachtbare Parameter der B/S-Formel ist die tatsächliche Volatilität σ des preisgenerierenden Prozesses. Grundsätzlich kann zwar die Prozeßvolatilität aus historischen Finanzmarktdaten geschätzt werden, siehe 6.3.4, jedoch ist allgemein bekannt, dass die grundlegende Annahme des B/S-Modells, eine geometrische Brown'sche Bewegung des zugrundeliegenden Finanztitels, in der Realität nur näherungsweise gegeben ist. Alternativ wird die B/S-Formel benutzt, um aus dem Marktpreis einer bestimmten Option den σ -Wert als implizierte Volatilität zu berechnen. Dies bedeutet jedoch nicht, dass die Marktteilnehmer die Annahmen des Black-Scholes-Verfahrens akzeptieren. Vielmehr nutzen Sie die B/S-Formel als eine bequeme Möglichkeit der Quotierung und Preisstellung von Optionen über diese Parameter.

Anhand der zu einem gegebenen Zeitpunkt für verschiedene Restlaufzeiten τ und Ausübungspreise K beobachteten implizierten Volatilitäten kann die Erwartung der Marktteilnehmer hinsichtlich der zukünftigen tatsächlichen Volatilität des zugrundeliegenden Finanztitels eingeschätzt werden. Hierbei muss beachtet werden, dass aus der implizierten Volatilität des B/S-Modells nicht direkt auf die tatsächliche Prozessvarianz des Kurses geschlossen werden kann. Die implizierte B/S-Volatilität reflektiert zwar eine Markterwartung, jedoch kann der theoretische Zusammenhang zwischen ihr und der tatsächlichen Volatilität nur auf der Grundlage bestimmter Annahmen hergestellt werden, siehe [Schönbucher \(1998\)](#), [Härdle and Hafner \(2000\)](#).

Implizierte Volatilitäten für ATM-DAX-Optionen werden für verschiedene Fristigkeiten von der Deutsche Börse AG kalkuliert. Eine detaillierte Beschreibung der Berechnungsweise des VDAX und seiner Subindizes findet sich bei [Redelberger \(1994\)](#). Dabei werden seit dem 18. März 1996 Optionslaufzeiten von ursprünglich 1, 2, 3, 6, 9, 12, 18 und 24 Monaten berücksichtigt. Zu diesem Zeitpunkt wurde an der EUREX der Handel in sogenannten "Long Term Options", d.h. in Optionen mit einer über 12 Monaten liegenden Laufzeit, aufgenommen. Unter Verwendung von Schlußnotierungen berechnet die Deutsche Börse AG börsentäglich insgesamt acht VDAX-Subindizes (VDAX = *Volatilitäts-DAX*) für die oben aufgeführten Laufzeiten. Diese spiegeln die implizierten Volatilitäten der jeweiligen DAX-Optionen "am

Geld” wider. Die Laufzeiten-Struktur für ATM DAX-Optionen kann börsentäglich anhand der VDAX-Indizes ermittelt werden. Abbildung 17.2 zeigt typische Strukturverläufe für 4 verschiedene Tage im Beobachtungszeitraum. Die Abbildung zeigt starke Veränderungen von Lage und Gestalt der Struktur im Zeitablauf.

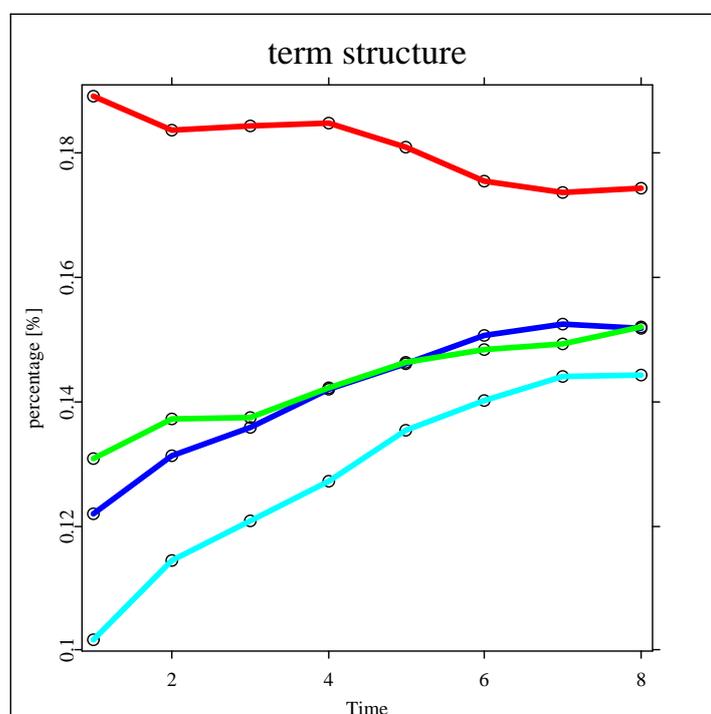


Abbildung 17.2.: Laufzeitstruktur von Implizierten DAX-Volatilitäten “am Geld”
[SFMVolaTermStructure.xpl](#)

Unsere hier durchgeführte Analyse beschränkt sich nicht nur auf die besonders liquiden Optionslaufzeiten, welche durch die ersten vier VDAX-Subindizes repräsentiert werden. Vielmehr beziehen wir aus folgenden Gründen alle acht Subindizes in die Analyse ein:

Zunächst findet an zahlreichen Handelstagen auch in den ”entfernten” Optionskontrakten (d.h. den Kontrakten mit Restlaufzeiten von mehr als einem Jahr) ein reger Handel statt, so dass die analytische Vernachlässigung der betreffenden Subindizes einen Informationsverlust zur Folge hätte. VDAX-Subindizes für lange Laufzeiten werden von der Deutschen Börse seit dem 18. März 1996 kalkuliert. Nach dem Verfalltagstermin vom 19. Dezember 1997 verschlechterte sich die Qualität der uns zur Verfügung gestellten Daten deutlich. Neben oft tagelang unveränderten Werten fehlten bei den entfernteren Subindizes häufig die entsprechenden Einträge. Vor diesem Hintergrund beschränken wir unsere Analyse auf den Zeitraum vom 18. März 1996 bis zum 19. Dezember 1997.

Die Einbeziehung relativ illiquider DAX-Optionen mit langen Laufzeiten erscheint auch aus einem weiteren Grund sinnvoll: Für unsere Analyse benötigen wir konstante Optionslaufzeiten, da die tägliche Verkürzung der Restlaufzeiten bei unbereinigten Daten zu erheblichen

Sub 1	Sub 2	Sub 3	Sub 4	Sub 5	Sub 6	Sub 7	Sub 8
20.8	9.06	6.66	6.84	4.29	2.48	2.11	1.38
9.06	9.86	6.67	4.44	3.21	1.72	1.11	0.92
6.66	6.67	6.43	3.87	2.63	1.49	1.01	0.53
6.84	4.44	3.87	4.23	2.66	1.39	1.38	0.68
4.29	3.21	2.63	2.66	2.62	1.03	1.02	0.51
2.48	1.72	1.49	1.39	1.03	2.19	0.63	0.33
2.11	1.11	1.01	1.38	1.02	0.63	1.76	0.43
1.38	0.92	0.53	0.68	0.51	0.33	0.43	152

Tabelle 17.1.: Empirische Kovarianz-Matrix $\hat{\Omega}$ der ersten Differenzen (alle Werte mit 10^5 multipliziert) [SFMVolaCov.xpl](#)

Verzerrungen der Analyseergebnisse führen kann. Dies gilt insbesondere für Optionen mit sehr kurzen Restlaufzeiten. Daher halten wir die Verwendung von interpolierten Volatilitäten mit jeweils konstanten Restlaufzeiten der zugrundeliegenden Optionen für zwingend notwendig. Wir verwenden in Anlehnung an das von der Deutsche Börse AG zur Berechnung des VDAX verwendete Verfahren die folgende lineare Interpolation:

Zu festen Laufzeiten von $\tau_1^* = 30, \tau_2^* = 60, \tau_3^* = 90, \tau_4^* = 180, \tau_5^* = 270, \tau_6^* = 360, \tau_7^* = 540, \tau_8^* = 720$ Kalendertagen berechnen wir täglich Volatilitätsindizes $\hat{\sigma}_t(\tau_j^*), j = 1, \dots, 8$, unter Verwendung der VDAX-Subindizes mit der jeweils nächsten kürzeren bzw. längeren Laufzeit $\hat{\sigma}_t(\tau_j^-)$ und $\hat{\sigma}_t(\tau_j^+)$ mittels

$$\hat{\sigma}_t(\tau_j^*) = \hat{\sigma}_t(\tau_j^-) \left[1 - \frac{\tau_j^* - \tau_j^-}{\tau_j^+ - \tau_j^-} \right] + \hat{\sigma}_t(\tau_j^+) \left[\frac{\tau_j^* - \tau_j^-}{\tau_j^+ - \tau_j^-} \right]. \quad (17.1)$$

Auf diese Weise erhalten wir 8 Volatilitätszeitreihen mit jeweils fester Restlaufzeit. Jede Zeitreihe repräsentiert ein gewichtetes Mittel zweier benachbarter VDAX-Subindizes und basiert auf $n = 441$ täglichen Beobachtungen der implizierten DAX-Volatilitäten “am Geld“.

17.3. Hauptkomponentenanalyse der VDAX-Dynamik

Zunächst untersuchen wir die Daten mit Hilfe des “Augmented Dickey-Fuller“-Tests (ADF-Test - vgl. (10.49)) auf Stationarität. Die Nullhypothese einer Einheitswurzel kann für die einzelnen VDAX-Subindizes $\hat{\sigma}_t(\tau_j^*)$ nicht einmal zum 10%-Niveau abgelehnt werden. Dieses Ergebnis legt die Verwendung erster Differenzen $x_{jt} = \Delta[\hat{\sigma}_t(\tau_j^*)] = \hat{\sigma}_{t+1}(\tau_j^*) - \hat{\sigma}_t(\tau_j^*), t = 1, \dots, n - 1$, der implizierten Volatilitätsindizes für die weitere Analyse nahe. Erneute ADF-Tests bestätigen die Annahme der Stationarität für die ersten Differenzen. [SFMadf.xpl](#)

Seien \bar{x}_j die jeweiligen Stichprobenmittel der ersten Differenzen x_{jt} . Tabelle 17.1 enthält die empirische Kovarianz-Matrix $\hat{\Omega}$ als Schätzer für die 8×8 -Matrix Ω der Kovarianzen

Hauptkomponenten	1	2	3	4	5	6	7	8
erklärter Varianzanteil	70.05%	13.06%	5.57%	3.11%	3.06%	2.12%	1.93%	1.10%
kumulierter Anteil	70.05%	83.12%	88.69%	91.80%	94.86%	96.97%	98.90%	100%

Tabelle 17.2.: Mittels Hauptkomponenten erklärte Stichprobenvarianz  SFMVolaPCA.xpl

$Cov(x_{it}, x_{jt}), i, j = 1, \dots, 8$. Mit Hilfe der Jordan-Zerlegung erhalten wir $\hat{\Omega} = \hat{\Gamma} \hat{\Lambda} \hat{\Gamma}^T$. Die Diagonalmatrix $\hat{\Lambda}$ enthält die Eigenwerte $\hat{\lambda}_k, k = 1, \dots, 8$ von $\hat{\Omega}$ als Diagonalelemente, $\hat{\Gamma}$ die Eigenvektoren. Zeitreihen der Hauptkomponenten erhält man mittels $Y = X_C \hat{\Gamma}$, wobei X_C die 440×8 -Matrix der zentrierten ersten Differenzen $x_{jt}^c = x_{jt} - \bar{x}_j, j = 1, \dots, 8, t = 1, \dots, 440$, darstellt. Die 440×8 -Matrix $Y = (Y_1, \dots, Y_8), Y_j = (y_{1j}, y_{2j}, \dots, y_{440,j})^T$ enthält die Hauptkomponenten.

Wie sehr bereits die ersten l Hauptkomponenten den Prozess der zentrierten ersten Differenzen bestimmen, kann anhand des erklärten Varianzanteils φ_l in Relation zur Gesamtvarianz der Daten gemessen werden. Der erklärte Varianzanteil entspricht dem relativen Anteil der betreffenden Eigenwerte, d.h.

$$\varphi_l = \frac{\sum_{k=1}^l \lambda_k}{\sum_{k=1}^8 \lambda_k} = \frac{\sum_{k=1}^l Var(y_{tk})}{\sum_{k=1}^8 Var(y_{tk})}, \quad (17.2)$$

wobei $\lambda_k, k = 1, \dots, 8$ die Eigenwerte der wahren Kovarianzmatrix Ω sind. Ein Schätzer für φ_l ist

$$\hat{\varphi}_l = \frac{\sum_{k=1}^l \hat{\lambda}_k}{\sum_{k=1}^8 \hat{\lambda}_k}.$$

In Tabelle 17.2 werden sowohl die individuellen erklärten Varianzanteile $\hat{\lambda}_l / \sum_{k=1}^8 \hat{\lambda}_k$ wie auch der über l Hauptkomponenten kumulierte Varianzanteil $\hat{\varphi}_l$ dargestellt. Deutlich wird dabei, dass die erste Hauptkomponente bereits 70% der Gesamtvarianz des verwendeten Datenvorrats erklärt. Mit der zweiten Hauptkomponente können weitere 13% der Gesamtvarianz im Untersuchungszeitraum erklärt werden. Zusammengefasst lässt sich ein Anteil in Höhe von 83% der Varianz der untersuchten ersten Differenzen unserer VDAX-Subindizes mit Hilfe der ersten beiden Hauptkomponenten darstellen. Offenbar nimmt die Erklärungsstärke der Hauptkomponenten ab der dritten Hauptkomponente deutlich ab.

Bei Abbildung der Eigenwerte in einer Grafik zeigt sich eine Gestalt mit einer starken Krümmung bei der zweiten Hauptkomponente. Gemäß dem bekannten "Ellenbogen"-Kriterium, sollten die ersten beiden Hauptkomponenten mit einem Erklärungsgehalt von über 80% der Gesamtvarianz zur Beschreibung des Datensatzes als ausreichend angesehen werden. Die verbleibende Varianz kann für analytische Zwecke als Auswirkung einer unsystematischen Störgröße aufgefasst werden. Abbildung 17.3 enthält die Faktorladungen der ersten beiden Hauptkomponenten. Aufgrund der Orthogonalität der Komponenten können

die Ladungen mittels Kleinste-Quadrate-Regression aus den Einzelgleichungen

$$x_{jt}^c = \sum_{l=1}^2 b_{jl} y_{lt} + \varepsilon_t, \quad (17.3)$$

geschätzt werden. Dabei ist ε_t ein unabhängiger Störterm.

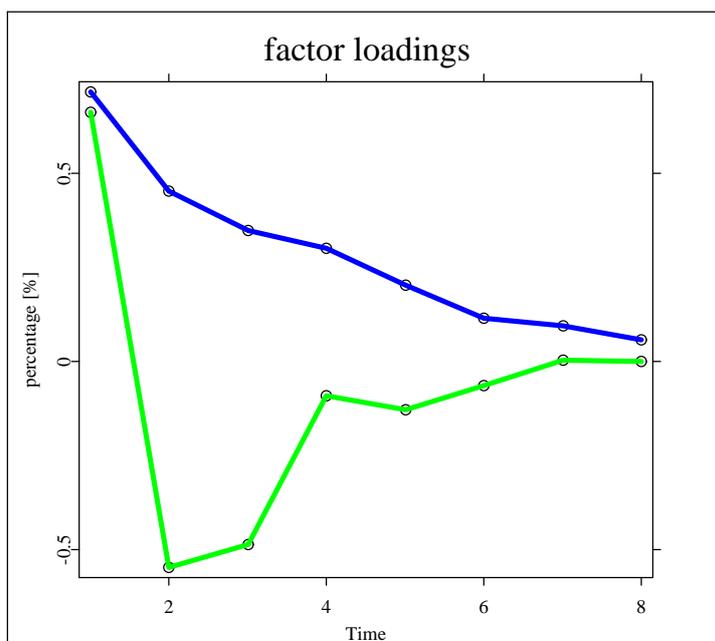


Abbildung 17.3.: Faktorladungen der ersten und zweiten Hauptkomponente [SFMPCA.xpl](#)

Anhand der Faktorladungen wird deutlich, dass ein Schock im ersten Faktor die implizierten Volatilitäten aller erfaßten Optionslaufzeiten auf ähnliche Weise beeinflusst bzw. eine nicht-parallele Verschiebung der Laufzeitstruktur bewirkt. Ein Schock in der zweiten Hauptkomponente bewirkt hingegen eine Verformung der Strukturkurve: Während bei der kürzesten Laufzeit eine positive Veränderung bewirkt wird, werden die längeren Laufzeiten negativ beeinflusst. Die absolute Größe des Effekts eines Schocks nimmt bei beiden Faktoren mit der Optionslaufzeit ab.

17.4. Stabilitätsanalyse der VDAX Dynamik

Für eine sinnvolle Anwendung der Hauptfaktoren zur Risikomessung von Portfolios müssen wir die Stabilität im Zeitablauf untersuchen. Wenn die Hauptkomponenten und die Faktorladungen im Zeitablauf starken Veränderungen unterliegen, so wäre eine globale Analyse nicht geeignet, um die künftige Varianz der implizierten Volatilitäten abzubilden und die Risiken von Optionsportfolios hinreichend genau beurteilen zu können.

Hauptkomponente:	1	2	3	4	5	6	7	8
Wochendaten								
18.03.96-19.12.97	73.85%	91.59%	95.09%	97.38%	98.80%	99.61%	99.88%	100%
Tagesdaten								
18.03.96-19.12.97	70.05%	83.12%	88.69%	91.80%	94.86%	96.97%	98.90%	100%
Subperiode 1								
18.03.96-05.02.97	83.36%	91.84%	94.65%	96.39%	97.76%	98.78%	99.52%	100%
Subperiode 2								
05.02.97-19.12.97	68.22%	82.21%	87.99%	91.35%	94.64%	96.93%	98.86%	100%

Tabelle 17.3.: Erklärte Varianzteile in verschiedenen Subperioden

Unser Vorgehen berücksichtigt zwei Aspekte: Zunächst ist der Zufallsanteil in Tagesdaten möglicherweise deutlich höher als in Wochendaten. Eine mögliche Ursache hierfür sind Asynchronitäten im Handelsverlauf, bedingt durch eine häufige Aktualisierung der Optionsquotierungen in den liquiden Kontrakten mit kurzer Restlaufzeit und "veraltete" Preise in den langlaufenden Kontrakten. Um den möglichen Einfluß dieses Effekts erkennen zu können, führen wir unsere Analyse analog auf der Basis wöchentlicher Daten durch. Bei hinreichender Stabilität der Hauptkomponentenanalyse sollte die Verwendung täglicher bzw. wöchentlicher Daten zu grundsätzlich ähnlichen Ergebnissen führen.

Des weiteren teilen wir unseren Datenvorrat in zwei nichtüberlappende Perioden gleicher Länge. Jede Teilperiode enthält $m = 220$ tägliche Beobachtungen des Differenzenprozesses. Wir führen für jede der beiden Teilstichproben wie oben beschrieben eine Hauptkomponentenanalyse durch und vergleichen die beiden jeweils größten Eigenwerte $\hat{\lambda}_k^i$, $k = 1, 2$, in den beiden Subperioden $i = 1, 2$.

Wie bereits erwähnt kann der Effekt von im Tagesverlauf anfallenden asynchronen Daten durch die Verwendung von Wochendaten eliminiert werden. Aus Tabelle 17.3 geht hervor, dass sich die Erklärungsgüte der ersten Hauptkomponenten bei wöchentlichen Daten etwas erhöht. Dies überrascht aufgrund des zu erwartenden höheren Störungsanteils in Tagesdaten nicht. Insgesamt bewegen sich die erklärten Varianzanteile bei Verwendung von Wochendaten in ähnlichen Größenordnungen. Dies spricht für die Stabilität des hier verwendeten Analyseverfahrens gegenüber Verzerrungen aufgrund von Asynchronitäten in den täglichen Daten.

Aus Tabelle 17.3 geht hervor, dass die durch die ersten beiden Hauptkomponenten erklärten Varianzanteile in der zweiten Subperiode abnehmen. Vor diesem Hintergrund ist ein Stabilitätstest geboten: Ein zweiseitiges Konfidenzintervall für die Differenz der Eigenwerte in den beiden Subperioden entspricht

$$\ln \hat{\lambda}_k^1 - 2q_\alpha \sqrt{\frac{2}{m-1}} \leq \ln \hat{\lambda}_k^2 \leq \ln \hat{\lambda}_k^1 + 2q_\alpha \sqrt{\frac{2}{m-1}}, \quad (17.4)$$

wobei q_α das α -Quantil einer Standardnormalverteilung kennzeichnet, siehe Härdle and Simar (2001). Hieraus folgt

$$|\ln \hat{\lambda}_k^1 - \ln \hat{\lambda}_k^2| \geq 2q_\alpha \sqrt{\frac{2}{m-1}} \quad (17.5)$$

als zweiseitiger Test für $H_0 : \lambda_k^1 = \lambda_k^2$. Unter der Nullhypothese sind die betreffenden Eigenwerte in beiden Perioden gleich. Die Nullhypothese wird verworfen, falls die Ungleichung für einen entsprechenden kritischen q -Wert erfüllt ist. Dies würde auf eine Instabilität der Hauptkomponentenanalyse im Zeitablauf hindeuten.

Kritische Werte für die Ablehnung der Nullhypothese sind 0.313 (Irrtumswahrscheinlichkeit 10%), 0.373 (Irrtumswahrscheinlichkeit 5%) und 0.490 (Irrtumswahrscheinlichkeit 1%). Die Differenzen der geschätzten Eigenwerte betragen 0.667 und 1.183. Beide Differenzen sind mit einer Irrtumswahrscheinlichkeit von 1% signifikant größer als Null. Dieses Ergebnis weist darauf hin, dass die Bestimmungsfaktoren der Volatilitätsdynamik im Zeitablauf Veränderungen unterliegen. Bei der Risikobestimmung von Optionsportefeuilles erscheint daher die Anwendung eines adaptiven Verfahrens der Hauptkomponentenanalyse geboten. Dabei wird die Schätzung über ein gleitendes historisches Datenfenster periodisch durchgeführt und die Fensterlänge adaptiv festgelegt, siehe [Härdle, Spokoiny and Teyssiere \(2000\)](#).

17.5. Messung des implizierten Volatilitätsrisikos

Der Marktwert P_t eines aus w unterschiedlichen Optionen bestehenden Portfolios ist abhängig von Änderungen der risikolosen Zinsen r_t , des Preises S_t des zugrundeliegenden Finanztitels, der Restlaufzeit τ und der individuellen implizierten Volatilitäten σ_I . Änderungen des Portfolioswertes können analytisch anhand der folgenden Taylor-Entwicklung approximiert werden, wobei von Optionen auf den selben zugrundeliegenden Titel ausgegangen wird.

$$\begin{aligned} \Delta P_t &= \sum_{u=1}^w \left\{ \frac{\partial V_{ut}}{\partial \sigma_I} \Delta \sigma_{I,t} + \frac{\partial V_{ut}}{\partial t} \Delta t + \frac{\partial V_{ut}}{\partial r} \Delta r_t + \frac{\partial V_{ut}}{\partial S} \Delta S_t + \frac{1}{2} \frac{\partial^2 V_{ut}}{\partial S^2} (\Delta S_t)^2 \right\} \\ \Delta P_t &= \sum_{u=1}^w \left\{ \frac{\partial V_{ut}}{\partial \sigma_I} \Delta \sigma_{I,t} + \Theta_u \Delta t + \rho_u \Delta r_t + \Delta_u \Delta S_t + \frac{1}{2} \Gamma_u (\Delta S_t)^2 \right\} \end{aligned} \quad (17.6)$$

Hier beschreibt V_{ut} den Preis der u -ten Option mit Restlaufzeit τ_u zum Zeitpunkt t , und $\Theta_u, \rho_u, \Delta_u, \Gamma_u$ sind die in Abschnitt 6.3 beschriebenen Kenngrößen der u -ten Option. In der Praxis gehen Optionshändler oftmals direkte "Vega"-Positionen ein. Dabei bauen sie Portfolios auf, deren Gewinn- und Verlustprofil ausschließlich durch Änderungen der implizierten Volatilitäten der jeweiligen Optionen bestimmt wird, siehe [Taleb \(1997\)](#). Derartige Portefeuilles bezeichnet man als (Δ, Γ) und Θ neutral. Die Empfindlichkeit des betreffenden Optionspreises gegenüber Volatilitätsänderungen wird anhand der Kennzahl \mathcal{V} ("Vega" - siehe (6.29)) gemessen.

Eine verbreitete Strategie zur Ausnutzung prognostizierter Veränderungen der Laufzeitenstruktur implizierter Volatilitäten besteht im Kauf und Verkauf sogenannter "Straddles" mit unterschiedlichen Restlaufzeiten. Der Aufbau eines Straddles erfolgt durch gleichzeitigen Kauf ("Long Straddle") oder Verkauf ("Short Straddle") einer gleichen Anzahl von ATM Kauf- und Verkaufsoptionen mit gleicher Restlaufzeit. Erwartet ein Händler einen relativ starken Anstieg der implizierten Volatilitäten in den kurzen Restlaufzeiten und einen relativ schwächeren Anstieg in den längeren Restlaufzeiten, so wird er kurzlaufende Straddles kaufen und langlaufende Straddles in einem geeigneten Verhältnis verkaufen. Das resultierende

Optionsportfolio ist (Δ, Γ) neutral und über kurze Zeiträume Θ neutral, d.h. insensitiv gegenüber Zeitwertverlusten. Die oben dargestellte Taylor-Reihe kann daher reduziert werden auf:

$$\Delta P_t \approx \sum_{u=1}^w \left\{ \frac{\partial V_{ut}}{\partial \sigma_I} \Delta \sigma_{impl,t}^{(t)} \right\} \quad (17.7)$$

Die ersten Differenzen der implizierten Volatilitäten werden nun als Linearkombination der Hauptkomponenten dargestellt. Bei stellvertretender Verwendung der den tatsächlichen implizierten Volatilitäten $\hat{\sigma}(\tau_u^*)$ zeitlich am nächsten gelegenen Volatilitätsindizes $\sigma_{I,t}$ erhält man folgende gemäß (17.3) Darstellung:

$$\Delta P_t \approx \sum_{u=1}^w \left\{ \frac{\partial V_{ut}}{\partial \sigma_I} \left(\sum_{l=1}^2 b_{jl} y_{lt} \right) \right\} \quad (17.8)$$

Die Zahl der im voranstehenden Ausdruck verwendeten Hauptkomponenten kann ohne gravierenden Informationsverlust auf die ersten beiden reduziert werden.

Das im folgenden vorzustellende Konzept des "Maximum Loss" (ML) beschreibt die Wahrscheinlichkeitsverteilung einer kurzfristigen Änderung des Portfoliowertes in Abhängigkeit von Veränderungen der wertbestimmenden Einflußgrößen. Die Wertänderung eines (Δ, Γ) -neutralen Optionsportfolios wird maßgeblich durch Veränderungen der implizierten Volatilitäten der im Portefeuille enthaltenen Optionen bestimmt. Zur Feststellung des "Maximum Loss" ist daher eine hinreichend genaue Abbildung der künftigen Verteilung der Volatilitätsänderungen von Optionen mit verschiedenen Restlaufzeiten erforderlich.

Der "Maximum Loss" ist definiert als der größtmögliche Verlust eines Portfolios, der über einen bestimmten Faktorraum A_t , entstehen kann und über eine bestimmte Halteperiode τ auftreten. Der Faktorraum A_t wird dabei durch eine abgeschlossene Menge mit $P(A_t) = \alpha$ bestimmt. Hierbei wird α auf 99% oder 99,9% gesetzt. Die ML-Definition ähnelt auf den ersten Blick der "Value-at-Risk"-Definition (siehe Kapitel 15). Dennoch besteht zwischen beiden Konzepten ein wesentlicher Unterschied: Während bei der Berechnung des "Value-at-Risk" die Renditeverteilung des gehaltenen Portfolios bekannt sein muss, ist der ML direkt über den Faktorraum definiert und besitzt somit einen zusätzlichen Freiheitsgrad, siehe [Studer \(1995\)](#).

In unserer Analyse haben wir die Laufzeitstruktur der implizierten Volatilitäten in zwei Hauptkomponenten zerlegt, die einen wesentlichen Teil der Variabilität der Strukturkurve erklären. Somit repräsentieren die ersten beiden Hauptkomponenten die im Rahmen einer ML-Modellierung zu verwendenden Risikofaktoren. Das Gewinn- und Verlustprofil des jeweils gehaltenen Portfolios wird in Abhängigkeit von den entsprechenden Änderungen der Risikofaktoren über ein geeignetes Bewertungsmodell ermittelt. Um dies zu erreichen, muß eine Bewertung des zugrunde liegenden Portfolios theoretisch für jeden Punkt des Faktorraums erfolgen. In der praktischen Anwendung wird der Faktorraum über ein hinreichend enges Gitter diskreter Datenpunkte $y_1^z (z = 1, \dots, N_1)$ abgetastet, wobei der jeweils andere Risikofaktor konstant gehalten wird. Wegen der Orthogonalitätseigenschaft der Hauptkomponen-

ten ist die Gewinn- und Verlustfunktion $PL()$ additiv mit $PL(y_1^{z_1}, y_2^{z_2}) = PL(y_1^{z_1}) + PL(y_2^{z_2})$.

Unter der Annahme multivariat normalverteilter Hauptkomponenten können Konfidenzbereiche für den "Maximum Loss" über die gemeinsame Dichte

$$\varphi_2(y) = \frac{1}{(2\pi)\sqrt{\det \Lambda_2}} \exp\left(-\frac{1}{2}y^T \Lambda_2^{-1}y\right), \quad (17.9)$$

mit $y = (y_1, y_2)^T$ konstruiert werden. Dabei entspricht die Matrix Λ_2 der (2,2) Diagonalmatrix der Eigenwerte λ_k , $k = 1, 2$. Die Zufallsvariable $y^T \Lambda_2^{-1}y = X_1^2 + X_2^2$ ist Chi-Quadratverteilt. Der Konfidenzbereich für ein bestehendes Portfolio resultiert mit $A_t = (y \mid y^T \Lambda_2^{-1}y \leq c_\alpha)$, c_α , wobei c_α das α -Quantil einer Chi-Quadratverteilten Zufallsvariable mit 2 Freiheitsgraden ist.

18. Nichtparametrische Schätzer für Kreditausfallwahrscheinlichkeiten

Das Schätzen von Ausfallwahrscheinlichkeiten auf der Grundlage von Informationen über Privatkunden oder Firmen ist ein wichtiger Bestandteil des Kreditscreening, d.h. der Beurteilung der Kreditwürdigkeit. Das Schätzen der Wahrscheinlichkeit, dass eine Firma innerhalb eines vorgegebenen Zeitraums von z.B. einem Jahr in Zahlungsschwierigkeiten gerät, ist auch eine wesentliche Information für die Erstellung eines Ratings oder für die Messung des Kreditrisikos. Auch hier erweisen sich nichtparametrische Ansätze als flexibles Handwerkszeug, um die gesuchten Ausfallwahrscheinlichkeiten ohne willkürliche Annahmen zu schätzen. In diesem Kapitel geben wir einen kurzen Überblick über verschiedene Ansätze, die nichtparametrisches oder semiparametrisches Schätzen bedingter Wahrscheinlichkeiten ermöglichen.

18.1. Logistische Regression

Zur Beurteilung der Kreditwürdigkeit eines Kunden stehen im allgemeinen eine Reihe von Daten zur Verfügung. Für einen Konsumentenkredit sind dies zum Beispiel (?): *Höhe des Kredits, Alter des Kunden, Dauer des Kredits* sowie die Informationen, ob der Kunde *arbeitslos* ist oder nicht und ob es früher schon *Probleme bei der Rückzahlung von Darlehen* gegeben hat. Für die Insolvenzprognose für ein kleines Unternehmen sind relevante Daten zum Beispiel (?): *Unternehmensalter, Umsatzentwicklung* in der jüngeren Vergangenheit, *Bildungsabschluss des Unternehmers, Gewerbe* und die Informationen über *Haftungsbeschränkungen*.

Ein Teil der Einflussgrößen sind wie Kreditvolumen oder Umsatzentwicklung quantitativ. Andere sind qualitativer Natur und müssen zum Schätzen der Ausfallwahrscheinlichkeit in Zahlen umgewandelt werden. Bei dichotomen Merkmalen (arbeitslos - nicht arbeitslos, beschränkte Haftung - unbeschränkte Haftung) werden hierzu Indikatorvariablen mit Werten 0 und 1 eingesetzt. Bei Merkmalen mit $d > 2$ möglichen und zudem ungeordneten Werten werden $d - 1$ Dummy-Variable eingeführt, die ebenfalls nur die Werte 0 oder 1 annehmen. Zur numerischen Kodierung des Merkmals *Gewerbe* mit den in Betracht gezogenen drei Ausprägungen *Handel, verarbeitendes Gewerbe, Sonstige* werden zwei Dummy-Variable Z_1, Z_2 benutzt, wobei $Z_1 = 1$ ($Z_2 = 1$) genau dann, wenn der Gewerbe *Handel* (*verarbeitendes Gewerbe*) ist. Ist $Z_1 = Z_2 = 0$, so gehört die betrachtete Firma zu einem der *sonstigen* Gewerbe, z.B. *Dienstleistungen*. Der Fall $Z_1 = Z_2 = 1$ kann nicht eintreten.

Sind die Werte eines qualitativen Merkmals hierarchisch geordnet, so bietet es sich an, es durch eine numerische Zufallsgröße zu repräsentieren. Der persönliche Eindruck des Sachbearbeiters einer Bank von der wirtschaftlichen Situation eines Unternehmens: *sehr gut, gut, befriedigend, schlecht, sehr schlecht* kann zum Beispiel in eine Notenskala: *1, 2, 3, 4, 5* übersetzt werden. Dabei ist aber zu berücksichtigen, dass jede monotone, d.h. ordnungserhaltende Transformation eine andere numerische Kodierung liefert, die erst einmal mit gleichem Recht benutzt werden kann. Statt *1, 2, 3, 4, 5* könnte man im Beispiel auch *0, 1, 3, 6, 10* wählen. Bei der Verwendung parametrischer Ansätze wie der logistischen Regression dieses Abschnitts sollte man diese Willkür bei der Festlegung von Notenskalen für hierarchische Merkmale bedenken. Durch eine monotone Transformation der Skala lassen sich eventuell bessere Ergebnisse beim Schätzen von Ausfallwahrscheinlichkeiten erzielen. Genügend flexible nicht- und semiparametrische Ansätze wählen dagegen automatisch eine geeignete Skalierung.

Um die Ausfallwahrscheinlichkeit eines Kredits, gegeben die bei der Vergabeentscheidung verfügbaren Information, zu schätzen, gehen wir von einer Stichprobe $(X_1, Y_1), \dots, (X_n, Y_n)$ von unabhängig, identisch verteilten Zufallsgrößen aus. $X_j \in \mathbb{R}^d$ steht für die bei Kreditvergabe vorhandenen Informationen über den j -ten Kunden, wobei bereits qualitative Merkmale wie oben diskutiert in numerische Werte übersetzt worden sind. $Y_j \in \{0, 1\}$ ist die Indikatorvariable des Kredits: sie nimmt den Wert 0 an, wenn das Darlehen ohne Probleme zurückgezahlt worden ist, bzw. den Wert 1, wenn der Kredit ganz oder teilweise ausgefallen ist. Die zu schätzende Ausfallwahrscheinlichkeit ist die bedingte Wahrscheinlichkeit, dass $Y_j = 1$, gegeben $X_j = x$:

$$\pi(x) = P(Y_j = 1 | X_j = x), \quad x \in \mathcal{X},$$

wobei $\mathcal{X} \subset \mathbb{R}^d$ für den Wertebereich von X_j steht.

Da $\pi(x)$ als Wahrscheinlichkeit nur Werte zwischen 0 und 1 annimmt, scheiden lineare Regressionsmodelle als Grundlage für Funktionsschätzer aus. Die Klasse der verallgemeinerten linearen Modellen (GLM = generalized linear models) eignet sich aber zum Schätzen von Wahrscheinlichkeiten. Dabei wird angenommen, dass

$$\pi(x) = G\left(\beta_0 + \sum_{i=1}^d x_i \beta_i\right) = G(\beta_0 + \beta^T x)$$

ist. $G: \mathbb{R} \rightarrow [0, 1]$ ist dabei eine bekannte Funktion, die nur Werte zwischen 0 und 1 annimmt, die reellwertigen Parameter β_0, \dots, β_d sind unbekannt und zu schätzen. Für den Spezialfall, dass G als logistische Funktion ψ gewählt wird:

$$G(t) = \psi(t) = \frac{1}{1 + e^{-t}},$$

erhalten wir das Modell der *logistischen Regression*: Gegeben X_1, \dots, X_n , sind die Kreditindikatoren Y_1, \dots, Y_n unabhängige Bernoulli-Zufallsgrößen mit Parametern $\psi(\beta_0 + \beta^T X_1), \dots, \psi(\beta_0 + \beta^T X_n)$. Die bedingte Likelihood-Funktion ist daher

$$L(\beta_0, \dots, \beta_d) = \prod_{j=1}^n [Y_j \psi(\beta_0 + \beta^T X_j) + (1 - Y_j) \{1 - \psi(\beta_0 + \beta^T X_j)\}].$$

Da Y_j nur die Werte 0 oder 1 annimmt, ist dementsprechend die bedingte log-likelihood-Funktion

$$\log L(\beta_0, \dots, \beta_d) = \sum_{j=1}^n [Y_j \log \psi(\beta_0 + \beta^T X_j) + (1 - Y_j) \log \{1 - \psi(\beta_0 + \beta^T X_j)\}].$$

Durch Maximieren von L oder $\log L$ erhält man die Maximum-Likelihood-Schätzer $\hat{\beta}_0, \dots, \hat{\beta}_d$ von β_0, \dots, β_d und damit auch den Maximum-Likelihood-Schätzer für die Ausfallwahrscheinlichkeit im logistischen Regressionsmodell:

$$\hat{\pi}(x) = \psi(\hat{\beta}_0 + \hat{\beta}^T x).$$

18.2. Semiparametrische Modelle für das Kredit scoring

Das logistische Regressionsmodell für das Schätzen bedingter Wahrscheinlichkeiten leidet unter denselben Einschränkungen wie das lineare Regressionsmodell beim Schätzen allgemeiner Funktionen. Um die Abhängigkeit von der speziellen parametrischen Modellform zu vermeiden und mehr Flexibilität bei der Funktionsschätzung zu gewinnen, liegt es nahe, $\pi(x)$ nichtparametrisch, zum Beispiel mit der durch (13.4) und (??) gegebenen LP-Methode, zu schätzen. Dabei ist allerdings nicht garantiert, dass der Funktionsschätzer zwischen 0 und 1 liegt. Um dies zu erzwingen, bietet sich wie im vorigen Abschnitt an, den Wertebereich der geschätzten Funktion durch eine gegebene Funktion G auf das Intervall $[0,1]$ zu transformieren:

$$\pi(x) = G(m(x))$$

wobei $m(x)$ eine beliebige reellwertige Funktion ist, die nichtparametrisch geschätzt werden kann. Für die Schätzung von Ausfallwahrscheinlichkeiten sind lokale Glättungsverfahren allerdings aus zwei Gründen wenig geeignet. Zum einen ist x in den Anwendungen oft so hochdimensional, z.B. nach Einführung der nötigen Dummy-Variablen 61-dimensional in dem von ? betrachteten Fall, dass selbst bei großen Stichprobenumfängen lokale Umgebungen von x , über die gemittelt wird, entweder zu wenige Beobachtungen enthalten oder zu groß sind, um eine verlässliche Schätzung von $m(x)$ zu gestatten. Dieses Problem ließe sich durch Einschränkung auf additive Modelle

$$\pi(x) = G\left(\sum_{i=1}^d m_i(x_i)\right)$$

beheben, wobei $m_1(u), \dots, m_d(u)$ beliebige Funktionen der eindimensionalen Variable u sind. Entscheidender ist allerdings, dass beim Kredit scoring viele Koordinaten von x nur zwei Werte 0 und 1 annehmen, da sie von vornherein dichotome Merkmale repräsentieren oder als Dummy-Variable für ungeordnete qualitative Merkmale eingeführt worden sind. Lokale Glättungsverfahren eignen sich von der ihnen zugrundeliegenden Philosophie her aber in erster Linie zum Schätzen von Funktionen kontinuierlicher Argumente.

Eine Kombination des nichtparametrischen mit dem parametrischen Ansatz bietet aber die Möglichkeit, die Flexibilität nichtparametrischer Verfahren auch beim Kredit scoring zu

nutzen (?). Hierfür werden die Einflussgrößen nicht zu einem Zufallsvektor X_j zusammengefasst, sondern auf zwei Zufallsvektoren $X_j \in \mathbb{R}^p, Z_j \in \mathbb{R}^q$ aufgeteilt. Die Koordinaten von Z_j repräsentieren einige ausgewählte ausschließlich quantitative Merkmale und eventuell hierarchische qualitative Merkmale mit genügend fein unterteiltem Wertebereich. Alle übrigen Merkmale, insbesondere dichotome und die Dummyvariablen ungeordneter qualitativer Merkmale, werden in X_j zusammengefasst. Als Grundlage für Schätzer der Ausfallwahrscheinlichkeiten wird ein verallgemeinertes partiell lineares Modell (GPLM = generalized partial linear model) betrachtet:

$$P(Y_j = 1 | X_j = x, Z_j = z) = \pi(x, z) = G(\beta^T x + m(z)).$$

G ist wieder eine bekannte Funktion mit Werten zwischen 0 und 1, zum Beispiel die logistische Funktion ψ . β_1, \dots, β_p sind unbekannte Parameter, m eine beliebige, unbekannte Funktion, die eine additive Konstante beinhalten kann und somit einen zusätzlichen Parameter β_0 überflüssig macht. ? haben in einer umfangreichen Fallstudie gezeigt, dass die zusätzliche Flexibilität durch den nichtparametrischen Teil $m(z)$ des Modells eine bessere Ausfallwahrscheinlichkeitsschätzung liefern kann als die rein parametrische logistische Regression.

Es gibt verschiedene Algorithmen zum Schätzen von β und $m(z)$, z.B. die Profile-Likelihood-Methode von ? und ? oder die auf [Hastie and Tibshirani \(1990\)](#) zurückgehende Backfitting-Methode. Im wesentlichen nutzen sie aus, dass bei bekannter Funktion $m(z)$ der Parametervektor β analog zur logistischen Regression durch Maximieren der log-likelihood Funktion

$$\log L(\beta) = \sum_{j=1}^n [Y_j \log G(\beta^T X_j + m(Z_j)) + (1 - Y_j) \log \{1 - G(\beta^T X_j + m(Z_j))\}]$$

geschätzt werden kann und bei bekanntem β die Funktion $m(z)$ durch lokale Glättung analog zur LP-Methode (13.4), (??). Diese beiden Optimierungsprobleme werden gemeinsam zu einem iterativen numerischen Algorithmus verknüpft.

BEISPIEL 18.1

Als Beispiel betrachten wir das bereits erwähnte Scoring von Konsumentenkrediten, das ? mit einem GPLM-Ansatz durchgeführt hat. Die Daten stellen einen Teil einer umfangreichen Stichprobe dar, die von ? detailliert beschrieben wird. Wir benutzen insgesamt $n = 564$ Datensätze, bei denen es in 24.3 % der Fälle Probleme bei der Rückzahlung des Kredits gegeben hat ($Y_j = 1$). Von den fünf betrachteten Einflussgrößen sind zwei dichotom; sie geben an, ob der Kunde arbeitslos ist oder nicht (X_{j1}) und ob er frühere Kredite problemlos zurückgezahlt hat oder nicht (X_{j2}). Die übrigen drei Einflussgrößen sind quantitativ: die Dauer des Kredits (X_{j3} mit Werten zwischen 4 und 72 Monaten), die Höhe des Darlehens (zwischen 338 und 15653 DM) und das Alter des Kunden (zwischen 19 und 75 Jahren). Die beiden letzten Variablen logarithmieren wir und transformieren sie anschließend linear so, dass sie Werte im Intervall $[0, 1]$ annehmen. Die Datenpunkte sind - wie in der Abbildung deutlich wird - so vergleichsweise homogen über einen Teil der Ebene verstreut, was das lokale

Glätten erleichtert. Die so transformierten Größen nennen wir Z_{j1} und Z_{j2} . Wir passen ein GPLM

$$P(Y_j = 1 | X_{j1} = x_1, X_{j2} = x_2, X_{j3} = x_3, Z_{j1} = z_1, Z_{j2} = z_2) = \psi\left(\sum_{k=1}^3 \beta_k x_k + m(z_1, z_2)\right)$$

an die Daten an und erhalten die Schätzwerte (in Klammern jeweils Schätzer für deren Standardabweichung)

$$\beta_1 = 0.965 (0.249), \beta_2 = 0.746 (0.237), \beta_3 = -0.0498 (0.0115).$$

 [SFMgplm.xpl](#)

Die Kreditausfallwahrscheinlichkeit steigt demnach, wenn der Kunde arbeitslos ist bzw. wenn es früher schon Probleme bei der Rückzahlung von Krediten gegeben hat. Sie sinkt mit der Dauer des Kredits. Die Abhängigkeit von transformiertem Kreditvolumen und Lebensalter schätzen wir nichtparametrisch. Aus der Abbildung ist ersichtlich, dass die geschätzte Funktion $\hat{m}(z_1, z_2)$ deutlich nichtlinear ist mit einem Maximum bei mittleren Werten von Kreditvolumen und Lebensalter. Das Sinken der Ausfallwahrscheinlichkeit bei hohem Kreditvolumen lässt sich eventuell so erklären, dass in der Stichprobe nur Kredite enthalten sind, die auch tatsächlich vergeben worden sind, und die Sachbearbeiter wesentlich zurückhaltender bei der Vergabe hoher Kredite waren, wenn ihnen die Kunden unzuverlässig erschienen sind. Dieser Effekt, dass durch das in der Vergangenheit, wenn auch nicht mit systematischen, modellbasierten Verfahren betriebene, Kreditscreening extreme Risiken bereits im Vorfeld ausgeschlossen worden sind und so in den Daten gar nicht mehr auftauchen, tritt beim Kreditscoring regelmäßig auf und muss bei Interpretation und Einsatz der Modelle berücksichtigt werden.

18.3. Kreditscoring mit neuronalen Netzen

Wie beim nichtparametrischen Anpassen von Finanzzeitreihenmodellen an Daten stellen auch beim Schätzen von Ausfallwahrscheinlichkeiten neuronale Netze eine Alternative zu lokalen Glättern wie der LP-Methode dar. Schon die logistische Regressionsfunktion $\pi(x) = \psi(\beta_0 + \beta^T x)$ ist nichts anderes als die durch ein neuronales Netz mit nur einem Neuron in einer verborgenen Schicht definierte Funktion, wenn die logistische Funktion ψ als Übertragungsfunktion gewählt wird. Durch Kombination mehrerer Neuronen in einer oder mehreren verborgenen Schichten lassen sich wie in der nichtparametrischen Regressionsanalyse Ausfallwahrscheinlichkeiten flexibel schätzen. Um Schätzwerte zwischen 0 und 1 zu erhalten, ist es nur nötig, die zum Beispiel durch (16.1) gegebene Funktion $\nu_H(x; \delta)$ durch eine Funktion G auf das Intervall $[0,1]$ abzubilden. Beschränken wir uns auf eine verborgene Schicht mit H Neuronen und wählen wir $G = \psi$, so hat die durch das neuronale Netz gegebene Ausfallwahrscheinlichkeit die Form

$$\pi_H(x; \vartheta) = \psi\left(v_0 + \sum_{h=1}^H v_h \psi\left(w_{oh} + \sum_{j=1}^d w_{jh} x_j\right)\right),$$

wobei ϑ wieder für den aus den v_h und w_{ih} , $0 \leq i \leq d$, $0 \leq h \leq H$, gebildeten Parametervektor steht. Zum Schätzen der Netzgewichte aus den Daten benutzen wir allerdings keinen Kleinste-Quadrate-Ansatz, der für Regressionsmodelle mit normalverteilten Residuen sinnvoll ist, sondern wir maximieren die Log-Likelihood-Funktion

$$\log L(\vartheta) = \sum_{j=1}^n [Y_j \log \pi_H(X_j; \vartheta) + (1 - Y_j) \log \{1 - \pi_H(X_j; \vartheta)\}]$$

entsprechend der Vorgehensweise bei der logistischen Regression. Durch Einsetzen des so gewonnenen Schätzers $\hat{\vartheta}_n$ erhalten wir als Schätzer für die Ausfallwahrscheinlichkeit

$$\hat{\pi}(x) = \pi_H(x; \hat{\vartheta}_n).$$

Um ein besonders einfaches Modell mit wenigen Parametern zu erhalten, modifiziert ? bei der Insolvenzprognose für kleine und mittlere Unternehmen den Ansatz geringfügig und geht von einer Ausfallwahrscheinlichkeit der Form

$$\pi_H^l(x; \vartheta) = \psi(\beta^T x + v_0 + \sum_{h=1}^H v_h \psi(w_{oh} + \sum_{i=1}^d w_{ih} x_i))$$

aus, was offensichtlich Ähnlichkeiten mit den verallgemeinerten partiellen linearen Modellen aufweist, außer dass hier ein Teil oder alle Einflussgrößen, d.h. die Koordinaten von x , sowohl im linearen wie im nichtparametrischen Teil auftreten können. Der lineare Term $\beta^T x$ lässt sich auch als Beitrag eines zusätzlichen Neurons interpretieren, dessen Übertragungsfunktion allerdings nicht die logistische Funktion $\psi(t)$, sondern die Identität $f(t) \equiv t$ ist. Das unter Einsatz von Modellselektionsverfahren resultierende Netz zum Schätzen der Insolvenzwahrscheinlichkeit ist überraschend einfach. Es enthält neben dem linearen Term nur ein einziges Neuron ($H = 1$), von den 6 Eingangsgrößen tragen nur 4 zum linearen Teil bei (Unternehmensalter, Umsatzentwicklung, Indikator für beschränkte Haftung, Dummy-Variable für verarbeitendes Gewerbe), d.h. die andere beiden Koeffizienten β_i sind 0, und nur 3 (Dummy-Variablen für verarbeitendes Gewerbe und für Handel, Indikatorvariable für Bildungsabschluss des Unternehmers) zum sigmoiden Teil, d.h. die entsprechenden anderen Gewichte w_{i1} sind 0. Bereits mit diesem einfachen Modell erzielt Anders in einem Validierungsdatensatz, der nicht zum Schätzen der Parameter benutzt worden ist, Trefferquoten von 83.3 % bei den insolventen und von 63.3 % bei den solventen Unternehmen.

Literaturverzeichnis

- Ango Nze, P. (1992). Critères d'ergodicité de quelques modèles à représentation markovienne, *Technical Report 315, ser. 1*, C.R. Acad. Sci. Paris.
- Artzner, P., Dealban, F., Eber, F.-J. and Heath, D. (1997). Thinking coherently, *risk magazine*.
- Black, F. (1976). Studies in stock price volatility changes, *Proceedings of the 1976 Meeting of the Business and Economic Statistics Section*, American Statistical Association, pp. 177–181.
- Black, F. and Scholes, M. (1973). The pricing of options and corporate liabilities, *Journal of Political Economy* **81**: 637–654.
- Bollerslev, T. P. (1986). Generalized autoregressive conditional heteroscedasticity, *Journal of Econometrics* **31**: 307–327.
- Bollerslev, T. P. and Wooldridge, J. M. (1992). Quasi maximum likelihood estimation of dynamic models with time-varying covariances, *Econometric Reviews* **11**: 143–172.
- Bossaerts, P. and Hillion, P. (1993). Test of a general equilibrium stock option pricing model, *Mathematical Finance* **3**: 311–347.
- Box, G. E. P. and Jenkins, G. M. (1976). *Time Series Analysis: Forecasting and Control*, Holden-Day, San Francisco.
- Brockwell, P. and Davis, R. (1991). *Time Series: Theory and Methods*, Springer Verlag.
- Campbell, J., Lo, A. and MacKinlay, A. (1997). *The Econometrics of Financial Markets*, Princeton University Press, Princeton, New Jersey.
- Carroll, R. J., Härdle, W. and Mammen, E. (1999). Estimation in an additive model when the components are linked parametrically, *discussion paper*, Sonderforschungsbereich, Quantification and Simulation of Economic Processes.
- Chan, K. and Tong, H. (1985). On the use of deterministic lyapunov functions for the ergodicity of stochastic difference equations, *Advanced Applied Probability* **17**: 666–678.
- Chen, R. and Tsay, R. S. (1993a). Functional-coefficient autoregressive models, *Journal of the American Statistical Association* **88**: 298–308.

- Chen, R. and Tsay, R. S. (1993b). Nonlinear additive ARX models, *Journal of the American Statistical Association* **88**: 955–967.
- Cleveland, W. S. (1979). Robust locally weighted regression and smoothing scatterplots, *Journal of the American Statistical Association* **74**: 829–836.
- Collomb, G. (1984). Propriétés de convergence presque complète du prédicteur à noyau, *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete* **66**: 441–460.
- Copeland, T. and Weston, J. (1992). *Financial Theory and Corporate Policy*, Addison-Wesley.
- Cox, J. C., Ingersoll, Jr., J. E. and Ross, S. A. (1985). A theory of the term structure of interest rates, *Econometrica* **53**: 385–407.
- Cox, J. C. and Ross, S. A. (1976). The valuation of options for alternative stochastic processes, *Journal of Financial Economics* **3**: 145–166.
- Davydov, Y. (1973). It mixing conditions for markov chains, *Theory of Probability and its Applications* **78**: 312–328.
- Dewynne, J., Howison, S. and Wilmott, P. (1993). *Mathematical Models and Computation*, Oxford University Press.
- Dickey, D. and Fuller, W. (1979). Distribution of the estimators for autoregressive time series with a unit root, *Journal of the American Statistical Association* **74**: 427–431.
- Diebolt, J. and Guégan, D. (1990). Probabilistic properties of the general nonlinear autoregressive process of order one, *Technical Report 128*, Université de Paris VI, Paris.
- Ding, Z., Granger, C. W. J. and Engle, R. F. (1993). A long memory property of stock market returns and a new model, *Journal of Empirical Finance* **1**: 83–106.
- Doukhan, P. and Ghindès, M. (1980). Estimation dans le processus $x_{n+1} = f(x_n) + \varepsilon_{n+1}$, *C.R. Acad. Sci. Paris, Sér. A* **297**: 61–64.
- Doukhan, P. and Ghindès, M. (1981). Processus autorégressifs non-linéaires, *C.R. Acad. Sci. Paris, Sér. A* **290**: 921–923.
- Duan, J.-C. (1995). The GARCH option pricing model, *Mathematical Finance* **5**: 13–32.
- Duan, J.-C. (1997). Augmented GARCH(p,q) process and its diffusion limit, *Journal of Econometrics* **79**: 97–127.
- Embrechts, P., Klüppelberg, C. and Mikosch, T. (1997). *Modelling Extremal Events*, Springer Verlag, Berlin.
- Engle, R. F. (1982). Autoregressive conditional heteroscedasticity with estimates of the variance of U.K. inflation, *Econometrica* **50**: 987–1008.

- Engle, R. F. and Gonzalez-Rivera, G. (1991). Semiparametric ARCH models, *Journal of Business and Economic Statistics* **9**: 345–360.
- Engle, R. F., Lilien, D. M. and Robins, R. P. (1987). Estimating time varying risk premia in the term structure: The ARCH-M model, *Econometrica* **55**: 391–407.
- Engle, R. F. and Ng, V. K. (1993). Measuring and testing the impact of news on volatility, *Journal of Finance* **48**: 1749–1778.
- Fama, E. F. (1965). The behavior of stock market prices, *Journal of Business* **38**: 34–105.
- Fan, J. and Gijbels, I. (1996). *Local polynomial modeling and its application – Theory and methodologies*, Chapman and Hall.
- Fengler, M., Härdle, W. and Villa, C. (2001). Common principal components analysis and the dynamics of implied volatilities, Discussion Paper, Sonderforschungsbereich 373, Humboldt University Berlin.
- Föllmer, H. and Schweizer, M. (1991). Hedging of contingent claims under incomplete information, *Applied Stochastic Analysis*, Gordon and Breach, London, pp. 389–414.
- Föllmer, H. and Sondermann, D. (1991). Hedging of non-redundant contingent claims, *Contributions to Mathematical Economics*, Amsterdam, North Holland, pp. 205–223.
- Forsythe, R., Palfrey, T. and Plott, C. (1982). Asset valuation in an experimental market, *Econometrica* **50**: 537–567.
- Friedmann, R. (1992). Zur güte von einheitswurzeltests, *Technical report*, unveröffentlichtes Manuskript.
- Glosten, L., Jagannathan, R. and Runkle, D. (1993). Relationship between the expected value and the volatility of the nominal excess return on stocks, *Journal of Finance* **48**: 1779–1801.
- Gouriéroux, C. (1997). *ARCH Models and Financial Applications*, Springer Verlag.
- Gouriéroux, C. and Monfort, A. (1992). Qualitative threshold ARCH models, *Journal of Econometrics* **52**: 159–199.
- Gouriéroux, C. and Monfort, A. (1996). *Time Series Models*, Economica.
- Gouriéroux, C., Monfort, A. and Trognon, A. (1984). Pseudo maximum likelihood methods: Theory, *Econometrica* **52**: 681–700.
- Gregory, A. (1989). A nonparametric test for autoregressive conditional heteroscedasticity: A markov chain approach, *Journal of Business and Economic Statistics* **7**: 107–115.
- Hafner, C. (1997). Estimating high frequency foreign exchange rate volatility with nonparametric ARCH models, *Journal of Statistical Planning and Inference* **68**: 247–269.

- Hafner, C. and Herwartz, H. (1998). Time-varying market price of risk in capm, approaches, empirical evidence and implications, *Finance* (19): 93–112.
- Hafner, C. and Herwartz, H. (1999). Testing for linear autoregressive dynamics under heteroskedasticity, *Technical report*, SFB 373, Humboldt–Universität zu Berlin.
- Härdle, W. (1990). *Applied Nonparametric Regression*, Cambridge University Press, Cambridge.
- Härdle, W. and Hafner, C. (2000). Discrete time option pricing with flexible volatility estimation, *Finance and Stochastics* 4: 189–207.
- Härdle, W., Lütkepohl, H. and Chen, R. (1997). A review of nonparametric time series analysis, *International Statistical Review* 65: 49–72.
- Härdle, W., Müller, M., Sperlich, S. and Werwatz, A. (2001). *Non- and Semiparametric Modelling*, MD*Tech.
- Härdle, W. and Simar, L. (2001). *Applied Multivariate Statistical Analysis*, Springer Verlag, Heidelberg.
- Härdle, W., Spokoiny, V. and Teyssiere, G. (2000). Adaptive estimation for a time inhomogeneous stochastic-volatility model, Discussion Paper, Sonderforschungsbereich 373, Humboldt University Berlin.
- Härdle, W. and Tsybakov, A. (1997). Local polynomial estimation of the volatility function, *Journal of Econometrics* 81: 223–242.
- Härdle, W., Tsybakov, A. and Yang, L. (1996). Nonparametric vector autoregression, *Journal of Statistical Planning and Inference* 68: 221–245.
- Harrison, M. and Kreps, D. (1979). Martingales and arbitrage in multiperiod securities markets, *Journal of Economic Theory* 20: 381–408.
- Harrison, M. and Pliska, S. (1981). Martingales and stochastic integrals in the theory of continuous trading, *Stochastic Processes Applications* 11: 215–260.
- Hassler, U. (1994). Einheitswurzeltests – ein überblick, *Allgemeines Statistisches Archiv* 78: 207–228.
- Hastie, T. J. and Tibshirani, R. J. (1990). *Generalized Additive Models*, Vol. 43 of *Monographs on Statistics and Applied Probability*, Chapman and Hall, London.
- Hastings, j. (1955). *Approximations for Digital Computers*, Princeton University Press, Princeton.
- He, C. and Teräsvirta, T. (1998). Properties of moments of a family of GARCH processes, *Journal of Econometrics* . forthcoming.
- Hill, B. (1975). A simple general approach to inference about the tail of a distribution, *Annals of Statistics* 3: 1163–1174.

- Hull, J. C. (2000). *Options, Futures and other Derivatives*, Prentice Hall.
- Hull, J. and White, A. (1987). The pricing of options on assets with stochastic volatilities, *Journal of Finance* **42**: 281–300.
- Ingersoll, Jr., J. E. (1987). *Theory of Financial Decision Making*, Rowman & Littlefield.
- Jarchow, H.-J. and Rühmann, P. (1994). *Monetäre Außenwirtschaft I. Monetäre Außenwirtschaftstheorie*, Vandenhoeck & Ruprecht, Göttingen.
- Jarrow (1992). *Finance Theory*.
- Jaschke, S. and Küchler, U. (1999). Coherent risk measures, Discussion paper, No. 64, SFB 373, Humboldt-Universität zu Berlin.
- Katkovnik, V. (1979). Linear and nonlinear methods for nonparametric regression analysis (in russian), *Avtomatika i Telemekhanika* pp. 35–46.
- Katkovnik, V. (1985). Nonparametric identification and data smoothing, Nauka.
- Kwiatkowski, D., Phillips, P., Schmidt, P. and Shin, Y. (1992). Testing the null hypothesis of stationarity against the alternative of a unit root, *Journal of Econometrics* **54**: 159–178.
- Lee, S. and Hansen, B. (1994). Asymptotic theory for the GARCH(1,1) quasi-maximum likelihood estimator, *Econometric Theory* **10**: 29–52.
- Lintner, J. (1965). Security prices, risk and maximal gains from diversification, *Journal of Finance* **20**: 587–615.
- Liptser, R. and Shirjaev, A. (1980). A functional central limit theorem for martingales, *Theory of Probability and its Applications* **25**: 667–688.
- Lubrano, M. (1998). Smooth transition GARCH models: A Bayesian perspective, *Technical report*, CORE, Louvain-la-Neuve.
- Lumsdaine, R. (1996). Consistency and asymptotic normality of the quasi maximum likelihood estimator in IGARCH(1,1) and covariance stationary GARCH(1,1) models, *Econometrica* **64**: 575–596.
- Maercker, G. (1997). *Statistical Inference in Conditional Heteroskedastic Autoregressive Models*, Shaker Verlag, Aachen.
- Mandelbrot, B. (1963). The variation of certain speculative prices, *Journal of Business* **36**: 394–419.
- McAllister, P. H. and Mingo, J. J. (1996). Bank capital requirements for securitized loan portfolios, *Journal of Banking and Finance* **20**.
- McKeague, I. and Zhang, M. (1994). Identification of nonlinear time series from first order cumulative characteristics, *Annals of Statistics* **22**: 495–514.

- Melino, A. and Turnbull, S. M. (1990). Pricing foreign currency options with stochastic volatility, *Journal of Econometrics* **45**: 239–265.
- Mills, T. (1993). *The econometric modelling of financial time series*, Cambridge University Press, Cambridge.
- Mokkadem, A. (1987). Sur un modèle autorégressif nonlinéaire. ergodicité et ergodicité géométrique, *Journal of Time Series Analysis* **8**: 195–204.
- Mossin, J. (1966). Equilibrium in a capital asset market, *Econometrica* **34**: 768–783.
- Nelson, D. B. (1990). Stationarity and persistence in the GARCH(1,1) model, *Econometric Theory* **6**: 318–334.
- Nelson, D. B. (1991). Conditional heteroskedasticity in asset returns: A new approach, *Econometrica* **59**: 347–370.
- Nummelin, E. and Tuominen, P. (1982). Geometric ergodicity of harris–recurrent markov chains with application to renewal theory, *Stoch. Proc. Appl.* **12**: 187–202.
- Rabemananjara, R. and Zakoian, J. M. (1993). Threshold ARCH models and asymmetries in volatility, *Journal of Applied Econometrics* **8**: 31–49.
- Redelberger, T. (1994). Grundlagen und konstruktion des vdax-volatilitätsindex der deutsche börse ag.
- Renault, E. and Touzi, N. (1996). Option hedging and implied volatilities in a stochastic volatility model, *Mathematical Finance* **6**: 277–302.
- Robinson, P. (1983). Non-parametric estimation for time series models, *Journal of Time Series Analysis* **4**: 185–208.
- Robinson, P. (1984). Robust nonparametric autoregression, in J. Franke, W. Härdle and Martin (eds), *Robust and nonlinear time series analysis*, Springer-Verlag, Heidelberg.
- Samaskij, A. (1984). *Theorie der Differenzenverfahren*, Akademische Verlagsgesellschaft Geest und Portig K.-G., Leipzig.
- Schlittgen, R. and Streitberg, B. (1995). *Zeitreihenanalyse*, Oldenbourg.
- Schönbucher, P. (1998). A market model for stochastic implied volatility.
- Sharpe, W. F. (1964). Capital asset prices: A theory of market equilibrium under conditions of risk, *Journal of Finance* **19**: 425–442.
- Shephard, N. (1996). Statistical aspects of ARCH and stochastic volatility models, in O.-N. D.R.Cox, D.V.Hinkley (ed.), *Time Series Models in econometrics, finance and other fields*, Chapman & Hall, London.

- Skiadopoulos, G., Hodges, S. and Clewlow, L. (1998). The dynamics of implied volatility surfaces, Working Paper, Financial Options Research Centre, Warwick Business School University of Warwick, Coventry.
- Stone, C. (1977). Consistent nonparametric regression, *Annals of Statistics* **5**: 595–645.
- Studer, G. (1995). Value at risk and maximum loss optimization, Discussion Paper, ETHZ, RiskLab: Technical Report.
- Taleb, N. (1997). *Dynamic Hedging: Managing Vanilla and Exotic Options*, John Wiley & Sons, New York.
- Teräsvirta, T. (1994). Specification, estimation, and evaluation of smooth transition autoregressive models, *Journal of the American Statistical Association* **89**: 208–218.
- Tong, H. (1983). *Threshold Models in Nonlinear Time Series Analysis*, Vol. 21 of *Lecture Notes in Statistics*, Springer Verlag, Heidelberg.
- Tsybakov, A. (1986). Robust reconstruction of functions by the local–approximation method, *Problems of Information Transmission* **22**: 133–146.
- Vieu, P. (1995). Order choice in nonlinear autoregressive models, Discussion Paper.
- Weiss, A. A. (1986). Asymptotic theory for ARCH models: Estimation and testing, *Econometric Theory* **2**: 107–131.
- White, H. (1982). Maximum likelihood estimation of misspecified models, *Econometrica* **50**: 1–25.
- Wiggins, J. (1987). Option values under stochastic volatility: Theory and empirical estimates, *Journal of Financial Economics* **19**: Journal of Financial Economics.
- Yang, L., Härdle, W. and Nielsen, J. (1999). Nonparametric autoregression with multiplicative volatility and additive mean, *Journal of Time Series Analysis* . in press.
- Zakoian, J. (1994). Threshold heteroskedastic functions, *Journal of Economic Dynamics and Control* **18**: 931–955.
- Zakoian, J. M. (1991). Threshold heteroskedastic models, *Technical report*, INSEE.

Index

- ”Vega”-Position, 279
- Abhängigkeit, 33
- Aktie, 13, 22
- Aktienkurs, 55
- Anleihen, 8
- äquivalentes Martingalmas, 222
- Arbitrage, 11
- Arbitragebeziehung, 11
- ARCH, 126, 171, 221
 - schwaches, 173
 - semi-starkes, 173
 - starkes, 173
- ARCH model
 - ARCH-M, 192
 - QTARCH, 191
 - TARCH, 191
 - TGARCH, 191
- ARCH Modell
 - EGARCH, 190
- ARCH-M model, 192
- ARIMA, 129, 151
- ARMA, 156, 185
- aus dem Geld, 16
- Autokorrelationsfunktion (ACF), 121
- Autokovarianzfunktion, 120
- autoregressiver Prozess, 153

- bedingte Verteilungsfunktion, 120
- bedingte Wahrscheinlichkeit, 35
- bedingter Erwartungswert, 35
- Bestandshaltete Kosten, 13
- Binomialmodell, 83
- Binomialprozess, 37
 - geometrischer, 43
- Black-Scholes, 59
 - Formel, 65
 - Modell, 59
- Black-Scholes-Formel, 67

- Black-Scholes-Gleichung, 61
- BLACK/SCHOLES
 - Modell, 137
- Bond, 8, 112
 - Zerobond, 8, 13
- Bondoption, 65, 112
- Bretton Woods, 171
- Brownsche Bewegung, 48
 - geometrische, 55, 59

- Capital Asset Pricing Model, 192
- CAPM, 133, 141, 192
- Cashflow, 61
- Cox, Ross, Rubinstein, 84
- Cox-Ingersoll-Ross Modell, 136

- DAX-Calls, Bewertung, 232
- DAX-Optionen, 233, 272
- Delta, 73
 - Delta-Hedge-Prozess, 27
 - deltaneutrale Position, 73
- Delta-Normal Modell, 237
- Devisenoption, 65
- Dickey-Fuller Test, 144
- Drift, 38, 53, 59
- Duplikationsportfolio, 12, 60
- dynamische Strategie, 23

- EGARCH Modell, 190
- einfache Rendite, 125
- Einheitswurzel, 144
- Einheitswurzeltest, 144
- EMA, 238
- erwartete Aktienrendite, 55
- erwarteter Shortfall, 239
- Erwartungswert, 31
- Exposure, 237

- fares Spiel, 124

Faktorladungen, 277
 Finanzderivate, 3
 flexible Volatilitätsschatzer, 221
 Floor, 22
 Forward-Preis, 4, 12
 Future, 15
 Future-Preis, 4

 Gamma, 75
 GARCH, 171, 184
 gedeckten Position, 71
 Generalisierte Fehlerverteilung, 191
 Geometrische Brownsche Bewegung, 140
 GLM, 284
 GPLM, 286

 Hauptkomponentenanalyse, 271
 Hedge, 22, 69
 Delta, 72
 dynamisches, 73
 Gamma, 76
 Hedge-Rate, 73

 im Geld, 16
 implizierte Volatilität, 271
 Innerer Wert, 16
 integrierter Prozess, 151
 Irrfahrt, 37, 122
 einfache, 37
 geometrische, 43, 84
 Hypothesen, 142
 symmetrische einfache, 38
 Itô-Integral, 50
 Itô-Prozess, 53
 Itô's Lemma, 57

 Korrelation, 33
 KPSS Test, 146
 Kreditscoring, 285
 Kursbaum, 38, 83
 Kurtosis, 33, 185

 Lag-Operator, 124
 Leverage Effekt, 226
 Likelihoodfunktion, 166, 179
 linear separierbar, 252
 linearer Prozess, 151

 log Rendite, 125
 logistische Regression, 284
 Lognormalverteilung, 30, 32
 long position, 9
 Long Straddle, 279
 LP-Methode, 200

 Markoff Kette, 192
 Markoff Prozess, 49, 124
 Markoff-Eigenschaft, 44
 Markteffizienz, 127, 132
 Marktpreis des Risikos, 115, 139, 141
 Marktrisiko, 22
 Martingal, 124
 Maximum-Loss, 280
 mean reversion, 137
 Mittelwertfunktion, 120
 Momentfunktion, 160
 Moneyness, 228, 271
 Monte-Carlo-Studie, 228
 Moving Average Prozess, 152

 News impact Funktion, 223
 News impact Kurve, 221, 222
 Nomineller Wert, 8
 Normalverteilung, 30

 Option, 4, 11, 16, 27
 amerikanische, 95, 102
 amerikanischer Call, 5
 amerikanischer Put, 5
 Average Strike, 109
 Barrier, 108
 Chooser, 108
 europäische, 65
 europäische Kaufoption, 5
 europäische Verkaufsoption, 5
 europäischer Call, 5
 europäischer Put, 5
 exotische, 5, 107
 Knock-out, 108
 Lookback, 111
 Standardoption, 5
 zusammengesetzte, 107
 Optionen
 asiatische, 109

pfadabhängig, 109
 Optionsbewertung, 221
 Optionsmanagement, 11
 Optionspreis, 66, 102
 OTC-Derivate, 5

 partielle Autokorrelation, 157
 Pay-off, 5
 perfekter Finanzmarkt, 11
 Persistenz, 228, 232
 Portefeuille, 9
 Portfolio, 9
 Duplikationsportfolio, 64
 Hedge-Portefeuille, 60
 Portefeuille-Versicherung, 22
 Portmanteau Statistik, 163
 Put-Call-Parität, 18, 24

 QTARCH model, 191

 RAROC, 246
 Rating, 283
 rationale Erwartungen
 RE-Gleichgewicht, 131
 Rebalanzieren, 73
 Rentenoption, 112
 Restlaufzeit, 4
 Rho, 77
 risiko-neutrale Bewertung, 228
 Risikofaktor, 280
 RMA, 238

 Schiefe, 33
 short position, 9
 Short rate, 136
 short rate, 9
 Short selling, 9
 Short Straddle, 279
 Shortfall
 erwarteter, 239
 Smile, 203, 271
 spot rate, 113
 Spot-Preis, 4
 Standardabweichung, 32
 Stationarität, 121
 strenge, 181

 statische Strategie, 23
 stetige Verzinsung, 8
 stochastische Differentialgleichung, 47, 53
 stochastische Integrale, 47
 stochastischer Prozess, 47, 120
 stetige Zeit, 47
 Verteilungsfunktion, 120
 stochastischer Prozeß, 37
 diskrete Zeit, 37
 Stop-Loss-Strategie, 71
 Straddle, 8

 TARCH model, 191
 Terminkontrakt, 3, 11
 Forward, 4
 Future, 4
 TGARCH model, 191
 Theta, 75
 threshold ARCH model, 191
 generalized, 191
 qualitative, 191
 Trend, 38
 Trinomialmodell, 102
 Trinomialprozess, 40

 unabhängige Zuwächse, 42
 ungedeckte Position, 71

 Value at Risk, 239
 Value-at Risk, 280
 VaR, 239
 Varianz, 31
 VDAX, 273
 VDAX-Dynamik, 275
 VDAX-Subindex, 273
 Vega, 77
 Versicherungsprämie, 22
 Verteilung, 29
 Verteilungsfunktion, 29
 Volatilität, 55, 59, 171
 historische, 78
 historische , 78
 implizierte, 78, 80
 implizite, 80
 Volatilitäts-Smile, 271
 Volatilitätsdynamik, 271

Volatilitätsfläche, 271
Volatilitätsindex, 273
Volatilität, 221

Wahrscheinlichkeit, 29
Wahrscheinlichkeitsdichte, 29
weißes Rauschen, 121
Wiener-Prozess, 47, 48
 allgemeinen, 53

Yield to Maturity, 136
Yule-Walker, 154

zentrale Grenzwertsatz, 31
Zerobond, 112
Zerobond-Bewertungsgleichung, 114
Zinsparitätentheorie, 134
Zufallsgröße, 29
 Bernoulli-verteilte Zufallsgröße, 31
 binomialverteilte Zufallsgröße, 31
Zufallsvektor, 33