

Skalierungsmodelle für das Risiko am Aktienmarkt – eine ökonomische Analyse

Zur Erlangung des akademischen Grades eines
Doktors der Wirtschaftswissenschaften
(Dr. rer. pol.)

von der Fakultät für
Wirtschaftswissenschaften
der Universität Fridericiana zu Karlsruhe

genehmigte
DISSERTATION

von Dipl. Wi.-Ing. Serguei Nemtsev

Tag der mündlichen Prüfung: 12. Januar 2006
Referent: Prof. Dr. Hermann Göppl
Korreferent: Prof. Dr. Kuno Egle

Karlsruhe, 2006



Danksagung

Die vorliegende Arbeit entstand während meiner Tätigkeit als Mitarbeiter am Institut für Finanzwirtschaft, Banken und Versicherungen der Universität Karlsruhe (TH). Sie wurde 2005 von der Fakultät für Wirtschaftswissenschaften als Dissertation angenommen.

Herrn Prof. Dr. Hermann Göppl, meinem Doktorvater, gilt mein herzlicher Dank für sein Vertrauen, Unterstützung und Förderung, die mir zugeteilt wurde und zu meiner Inspiration entscheidend beigetragen hat. Herrn Prof. Dr. Kuno Egle danke ich für die Übernahme des Korreferats und für viel Zeit, die er sich genommen hat, um manche wichtigen Klarstellungen zu machen. Die verbleibenden Fehler gehen natürlich einzig und alleine auf mein Konto.

Bei meinen Kollegen und Freunden bedanke ich mich für eine gute Zusammenarbeit. Dies gilt vor allem für Liang Ding, Andreea Hannich, Matthias Häußler, Dr. Torsten Lüdecke, Monika Bachofner, Dr. Ralf Herrmann und Martin Wilhelm. Einen besonderen Dank gilt hierbei Dr. Thomas Burdelski und Joachim Moser. Unserer Sekretärin Frau Fürniss danke ich für ihre jederzeitige Hilfsbereitschaft und eine tatkräftige Unterstützung.

Last but not least, danke ich meiner Frau und meinen Kindern für die gemeinsame Zeit, auf die sie verzichten mussten, während diese Arbeit entstand. Ich hoffe, sie ist es Wert.

Karlsruhe, im März 2006

Serguei Nemtsev



Inhaltsverzeichnis

1. Einleitung und Motivation	1
1.1 Skalierung – was steckt dahinter?	1
2. Zur Markteffizienz	15
2.1 Die Paradigmen	15
3. Erweiterungen der Markteffizienzhypothese	35
3.1 Theorie vs. Empirie	35
3.2 Das Skalierungskonzept	41
3.3 Die Fraktale Market Hypothese	48
4. Regeln und Ausnahmen I: typische Ereignisse	53
4.1 Entropie als inverses Informationsmaß	53
4.2 Die „Nicht-Shannon“-Entropien	76
5. Regeln und Ausnahmen II: große Abweichungen	87
5.1 Large Deviations	87
6. Modellierung der Langzeitabhängigkeiten	101
6.1 Das Langzeitgedächtnis	101
6.2 Selbstähnliche Zufallsprozesse	112
6.3 Gebrochene Brownsche Bewegung	115
6.4 Problematik stochastischer Integration	118
6.5 Überlegungen zur Arbitragefreiheit des Marktes	126

II	Inhaltsverzeichnis	
	6.6 Vorhersage der gebrochenen Brownschen Bewegung ...	129
	6.7 Überlegungen zur Arbitragefreiheit des Marktes. P.S.	131
7.	Skalierung anschaulich: Fraktale und Multifraktale	137
	7.1 Fraktale	137
	7.2 Multifraktale	154
	7.3 Das Multifraktale Modell of Asset Returns	174
8.	Analyse der Ergebnisse	179
	8.1 Skalierung: Modell oder Hypothese?	179
	8.2 Die Skalenrisiken	183
	Anhang	191
A.	Hilfsmittel der Stochastik	193
	A.1 Das (Wahrscheinlichkeits-) Maß	195
	A.2 Charakterisierung der Wahrscheinlichkeitsverteilungen .	205
	A.3 Grenzwertsätze	214
	A.4 Stochastische Prozesse	217
	A.5 Stabile Verteilungen	231
	A.6 Lévy Prozesse	237
	A.7 Spezialfunktionen	241
	A.8 Konvergenzbegriffe in der Stochastik	247
	Literaturverzeichnis	249

1. Einleitung und Motivation

“The most beautiful thing we can experience is the mysterious. It is the source of all true art and all science. He to whom this emotion is a stranger, who can no longer pause to wonder and stand rapt in awe, is as good as dead: his eyes are closed.”

– Albert Einstein

1.1 Skalierung – was steckt dahinter?

Auf der Suche nach kausalen Zusammenhängen in der Natur und der Gesellschaft kommt kaum ein Forscher an *Skalierungseigenschaften* vorbei. In dieser Arbeit interessieren wir uns hauptsächlich (aber nicht ausschließlich) für die finanzwirtschaftlichen Auswirkungen der Skalierung auf die Beschreibung der Dynamik der Wertpapierpreise/Renditen; vor allem aber auf die Möglichkeiten und Notwendigkeiten einer erweiterten Interpretation eines Begriffes, der im Finance mindestens so wichtig ist wie das Wort Finance selbst: das *Risiko*.

Ein triviales Beispiel der Skalierung sind *Economics of Scale*. Allein die Veränderung der Kostenstruktur eines großen Projektes oder gar eines Großunternehmens im Vergleich zu einem kleinen Projekt kann dabei Grundlage für zusätzliche ökonomische Vorteile bilden. Aus der BWL lernen wir, dass die Fixkosten bei der Berechnung des Grenzgewinnes keine Rolle spielen, da sie per Definition von der produzierten Menge unabhängig sind. Es liegt auf der Hand, dass ein *homoeconomicus* solange immer mehr Grenzgewinn erwirtschaften wird, wie sich die Grenzkosten durch die Verlagerung des variablen Anteils der Kosten auf den Fixkostenbereich senken lassen. Wir hätten wahrscheinlich

nur noch riesige Betriebe, gäbe es keine wirtschaftlichen Faktoren in Form einer Art „negativer Rückkopplung“ wie eingeschränkte Gesamtnachfrage oder unzureichende Flexibilität der Großbetriebe bei der Umstellung auf neue Produktion.

Auf ähnliche Weise können wir auf die Losgrößenvorteile bei den Banken eingehen, Portfoliodiversifikation, eine langfristige Finanzanlage im Vergleich zu einer kurzfristigen nach Rendite/Risiko-Kriterium (oder Kriterien, man kann mehrere davon heranziehen)... Übrigens, wie es sich herausstellt, ist das letzte Beispiel (lang/kurzfristige Finanzanlage) gar nicht so trivial. Es ist gerade der Fall, in dem die *Skalierung in der Zeit* sowohl linear als auch nichtlinear wirken kann. Und das hängt davon ab, welche Kennzahl wir für das Risiko verwenden bei einer sehr populären Behauptung „eine langfristige Anlage wäre allein auf Grundlage des längeren Zeithorizontes weniger riskant, als eine kurzfristige“. Wenn das Risiko von der Zeit nichtlinear abhängig ist, wird die Situation nicht einfacher, sondern schwieriger und schon auf keinen Fall eindeutig. Jede langfristige Anlage wird irgendwann zur kurzfristigen, bevor sie ganz aufgelöst wird. Ohne zusätzlichen Freiheitsgrad, den Zeitpunkt der Auflösung flexibel zu wählen, wird womöglich nicht jeder Anleger oder gar institutioneller Investor mit seiner Anlageperformance zufrieden sein.

Merkwürdig: die spezifische Art, Eigenschaften der Zielobjekte aus der Sicht von unterschiedlichen Skalen zu untersuchen, ist alles andere als intuitiv. Die Ursache scheint darin zu liegen, dass ein Mensch nicht gewohnt ist, „multiplikativ zu denken“; man kommt eher mit additiven Merkmalen klar. Skalierung operiert dagegen mit multiplikativen Größen, was entsprechend die Additivität auf logarithmischer Ebene impliziert. Wir bilden mühelos einfache Durchschnitte wie die Gesamtnote in einem Abschlusszeugnis. Das Ergebnis kann zufriedenstellend sein oder auch nicht; es ist aber verständlich und leicht nachvollziehbar. Lädt sich aber ein Erwartungsnutzen eines Repräsentativinvestors mit logarithmischer Nutzenfunktion ebenso leicht errechnen? Dabei geht es lediglich um ein Lehrbuchbeispiel zu einer konstanten relativen Risikoaversion (CRRA): ein solcher Investor ist mit einer konstanten Wachstumsrate seines Vermögens „höchst zufrieden“. Im Schatten bleibt natürlich seine absolute Risikoaversion, die, *multipliziert* mit der Höhe des Vermögens, ebenfalls eine Konstante ergibt. Diese Größe scheint bei logarithmischer Nutzenfunktion weniger geeignet zu sein, um das ökonomische Verhalten des Investors zu charakterisieren.

Schauen wir uns den berühmten Goldenen Schnitt an, wo es darum geht, eine Einheit bestimmter Länge im *Verhältnis*

$$\frac{\text{groß}}{\text{klein}} = \frac{\text{gesamt}}{\text{groß}}$$

aufzuteilen. Nach einer kleinen Umformung ist es leicht zu sehen, dass wir es hier mit dem Mittelwert einer logarithmischen Größe zu tun haben:

$$\ln(\text{groß}) = \frac{1}{2} [\ln(\text{klein}) + \ln(\text{gesamt})] = E \ln(\cdot).$$

Wir hören uns Musik an, die aus Noten besteht; Abstände zwischen den Noten (gemessen an deren Frequenz) sind logarithmisch geordnet. Wir schauen uns die wunderbaren Werke von Leonardo da Vinci an, der neben Schönheit sehr viel Wert auf Harmonie gelegt hat und sich mit solchen Fragestellungen wie Größenverhältnissen des Kopfes oder Studien von Armen und Händen beschäftigte. Nicht mal ein „hindernispassierendes und herabstürzendes Wasser“ ist ihm entkommen. Leonardo da Vinci stellte die Regeln der Harmonie auf. Die meisten davon, wie auch der Goldene Schnitt, sind multiplikativ.

Ein „additives Denken“ ist sehr bequem. Es entspricht am ehesten einer natürlichen menschlichen Denkweise. Die Natur scheint hingegen multiplikativ zu sein. Die Werke des Menschen sind geordnet: quadratisch, rund, glatt. . . Die Werke der Natur sind ebenfalls geordnet, doch sieht diese Ordnung anders aus: sie ist gebrochen, krumm, rau, *selbstähnlich*, logarithmisch. Dabei bleibt der Mensch ein Kind der Natur! Liegt hier ein Widerspruch vor?

Wie steht es um den Finanzmarkt? Er wird doch von den Menschen organisiert. Die Preise werden ebenfalls von den Menschen gemacht, wenn man den Rechnern auf der Börse nichts weiter als die Zusammenführung (*matching*) der eingegebenen Kauf- oder Verkaufsaufträge zutraut. Benoît Mandelbrot ist in seinen Büchern und Artikeln nach 40 Jahren Forschung zum Schluss gekommen, dass sowohl die Natur [Mandelbrot 1982] als auch der Finanzmarkt [Mandelbrot 1997] über multiplikative Eigenschaften verfügen, die am ehesten durch die *Skalierungsgesetze* beschrieben werden können. Dabei nimmt die Skalierung nicht selten ihre einfachste Form ein, die als *Fraktal* bekannt ist. Die charakteristischen Eigenschaften des Untersuchungsgegenstandes sind in diesem Fall invariant zur Änderung der Beobachtungsskala und weder Vergrößerung noch Verkleinerung des „Objektes“ Schneeflocke

lassen uns von der Meinung abbringen, dass es sich immer noch um eine Schneeflocke handelt. Manchmal lassen sich mehrere Skalen mit jeweils ähnlichen Eigenschaften erkennen, die ihrerseits wiederum deterministischer oder stochastischer Natur sind. Dann spricht man über *Multifraktale*.

Fraktale und Multifraktale stellen ein anschauliches Beispiel aus der Natur dar, das uns hilft, die abstrakten Zusammenhänge mit verständlichen Mitteln zu beschreiben. Die Kategorie „Einfachheit“ bekommt in dieser Hinsicht eine besondere Bedeutung. Damit versuchen wir einen Widerspruch zwischen einer „einfachen Welt“ und den „komplizierten Zusammenhängen“, die in dieser Welt ablaufen, aufzulösen. Wir *erfinden* uns eine *künstliche*, möglicherweise, komplizierte Welt so, dass die dahin transformierten Zusammenhänge sich relativ einfach beschreiben lassen. Wir definieren Gesetze, nach denen diese komplizierte Welt funktioniert, lösen dort unsere Aufgabe und transformieren die Lösung zurück. Die Rollen tauschen sich. Gehen wir davon aus, dass die von uns beobachteten charakteristischen Eigenschaften allgemeiner Natur sind und in ihrem Erscheinen als *typisch* bezeichnet werden können, ist es zwecks Modellierung einfacher, sie den Gegenständen oder Vorgängen zu „entnehmen“ und auf die Umgebung zu übertragen. Dadurch werden die Zusammenhänge einfacher und die Umgebung komplizierter. Irgendwann könnte es plötzlich vorkommen, dass dieses *Modell* uns ganz natürlich erscheint. . . Gelegentlich gehen wir noch weiter und schaffen aus einem *positiven* Modell, das beispielsweise beschreibt, wie sich ein rationaler Investor verhalten *sollte*, ein *normatives* Modell. Unter der Annahme, dass wir darüber einig sind, was die *Rationalität* im ökonomischen Sinne bedeutet, versetzt das normative Modell das Verhalten eines rationalen Investors in gewisse engere Rahmen, die nicht mehr verletzt werden dürfen.

Weitere Skalierungsbeispiele aus dem Kapitalmarktbereich liegen auf der Hand. Kaum eine andere Kennzahl kann sich mit der Aktienrendite, definiert als $R_t = (S_t - S_{t-1}) / S_{t-1}$ im diskreten und $R_t = \ln(S_t / S_{t-1})$ im sogenannten kontinuierlichen Fall¹ nach Wichtigkeit und Bedeutung messen lassen. In Abhängigkeit davon, was wir unter der Zeitvariablen t verstehen, lassen sich die Renditen nach täglichen, wöchentlichen, monatlichen usw. unterscheiden. Noch nicht so lange gibt es Renditen, die auf Intraday-Basis berechnet werden, entweder syn-

¹ Mit S_t bezeichnen wir, wie gewöhnlich, einen Aktienkurs zum Zeitpunkt t .

chron in den festen Zeitabständen wie 15 Sek. oder 5 Min., oder sogar Tick-by-tick-Weise: dabei wird jede einzelne Transaktion in die Berechnung miteinbezogen². Eine inzwischen sehr verbreitete Frage lautet: gegeben die Renditeplots, die sich auf unterschiedliche Zeitbasen (Zeitfenster oder Zeithorizonte) beziehen, ist es möglich, das Bezugszeitfenster, die Skala, visuell zu bestimmen? Und mit Hilfe statistischer Methoden? Welche Methoden sind dafür geeignet? Welche charakteristischen Eigenschaften würden darauf hindeuten, dass es sich um eine bestimmte Zeitskala handelt bzw. dass die Aufteilung auf die möglicherweise unabhängigen Skalen überhaupt sinnvoll und möglich ist? All dies sind Fragen, die mit der sogenannten *Selbstähnlichkeit* (oder *Selbstaffinität* im engeren Sinne) zu tun haben. Manchmal ist es interessant festzustellen, dass eben *kein* Unterschied in den Eigenschaften vorliegt, unabhängig davon, ob man die Zeitintervalle in Sekunden oder Monaten misst.

Was ist der Hauptgrund, kontinuierliche statt diskreter Renditen zu verwenden? Die ersten lassen sich einfach addieren; die letzteren eben nicht! Der Preis dieser Bequemlichkeit schlägt sich in den Abweichungen nieder. In den meisten Fällen können sie unbeachtet bleiben; hin und wieder gibt es aber Probleme, wenn die Abweichungen *zu groß* werden. Betrachten wir³ eine n -periodige Rendite $\prod_{t=1}^n R_t$ für großes n . Die einperiodige Rendite R_t sei eine Realisierung aus der Serie von unabhängig identisch verteilten Zufallsvariablen X_t mit dem Erwartungswert $\mu = E R_1$. Es sei weiter $\alpha = E[\ln R_1]$. Wegen Unabhängigkeit und identischer Verteilung gilt

$$E \left(\prod_t R_t \right) = \mu^n. \quad (1.1.1)$$

Nach dem starken Gesetz der Großen Zahlen genügt gleichzeitig die Summe der Log-Renditen der Relation

$$\mathbb{P} \left(\frac{1}{n} \sum_{t=1}^n \ln R_t \rightarrow \alpha \right) = 1. \quad (1.1.2)$$

² Zur Vermeidung einer unnötigen Komplexität lassen wir die Bid/Ask-Preise und sich daraus ergebende Problematik der Renditeberechnung an dieser Stelle außer Betracht.

³ Dieses Beispiel führen wir zum Teil in Anlehnung auf Aoki (1996), S. 63 bzw. Schwartz und Weiss (1995) auf. Um technische Schwierigkeiten zu vermeiden, soll $1 \leq R_t \leq 2$ für alle t gelten.

Die Anwendung der Jensen-Ungleichung auf eine (konkave) Log-Funktion liefert schließlich für $t \gg 1$

$$\ln E[R_t] \geq E[\ln R_t], \quad (1.1.3)$$

$$\mu = \exp(\ln E[R_t]) \geq \exp(E[\ln R_t]) = e^\alpha. \quad (1.1.4)$$

Der Bezug zur Skalierung lässt sich natürlich auch hier erkennen. Die Gesamtrendite über einen aggregierten Zeitraum errechnet sich entweder durch eine *Multiplikation* im Falle diskreter Renditen, oder durch eine *Aufsummierung* andernfalls. Was sagt uns die Gleichung (1.1.4)? Heißt es, dass die Renditen (und folglich auch die Aktienkurse) im Durchschnitt steigen (wir legten ja $\mu = E R_1$ fest)? Dies würde automatisch bedeuten, dass eine buy-and-hold Strategie allen anderen Strategien überlegen ist. Aus der Mathematik kennen wir eine berühmte Relation

$$\lim_{x \rightarrow 0} \frac{\ln(1+x)}{x} = 1.$$

Es ist nicht schwer, diese Formel im Zusammenhang mit diskreten r_d und kontinuierlichen r_c Renditen zu betrachten: im einperiodigen Fall gilt $1 + r_d = e^{r_c}$. Durch Approximation diskreter Renditen durch die kontinuierlichen lassen wir den lim-Operator außer Acht. Dies ist für kleine Argumente zulässig. Darunter verstehen wir entweder die *kleinen Renditen* im absoluten Betrag oder die *kleinen Zeitabstände* zwischen den Messungen. Sind die Argumente nicht mehr klein, muss man sich etwas einfallen lassen, damit die entstehenden Abweichungen korrigiert werden können.

Nun geht der klassische Erklärungsansatz von einem Korrekturfaktor für die Konvexität aus. Eine Sequenz der Rendite-Prozesse (Random Walks) konvergiert schwach⁴ gegen eine arithmetische Brownsche Bewegung mit Drift. Bezogen auf die Aktienkurse resultiert daraus eine geometrische Brownsche Bewegung. Der Korrekturfaktor ($-\sigma^2/2$) ist dann nichts weiter als ein Ergebnis der Anwendung der Itô-Formel auf die Log-Funktion. Er kommt beim Übergang vom *Diskreten* (Random Walk) ins *Kontinuierliche* (Brownsche Bewegung) dazu und bringt die Welt wieder in Ordnung. Man kann aus unterschiedlichen Verfahren der Renditeberechnung kein zusätzliches Kapital schlagen und es ist gleichgültig, wann man in den Markt mit einer Handelsstrategie einsteigt. Nur die Dauer der Investition spielt eine Rolle. *In der Grenzwertbetrachtung bleibt die erwartete Rendite pro Zeiteinheit immer gleich.*

⁴ Billingsley (1986), S. 327.

Die Skalierung ist linear und deterministisch in der Zeit: eine doppelte Haltedauer verspricht *ceteris paribus* eine Verdoppelung der erwarteten Rendite. Das Market-Timing ist in einer solchen Umgebung bestenfalls (wenn man von Transaktionskosten absieht) sinnlos. Die Dynamik der Preisbewegungen ist trivial und reduziert sich auf die Betrachtung einer einzigen Periode. Die Erweiterung auf mehrere Zeitintervalle ist durch Multiplikation mit der entsprechenden Kennzahl für die Haltedauer problemlos möglich.

Eine ähnliche Erklärung liefert die Large Deviations Theorie⁵. Danach besagt die Gleichung (1.1.4), dass die erwartete Wachstumsrate über n Perioden durch den Term $e^{n\alpha}$ zum Ausdruck kommt und nicht mit dem größeren Betrag $e^{n\ln\mu}$ übereinstimmt. Die *typischen* Aktienkursbewegungen über n Perioden resultieren in Renditen von der Ordnung $O(n\alpha)$. Manchmal erfolgen aber größere Ausschläge, so dass die Renditen von der Ordnung $O(n \ln\mu)$ realisiert werden können. Solche Verläufe gelten als extrem (hier: exponentiell) *unwahrscheinlich*. Sie kommen selten vor, klingen gegebenenfalls ebenso schnell ab und gelten als *Ausnahme*. Je länger der Zeithorizont, desto besser sind die Bedingungen für die Anwendung des starken Gesetzes der Großen Zahlen und desto höher ist die Performance der auf (1.1.2) bezogenen Handelsstrategie. Kurzfristig kann natürlich ein exponentiell unwahrscheinliches Ereignis zu mehr (oder weniger) Performance verhelfen.

Es lässt sich vermuten, dass der oben erwähnte Korrekturprozess in diesem Fall nicht mehr so einfach verläuft wie vorher. Die Abhilfe schafft eine Denkweise in Kategorien der *kumulierten oder integrierten Größen*, wie zum Beispiel die *Gesamtkorrektur über $t = 1$ bis $T > t$* . Ohne den einzelnen Zeitpunkt vorhersagen zu können, geht man davon aus, dass die Summe der Abweichungen vom Mittelwert innerhalb eines vorgegebenen Zeitabschnittes in beide Richtungen (nach unten und nach oben) beschränkt ist. Zur besseren Anschaulichkeit kann man sich eine *Fläche* (als integrierte Größe) unter der Trajektorie vorstellen, die in einem Laufband gefangen bleibt. Die „Rückkehr“ zum Erwartungswert erfolgt exponentiell schnell und wird formal durch die sogenannte *Raten-Funktion* beschrieben⁶. Da die oben beschriebene Beschränkung nicht deterministischer, sondern *stochastischer* Natur angenommen wird, bezeichnet man solche Ereignisse, die

⁵ Mehr dazu in Kapitel 5.

⁶ Wir geben einen wiederholten Verweis auf Kapitel 5.

über das Laufband „hinausschießen“, nicht als *unmöglich*, sondern als *untypisch*.

Jetzt bleibt nur noch der letzte kleine Schritt, bis die Skalierung wieder im Spiel ist. Die kumulierten Größen lassen sich mathematisch sehr bequem durch ihre Momente beschreiben. Gewöhnlich benutzt man dafür die Moment- oder Kumulanterzeugende Funktion. Eine breite Verwendung findet ebenfalls die charakteristische Funktion, die vor den letzten beiden einen zusätzlichen Vorteil hat, dass sie immer existiert. Bei der Normalverteilung interessiert man sich nur für die ersten zwei Momente, den Erwartungswert und die Varianz. Alle anderen ungeraden Momente verschwinden wegen Symmetrie; die geraden lassen sich auf die Varianz zurückführen. Die Kumulanterzeugende Funktion der Normalverteilung ist eine quadratische Funktion oder einfach eine Parabel. Nach zweifachem Differenzieren⁷ bekommen wir eine Konstante, und genau das ist der Grund, warum der Korrekturfaktor $(-\sigma^2/2)$ so einfach aussieht und vor allem *unabhängig von der Zeit bzw. Anzahl der Perioden* ist⁸. Was hat aber das Ganze mit der Skalierung zu tun?

Stellen wir uns vor, die zweite Ableitung der Kumulanterzeugenden Funktion wäre keine Konstante, sondern könnte durch eine Formel angegeben werden, die in irgendeiner Form die Zeit als Variable enthielte. Dann hätten wir einen unangenehmen Effekt, dass der Korrekturfaktor (im finanzwirtschaftlichen Kontext können wir über *Risikoprämie* sprechen) nicht mehr konstant, sondern variabel ist. Man muss sich im Klaren sein, dass dieser Schritt ganz bewusst aus dem Geltungsbereich der Normalverteilung hinausführt. Andererseits ist das Vorgehen insoweit plausibel, dass, unterstützt durch die empirischen Erkenntnisse, die sogenannten *stylized facts*, solche Wege in der Theorie *gezielt gesucht* werden. Wie können wir an dieses Problem herangehen?

Als erstes sehen wir jetzt beim Korrekturfaktor einen Ausdruck, der aus zwei Funktionen besteht, die miteinander multiplikativ verbunden sind, sagen wir $f(\sigma)$ und $f(t)$. Es spricht nichts dagegen, die beiden Größen sowohl deterministisch als auch stochastisch zu modellieren. So nebenbei bewegen wir uns jetzt zum *market time*-Konzept (stochas-

⁷ Die Momente lassen sich u.a. durch das Ableiten der Kumulanterzeugenden Funktion im $t = 0$ bestimmen.

⁸ Zur Vereinfachung machen wir eine Weile keinen Unterschied zwischen den Begriffen *Zeit* und *Anzahl der Perioden*.

tische Zeit) bzw. *stochastischer Volatilität* (andernfalls). Beides hat unmittelbar mit Skalierung zu tun: sie erfolgt jetzt aber *nichtlinear*, und zwar in der *Zeit* bzw. *Amplitude* oder absoluten Höhe der Renditen. Als weiteres beobachten wir, dass sich die Risikoprämie *ungleichmäßig* zwischen den kurz- und langfristigen Renditen bzw. zwischen (relativ) großen und (relativ) kleinen Renditen aufteilt. Die mathematische Beschreibung dieser Effekte läuft letztendlich durch Untersuchung der Momente der Renditeverteilung entweder auf anomale zeitliche Abhängigkeiten in den Renditezeitreihen wie *gebrochene Brownsche Bewegung* (Kapitel 6.3) hinaus oder man ist auf die Wahrscheinlichkeitsverteilungen angewiesen, die über extreme Eigenschaften wie *nicht-existierende* bzw. *unendliche* Momente, vor allem die unendliche Varianz, verfügen (stabile Verteilungen, Kapitel A.5). Mischformen und ähnliche Ansätze jeglicher Art wie *Subordination* durch die schon erwähnte *market time*, Mandelbrot's *Multifraktales Modell of Asset Returns*, Lévy's *unendlich teilbare Verteilungen* etc. etc. sind möglich und befinden sich zur Zeit in den Phasen zwischen intensiver Forschung und intensiver Anwendung im quantitativen Finance, vor allem bei der Derivatebewertung, Risk Management und im dynamischen Portfoliomanagement.

Viele dieser Modelle bilden ein mathematisches Gerüst für die empirischen Beobachtungen aus dem Kapitalmarkt. Nur wenige davon versuchen, eine ökonomische Basis für die *Erklärung* der beobachteten Skalierungsphänomene zu schaffen. Dies ist auch nicht weiter verwunderlich. Es ist kein Geheimnis, dass das Feld *quantitativen Finance* zum großen Teil von Physikern geprägt wird. Statt „Marktteilnehmer“ spricht man von „Agenten“, bei der Informationsverarbeitung denkt man an die Teilchen, die sich in einem porösen Medium verbreiten. Das Gute daran ist, dass in vielen Fällen das Rad schon längst erfunden ist; es nimmt nur unterschiedliche Gestalten auf. *Die Modellierungswerkzeuge sind gleich, die Interpretation ist anders*. Viele Ideen, aus denen sich die Finanztheorie Steinchen für Steinchen zusammenlegte, kamen von den Menschen, die sich anfangs mit einem anderen Gebiet der Wissenschaft beschäftigt haben. Man denke an den aus der Physik stammenden Fisher Black. Man denke an den Mathematiker Benoît Mandelbrot. Man denke an Andrei Nikolaevitsch Kolmogorov, der zwar keine ökonomischen Theorien, dafür aber die moderne Wahrscheinlichkeitstheorie aufgebaut hat. Ohne diese Theorie wäre eine Risikomessung nach derzeitigem Wissenstand unvorstellbar. Er

arbeitete an der Modellierung der Turbulenz. In seinen späten Arbeiten beschäftigte sich Kolmogorov mit der Formalisierung des Begriffes „Komplexität.“

Skalierung ist ein Merkmal zunehmender Komplexität. In komplexen Systemen gehen wir davon aus, dass das Verhalten eines ganzen Systems aus einer Vielzahl nichttrivialer Interaktionen zwischen den unterschiedlichen Skalen resultiert. Damit erfordert die Beschreibung des ganzen Systems ein qualitativ anderes Vorgehen, als beispielsweise die Vorgabe eines konstanten Skalierungsfaktors. Die bekanntesten Beispiele komplexer Systeme sind die Turbulenz oder das Gehirn. Jeder Versuch, das Verhalten eines solchen Systems durch eine einfache Interpolation von deren Einheiten vorherzusagen, führt offensichtlich ins Leere. Der Nobelpreisträger in Physik Richard Feynman schrieb in seinen „Sechs physikalischen Fingerübungen“ (Feynman (2002), S. 120-121):

*„Und schließlich gibt es ein . . . sehr altes Problem, das jedoch noch immer nicht gelöst ist. Es geht nicht darum, neue Elementarteilchen zu entdecken, sondern um etwas, das seit langer Zeit – seit mehr als hundert Jahren – schlicht liegengeblieben ist. Kein Physiker war wirklich in der Lage, es mathematisch zufriedenstellend zu analysieren, obwohl es für die Schwesterwissenschaften sehr wichtig ist. Es handelt sich um die Analyse **zirkulierender oder turbulenter Flüssigkeiten**⁹. . .*

Am einfachsten lässt dieses Problem sich folgendermaßen veranschaulichen: Wir nehmen eine sehr lange Röhre und pumpen mit hoher Geschwindigkeit Wasser hindurch. Und dann fragen wir: Welchen Druck müssen wir ausüben, um eine gegebene Menge Wasser durch die Röhre zu pumpen? Niemand kann dies ausgehend von Grundprinzipien und den Eigenschaften von Wasser analysieren. Fließt das Wasser sehr langsam oder haben wir eine zähflüssige Masse wie Honig, dann geht das ohne weiteres. Das können Sie in Ihrem Lehrbuch nachlesen. Doch es gelingt uns einfach nicht, diese Frage zu lösen, wenn es um echtes nasses Wasser geht, das durch eine Röhre fließt. Das ist das Kernproblem, das wir eines Tages lösen sollten, aber bislang nicht geschafft haben.“

Einer der inzwischen vielen verbreiteten Lösungsansätze besteht darin, an die Komplexität der Interaktionen schrittweise heranzugehen. Kein Wunder, dass solche Schritte „log-skaliert“ sind: es liegt in der Natur

⁹ Originale Hervorhebung.

der Sache, ob man ein Mikroskop oder ein Teleskop braucht, um sich auf einer spezifischen Skala eine bessere Übersicht zu verschaffen.

Im ökonomischen Kontext kann man sich vorstellen, dass die Informationsverarbeitung, die in Bezug auf die Kursbewegungen gewöhnlich als kausal angesehen wird, *parallel auf mehreren Zeitskalen* erfolgt. Die ökonomischen Zusammenhänge sind komplex: gegeben die Tatsachen, braucht die Erkenntnis über deren Bedeutung oft eine gewisse Zeit. Die volkswirtschaftlichen Informationen entfalten ihre Wirkung nicht unbedingt innerhalb weniger Stunden, seitdem sie bekannt werden. Gewöhnlich vergehen Monate, bis sich beispielsweise die Zinsänderungsschritte der FED oder EZB auf das Wirtschaftswachstum und dadurch auch auf ein bestimmtes Unternehmen auswirken. Die vorweggenommene Veränderung der Geschäftslage schlägt sich aber in den Aktienkursen womöglich schon viel früher nieder. Ist das nicht ein turbulenz-ähnlicher Zustand?

Die größeren Renditeauschläge sowie die vermutlichen Trends ließen sich unter diesen Bedingungen relativ einfach und *ohne Einbezug von exponentiell unwahrscheinlichen Ereignissen* erklären: *die Ereignisse aus den unterschiedlichen Zeitskalen überlagern sich!* Dieser Ansatz weist einige spezifische qualitative Merkmale auf: eine längere Renditefolge des gleichen Vorzeichens oder ein Market-Crash stellen dabei *typische* Ereignisse dar. Die Unsicherheit bekommt einen zusätzlichen Freiheitsgrad: die Skala. 20% Rendite an einem Tag sind viel; bezogen auf drei Jahre scheint es nicht mehr so ungewöhnlich zu sein (ein Ereignis aus der längeren Zeitskala ist eingetreten). Vier Monate lang jeden Tag nur im Plus geschlossen? Dies könnte eine (zufällige?) Überlagerung verschiedener Informationsskalen sein! Da wir den, sagen wir, *Generator der Unsicherheit* weder kennen noch beobachten können, ist es schwierig einzuschätzen, wie sich die kursrelevanten Informationen zwischen den „ruhigen“ Skalen am Rande und „wildem“ Strömungen irgendwo in der Mitte einer turbulenten Börsenzeit zur Verarbeitung aufteilen. Insoweit erübrigt sich die Problematik der Vorhersagbarkeit der Aktienkurse: gehen wir davon aus, dass die Börsenwelt eine turbulente Welt ist, wie ein „echtes nasses Wasser,“ nicht mal Richard Feynman könnte ein solches Wissen gewinnbringend ausnutzen!

Gliederung der Arbeit

Die Arbeit beginnt mit der Übersicht des grundlegenden und gleichzeitig nicht unumstrittenen Konzeptes der informationseffizienten Kapitalmärkte (EMH). Eine Vielzahl kontroverser Aussagen darüber, ob der Kapitalmarkt informationseffizient ist oder nicht, führt dazu, dass in dieser Frage auch nach mehreren Jahrzehnten immer noch keine Einigung in der wissenschaftlichen Gemeinde festzustellen ist. Obgleich die EMH ein notwendiges Fundament für die nachfolgenden Bewertungsmodelle ist, hat es sich nicht immer als guter Ausgangspunkt für die Erklärung empirischer Zusammenhänge erwiesen.

Im zweiten Kapitel werden alternative Konzepte der Modellierung von Aktienkursbewegungen diskutiert, die wegen einer höheren Komplexität der dazugehörigen quantitativen Verfahren (noch) keine ausreichende Akzeptanz in einem breiten Kreis der Anwender gefunden haben. Die sogenannten *stylized facts* bringen die Schwächen herkömmlicher Modelle ins Licht und bewegen zum Umdenken bei der ökonomischen Interpretation empirischer Befunde und ihrer Klassifizierung als Regeln oder Ausnahmen.

Das Thema „Regeln und Ausnahmen“ wird in den folgenden zwei Kapiteln weiterentwickelt. Der Grundgedanke dabei ist, dass die Prinzipien der Informationsverarbeitung, die in verschiedenen wissenschaftlichen Gebieten angewendet werden, ähnlich sein sollen. Die Anforderung nach Rationalität der Marktteilnehmer setzt voraus, dass die *Spielregeln* formuliert werden können. Das Konzept der Entropie bildet eine ausgezeichnete Basis für die Unterscheidung *typischer* und *untypischer* Ereignisse (Kapitel 4). Diese Unterscheidung kann zur Modellierung des asymptotischen Verhaltens stochastischer Prozesse verwendet werden. Weiterhin lassen sich auf dieser Grundlage Schranken für die großen Abweichungen von den typischen Werten herleiten (Kapitel 5).

Die zeitlichen Abhängigkeiten in den Aktienrenditen sind im Rahmen der EMH nur eingeschränkt zugelassen. Sämtliche preisrelevanten Informationen aus der Vergangenheit sollen im gegenwärtigen Aktienkurs vollständig und korrekt enthalten sein und dürfen die erwartete zukünftige Rendite nicht beeinflussen. Dies gilt automatisch, wenn den Aktienrenditen eine Semimartingaleigenschaft unterstellt wird. Leider kann die theoretisch „erlaubte“ (und empirisch beobachtete)

zeitliche Abhängigkeit in der Volatilität durch Semimartingale nicht nachgebildet werden. Lässt sich die Martingalannahme lockern? Ist eine fehlende Semimartingaleigenschaft hinreichend dafür, dass der Markt nicht mehr arbitragefrei ist? Nach aktuellem Kenntnisstand gibt es in der Literatur keine eindeutige Antwort auf die letzte Frage. Im Kapitel 6.3 werden am Beispiel eines speziellen stochastischen Prozesses, der gebrochenen Brownschen Bewegung, die Vor- und Nachteile unterschiedlicher Lösungswege aufgezeigt.

Ein zentrales Hilfsmittel bei der Beschreibung von Skalierungseigenschaften der Aktienkurse/Renditen bilden singuläre Verteilungen, die sehr anschaulich mittels Fraktale und Multifraktale nach B. Mandelbrot dargestellt werden können. Nach einer Einleitung in Kapitel 7 werden im folgenden Kapitel Modelle dargestellt, die ihren Ursprung teilweise in der Physik haben [Hasley u.a. 1986], aber dennoch eine hervorragende ökonomische Interpretation zulassen, indem durch ein Zufallsmaß ein unbekannter Generator der Unsicherheit modelliert wird. Das von B. Mandelbrot und Kollegen vorgeschlagene Multifraktale Modell of Asset Returns vereinigt eine mögliche zeitliche Abhängigkeit von Renditen mit einer gemäßigten Auswirkung von größeren Ausschlägen und eignet sich damit wesentlich besser zu einer realistischen Abbildung von empirischen Eigenschaften der Aktienkursbewegungen als die standardmäßig verwendeten Modelle.

Da die asymptotischen Grenzwerte auf der Log-Skala nur grobe Annäherungen sind, lassen sich auch durch die vorgeschlagenen Modelle keine Wunder bewirken, d.h. die Informationen aus dem Randbereich einer Verteilung (bzw. Informationen über die Gewichte einer unbekannter Mischung der Verteilungen) bleiben solange nicht zugänglich, bis sich im Laufe der Zeit genug relevante Ereignisse realisiert haben. Ein entscheidendes Merkmal der Multifraktale ist dabei ein Kompromiss zwischen der Häufigkeit des Auftretens eines Ereignisses und seiner Auswirkung im Falle des Auftretens.

In der Zusammenfassung werden schließlich die wichtigsten Ergebnisse der Arbeit diskutiert sowie Anregungen für die zukünftige Forschung gegeben. Im Anhang findet man zusätzliche Hilfsmittel, die für das Verständnis einiger mathematischer Zusammenhänge vom Nutzen sind.



2. Zur Markteffizienz

„Scientists have thick skins. They do not abandon a theory merely because facts contradict it. They normally invent some rescue hypothesis to explain what they then call a mere anomaly or, if they cannot explain the anomaly, they ignore it . . .

– Lakatos (1978), S.4.

2.1 Die Paradigmen

Die Aktienkursbewegungen (oder allgemein die Veränderungen der Wertpapierpreise wie Aktien, Währungen, Anleihen usw.) werden in der Finanzwirtschaft durch zwei zentralen Paradigmen¹ erklärt: die *These vom informationseffizienten Kapitalmarkt (Efficient Market Hypothesis, EMH)* und das Konzept der *rationalen Erwartungsgleichgewichte (RE)*. Dabei handelt es sich um die folgenden grundlegenden Ideen.

Erstens, der Preis oder der Marktwert eines Wertpapiers soll seinem Fundamentalwert entsprechen. Im allgemeinen versteht man unter Fundamentalwert die Summe künftig erwarteter diskontierter Cash Flows. Durch den Aktienbesitz erhalten die Eigentümer Ansprüche auf

¹ Eine der möglichen Definitionen des Begriffes „Paradigma“ entnehmen wir Kuhn (1970a): *“a paradigm is what the members of a scientific community share”* und *“a scientific community consists . . . of the practitioners of a scientific speciality.”* Auf die Darstellung der grundlegenden Konzepte von Thomas Kuhn und Karl Popper bezüglich der Ziele und des Verlaufs der wissenschaftlichen Forschung müssen wir in dieser Arbeit leider verzichten. Speziell zur Anwendung auf die EMH siehe Frankfurter und McGoun (2000) und Du (2004).

Dividendenzahlungen. Diese können in erster Näherung als Stellvertreter der Cash Flows betrachtet werden. Darum erwartet man, dass der Aktienpreis den aggregierten Wert der Dividenden wiedergibt.

Der zweite Punkt verdeutlicht die Vorgehensweise, die zur Übereinstimmung der beiden Größen – des Marktwertes und des Fundamentalwertes – führen soll. Die Grundannahme ist, dass sämtliche verfügbare Informationen über den Fundamentalwert jedem Investor zugänglich sind. Im Idealfall sind Investoren auch nicht daran gehindert, auf Grundlage der neu ankommenden Informationen zu handeln, wodurch diese Informationen in die Preise weitergeleitet werden. Wegen der Konkurrenz auf dem Markt verläuft dieser Prozess sehr schnell und es bleibt keine Möglichkeit für einige Investoren, einen Informationsvorsprung zu bekommen und diesen gewinnbringend (zur Erzielung von Überrenditen) auszunutzen. Die Überrenditen können allein auf Grundlage eines Informationsvorsprungs erzielt werden, was hier gerade nicht zulässig ist. Die Preise ändern sich nur auf Grundlage der neu ankommenden Informationen. Damit sind sie *effizient* und unvorhersagbar (EMH). Das Risiko bleibt dabei ein exogener Faktor; es werden keine Aussagen bezüglich der Rendite/Risiko-Beziehung getroffen.

Die *semantische* Informationskomponente, also die Interpretation der preisrelevanten Informationen seitens der Marktteilnehmer, ist Gegenstand der rationalen Erwartungen. Nach Muth (1961) wird davon ausgegangen, dass die Investoren perfekt rational agieren: sie sind in der Lage, beliebig komplizierte Zusammenhänge zu erkennen, diese mittels eines allgemein bekannten Modells zu verarbeiten und Handlungsentscheidungen zu treffen. Alle neu ankommenden Informationen werden zunächst nach dem Prinzip „preisrelevant“ vs. „nicht preisrelevant“ aufgeteilt. Zu den ersten zählen die ökonomischen Faktoren, die die Cash Flows beeinflussen oder in Zukunft beeinflussen können und daher für den Fundamentalwert von Bedeutung sind. Bei der Verarbeitung dieser Informationen entstehen *rationale Erwartungen* über die zukünftigen Cash Flows. Die restlichen Informationen wie die Chartanalyse oder die Stimmung der Marktteilnehmer (*behavioral aspects*) sollen dabei zu keiner Erwartungsbildung führen; tun sie es doch, heißen solche Erwartungen *irrational*. Das Marktgeschehen kann damit als ein permanenter Transfer von Erwartungen in die Preise gesehen werden. Dieser Prozess soll nach den Regeln verlaufen, die von den beiden Paradigmen aufgestellt sind.

Da die Erwartungen zukunftsgerichtete Größen darstellen, werden sie durch den Informationsfluss verfeinert und präzisiert. Zu gegebener Zeit werden sie zu Nachrichten. Dies gilt insbesondere für die Berichtsaison, wenn Unternehmen ihre Quartalszahlen veröffentlichen. Dann hört man, dass die Erwartungen der Investoren *getroffen* oder sogar *übertroffen* sind, wobei die anderen Unternehmen auch *enttäuschen* können. Es ist üblich, dass sich die Unternehmen gleichzeitig mit der Bekanntgabe der Quartalszahlen in einer Presse-Konferenz zu den Geschäftsaussichten für das nächste Quartal äußern. Diese Schätzungen werden von der Vielzahl der Marktteilnehmer aufgefangen und durch das ganze laufende Quartal getragen und angepasst. So entstehen neue Erwartungen und der Kreis schließt sich. Natürlich gibt es eine ganze Menge anderer Faktoren, die kursrelevant sind oder sein können, wie zum Beispiel volkswirtschaftliche Faktoren und Kennzahlen über die Beschäftigung, über die Produktion, desweiteren politische Faktoren, die regulierenden Eingriffe der Zentralbanken u.v.m. All diese Faktoren sollen laufend berücksichtigt und eingepreist werden (EMH), und zwar unter dem Gesichtspunkt eines rationalen Investors (RE).

Eugene Fama als anerkannter „Vater“ der EMH gibt in einer seinen ersten Arbeiten zu diesem Thema [Fama 1970] die Kernidee eines effizienten Marktes wieder: *„Auf einem effizienten Markt sind die preisrelevanten verfügbaren Informationen in den Preisen vollständig abgebildet.“* Die hinreichenden Bedingungen dafür sind:²

- i. Es gibt keine Transaktionskosten in handelbaren Wertpapieren;
- ii. Informationen sind kostenlos für alle Marktteilnehmer verfügbar;
- iii. Alle Investoren sind sich über die Auswirkungen der vorhandenen Informationen auf den gegenwärtigen Preis sowie auf die Verteilung der zukünftigen Preise jedes Wertpapiers einig.

Wahrscheinlich konnte damals kaum jemand vermuten, dass diese Sätze im Laufe der Zeit so viele kontroverse Diskussionen auslösen würden. Es hat sich ergeben, dass die Verfeinerung, Weiterentwicklung und die empirischen Tests dieser These zur Fama's Lebensaufgabe wurden. Die EMH gibt eine plausible Beschreibung dafür, wie die Preisbewegungen auf dem Kapitalmarkt erfolgen. Die Annahme, dass Preis-

² Fama (1970), S. 384

BEISPIEL 2.1 Weitere Definitionen der Markteffizienz

Sharpe u.a. (1999): "A market is efficient with respect to a particular set of information if it is impossible to make abnormal profits by using this set of information to formulate buying and selling decisions."

Frankfurter und McGoun (2000): "Market efficiency in its vernacular form states that there are no unexploited profit opportunities lying around out there that the astute rational investor can consistently grab."

änderungen durch Bekanntgabe neuer, unvorhersagbarer Informationen hervorgerufen werden, lässt sich mit stochastischen Modellen sehr gut vereinbaren. Es wird eine logische Brücke zur mathematisch-statistischen Modellierung ökonomischer Zusammenhänge aufgebaut.

Alle Investoren sind . . . einig. Sind sie es?

Sind die preisbeeinflussenden Faktoren nicht vorhersagbar, dann gilt dies auch für die Preise selbst. In vielen Fällen wird aber eine *relative* Bewertung ermöglicht, indem man einen bestimmten Zusammenhang zwischen den unterschiedlichen Wertpapieren oder Wertpapierarten unterstellt, der auf dieselben möglicherweise stochastischen Eingangsgrößen zurückzuführen ist. Eines der bekanntesten Beispiele dafür ist die *Arbitrage Pricing Theory*³ (APT), die eine Bewertung unsicherer zukünftiger Zahlungsströme mittels einer *unbestimmten Anzahl der a-priori nicht bekannten* orthogonalen Faktoren ermöglicht.

Die ökonomische Argumentation der APT ist perfekt: lassen sich ökonomieweit einige wenigen Faktoren finden, die als „Generator der Unsicherheit“ oder Risikoquelle interpretiert werden können, sind die erwarteten Renditen in der Regel eindeutig bestimmt. Die Einigkeit der Investoren über die Auswirkungen der vorhandenen Informationen auf den gegenwärtigen Preis ist irrelevant, da die preisbeeinflussenden Faktoren exogene Größen sind. Jede erzielbare Rendite kann als die Summe der Projektionen auf die Unsicherheitsfaktoren dargestellt werden, die Rendite/Risiko-Beziehung ist linear und eindeutig. Die „Renditequelle“ ist fest an die bekannte Risikoquelle gebunden. Eine Abweichung von dieser Regel würde bedeuten, dass es ein oder

³ Ross (1976), empirische Untersuchung für deutschen Aktienmarkt vgl. Sauer (1994).

BEISPIEL 2.2

Unbeachtet der Tatsache, dass die Anzahl der Faktoren weder von der APT selbst noch durch das formale Verfahren der Faktoranalyse vorgegeben wird, erscheint ein weiteres Detail erwähnenswert zu sein: die Kausalität der Beziehung zwischen den erklärenden und den zu erklärenden Variablen, der rechten und der linken Seite der Gleichung. Karl Popper in seiner Arbeit „Die Zielsetzung der Erfahrungswissenschaft“ geht sehr ausführlich auf das Problem der Kausalität allgemein ein: „*Ich nehme an, dass es das Ziel der empirischen Wissenschaft ist, befriedigende Erklärungen zu finden für alles, was uns einer Erklärung zu bedürfen scheint. Mit einer Erklärung ist eine Klasse von Sätzen gemeint, von denen einer den Sachverhalt beschreibt, der erklärt werden soll (das explicandum), während die anderen, die erklärenden Aussagen, die „Erklärung“ im engeren Sinne des Wortes bilden (das explicans des explicandums).*“ Und weiter: „*Das explicans... wird in der Regel nicht bekannt sein; es muß entdeckt werden. Daher wird die wissenschaftliche Erklärung, wenn immer sie eine Entdeckung ist, die Erklärung des Bekannten durch das Unbekannte sein... Damit das explicans nicht ad hoc sei, muß es reich an Gehalt sein; es muß eine große Zahl prüfbarer Folgerungen enthalten und unter ihnen, vor allem, prüfbare Folgerungen, die von dem explicandum ganz verschieden sind.*“ [Popper 1972], S. 29-30.

mehrere zusätzliche, bisher unbekannte preisbeeinflussende Faktoren gibt, was der Ausgangssituation widerspricht.

Die praktische Anwendung der APT zieht einige interessanten Konsequenzen mit sich. Sollen die Faktoren ein ökonomisches Gehalt haben, gibt es a-priori keinen Grund für sie, orthogonal oder unabhängig zu sein. Sollten beispielsweise die preisbeeinflussenden Größen volkswirtschaftlicher Natur sein,⁴ vermutet man unweigerlich Wechselwirkungen zwischen ihnen. Dies führt aber zu einem Widerspruch, weil die APT von orthogonalen Faktoren ausgeht. Sollen die nach der APT bestimmten Faktoren als Risikoquelle akzeptiert werden, welchen ökonomischen Gehalt haben sie dann? Bleibt diese Frage unbeantwortet, entsteht sofort die nächste: wie begründet man *kausale Zusammenhänge* zwischen den Wertpapierpreisen und den Faktoren, die ökonomisch schwierig zu interpretieren sind? Es scheint ein philosophisches Problem der *Erklärung des Bekannten durch das Unbekannte* zu sein.⁵

Angenommen, die Risikofaktoren sind allein durch die neu ankommenden preisrelevanten Informationen repräsentiert. Sind die zustande gekommenen Preise korrekt im Sinne der EMH, dann impliziert die

⁴ Vgl. Beispiel 2.3. Der Bezug auf den FOREX ist natürlich nur veranschaulichend.

⁵ Vgl. Beispiel 2.2.

BEISPIEL 2.3 Prof. Dr. Norbert Walter, Chef-Volkswirt der Deutschen Bank, schreibt in „Börse am Sonntag“ [Walter 2004]:

„Von den bedeutenden Hypothesen, wodurch der Wechselkurs beeinflusst wird, sind Wachstums- und Zinsdifferenzen und große Leistungsbilanzungleichgewichte zu nennen. Die ersten beiden Hypothesen richten sich auf die Renditerelationen zwischen Anlagen in verschiedenen Währungen; die letztere spiegelt die Optimierungsstrategie bezüglich der Währungszusammensetzung der Portfolios wider.

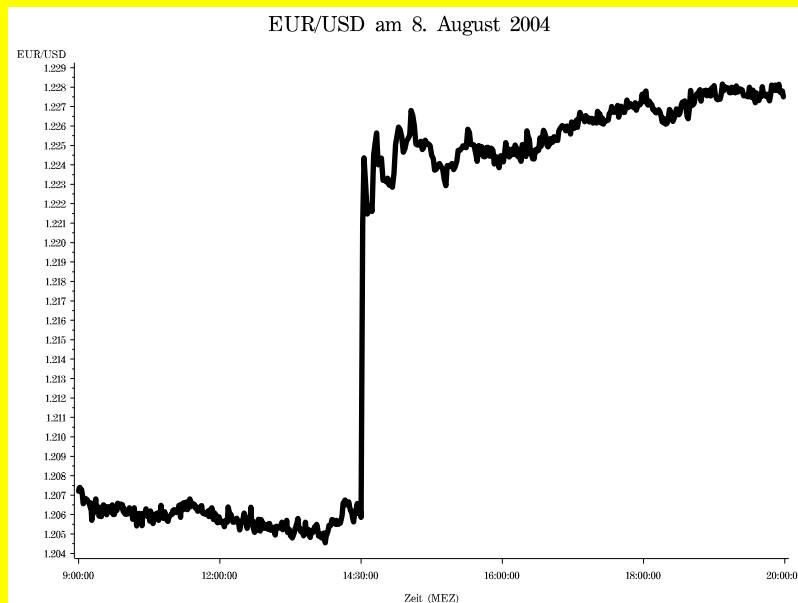
Wäre die Wachstumsdifferenz relevant, folgt hieraus derzeit eine klare Aufwertungs-tendenz für den Dollar.

Bei der Zinsdifferenz liegen die Dinge nicht so einfach. . . . die Kapitalmarktzinsdifferenzen sprechen schon jetzt eindeutig für eine Dollaraufwertung.

Die anhaltend defizitäre Leistungsbilanz der USA in Höhe von 5% des Sozialprodukts oder gut 500 Mrd. Dollar hingegen liegt so weit über jeder vertretbaren Defizitgröße, dass . . . der Dollar erst einmal kräftig unterbewertet sein muss. Also, für's erste schwächerer Dollar angesagt!

Da die skizzierten Erklärungshypothesen zu unterschiedlichen Trends für den Wechselkurs führen, ist eine Prognose wohl unmöglich. . . .“

Wie man auf dem folgenden Bild sieht, können die Anpassungen an die neuen Informationen recht heftig erfolgen. Am 8. August 2004 wurden in den USA die Arbeitsmarktdaten veröffentlicht, die die Erwartungen bei weitem verfehlt haben (Es entstanden im Juli 2004 32 000 neuer Jobs anstatt der prognostizierten 240 000). Es folgte sofort ein Mini-Crash der US-Währung!



Einigkeit der Investoren (weltweit, ohne jede Absprache, man muss ja schnell reagieren) über die Auswirkungen der vorhandenen Informationen auf den gegenwärtigen Preis nur eins: diese Auswirkungen sind trivial. Der triviale Fall wurde schon im Rahmen der APT betrachtet, wobei die Einigkeit der Investoren irrelevant ist, aber nicht ausgeschlossen werden kann. Der nichttriviale Fall kann a) die unbekanntenen Auswirkungen oder b) die fehlende Einigkeit der Investoren einschließen. Bei a) erfolgt eine vollständige Entkoppelung der Marktpreise von den Fundamentaldaten, was der Grundannahme der EMH widerspricht und b) führt zwar nicht unmittelbar zum Widerspruch mit der EMH (eine hinreichende Bedingung ist verletzt), garantiert aber nicht mehr ihre automatische Gültigkeit, da es auch durch den Zufall zu den korrekten Preisen kommen kann.

Wie das Beispiel 2.3 zeigt, bedeuten die bekannten Auswirkungen der ökonomischen Faktoren noch lange keine Einigkeit der Investoren über das Zusammenspiel dieser Faktoren bei der Entstehung der Wertpapierpreise. Die dritte hinreichende Bedingung nach Fama (1970) ist damit nicht automatisch erfüllt und die EMH bleibt eine Hypothese, die einer Vielzahl empirischer Tests unterzogen worden ist. Das Testen der *prüfbar*en Folgerungen ist natürlich nur indirekt möglich. Ein *direkter* Test der Hypothese solcher Größenordnung wäre einfach nicht vorstellbar. Er müsste gegebenenfalls mit einer „vernünftigen“ Anzahl der Marktteilnehmer durchgeführt werden, wobei jeder Beteiligte alle preisrelevanten Informationen zur Verfügung gestellt bekäme und danach gefragt würde, welche Informationen er oder sie bei ihren Entscheidungen berücksichtigt bzw. ob und in welchem Ausmaß die gegebenen Informationen eine Wirkung auf die Wertpapierpreise haben sollten.⁶

Ein indirektes Testen führt zu einem *joint hypothesis*-Problem, weil eine *bedingte* Aussage getestet wird: gegeben alle preisrelevanten Informationen, müssen sich die Wertpapierpreise auf eine bestimmte Weise verhalten. Dieses Verhalten wird von einem weiteren Modell vorgegeben. Insoweit lässt sich weder ein positiver noch ein negativer Ausgang des Tests allein auf die EMH zurückführen, wodurch sich seine Aussagekraft entsprechend wesentlich reduziert.

Nimmt man diese Problematik in Kauf, ermöglichen schon die einfachsten stochastischen Modelle wie ein Random Walk die Formulie-

⁶ Frankfurter und McGoun (2000), S. 203.

rung und die Überprüfung testbarer Hypothesen, die gegebenenfalls verworfen werden können. Die Ergebnisse lassen sich dann eingeschränkt für oder gegen die EMH interpretieren.

Doch schon lange bevor die EMH formuliert worden ist, hat man ein großes Interesse an empirischen Untersuchungen auf dem Bereich Aktienkursentwicklung gezeigt, wobei es in erster Linie um die Vorhersagbarkeit der Aktienkurse ging. Diese historische Periode vor EMH bezeichnet man im Sinne von Kuhn (1970b) als *pre-paradigm stage*.

Vom Random Walk zum Martingal

Eine der ersten Arbeiten auf dem Bereich Aktienkursentwicklung war die (inzwischen) berühmte Ph.D.-These vom Louis Bachelier “Théorie de la Spéculation” aus dem Jahre 1900⁷. Aufgrund seiner Beobachtung, dass “*past, present and even discounted future events are reflected in market price, but often show no apparent relation to price changes,*” ging er von der Unkorreliertheit und der Stationarität der Zuwächse aus und kam zu einer eindimensionalen Diffusionsgleichung, einem Modell der Brownschen Bewegung, das später (ohne über Bachelier’s Ergebnisse zu wissen) von Einstein (1905) vorgeschlagen war.

Es ist bekannt, dass eine solche Konstruktion zum Wiener Prozess führt. Die Zuwächse des Wiener Prozesses sind unabhängig und normalverteilt. Damit besteht auch ein enger Zusammenhang mit dem zentralen Grenzwertsatz. Wird ein Zeitintervall $[t, t + \Delta t]$ in eine große Anzahl von Subintervallen aufgeteilt, stellt sich heraus, dass die Zuwächse⁸ über die längeren Zeitintervalle $P(t + \tau) - P(t)$ durch eine Summe der Zuwächse aus den Subintervallen darstellbar sind. Sind die letzten unabhängig, dann ist ihre Summe asymptotisch normalverteilt.

Im diskreten Fall wird durch eine regressionsähnliche Gleichung für die Log-Preise $\{R_t \equiv \ln P_t : t = 1, 2, \dots\}$ ein *Random Walk* mit einem Driftparameter μ und einem zufälligen Störterm ε_t ($E\varepsilon_t = 0$ für alle t) spezifiziert:

⁷ Eine exzellente Übersicht dazu gibt Taquu (2001); siehe auch Dimson und Mussavian (2000).

⁸ Im ökonomischen Kontext handelt es sich bei den Zuwächsen natürlich in erster Linie um Zuwächse der Aktienkursen, d.h. Renditen. Das P steht dann logischerweise für den *Preis*.

$$R_{t+1} = \mu + R_t + \varepsilon_{t+1}. \quad (2.1.1)$$

Der Zeitindex t soll den *dynamischen* Charakter der Gleichung betonen. Dazu kommt eine weitere Bedingung $E[\varepsilon_t - \varepsilon_{t-k}] = 0$ für alle $t > k > 0$, wodurch sich die Unabhängigkeit der Zuwächse über die Zeit mit einem Markov-Prozess modellieren lässt.⁹ In der kontinuierlichen Approximation ergibt sich eine *Brownsche Bewegung*:

$$\ln[P(t+dt)] - \ln[P(t)] = \ln[P(t+dt)/P(t)] \sim N(\mu dt, \sigma^2 dt). \quad (2.1.2)$$

Die Unmöglichkeit einer systematischen Vorhersage der Aktienpreise wurde damit von Louis Bachelier postuliert und formalisiert: “*the mathematical expectation of the speculator is zero.*”¹⁰ Der Random Walk kann noch nicht als Theorie bezeichnet werden, es stellt lediglich ein abgeschlossenes Modell für die Preisbewegungen am Aktienmarkt dar. Es sind Jahre vergangen, bevor man über eine Theorie im ökonomischen, analytischen, mathematischen und statistischen Sinne sprechen konnte. Das ursprüngliche Konzept der Unmöglichkeit einer systematischen Vorhersage bildete nach wie vor den Kernpunkt. Dazu kam das Konzept der Information. Es wurde präzisiert und seine Bedeutung wurde geklärt. Es folgte eine Verallgemeinerung zu Martingalen und es wurden unzählige statistische Tests entwickelt, um eine berühmte Frage zu beantworten, die Alfred Cowles schon vor siebzig Jahren gestellt hat: “*Can stock market forecasters forecast?*”¹¹

Zu den ersten empirischen Untersuchungen zu Gunsten des Random Walk zählen Cowles (1933, 1944), Cowles und Jones (1937), die sich mit Aktienkursen beschäftigten. Working (1934) betrachtet die Warenpreise. Kendall (1953) versucht Zyklen in den Preisbewegungen zu finden. Seine Arbeit kann als bahnbrechend bezeichnet werden, auch wenn sie rein statistischer Natur ist und keinen tiefgehenden ökonomischen Bezug aufweist. “*How do market prices behave, what stochastic processes can be used to describe their dynamics?*” sind die Kernfragen.

⁹ Der Kernpunkt dieses Modells liegt in der Überlegung, dass keiner der Summanden eine dominierende Rolle in der Summe annehmen darf bzw. dass jedes Glied in der Summe asymptotisch vernachlässigbar ist (*die Lindeberg-Bedingung*). Haben alle Subintervalle identische Länge und geht man von der gleichen Verteilung auf jedem Subintervall aus, dann impliziert die Lindeberg-Bedingung eine endliche Varianz elementarer Zuwächse.

¹⁰ Es muss beachtet werden, dass Louis Bachelier nicht mit den Log-Preisen gearbeitet hat, sondern mit den Preisen selbst. Die daraus entstehenden Probleme der Nichtnegativität und der Asymmetrie der Preise hat er erkannt, aber ignoriert.

¹¹ Cowles (1933). Seine eigene Schlussfolgerung: “It is doubtful.”

Auf der Basis von wöchentlichen Kursen von 19 Aktien aus den Jahren 1928–1938, monatlichen durchschnittlichen Preisen für den Weizen (Börse Chicago, 1883–1934), Preisen für Baumwolle auf NYME (New York Mercantile Exchange, 1816–1951) stellte er keine Zyklen, Trends oder Rhythmen fest. Nach seiner Auffassung sahen die Daten so aus, als ob „... *the Demon of Chance drew a random number... and added it to the current price to determine the next... price.*“ In der Literatur schreibt man sehr oft dieser Arbeit die Aufstellung des Random Walk-Modells zu. Roberts (1959) argumentiert ebenfalls zu Gunsten von Random Walk. Er zeigte heuristisch, dass eine Zeitreihe, die aus einer Folge von Zufallszahlen erzeugt ist, nicht von einer Zeitreihe der US-Aktienpreise unterschieden werden kann. Osborne (1959) nimmt eine wichtige Erweiterung des Random Walk-Modells vor: die logarithmische Transformation der Preise. Dies führt zu einer Lognormalverteilung der Aktienkurse. Durch die Anwendung bekannter Verfahren statistischer Mechanik auf den Aktienmarkt führt Osborne eine detaillierte Analyse der Fluktuationen von Aktienkursen aus der Sicht eines Physikers durch. Diese Idee wurde von Samuelson (1964) weiterentwickelt. Er führte die *geometrisch Brownsche Bewegung*

$$P_t = P_0 \exp\left\{(\mu - \sigma^2/2)t + \sigma W_t\right\}, \quad t \geq 0 \quad (2.1.3)$$

im Finance ein und bezeichnete sie als *“economic Brownian motion.”* Inzwischen ist es üblich, die Formel (2.1.3) als „Samuelson-Modell“ zu bezeichnen. Die Annahme, dass der heutige Aktienkurs alle preisrelevanten Informationen beinhaltet und damit der beste Schätzer seines zukünftigen Wertes ist, führt nach Samuelson (1965) *“Proof that properly anticipated prices fluctuate randomly”* zur *Martingaleigenschaft*: $E_t[P_{t+1} - P_t | \mathcal{F}_t] = 0$. Diese Formulierung bedeutete Abschwächung einer sehr strengen Random Walk-Restriktion. Lediglich eine Abhängigkeit des ersten Momentes ist dabei ausgeschlossen. Es wird keine Aussage über das Verhalten der Momente höherer Ordnungen, vor allem der Varianz, getroffen. Seine ökonomische Motivation ist die Rationalität der Marktteilnehmer: *“in competitive markets there is a buyer for every seller. If one could be sure that a price would rise, it would have already risen.”* Damit wurden die letzten wichtigen Meilensteine für das Konzept der Markteffizienz gelegt, obwohl der Nobelpreisträger Paul Samuelson damals noch nicht zu einer formalen Definition gekommen war. Das *pre-paradigm stage* endete.

Die Aufstellung der Effizienzthese

Die Arbeit von Fama, Fisher, Jensen und Roll (1969) (FFJR) wird in der Literatur¹² als *die Arbeit* bezeichnet, in der der Begriff „effizienter Markt“ zum ersten mal erwähnt wird. FFJR bezeichnen einen effizienten Markt als *“a market that adjust rapidly to new information.”* Vor dieser Zeit existierte lediglich eine heuristische Verbindung zwischen statistischer und ökonomischer Beschreibung des Marktgeschehens. Der entscheidende Fortschritt erfolgt 1970, wenn Eugene Fama die EMH als *Hypothese* formuliert und in drei Formen aufteilt:

- In ihrer *strengen* Form postuliert die Efficient Market Hypothesis, dass *sämtliche verfügbare Informationen, sowohl öffentliche als auch private, zu jedem Zeitpunkt in den Wertpapierpreisen enthalten sind.* Diese Form der Markteffizienz lässt keine Insiderinformationen zu. Weder hat ein Investor einen monopolistischen Zugang zu den „preisempfindlichen“ Informationen noch besitzt er irgendeine Art extraordinärer Fähigkeiten und kann deswegen nicht durch den Handel mit solchen Informationen Überrenditen erzielen. Es wird angenommen, dass die Preise sich unmittelbar an die neuen Informationen anpassen; damit werden private Informationen wertlos. Gilt diese Form, gibt es weder eine Erklärung noch eine Möglichkeit für systematische Überrenditen.
- Die Grundidee der *schwachen* Form besteht darin, dass *in den gegenwärtigen Preisen nur alle relevanten Finanzmarktinformationen enthalten sind.* Dazu gehören historische Preise, Renditen, Volatilitäten usw., alle Informationen, die aktuell durch den Handel auf dem Markt entstanden sind. Investoren können Überrenditen erzielen, falls sie im Besitz von Privatinformationen sind, die nicht gerade am Marktplatz bekannt sind. Sollten diese Privatinformationen öffentlich bekannt werden, bleibt es immer noch möglich, davon überdurchschnittlich zu profitieren, weil die Verarbeitungsgeschwindigkeit dieser Informationen durch die restlichen Marktteilnehmer unzureichend ist.
- Die *halb-strenge* Form der Markteffizienz unterscheidet sich geringfügig von der schwachen in dem Punkt, dass *die Anpassung der Preise an die öffentlichen Informationen schnell genug erfolgen, um auf dieser*

¹² Ball (1989).

Grundlage keine Überrenditen zu ermöglichen. Die Investoren können zusätzliche Vorteile erzielen indem sie Privatinformationen suchen und sich vor der Bekanntgabe dieser Informationen auf dem Markt entsprechend positionieren.

Zusammenfassend lautet die Grundidee in Fama (1970): ein effizienter Markt lässt keine Überrenditen auf Grundlage eines Informationsvorsprunges zu. Diese Aussage impliziert eine versteckte Annahme, dass der existierende Preis korrekt ist. Sechs Jahre später (1976) präzisiert Fama unter Berücksichtigung dieser Anmerkung von Le Roy die Grundidee des effizienten Marktes: der Markt ist Informationseffizient,¹³ wenn er die gegebenen Informationen zur Preisbildung „korrekt“ verwendet.

Die Anforderung nach einer korrekten Preisbildung ohne nähere Spezifikation verlangt nach einem weiteren Modell, in dem die Rendite/Risiko-Beziehung auf eine eindeutige Weise zum Ausdruck kommt und die (normalen) Renditen von den Überrenditen unterschieden werden können. Ein solches Modell kann beispielsweise das CAPM oder die APT sein. Gleichzeitig entsteht ein schon erwähntes Problem einer zusammengesetzten Hypothese,¹⁴ wodurch das Testen der EMH erheblich erschwert wird.

Die Tests

Die Literatur zum Thema „Testen der EMH“ ist sehr umfangreich. An erster Stelle ist selbstverständlich die historische Reihe von Fama (1970, 1991, 1998) zu erwähnen. In einer populären Form empfiehlt sich Bernstein (1992), sehr gute Übersichten geben außerdem Dimson und Mussavian (2000) bzw. Dimson, Marsh und Staunton (2001).

In Abhängigkeit davon, welche Form der Markteffizienz getestet werden soll, lassen sich die Tests ebenfalls in drei Kategorien unterteilen. Beim Testen der schwachen Form der Markteffizienz geht es darum, ob die Marktpreise alle verfügbare Informationen enthalten (*How well do past returns predict future returns?*). Man untersucht die Profitabilität der Handelsstrategien unter Berücksichtigung der historischen Kursverläufe und erwartet u.a. keine Autokorrelation in den Zeitreihen

¹³ Die Effizienz im Sinne der EMH ist immer die *Informationseffizienz*.

¹⁴ Campbell u.a. (1997), S. 24, Fama (1991), S. 1575-1576.

der Preise/Renditen, auf deren Grundlage gewinnbringende Handelsstrategien aufgebaut werden können. Auf einem schwach-effizienten Markt sind die zukünftigen Renditen unvorhersagbar und die technische Analyse ist sinnlos.

Die halb-strenge Form der Markteffizienz glaubt man durch die sogenannten "event-studies" falsifizieren zu können. Dabei wird der Grad der Reaktion der Marktteilnehmer auf News analysiert. (*How quickly do security prices reflect public information announcements?*) Schließlich konzentriert man sich beim Testen der strengen Form zum größten Teil auf die Performance der Investmentfonds. (*Do any investors have private information that is not fully reflected in market prices?*)¹⁵ Wie viele Fondsmanager schaffen es, die Benchmark zu schlagen? Ist das Rendite/Risiko-Verhältnis (gemessen gewöhnlich an Sharpe-Ratio) besser als bei der Benchmark? Nach Fama (1991) sind das "tests for private information."

Unter dem Strich steht fest, dass die Markteffizienz in ihrer schwachen Form durch die meisten empirischen Untersuchungen nicht widerlegt werden kann. In seiner Übersicht 1970 schließt Fama, dass "*the results are strongly in support*" der schwachen Form, bei den anderen beiden Formen "*the evidence in support of the efficient market model is extensive, and contradictory evidence is sparse*", aber "*much remains to be done.*" Eine deutliche Unterstützung bekommt die EMH bei Jensen (1978): "*There is no other proposition in economics which has more solid empirical evidence supporting it than the Efficient Market Hypothesis. . . In the literature of finance, accounting and the economics of uncertainty, the Efficient Market Hypothesis is accepted as a fact of life.*" Allerdings ruft Jensen gleich in dieser Arbeit zur Vorsicht auf: "*As better data become available (e.g. daily stock price data) and as our econometric sophistication increases, we are beginning to find inconsistencies that our cruder data and techniques missed in the past. It is evidence which we will not be able to ignore.*"

Und das ist nur der Anfang. Fama selbst berichtet über eine leichte serielle Korrelation in den täglichen Aktienrenditen sowie über andere Anzeichen für die Verletzung der statistischen Unabhängigkeit in den Renditezeitreihen. Er konnte sich nicht entscheiden, ob diese Effekte tatsächlich vorhanden wären oder nur Ergebnis seiner beschränkten statistischen Modelle darstellten. "*. . . We can probably never hope to*

¹⁵ Die Sätze in Klammern sind nach Fama (1970) zitiert.

find a time series that is characterized by perfect independence . . . For practical purposes, however, we may be willing to accept the independence assumption of the model . . . [at] some 'minimum acceptable' level. What constitutes a 'minimum acceptable' level of dependence depends, of course, on the particular problem that one is trying to solve. . ." [Fama 1965]. Einige Jahre später [Fama 1991] wird bei der schwachen Form nicht mehr die Unabhängigkeit der Renditen getestet; man konzentriert sich vollständig auf ihre Unvorhersagbarkeit.

LeRoy (1973) und Lucas (1978) haben gezeigt, dass das Random Walk-Modell weder notwendig noch hinreichend für rationale Preise ist. In anderen Worten, aus der Unvorhersagbarkeit der Preise folgt nicht die Existenz eines funktionierenden Finanzmarktes mit rationalen Investoren, und die vorhersagbaren Preise bedeuten noch lange nicht die Irrationalität der Investoren¹⁶. Unter Einbeziehung von Risiko können die Investoren einen bestimmten Grad an Risikoaversion aufweisen und eine zusätzliche Rendite für die Risikoübernahme verlangen bzw. bereit sein, eine Versicherungsprämie für die Vermeidung des Risikos zu zahlen. Ist die Information über das Risikoverhalten der Investoren in den Preisen enthalten, brauchen letztere weder einem Random Walk folgen noch ein Martingal sein.

Grossman (1976), Grossman und Stiglitz (1980) zitierten und kritisierten Fama's 1970 Definition, dass zu jedem Zeitpunkt alle preisrelevanten Informationen in den Preisen enthalten sind. Sie zeigten, dass es in einem solchen Markt keinen Anreiz für die Investoren geben würde, kostspielige Informationen zu gewinnen, die für die anderen Marktteilnehmer kostenlos verfügbar wären. Sie schlugen eine Umformulierung des Begriffes „Effizienz“ im *noise context* vor.¹⁷ Das Gleiche gilt für die Fähigkeit der Marktteilnehmer, Informationen aus Preisen zu extrahieren. Als Folge unterscheidet man informierte und uninformierte Trader und es ist möglich, dass einige private Informationen nicht in den Preisen enthalten sind: es handelt sich um *asymmetrische Information*. Desweiteren argumentierten Grossman und Stiglitz (1980), die EMH sei inkonsistent mit der Realität einer kostspieligen Arbitrage. Sie entwickelten ein einfaches Modell mit einer CARA Nutzenfunktion¹⁸ und zeigten, dass kostenlose Informationen notwendig und hinreichend sind, damit die Preise alle verfügbaren Informationen

¹⁶ Lo und MacKinlay (2002), S. 5.

¹⁷ Man spricht von *noisy rational expectations*.

¹⁸ CARA ist eine Abkürzung für *Constant Absolute Risk Aversion*.

vollständig enthalten. Kostenlose Informationen als hinreichende Bedingung für die Markteffizienz sind inzwischen generell akzeptiert. Allerdings ist es nicht klar, welche Bedingung notwendig ist. Die Kosten für die Beschaffung und Entscheidung aufgrund neuer Informationen würden dazu führen, dass die Renditen in einem Markt unter vollständiger Konkurrenz nicht einem Random Walk zu folgen bräuchten.

Lo und MacKinlay (1988) verwerfen empirisch die Random Walk-Hypothese für den US Aktienmarkt. Malkiel (2000) bleibt davon unbeeindruckt und verteidigt sogar im Jahre 2000 den Random Walk. In einem vierhundertseitigen Buch kommen Lo und MacKinlay (2002) nach einer Vielzahl empirischer Untersuchungen zum Schluss: die Märkte sind effizient, ohne dass die Möglichkeit abnormaler Renditen vollständig eliminiert ist:¹⁹ *“Financial markets are predictable to some degree, but far from being a symptom of inefficiency or irrationality, predictability is the oil that lubricates the gears of capitalism.”* Kann das als entscheidendes Argument gegen die EMH interpretiert werden? Wohl kaum. Stattdessen wird lediglich die „totale Unabhängigkeit“ aufgegeben. Ende siebziger – Anfang achtziger Jahren kommt es erneut zu heftigen Diskussionen. Es werden Anomalien gefunden. So nennt man Ausnahmen aus der Regel, die in der Gesamtheit noch nicht ausreichen, um die Regel zu ändern.²⁰ Die Unternehmen mit relativ kleinem Kurs/Gewinn-Verhältnis weisen dauerhaft höhere Renditen auf. Die Kalendereffekte werden entdeckt: Aktien kleiner Unternehmen tendieren dazu, im Januar überdurchschnittliche Renditen zu erwirtschaften bzw. Montags-Renditen scheinen im Vergleich mit den anderen Wochentagen niedriger auszufallen. Fama und French (1988) und Poterba und Summers (1988) berichten über ein langfristiges mean-reverting Verhalten auf den US Aktienmärkten. Laut Fama und French sind zwischen 25% und 40% der Portfolio-Variabilität aus der Sicht von drei bis fünf Jahren vorhersagbar.²¹

Die Anomalien lassen sich aber im Rahmen der Paradigmen erklären. Eugene Fama hat keinen Zweifel an Markteffizienz. *“... when Fama and French delete the 1926-1940 period from the tests, the evidence of strong negative autokorrelations in 3- to 5-year returns disappears.”* (Fama (1991), S. 1581). *“The univariate tests on long-horizon returns*

¹⁹ Diese These stammt schon aus ihren früheren Arbeiten, vgl. z.B. Campbell, Lo und MacKinlay (1997).

²⁰ Es bleibt zu klären, ob die Anzahl der Ausnahmen von Bedeutung ist.

²¹ Eine ausführliche Übersicht der Anomalien findet man in Schwert (2002).

BEISPIEL 2.4 Ormerod (1998), S. 65–66:

“Some may object that even ... limited amount of information processing is far too difficult a task for many. The Basic Skills Agency carried out a survey of 6,000 men and women in seven developed economies. Without access to a calculator, they were asked to solve twelve elementary problems in arithmetic, such as ‘subtract 1.78 from 5’ and ‘multiply 6 by 21’. Those in Japan scored highest, but even then only 43 per cent of all Japanese got all the questions right. In Britain, only 20 per cent succeeded in giving correct answers to them all. A rather more esoteric finding by *Newsweek* also casts doubt on the general ability to process information rationally: almost one-third of American citizens believe that the US government is in regular contact with aliens. ... Individuals’ ability, or rather lack of it, to process information efficiently about future consequences creates a very serious internal problem for the theory of competitive equilibrium.”

BEISPIEL 2.5 Ormerod (1998), S. 14:

“In the early 1980s, comments began to circulate that the dollar was over-valued, and the trickle turned into a tidal wave. For most of 1983, the whole 1984, and into 1985, it was scarcely possible to pick up the *Wall Street Journal* or the *Financial Times* without reading about a stockbroker or dealer expressing this view. Yet, far from selling dollars, which seems the rational thing to do in these circumstances, dealers bought more and more of the currency. The dollar did, of course, eventually crash rather spectacularly in 1985, but for the best part of two years, dealers had exhibited clear signs of schizophrenia. They both knew and said that it was over-valued, yet still they continued to buy.”

of Fama und French (1988) und Poterba und Summers (1988) are a statistical power failure.” (Fama (1991), S. 1582). Nach wie vor wiederholt er die Problematik zusammengesetzter Hypothese: “*The joint-hypothesis problem is more serious. Thus, market efficiency per se is not testable. It must be tested jointly with some model of equilibrium, an asset-pricing model. ... We can only test whether information is properly reflected in prices in context of a pricing model that defines the meaning of “properly.” As a result, when we find anomalous evidence on the behavior of returns, the way it should be split between market inefficiency or a bad model of market equilibrium is ambiguous.*” (Fama (1991), S. 1575-1576). Spätestens jetzt stellt sich die Frage: wie soll es zum Gleichgewicht kommen, sollte der Markt ineffizient sein?

Es kommt immer wieder zum Gedanken, die EMH per se sei keine wohldefinierte und empirisch falsifizierbare Hypothese. Um ihre Aussagekraft zu stärken muss eine zusätzliche Struktur spezifiziert werden, u.a. die Informationsmenge.²² Die zufälligen Kursbewegungen

²² Lo (1997), S. 239.

BEISPIEL 2.6 Schleifer (2000):

“Investors follow the advice of financial gurus, fail to diversify, actively trade stocks and churn their portfolios, sell winning stocks and hold on to losing stocks thereby increasing their tax liabilities, buy and sell actively and expensively managed mutual funds, follow stock price patterns and other popular models. In short, investors hardly adopt the passive strategies expected of uninformed market participants by the efficient markets theory.”

BEISPIEL 2.7 Farmer und Lo (1999):

“The hypothesis that investors are fully rational agents that instantaneously and correctly process all available information is clearly unrealistic – rationality is difficult to define, human behavior is often unpredictable, information can be difficult to interpret, technology and institutions change constantly, and there are significant “frictional” costs to gathering and processing information and transacting. But how can we take all the complexities of the real world into account?

One new direction is to treat the EMH as an idealization that provides a useful reference point. What, then, can we conclude about the EMH? Amazingly, there is still no consensus among financial economists.”

BEISPIEL 2.8 George Soros, Soros Funds Management, New York, NY. 1994.

: “This [efficient market theory] interpretation of the way the financial market operates is severely distorted . . . It may seem strange that a patently false theory should gain such a widespread acceptance.”

BEISPIEL 2.9 Eugene Fama:

“. . . a ubiquitous problem in time-series tests of market efficiency, with no clear solution, is that irrational bubbles in stock prices are indistinguishable from rational time-varying expected returns.” (Fama (1991), S.1581).

“The SLB model (*Gemeint ist hier Sharp (1964), Lintner (1965) und Black (1972), S.N.*) is just a model and so surely false.” (Fama (1991), S. 1590).

“The empirical work on the consumption model often jointly tests its time-series and cross-section predictions . . . The tests usually reject. This is not surprising since we know all models are false.” (Fama (1991), S. 1596).

“Attacks on efficiency belong, of course, in the camp of the devil.” (Fama (1991), S. 1602).

“. . . the alternative hypothesis is . . . unacceptable.” (Fama (1998), S. 284).

werden durch die neu ankommenden Informationen verursacht. Können diese Informationen, die aus Shocks in den Fundamentalfaktoren hervorgehen und klassisch durch Dividendenänderungen erklärt werden, auch das Ausmaß der beobachteten Volatilität erklären? Oder sind die anderen Faktoren für die Varianz zuständig? Können sich die *fundamentals* so schnell ändern, dass es zu einem Crash kommt? Nach Ansicht von Robert Shiller (1989) sind die Preise dafür *zu volatil*. Man spricht von der Spekulation außerhalb rationaler Erwartun-

gen unter dem Stichwort *noise traders*.²³ Die Investoren sind nicht *smart* genug und folgen trends und fashions. Die Preise seien Ergebnis von “*indifferent thinking by millions of people*”,²⁴ die es sehr schwierig haben, Wahrscheinlichkeiten zukünftiger Ereignisse abzuschätzen. Es wird die *Rationalität* der Investoren in Frage gestellt.²⁵

Abgesehen davon nehmen wir an, die für die Interpretation der Informationen zuständigen Analysten seien intelligent, perfekt ausgebildet und haben sehr große Erfahrungen in qualitativen und quantitativen Methoden. Man kann kaum erwarten, dass sie weder die ökonomischen Daten noch ihre Auswirkungen auf die Aktienkurse eindeutig interpretieren werden, schon allein aus dem Grund, dass *es mehrere unterschiedliche Möglichkeiten gibt, die selben Daten zu interpretieren*.

Welche Antwort gibt Eugene Fama?

The alternative hypothesis . . . is unacceptable (?)

Mit dem Erscheinen des 1998-en Artikels von Eugene Fama beginnt eine neue Ära in der Geschichte der Markteffizienz: *Schutz vor Falsifikation*. Eine der Kernaussagen, die diese These bestätigen, gehört paradoxerweise Eugene Fama selbst:

“*A problem. . . on long-term return studies is that they rarely test a specific alternative to market efficiency. Instead, the alternative hypothesis is vague, market inefficiency. This is unacceptable. Like all models, market efficiency. . . is a faulty description of price formation. Following the standard scientific rule, however, market efficiency can only be replaced by a better specific model of price formation, itself potentially rejectable by empirical tests.*”²⁶

Dies steht sicherlich zum Teil im Einklang mit Aussagen von Thomas Kuhn oder Hans Albert.²⁷ Das Hauptproblem liegt aber darin, dass die Möglichkeit eines alternativen Paradigmas nicht einmal in Frage kommt. Weiterhin betont Fama, dass jedes alternative Modell falsifi-

²³ Black (1986), Summers (1986), Shiller (1989).

²⁴ Schleifer (2000), S. 203.

²⁵ Vgl. Beispiele 2.4–2.9 sowie Rubinstein (2001), Taleb (2004) und Referenzen dort. In dieser Arbeit gehen wir auf die Aspekte von *behavioral finance* nicht weiter ein.

²⁶ Fama (1998), S. 284.

²⁷ Vgl. Beispiele 2.10–2.11.

BEISPIEL 2.10 Kuhn (1970b), S. 79:

“Once a first paradigm through which to view nature has been found, there is no such thing as research in the absence of any paradigm. To reject one paradigm without simultaneously substituting another is to reject science itself.”

BEISPIEL 2.11 Albert (1972), S. 8–9:

„Wir haben heute allen Anlaß, an die Möglichkeit zu denken, daß keine einzige der bisher vorliegenden wissenschaftlichen Theorien im strengen Sinne des Wortes wahr, d.h. ohne jede falsche Konsequenz, ist und daß dasselbe auch für spätere Systeme gilt. Das braucht uns aber nicht davon abzuhalten, davon auszugehen, daß sie auf Grund ihres verschieden großen Wahrheitsgehalts einer adäquaten Erfassung der Realität mehr oder weniger nahekommen, auch wenn die Feststellung der unterschiedlichen Wahrheitsnähe theoretischer Aussagen und Systeme nie Gewißheit für sich beanspruchen kann.

Das bedeutet unter anderem, daß es nicht nur möglich sein muß, falsche Theorien zur Erklärung bestimmter Tatbestände zu verwenden, sondern darüber hinaus, daß auch die Kenntnis ihrer Falschheit kein Hindernis für eine solche Verwendung zu sein braucht, wenn der Wahrheitsgehalt der betreffenden Theorie für die Erklärung ausreichend ist und keine bessere Alternative zur Verfügung steht. Eine Theorie von einiger Erklärungskraft wird im allgemeinen, auch wenn sie mit bestimmten mit ihr unvereinbaren Tatbeständen nicht fertig wird, erst dann aufgegeben werden, wenn eine alternative Theorie zur Verfügung steht, die mehr für die Erklärung der in Betracht kommenden Tatbestände leistet.“

zierbar sein muss. *“Does the new model produce rejectable predictions that capture the menu of anomalies better than market efficiency?”*²⁸

Es stellt sich aber die Frage, ob die EMH selbst falsifizierbar ist. Bei einer möglichen negativen Antwort würde die EMH ernsthaft an wissenschaftlicher Bedeutung verlieren.²⁹ Solange die sogenannten Anomalien als Ausnahmen aus der existierenden Regel betrachtet werden und nicht als Grund, eine neue Regel aufzustellen, bleibt ihre Anzahl ohne Bedeutung: *“... in spite of hundreds of known anomalies we do not regard [a specific theory] as falsified (that is, eliminated) until we have a better one.”*³⁰

Laut Fama können (und sollen) die Untersuchungsverfahren geändert werden, damit die Anomalien verschwinden. Leider gilt zu dieser Aussage auch der Umkehrschluß: wenn die Anomalien durch eine geeignete Änderung der Untersuchungsverfahren verschwinden, dann können sie durch die Anwendung alternativer Verfahren wiederum entste-

²⁸ Fama (1998), S. 291.

²⁹ Frankfurter und McGoun (2000), S. 207.

³⁰ Lakatos (1970), S. 121.

BEISPIEL 2.12 Frankfurter und McGoun (2000)

“There are no facts that speak for themselves; rather, we find the facts that say whatever it is we want to say.” (S. 205).

“The third step is an obvious immunization of the EMH against falsification, but is certainly not presented as such. Essentially, Fama is saying that, if our methods throw up anomalies that we cannot eliminate through **reasonable** changes in method, then we can do one of three things: find a rational explanation for them consistent with the EMH, relegate them to a backwater of the market in which the EMH's applicability is of little or no interest (i.e., small stocks), or try a little harder and eventually penetrate the illusion. Essentially, then, nothing will ever falsify the EMH.” (S. 208).

hen. Letztendlich scheint es bei den empirischen Untersuchungen darauf anzukommen, auf welcher Seite man steht. Shefrin (2001) macht eine wichtige Beobachtung: Fama (1998) verlangt nicht mehr, dass der Preis mit dem Fundamentalwert immer übereinstimmt, sogar unter Berücksichtigung historischer Preise und öffentlich zugänglicher Informationen. Ist das ein Schritt zurück?

Die Schlussfolgerung aus all dem könnte sein, bei der Frage der Markteffizienz keine triviale Aufteilung auf schwarz und weiß vorzunehmen. Kann nicht eine Kompromisslösung gefunden werden? Die Idee alternativer Untersuchungsverfahren findet unsere volle Zustimmung. Schließlich “... *it is remarkably hard to profit from even the most extreme violations of market efficiency.*” [Roll 1994].

3. Erweiterungen der Markteffizienzhypothese

*“There are many things you can do with problems besides solving them. First you must define them, pose them. But then of course you can also **refine** them, **depose** them, or **expose** them or even **dissolve** them! A given problem may send you looking for analogies, and some of these may lead you astray, suggesting new and different problems, related or not to the original. Ends and means can get reversed. You had a goal, but the means you found didn't lead to it, so you found a new goal they did lead to. It's called play.”*

– D. Hawkins (1988), S. 44.

3.1 Theorie vs. Empirie

Als ökonomisches Konzept stellt die EMH zweifellos einen sehr starken Ansatz dar, wenngleich es einige Zusammenhänge, die empirisch beobachtet werden, nur unzureichend erklärt. Steht allein das Ökonomische in Vordergrund, lassen sich (unter Vorbehalt der Markteffizienz) schöne und einfache Gleichgewichts- oder Arbitragemodelle für die Bewertung originärer und derivativer Finanztitel und ganzer Unternehmen erstellen, Beziehungen zwischen Rendite und Risiko herleiten und Entscheidungen unter Unsicherheit begründen.

Solange die Modellannahmen nicht verletzt sind, funktionieren die meisten Modelle auch recht gut, da sie in der Regel keine internen

BEISPIEL 3.1

Die Wurzeln der *story-telling*-Funktion der Modelle gehen auf einen klassischen Artikel von Gibbard und Varian (1978) zurück: "A model . . . is a *story* with a specified structure. The *structure* is given by the logical and mathematical form of a set of postulates, the *assumptions* of the model. The structure forms an uninterpreted system, . . . Although the term 'model' is often applied to a structure alone, we shall use it in another sense. In economists' use of models, there is always an element of interpretation: the models always tells a story."

In einer alternativen Darstellung von McCloskey (1990) (S.61) erfolgt eine Unterscheidung zwischen Metaphoren und Geschichten: "When a biologist is asked to explain why the moulting glands of a crab are located just as they are he has two possibilities. Either he can call on a model – a metaphor – of rationality inside the crab, explaining that locating them just *there* will maximize the efficiency of the glands in operation; or he can tell a story, of how crabs with badly located glands will fail to survive. If he is lucky with the modeling he will discover a mathematical model with analytical solutions. If he is lucky with the storytelling he will discover a true history of some maladapted variety of crabs, showing that it is dying out. Metaphors and stories, models and histories, are the two ways of answering 'why'".

Inkonsistenzen aufweisen. Die Annahmen sind für ein Modell per Definition *exogene Größen*. Welche Kriterien werden herangezogen, um die Modellannahmen vernünftig zu formulieren? Oft sind Ergebnisse eines Modells lediglich Verfeinerungen der Annahmen durch Formeln und Koeffizienten, die eine praktische Anwendung eines Modells ermöglichen. Manchmal steckt hinter einem Modell eine Metaphor oder sogar eine ganze Geschichte (vgl. Beispiel 3.1). Dann bilden die Annahmen den Ausgangspunkt dieser Geschichte. *Welche Story steckt hinter dem Random Walk bzw. hinter der geometrischen Brownschen Bewegung als Modell der Aktienkursbewegungen?*

Robert Brown beobachtete 1828 eine irreguläre Bewegung der Pollen in Mikroskop, konnte aber nicht erklären, warum sich die Teilchen von allein bewegen. Die Antwort kam vom Albert Einstein 1905, der aber keinen Bezug auf die Brownsche Bewegung nahm, da ihm der Artikel von Brown unbekannt war. Einstein's Ergebnisse ermöglichten die Berechnung der Avogadro Zahl und eine *physikalische Erklärung* der Brownschen Bewegung, die später von Jean Perrin experimentell bestätigt wurde und ihm 1926 den Nobelpreis in Physik einbrachte. Seit 1905 ist die Brownsche Bewegung zum Prototyp von Zufallsprozessen geworden, obwohl einige ihrer grundlegenden Eigenschaften schon um 1900 von Louis Bachelier hergeleitet wurden. Bachelier hat

in seiner Ph.D.-These u.a. einen Ausdruck für die Wahrscheinlichkeit $\mathbb{P}(x, t)$ angegeben, mit der sich ein Random Walker zum Zeitpunkt t am Punkt x befindet, wenn er seine Irrfahrt in $x = 0$ und $t = 0$ beginnt. Die Gleichung für $\mathbb{P}(x, t)$ in einer Dimension ist gegeben durch¹

$$\frac{\partial \mathbb{P}(x, t)}{\partial t} = D \frac{\partial^2 \mathbb{P}(x, t)}{\partial x^2}. \quad (3.1.1)$$

Die Lösung ist durch die Dichtefunktion einer Normalverteilung

$$\mathbb{P}(x, t) = \frac{1}{\sqrt{4\pi Dt}} e^{-x^2/4Dt} \quad (3.1.2)$$

mit der mittleren quadratischen Abweichung

$$\mathbb{E}[x^2(t)] = 2Dt \quad (3.1.3)$$

darstellbar. Als diskrete Approximation der Brownschen Bewegung² ist der Random Walk eine Brücke zwischen der mikroskopischen Dynamik kleiner Atome und den makroskopischen Größen wie Diffusionskoeffizient oder Avogadro Zahl.³ Die Diffusionsgleichung (3.1.1) war schon lange als Wärmeleitungsgleichung bekannt; Bachelier's Annahme war lediglich, dass sich die Wahrscheinlichkeit auf ähnliche Weise verbreiten kann. Eine rigorose mathematische Darstellung der Brownschen Bewegung ist Norbert Wiener (1923) zu verdanken. In der Diffusionsgleichung sind keine Informationen über die Form einer typischen Teilchenlaufbahn enthalten. Die Diffusionskonstante D ist durch den Grenzwert $(\Delta x)^2 / \Delta t$ gegeben, wenn sowohl Δx als auch Δt infinitesimal klein sind. Dieser Grenzwert ist *per Definition* konstant. Daraus folgt, dass die Geschwindigkeit des Brownschen Teilchens $\Delta x / \Delta t$ unendlich groß sein muss. Da man unter Geschwindigkeit gewöhnlich die Ableitung der Laufbahn nach der Koordinaten versteht, ist diese Laufbahn unendlich ausgezackt. Wiener zeigte, dass die Trajektorie der Brownschen Bewegung fast überall kontinuierlich, aber nirgendwo differenzierbar ist.

¹ Shlesinger u.a. (1999).

² Zur Vereinfachung betrachtet man oft eine diskrete Version der Brownschen Bewegung auf dem periodischen Gitter, wobei die Schritte mit einer regulären Rate erfolgen. In der Grenzwertbetrachtung geht die Anzahl der Schritte gegen unendlich bzw. die Schrittlänge gegen Null und der Random Walk lässt sich annähernd durch die Gleichungen (3.1.1 – 3.1.3) beschreiben.

³ Metzler und Klafter (2000).

Die Grundeigenschaften der Brownsche Bewegung, die Unabhängigkeit und die Normalverteilung der Zuwächse, haben aus *ökonomischer* Sicht mehrere Vorteile. Auf dieser Basis lassen sich zahlreiche Bewertungsmodelle entwickeln, die einem breiten Anwenderkreis auch ohne tiefere mathematische Kenntnisse verständlich sind. In vielen Fällen können die Lösungen in geschlossener Form angegeben werden; ansonsten sind leistungsfähige numerische Verfahren verfügbar. Dies ermöglicht eine effektive Sensitivitätsanalyse abhängiger Variablen durch die Veränderung der Eingangsgrößen. Sowohl eine qualitative als auch eine quantitative Analyse auf Grundlage der Wiener Prozesse führt unausweichlich zur Normalverteilung der Renditen und kann mit einem relativ geringen Aufwand weiter vermittelt werden. Mit der Unterstützung der Normalverteilungshypothese von der empirischen Seite wäre das Weltbild einfach perfekt. Leider lässt sich der Abstand zwischen *wäre* und *ist* nicht verkürzen.

Die empirische Kapitalmarktforschung beschäftigt sich seit Jahrzehnten mit den Verteilungseigenschaften der Aktienkursen (oder "assets" im Allgemeinen) bzw. deren Renditen. Dies ändert aber nichts in der Tatsache, dass es immer noch keine generelle Vereinbarung dafür gibt, wie die statistische Verteilung der Renditen aussieht. Einer weiten Akzeptanz unterliegt ein Paradigma der Normalverteilung der Renditen bzw. der Lognormalverteilung der Aktienkurse. Sowohl die moderne Portfolio-Theorie als auch die Black/Scholes Methodologie der Derivatenbewertung bauen auf dieses Fundament.

Schlagen wir das Lexikon – "New Palgrave Dictionary of Money and Finance"⁴ – auf. Gerade in Bezug auf die Portfolio Analyse können wir folgendes lesen: "*Although normally distributed returns are a poor approximation of actual returns in a world of limited liability. . . the mean-variance model is by far the most widely used. . .*" Exakt sieben Zeilen später steht: "*Finally, the normality assumption appears to provide a reasonable approximation of the returns for well diversified portfolios in many cases. . .*" Man kann wohl kaum deutlicher einen Unterschied zwischen dem *Sollzustand* und dem *Istzustand* zum Ausdruck bringen.

Ausgehend von den ersten empirischen Tests der Markteffizienzhypothese (FFJR, Fama u.a. (1969)) zeigen zahlreiche empirische Untersuchungen, dass die statistischen Eigenschaften der Wertpapierrenditen im allgemeinen eine Reihe von Besonderheiten, der sogenannten

⁴ Eintrag "Portfolio Analyse", Hakansson (1992), S. 152.

Tabelle 3.1. Deskriptive Statistik der Renditeverteilungen von wichtigen Weltindizes, 1. April 1996 – 30. April 2004.

Index	# Obs.	Mittelw.	Std.	Schiefe	Kurtosis	%Quantile/Std.	
		p.a.	p.a.	tgl.	tgl.	0,5%	99,5%
Die Normalv.		–	–	0	0	–2,58	2,58
CAC 40	1982	7,39	24,36	–0,07	1,93	–2,98	3,29
DAX	2033	5,78	27,67	–0,24	2,19	–3,31	3,15
DJIA	1941	7,84	20,01	0,10	6,20	–3,38	3,25
HANG SENG	1895	0,11	29,84	0,18	9,26	–3,22	3,28
MIB 30	1982	9,21	24,26	–0,09	1,89	–3,07	2,98
MSCI WORLD	1945	4,41	16,01	–0,15	2,34	–3,44	3,09
NASD. COMP	1912	7,95	32,24	–0,04	3,09	–3,12	3,12
Nikkei 225	1923	–7,88	24,50	0,03	1,61	–3,33	2,97
S & P 500	1927	6,91	19,90	–0,16	2,61	–3,11	3,12
SMI	1976	5,81	21,73	–0,11	3,39	–3,52	3,70

stylized facts, aufweisen.⁵ Die Rendite-Verteilungen kennzeichnen sich durch starke Ränder oder *heavy tails*. Darunter versteht man mehr Verteilungsmasse in den Rändern der Verteilung und eine gewisse Spitzgipfligkeit. Dies lässt sich auch aus den Q-Q-Plots empirischer Renditeverteilungen der wichtigsten Weltindizes und der dazugehörigen Statistik ablesen (Abb. 3.1-3.2, Tab. 3.1).⁶

Zu weiteren statistischen Besonderheiten empirischer Zeitreihen von Aktienkursen/Renditen zählt u.a. die fehlende oder insignifikante lineare Autokorrelation der Renditen, wobei die Clusters in der Volatilität im Gegenteil durch eine positive Autokorrelation sehr deutlich ausgeprägt sind. Die positive Autokorrelation in den absoluten Renditen klingt langsamer als exponentiell ab, was gewöhnlich als Zeichen für das Langzeitgedächtnis interpretiert wird. Durch diese zwei Eigenschaften, *heavy tails* und Langzeitgedächtnis, wird das traditionelle Bild von unabhängigen und normalverteilten Renditen wesentlich getrübt. Gleichzeitig gewinnen die Alternativen wie beispielsweise Skalierungsmodelle immer mehr an Bedeutung.

⁵ Eine Liste der Arbeiten zu diesem Thema ist extrem lang. Eine gute Zusammenfassung geben Pagan (1996) und Cont (2001), sehr ausführlich ist z.B. Campbell u.a. (1997).

⁶ Die Statistik umfasst einen Zeitraum von ca. acht Jahren vom 2. April 1996 bis 30. April 2004.

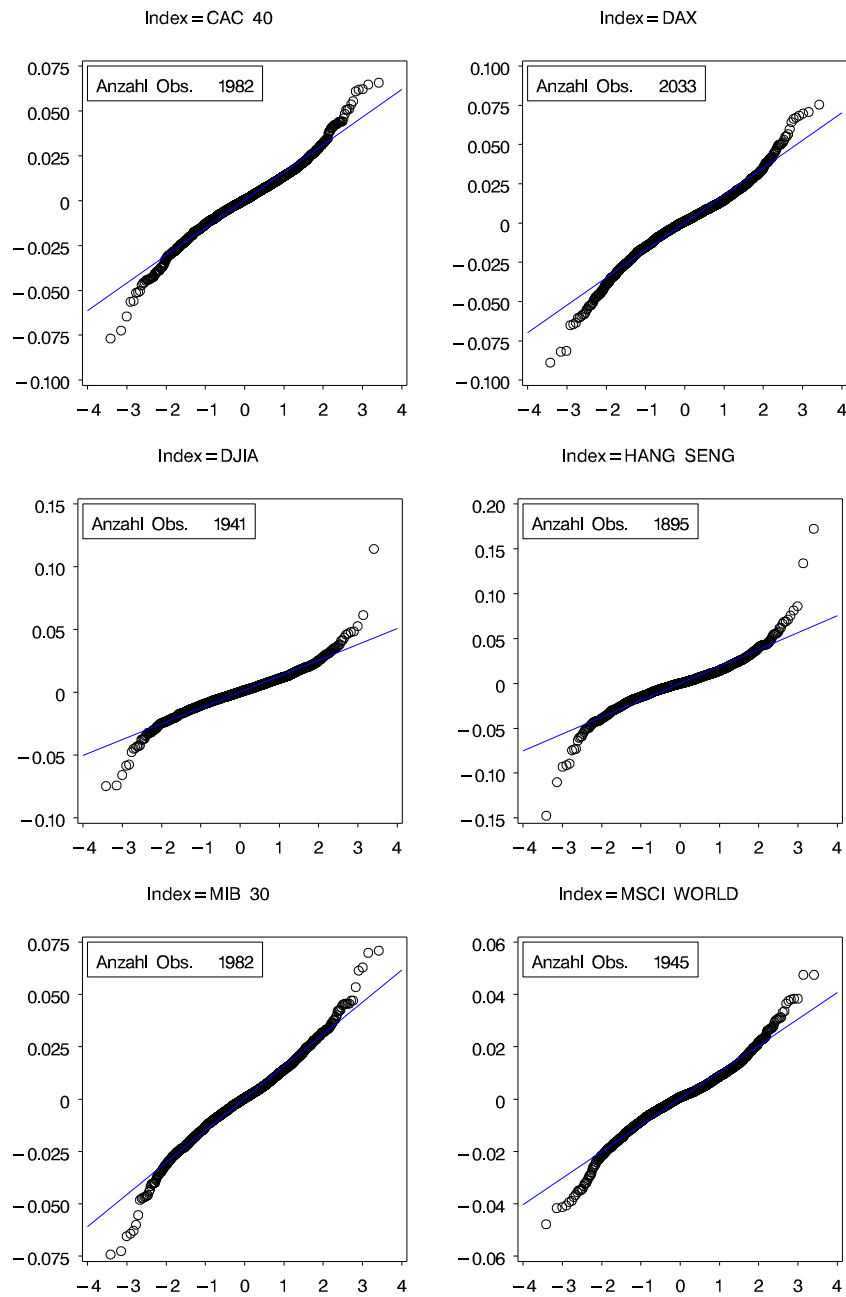


Abb. 3.1. In den Q-Q Plots empirischer Renditeverteilungen der wichtigsten Weltindizes auf täglicher Basis gegenüber der geschätzten Normalverteilung sind die fat-tails (Abweichungen links nach-unten und rechts nach-oben) deutlich zu erkennen.

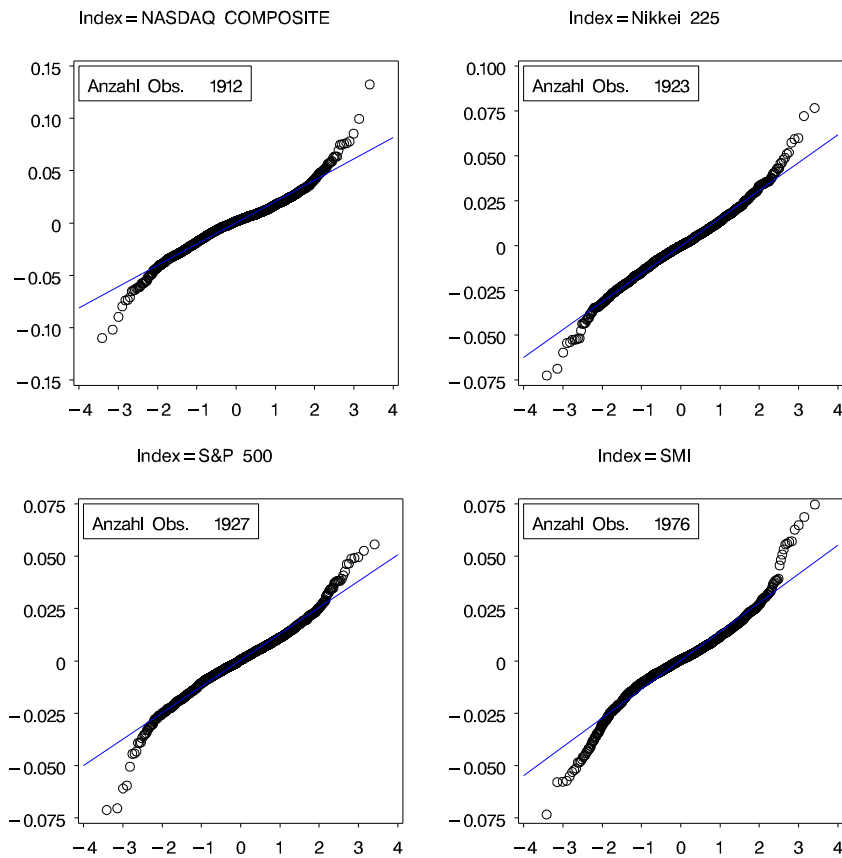


Abb. 3.2. Die Q-Q Plots empirischer Renditeverteilungen der wichtigsten Weltindizes (Fortsetzung).

3.2 Das Skalierungskonzept

Das Skalierungskonzept in seiner Anwendung auf den Finanzmarkt wurde vom Mandelbrot (1963b, 1997) initiiert. In der Literatur⁷ werden gewöhnlich zwei Arten der Skalierung unterschieden:

1. Einerseits wird ein bestimmter Zusammenhang zwischen dem Verhalten von einigen Formen der Volatilität (Varianz der Renditen, absoluter Wert der Renditen) und dem *Zeitintervall*, an dem die Renditen gemessen werden, unterstellt. Dies führt zur Bestim-

⁷ Dacorogna u.a. (2001), Matteo u.a. (2004, 2005).

BEISPIEL 3.2 Dr. Gary Hirst, Founder & Chairman, Hirst Investment Management Inc., Zitat nach Mauldin (2004), S. 210:

“I had heard about technical analysis and chart patterns, and looking at this stuff I ... was very, very skeptical that technical analysis had value. So I used the computers to check it out, and what I learned was that there was, in fact, no useful reality there. Statistically and mathematically all these tools – stochastics, RSI, chart patterns, Elliott Wave, and so on – just don't work. If you code any of these rigorously into a computer and test them they produce no statistical basis for making money; they're just wishful thinking. But I did find one thing that worked. In fact, almost all technical analysis can be reduced to this one thing, though most people don't realize it: The distributions of returns are not normal; they are skewed and have 'fat tails.' In other words, markets do produce profitable trends. Sure, I found things that work over the short term, systems that work for five or ten years but then fail miserably. Everything you made, you gave back. Over the long term, trends are where the money is.”

mung eines Skalierungsexponenten wie den Hurst-Exponenten, dessen Wert die zeitliche Abhängigkeit quantifizieren soll.

2. Darüber hinaus ist das Verhalten der Tails der Renditeverteilung als Funktion der *Größe der Bewegungen* bei einem fixierten Zeitintervall von Interesse. Dies führt zur Bestimmung eines Tail-Indexes der Verteilung als Kennzahl der Geschwindigkeit, mit der sich die Flanken der Verteilungsdichte der Abszisse nähern.

Die Frage, ob die Skalierung eine allgemeine Eigenschaft von Kapitalmarktdaten ist oder einfach ein „Artefakt“ der Daten darstellt, scheint immer noch nicht endgültig beantwortet zu sein.⁸ Es steht lediglich fest, dass eine breite Akzeptanz der Skalierung gleichzeitig eine Akzeptanz und eine Notwendigkeit einer wesentlich komplexeren Analyse sowohl in der Theorie als auch in der Praxis impliziert. Kann das gerechtfertigt werden? Lohnt es sich überhaupt?

Wenn im Laufe der Zeit immer mehr empirische Untersuchungen die Normalverteilungsannahme der Renditen widerlegen, stellt sich ganz natürlich die Frage nach einer alternativen Erklärung. Was kann der Grund dafür sein, dass viel mehr sehr kleine sowie viel mehr sehr große Werte in den Zuwächsen beobachtet werden, als man bei der Normalverteilung erwarten würde? Diese Eigenschaft (*Leptokurtosis* genannt) ist für die Normalverteilung nicht typisch. Mandelbrot

⁸ Vgl. LeBaron (2001), Lux (2001), Mandelbrot (2001d), Stanley und Plerou (2001) und Beispiel 3.2.

(1963a, 1963b) sieht die unendliche Varianz als eine mögliche Ursache. Der zentrale Grenzwertsatz ist sehr allgemein und die Tatsache, dass sich die Daten nicht daran halten, kann nur in der Verletzung der Annahmen liegen. Geht man davon aus, dass die Summanden identisch verteilt sind, scheint die Endlichkeit der Varianz die einzige Bedingung zu sein, die verletzt werden kann. Dies führt entweder zum Langzeitgedächtnis, wobei die Unendlichkeit der Varianz durch die Divergenz der Kovarianz verursacht wird, oder (unter Annahme von unabhängigen Zuwächsen) zur Klasse stabiler Verteilungen mit der Konsequenz, dass elementare Zuwächse innerhalb eines beliebig kleinen Zeitintervalls beliebig groß werden dürfen.⁹

Ein Standardmodell zur Modellierung des Langzeitgedächtnisses ist die gebrochene Brownsche Bewegung,¹⁰ wobei die Zuwächse normalverteilt, aber korreliert sind. Die Normalverteilung kann als Spezialfall aus der Klasse stabiler Verteilungen gesehen werden, die ihrerseits zu den Lévy-Verteilungen gehört (vgl. Beispiel 3.3). Sind aber die auf dem Kapitalmarkt ankommenden Informationen tatsächlich so bedeutend und unerwartet, dass sich die Aktienkurse sprunghaft ändern müssen? Oder ist die Verarbeitungsgeschwindigkeit der neu ankommenden Informationen so langsam, dass Trends entstehen?

Als ökonomischer Hintergrund für die Kontinuität in den Preisen gab Alfred Marshall zu jeder Ausgabe¹¹ seiner *“Principles of Economics”* das Motto: *“Natura not facit saltum”*. Benoît Mandelbrot¹² ist anderer Meinung: *“The only reason for assuming continuity is that many sciences tend, knowingly or not, to copy the procedures that prove successful in Newtonian physics. . . But prices are different: mechanics involves nothing comparable.”*

Welche Story steckt hinter dem Skalierungskonzept in den Wirtschaftswissenschaften? Der italienische Ingenieur, Volkswirtschaftler und Soziologe Wilfredo Pareto (1848–1923) hat schon vor über 100 Jahren beobachtet, dass die Anzahl der Menschen mit einem sehr großen Einkommen einem einfachen Potenzgesetz folgt.¹³ Ausgehend von dieser

⁹ Vgl. Mandelbrot (1963b, 1967), Fama (1963a, 1963b, 1965), Akigray und Booth (1988), Lux (1996), Mittnik und Rachev (1993, 2000), Taquq (2001), Bradley und Taquq (2003) und Beispiel 3.3.

¹⁰ Mandelbrot und van Ness (1958) sowie Kapitel 6.3 dieser Arbeit.

¹¹ Die erste Ausgabe erschien in 1890.

¹² Mandelbrot (1982), S. 335.

¹³ Vgl. Beispiel 3.4.

BEISPIEL 3.3 Mathematik der „Babuschka's“

Der zentrale Grenzwertsatz von De Moivre/Laplace besagt in seiner Endformulierung nach Paul Lévy, dass die Summe S_n von n unabhängigen und identisch verteilten Zufallsvariablen X_j ($1 \leq j \leq n$) mit endlicher Varianz bei $n \rightarrow \infty$ *asymptotisch normalverteilt* ist. Die Normalverteilung zählt ihrerseits zu einer breiteren Klasse der sogenannten *stabilen Verteilungen*.

Welche Form haben die möglichen Grenzwertgesetze der zentrierten und normierten Summen $(S_n - a_n)/b_n$ mit den Skalierungsfolgen a_n und b_n , wenn man die Anforderung nach endlicher Varianz fallen lässt? Im allgemeinen braucht der Grenzwert nicht existieren. Falls er existiert, dann stammt die einzig mögliche Grenzverteilung notwendig aus einer Klasse *stabiler Verteilungen* \mathcal{L} : für $n \geq 1$ und i.i.d. X_j mit einer Verteilung \mathcal{L} existieren a_n und $b_n > 0$ derart, dass $(S_n - a_n)/b_n$ ebenfalls die Verteilung \mathcal{L} hat. Die charakteristische Funktion einer symmetrisch stabilen Verteilung schreibt sich in ihrer einfachsten Form als $\exp\{-b|t|^\alpha\}$ mit Parametern $b > 0$ und $0 < \alpha \leq 2$. Das α heißt Index der Verteilung. Der Fall $\alpha = 2$ entspricht der Normalverteilung. Für alle $0 < \alpha < 2$ ist die Varianz einer stabilen Verteilung unendlich.

Weiter geht es wirklich wie bei den russischen „Babuschka's“: es stellt sich heraus, dass die Klasse stabiler Verteilungen ihrerseits auch zu einer breiteren Klasse der *unbeschränkt (oder unendlich) teilbaren Verteilungen* gehört. Das sind die möglichen Grenzwertgesetze von partiellen Summen von unabhängigen Zufallsvariablen, die nicht notwendig identisch verteilt, dafür aber unter bestimmten Bedingungen aufsummierbar sind. Eine Verteilung \mathcal{L} ist unbeschränkt teilbar, wenn sie für jedes n als eine n -malige Konvolution $(\mathcal{L}_n)^{\otimes n}$ einer Wahrscheinlichkeitsverteilung \mathcal{L}_n darstellbar ist. Die stabilen Verteilungen, aber auch viele andere, wie z.B. die Gamma-Verteilung, besitzen diese Eigenschaft und gehören damit zu den unbeschränkt teilbaren Verteilungen. Dabei besitzen nur die stabilen Verteilungen einen eigenen Anziehungsbe- reich („domain of attraction“): alle \mathcal{L}_n und \mathcal{L} gehören bis auf die Parameter zur selben Verteilung. Die *Lévy-Khinchin Darstellung* der charakteristischen Funktion zeigt eine unbeschränkt teilbare Verteilung als eine Verteilung des Grenzwertes der Summe von unabhängigen zusammengesetzten (compound) Poisson'schen Sprüngen unterschiedlicher Größe mit einer möglichen Gausschen Komponente.

Beobachtung lassen sich viele weitere Beispiele finden, die alle durch eine gemeinsame Idee verbunden sind: ein Effekt wird durch die Unausgewogenheit zwischen der Menge (die, bezogen auf die Gesamtmenge, immer relativ klein ist) und der Wirkung (die, bezogen auf die aktuelle Menge, immer relativ groß ist). In einem Verkaufsgeschäft sind wenige Artikel für den Großteil des Umsatzes verantwortlich. Aus 20–30 Innovationen sind fünf wertlos, fünfzehn interessant und zwei bis drei einfach genial. Zipf (1949) sah die Potenzgesetze als ein natürliches Muster in den Sozialwissenschaften.

Nach Brock (1999) ist die Aufdeckung und die darauf folgende empirische Untersuchung solcher Zusammenhänge wie Skalierungsbezie-

hungen aus mehreren Gründen nutzvoll. Dies bringt einen gewissen Grad an Disziplin in die Wissenschaft, denn die theoretischen Aussagen bleiben mit den empirischen Ergebnissen konsistent. Die empirischen Ergebnisse bilden eine natürliche Grundlage für eine spezifische angewandte Interpretation, ein Fundament für die Story. Schließlich ermöglichen sogar die einfachsten *scaling plots* zusammen mit der konventionellen statistischen Analyse einen tieferen Blick in die Eigenschaften eines möglichen zugrundeliegenden stochastischen Prozesses, der kontextabhängig auch als Datengenerator oder Unsicherheitsgenerator bezeichnet wird. *“A good example to discuss the potential usefulness of ‘scaling plots’ is asset returns. We are interested in characterization of the stochastic process of asset returns. We are also interested in the economic processes that lie behind the stochastic process generating asset returns. We are interested in this not only for out-of-sample prediction of features of the conditional distribution such as the conditional mean and conditional variance, but also for prediction of system response to ‘policy’ interventions or ‘outside shocks.’”* [Brock 1999]. Wir sind auch daran interessiert, eine der wichtigsten Fragen der empirischen Kapitalmarktforschung zu beantworten: *If asset returns do not follow random walk what are they?*

Eine der möglichen *Stories*, die die Modellierung von Aktienkursbewegungen anschaulich darstellen können, ist die Story über Fraktale. Robert Brown und die Molekularbewegung der Teilchen, Benoît Mandelbrot und (Multi) Fraktale: bestimmte Assoziationen werden mit der Zeit einfach untrennbar.¹⁴ Was ändert sich grundsätzlich beim Übergang von dem Random Walk zum Fraktal? Unter fraktalen Eigenschaften versteht man in erster Linie die Selbstähnlichkeit auf unterschiedlichen Skalen. Kann ein menschliches Auge ohne weiteres die Charts der täglichen Renditen von den der wöchentlichen unterscheiden? Wohl kaum. Mit der fraktalen Analyse hat Mandelbrot eine Theorie geschaffen, die die quantitativen Aussagen über die lokale und die globale Regularität ermöglicht. Eine Grundlage dafür bilden die Kennzahlen wie der Hölder-Exponent, die verallgemeinerte fraktale Dimension und das multifraktale Spektrum. Die multifraktalen Prozesse¹⁵ sind Verallgemeinerungen der traditionellen Modellierungswerkzeuge wie

¹⁴ Walter (2001) gibt eine historische Übersicht über die Entwicklung der empirischen Kapitalmarktforschung in Bezug auf die Skalierungseigenschaften. Siehe auch Everetsz u.a. (1999), Mandelbrot (1999c) und Clarkson (2002).

¹⁵ Kapitel 7.

BEISPIEL 3.4 Das Pareto-Gesetz

Das Pareto-Gesetz geht von einem Zusammenhang für die Wahrscheinlichkeitsverteilung in Form

$$\ln[\mathbb{P}(X > x)] = A - a \ln(x),$$

wobei a (der *Tail-Exponent*) und A Konstanten sind. Wäre X normalverteilt, würde $\ln \mathbb{P}(X > x)$ mit steigendem x *schneller* abklingen, als $\ln(x)$. Existieren dagegen positive Konstanten α und c derart, dass für $\bar{F}(x) = 1 - F(x) = \mathbb{P}\{X > x\}$

$$\lim_{x \rightarrow \infty} x^{-\alpha} \bar{F}(x) = c \quad (3.2.1)$$

gilt, gehört die Verteilung $F(x)$ zur Klasse der Pareto-Gesetze und wird als *heavy tailed distribution* bezeichnet. Dies impliziert natürlich ein unendliches Limit

$$\lim_{x \rightarrow \infty} e^{\alpha x} \bar{F}(x) \rightarrow \infty. \quad (3.2.2)$$

und deutet daraufhin, dass Ereignisse, die *relativ weit* von dem Erwartungswert liegen, zwar *relativ selten* vorkommen, aber in ihrer *Auswirkung* nicht vernachlässigt werden dürfen. Eine Zwischenstellung in dieser Klassifizierung nehmen die Verteilungen an, deren Tails schwerer als die der Normalverteilung, aber nicht so schwer, als die der Potenzgesetze sind. Dazu zählen z.B. die Lognormal- und die Weibullverteilung. Die einparametrische Version der Paretoverteilung bzw. ihre Dichtefunktion ist definiert als

$$F(x) = 1 - (1+x)^{-\alpha}, \quad f(x) = \alpha(1+x)^{-\alpha-1}, \quad x \geq 0. \quad (3.2.3)$$

Ein spezifisches Merkmal der Potenzgesetze ist, dass ihre Momente ab einer bestimmten Größenordnung unendlich werden. Tatsächlich, es ist einfach zu sehen, dass die Existenz (oder Endlichkeit) eines k -ten Momentes $\mathbb{E}[X^k]$ nur für $\alpha > k$ gesichert ist. Für $\alpha > 1$ bzw. $\alpha > 1$ existiert entsprechend der Erwartungswert bzw. das zweite Moment

$$\mathbb{E}[X] = (\alpha - 1)^{-1}, \quad (3.2.4)$$

$$\mathbb{E}[X^2] = 2[(\alpha - 1)(\alpha - 2)]^{-1}. \quad (3.2.5)$$

Im Allgemeinen kann die Paretoverteilung für ein positives endliches α *nie* alle Momente haben (und damit auch keine Laplace-Transformation):

$$\alpha = \sup\{a > 0 : \mathbb{E}|X|^a < \infty\} \quad (3.2.6)$$

der Random Walk und das Martingal mit einem zusätzlichen Freiheitsgrad: der variierenden Regularität. Dieser Freiheitsgrad erfüllt gleichzeitig zwei Aufgaben: damit können sowohl schwere Tails als auch das Langzeitgedächtnis nachgebildet werden. Das Multifraktale Modell of Asset Returns (MMAR)¹⁶ postuliert, dass die Renditedynamik

¹⁶ Calvet u.a. (1997), Fisher u.a. (1997), Mandelbrot u.a. (1997), Calvet und Fisher (2002).

durch einen zusammengesetzten Prozess beschrieben werden kann: eine (möglicherweise gebrochene) Brownsche Bewegung in der multifraktalen Zeit. Die Zeitskala ist nicht mehr deterministisch, sondern sie stellt selbst Ergebnis einer stochastischen Transformation der *Kalenderzeit* auf die *Marktzeit* dar, die einem multifraktalen Prozess folgt.

Die Idee einer Zeittransformation im Allgemeinen von *clock time* auf die *market* oder *intrinsic time* durch einen untergeordneten (*subordinated*) Prozess geht auf Mandelbrot und Taylor (1967) zurück. Sie verwendeten eine einseitige stabile Verteilung mit dem Index $\alpha/2$ (mit $\alpha < 2$), um die Verteilung der Marktzeitwüchse dt (gemessen an der Anzahl der Transaktionen in der Kalenderzeit) zu modellieren.¹⁷ Der resultierende stochastische Prozess für die Renditen in der Kalenderzeit ist eine symmetrische stabile Verteilung mit dem Index α . Einerseits ist dieses Modell theoretisch sehr attraktiv, andererseits hat es zwei wesentliche Einschränkungen: der Index α kann nicht variieren und das Langzeitgedächtnis kann wegen der Unabhängigkeitsannahme in Renditen nicht modelliert werden.

Als nächster verwendete Clark (1970, 1973) einen untergeordneten Prozess zur Modellierung einer irregulären Intensität des Wertpapierhandels in der Zeit. Anstatt eines gewöhnlichen Wiener Prozesses schlug er einen untergeordneten Wiener Prozess vor, wobei $T(t)$ ein Prozess mit nichtfallenden Pfaden beginnend bei Null ist. Die Irregularität in der Handelsintensität bzw. in der Marktzeit sollte mit dem Handelsvolumen (Anzahl der gehandelten Wertpapiere) in Zusammenhang stehen, deren kumulierte Wert die Marktzeit nachbildete.

Nach der Arbeit von Clark folgten zahlreiche weitere Modelle der stochastischen Zeittransformation mit den Verfeinerungen sowohl von der statistischen als auch von der ökonomischen Seite. Es wurden unterschiedliche Verteilungen zur Modellierung der Marktzeit vorgeschlagen und weitere Versuche unternommen, die Marktzeit ökonomisch zu interpretieren (sehr oft wird beispielsweise die Anzahl der Transaktionen als Maß der Handelsintensität verwendet).¹⁸

Mathematisch lässt sich die stochastische Zeittransformation sehr gut in die existierende Umgebung der Semimartingale einbinden. Monroe

¹⁷ Eine einseitige $\alpha/2$ -stabile Verteilung ($\alpha < 2$) hat einen Träger auf \mathbb{R}^+ .

¹⁸ Eine detaillierte Übersicht über die Modellierung mittels untergeordneten Prozesse geben Hurst, Platen und Rachev (1997). Siehe auch Rachev und Samorodnitsky (1993), Hurst u.a. (1999), Marinelli u.a. (1999) Ané und Geman (2000), Geman, Madan und Yor (2000).

(1978) hat gezeigt, dass jedes Semimartingal durch eine Brownsche Bewegung in der stochastischen Zeit darstellbar ist. D.h., es existiert eine Brownsche Bewegung $W(t)$ und eine stochastische Zeittransformation $T(t)$, wobei $T(t)$ ein steigender stochastischer Prozess ist, derart, dass

$$\ln P(t) = \ln P(0) + W(T(t)).$$

Dies impliziert, dass die Untersuchung der Preisfluktuationen auf die Untersuchung der Zeittransformation für die Brownsche Bewegung zurückgeführt werden kann.¹⁹

Eine weitere Idee für die Zeittransformation liegt in der Verwendung von Zufallssummen. Zu den Annahmen des zentralen Grenzwertsatzes gehören u.a. eine deterministische Anzahl der Summanden n : die Zeitintervalle zwischen den einzelnen Trades werden deterministisch angenommen. Ist dagegen die Anzahl der Summanden eine stochastische Größe, so wird die resultierende Verteilung der Summe keine Normalverteilung mehr, sondern eine Mischung aus mehreren Normalverteilungen. Es kann gezeigt werden, dass die Zufallsindexierung zu den höheren Kurtosis-Werten führt.²⁰ Rachev und Rüschendorf (1994) wenden das Prinzip der Zufallsindexierung zur rationalen Bewertung europäischer Optionen an.

Zusammenfassend ist die Idee der stochastischen Zeittransformation ein wichtiger qualitativer Schritt in Richtung realistischer Modellierung der Renditen von Finanzmarktdaten. Wird dieser Schritt gemacht, stellt sich allerdings die Frage: wenn ökonomische Prozesse in der Marktzeit verlaufen, wie kommt es dann zur Markteffizienz in der Kalenderzeit?

3.3 Die Fraktale Market Hypothese

Offensichtlich muss die klassische Behauptung, dass alle preisrelevanten Informationen zu jedem beliebigen Zeitpunkt in den Preisen enthalten sind, abgeschwächt werden. Der Ausgangspunkt dafür ist die Annahme, dass die Informationsverarbeitung auf unterschiedlichen Zeitskalen erfolgt. Es geht nicht darum, die These der Markteffizienz zu widerlegen, sondern darum, sie zu modifizieren.

¹⁹ Ané und Geman (2000), Geman u.a. (2000).

²⁰ Gnedenko und Korolev (1996), S. 79.

In seiner Fraktalen Market Hypothese (FMH) versteht Peters (1991, 1994) unter *fraktal* nicht die Auswirkungen der Informationsverarbeitung auf das Verhalten der Aktienkurse/Renditen, sondern das Verhalten der Marktteilnehmer selbst. Ein klassischer Repräsentativinvestor ist ein rationaler Preisnehmer; er zieht mehr Wohlstand weniger Wohlstand vor und ist immer fähig, eine korrekte Bewertung von beliebig komplexen Informationen vorzunehmen. Unter diesen Annahmen erfolgt keine Trennung nach der Art der zu verarbeitenden Informationen und nach der entsprechenden Art der Investoren. Genau das ist nach Peters die Schwachstelle der traditionellen Denkweise.

Peters (1994) schlägt vor, das Verhalten der Marktteilnehmer *in Bezug auf Investmenthorizont heterogen* zu modellieren, wobei die kurzfristig und die langfristig orientierten Investoren für die Verarbeitung der preisrelevanten Informationen ausschließlich auf der entsprechenden Zeitskala zuständig sind. Im Einzelnen gilt:

- Der Markt besteht aus einer großen Anzahl von Investoren mit unterschiedlichen Investmenthorizonten; dies ist ein stabiler Zustand des Marktes.
- Der Einfluss von Informationen auf die Wertpapierpreise lässt sich in Bezug auf den Investmenthorizont unterscheiden. Aus kurzfristiger Sicht gewinnen Informationen bezüglich der Marktstimmung und technische Indikatoren an Relevanz. Mit steigendem Investmenthorizont dominiert die Fundamentalanalyse. Mit anderen Worten: die Preisänderungen reflektieren nur Informationen über den entsprechenden Investmenthorizont.
- Die Marktstabilität ergibt sich durch eine ausreichende Liquidität (den Ausgleich von Angebot und Nachfrage). Die Liquidität entsteht infolge Transaktionen der Marktteilnehmer mit unterschiedlichen Investmenthorizonten. Erfolgt ein Ereignis, das die Gültigkeit der fundamentalen Informationen in Frage stellt, folgen die Langzeitinvestoren einer der beiden Strategien: entweder stellen sie ihre Positionen glatt und verlassen den Markt oder sie entscheiden sich, auf die Seite von kurzfristig orientierten Investoren zu wechseln und versuchen, von den kurzfristigen Informationen zu profitieren. Tun das immer mehr Marktteilnehmer, schrumpft der Investmenthorizont des Marktes zusammen, weil keine Langzeitinvestoren mehr da sind, die die Liquidität für den kurzfristigen Handel gewährleisten können. Der Markt wird instabil.

- Die Preise entstehen auf Grundlage der Interaktionen zwischen dem kurzfristigen technischen Handel und der langfristigen fundamentalen Bewertung. Die kurzfristigen Preisbewegungen neigen im Vergleich mit der langfristigen Alternative dazu, volatil zu sein und enthalten mehr Rauschen. Die langfristigen Trendbewegungen können durch die Änderungen in den erwarteten Gewinnen der börsennotierten Unternehmen verursacht werden. Die letzten sind ihrerseits Ergebnis einer veränderten ökonomischen Situation. Die kurzfristigen Trends scheinen dagegen Ergebnis von Massenverhalten (*crowd behavior*) zu sein. Weiterhin gibt es keine Anhaltspunkte für einen Zusammenhang zwischen den kurzfristigen und den langfristigen Trends.
- Die Wertpapiere von Unternehmen, die kein Bezug zu einem ökonomischen Zyklus aufweisen, sind für die langfristigen Trends nicht anfällig. In solchen Fällen dominieren der Handel, die Liquidität und die kurzfristigen Informationen.

Nach einer Vielzahl von empirischen Untersuchungen auf dem Devisenmarkt kommen Olsen & Associates²¹ zum Schluss, dass die Markteffizienz in der schwachen Form nicht gegeben ist. Sie betonen die Relevanz der Zeitskalen für die Marktteilnehmer und schlagen eine alternative Hypothese, *The Heterogeneous Market Hypothesis (HMH)* vor. Sie besteht aus zwei wichtigsten Aussagen:

- Alle Marktinformationen müssen den Entscheidungsträgern zugänglich sein.
- Es muss Marktteilnehmer mit unterschiedlichen Zeitskalen und heterogenen Erwartungen geben. Aus deren Handel zwischeneinander werden die Friktionen sowie Transaktionskosten minimal.

“There is a broad spectrum of market agents with different time horizons. These horizons range from one minute for short-term traders to several years for central bankers and corporations. The reactions of market agents to an outside event depend on the framework of his/her time horizon. Because time horizons are so different and vary by a factor of one million, economic agents take different decisions.” [Olsen 1994]

Welche Auswirkungen haben die vorgeschlagenen Erweiterungen auf die Markteffizienz? Unter der Annahme, dass die Informationsverar-

²¹ Dacorogna u.a. (2001), Olsen u.a. (1992), Müller u.a. (1993), Olsen (1994).

BEISPIEL 3.5 Olsen u.a. (1992):

“We believe the most promising approach is to differentiate the expectations according to their time dimension because we consider the *different time scales* of the market participants the key characteristic of the market.

... The broad set of market agents with different time horizons lends a *fractal* structure to the market... The interaction of components with different time scales gives rise to ... effects such as certain properties of volatility clusters, trend persistence, lag between interest rate adjustment and FX rate adjustment. The latter is a good example of what conventional theory considers an inefficiency while we see it as an effect arising from different time scales involved in the market.”

beitung auf unterschiedlichen Zeitskalen erfolgt, werden Semimartingale zu einem Spezialfall eines viel allgemeineren (multi)fraktalen Prozesses. Dies impliziert, dass die EMH eine spezielle und keine allgemeine Hypothese ist: in der Formulierung nach Fama muss sie nicht immer gelten, weil die Heterogenität der Marktteilnehmer bezüglich des Investmenthorizonts unberücksichtigt bleibt. Die Konsistenz mit den empirischen Beobachtungen kann in dieser Hinsicht nur durch Skalierungsmodelle hergestellt werden. Einige dieser Modelle werden im Folgenden detailliert betrachtet.

Der prinzipielle Unterschied zu den herkömmlichen Modellen besteht daran, dass die Regeln und die Ausnahmen umdefiniert werden müssen. Gewöhnlich lassen sich die typischen und die untypischen Ereignisse mit Hilfe statistischer Gesetze scharf trennen. In der Regel ist auch ein qualitativer Zusammenhang zwischen der Häufigkeit des Vorkommens und der Auswirkung gegeben: die seltenen Ereignisse können in ihrer Wirkung relativ zur Gesamtheit der Ereignisse vernachlässigt werden, da sie *nicht informativ* sind. Mit steigender Wirkung untypischer Ereignisse wird es immer schwieriger, ihre asymptotische Vernachlässigung zu begründen. Im Extremfall können die „Regelereignisse“ im Vergleich mit den Ausnahmen bedeutungslos werden. Sind dann die Ausnahmen immer noch nicht informativ?

Die Annahme, dass die Informationsverarbeitung auf dem Kapitalmarkt auf unterschiedlichen Zeitskalen erfolgt, impliziert, dass Ereignisse in ihrer Wirkung auf einer gegebenen Zeitskala immer beschränkt sind. Der typische Zustand lässt sich durch die Ausweitung der Wirkungen über mehrere Zeitskalen beschreiben. Untypisch dagegen ist das Schrumpfen der Heterogenität der Zeitskalen, was gleichzeitig zur Instabilität des Marktes führt.



4. Regeln und Ausnahmen I: typische Ereignisse

“For more than a century, financiers and economists have been striving to analyze risk in capital markets, to explain it, to quantify it, and, ultimately, to profit from it. I believe that most of the theorists have been going down the wrong track. The odds of financial ruin in a free, global-market economy have been grossly underestimated.”

– Mandelbrot und Hudson (2004), S. 4.

4.1 Entropie als inverses Informationsmaß

Motivation und historischer Überblick

Der Grundgedanke hinter der Anwendung der Entropie als Informationsmaß kann als alternative (zu den Grenzwertsätzen) Charakterisierung der Zusammenhänge zwischen Theorie und Empirie gesehen werden. Ähnlich wie bei den Grenzwertsätzen versucht man damit, eine Grundlage für die Vergabe der Wahrscheinlichkeiten an die Ereignisse unter Berücksichtigung der Erfahrungswerte über die vergangene Entwicklung zu schaffen. Historisch gab es diesbezüglich zwei grundsätzliche Fragestellungen. Zum einen entwickelten schon im achtzehnten Jahrhundert James Bernoulli¹, später auch Pierre Simon de Laplace (1814) und Thomas Bayes (1763) die ersten Ansätze dafür, wie die empirisch messbaren oder gemessenen „objektiven“

¹ James Bernoulli wird im Deutschen auch Jacob genannt. Sein Buch *“Ars Conjectandi”* (*“Principle of Insufficient Reason”*) wurde im Jahre 1713, acht Jahre nach seinem Tod, herausgegeben.

Größen in Form von Erwartungswerten (oder der likelihood Funktion) zur Aktualisierung „subjektiver Wahrscheinlichkeiten“ sinnvoll eingesetzt werden können.² Der Anfangs- oder *a-priori*-Zustand beinhaltet in dieser Darstellung die höchste Unsicherheit. Das berühmte Bernoulli-Prinzip lautet: *“If given an enumeration of mutually exclusive possibilities X_j and no further information, then assign the same probability to all.”* Das Wissen aus dem Experiment wird in das Modell mittels Aktualisierung der Wahrscheinlichkeiten nach und nach integriert. Dies ist der Ursprung für das Verfahren *Bayes’scher Inferenz*. Auf dieser Basis entwickelte Jeffreys (1961) weitere Ansätze zur Charakterisierung der Eigenschaften (wie z.B. einer Wahrscheinlichkeitsverteilung) eines empirischen Systems unter unvollständiger Information.

Zum anderen stellt sich dadurch, dass gewisse globale Eigenschaften *exogen (theoretisch, von einem Modell)* vorgegeben sind, ein inverses Problem: wie beschreibt man die unbekannt restlichen Eigenschaften? Diesem Gedanken folgten die Arbeiten von James Clerk Maxwell, Ludwig Boltzmann und Josiah Willard Gibbs im späteren 19-ten Jahrhundert. Den entscheidenden Durchbruch schaffte erst Claude Shannon (1948) mit seiner Arbeit über mathematische Verfahren zur statistischen Modellierung aus der Sicht der *Informationstheorie* sowie über eine effiziente Kodierung und Übertragung der Nachrichten. Mit dem von Jaynes (1957) entwickelten Maximum-Entropie-Verfahren hat die Philosophie des Schlussfolgerns auf Basis unvollständiger Information einen weiteren Impuls bekommen. Damit scheint es gelungen zu sein, eine mathematische Beziehung zwischen Entropie und Information in der Form eines Optimierungsproblems unter Nebenbedingungen darzustellen.

Aus physikalischer Sicht beschreibt Entropie die Beziehung zwischen den Mikro- und Makrozuständen in der Natur und bestimmt das Verhalten von makroskopischen Systemen im Gleichgewicht (oder nahe zum Gleichgewicht). Das Wort selbst entstammt dem griechischen $\tau\rho\omicron\pi\eta$ (Transformation) und soll vom deutschen Physiker Rudolf Clausius im 1865 ins Leben gerufen sein. Das “en” vorne diene vermutlich der Ähnlichkeit mit “energy”.

² Wir möchten hier die Aufteilung auf „objektiv“ und „subjektiv“ nicht weiter präzisieren und lassen deshalb die Anführungszeichen stehen. Vor allem soll ein möglicher subjektiver Einfluss bei der Datenauswahl in empirischen Studien außer Betracht bleiben.

Eine große Entdeckung vom Ludwig Boltzmann im Jahre 1877 war eine Formel für die physikalische oder thermodynamische Entropie $S = k \ln W$. Diese Formel stellte einen Zusammenhang zwischen einer Zustandsvariable oder Entropie S und dem Grad der Unordnung, gemessen an der Anzahl der Mikrozustände (“elementary complexions”) W , in einem physikalischen System dar (das k ist eine Konstante).

Josiah Willard Gibbs, ein Professor der mathematischen Physik am Yale College, begründete eine neue Forschungsrichtung in der Physik namens statistische Mechanik, um das Verhalten einer Substanz durch Ordnung und Unordnung der kleinsten Bestandteile, der Atome und Moleküle, zu beschreiben. Im letzten Drittel des neunzehnten Jahrhundert leitete er eine andere Entropie-Formel her, die sich zu $k \ln W$ reduziert, falls alle Mikrozustände im System gleichwahrscheinlich sind.

Bei physikalischen Systemen ist die Interpretation der Entropie als inverses Maß der Entfernung vom Gleichgewicht ein Standard. Nach dem Zweiten Hauptsatz der Thermodynamik strebt ein abgeschlossenes thermodynamisches System irreversibel zum Gleichgewicht, indem es sich so verhält, dass die Entropie der Mikrozustände steigt oder konstant bleibt. Im Gleichgewicht unterscheiden sich die Mikrozustände am wenigsten voneinander, im Idealfall sind sie gleichwahrscheinlich und die Entropie nimmt ihren maximalen Wert an. Das Maximum-Entropie Verfahren ermöglicht damit eine Schlussfolgerung auf eine *typische* Konfiguration (das Gleichgewicht) der Mikrozustände aus der „Vogelperspektive“, das heißt, auf Grundlage einiger bekannter oder gemessener globaler Größen.

In der Kommunikationstheorie ist die Ausgangssituation etwas anders. Es kann wohl kaum die Rede vom Gleichgewicht sein. Da jede Nachricht in der Praxis eine endliche Länge besitzt, kann die Entropie als inverses Informationsmaß nicht beliebig groß werden und muss sogar in Folge der Kommunikation abnehmen.³ Trotzdem versucht man auch hier eine Art *typische* Konfiguration zu ermitteln, indem eine

³ In der Kommunikationstheorie kann zwischen Entropie der Quelle und Entropie der Nachricht unterschieden werden; die Summe von beiden ist konstant und endlich. Die Abnahme der Entropie bezieht sich in diesem Kontext auf die Entropie der Quelle. Im Laufe der Kommunikation wird die *Unkenntnis* über die Quelle nach und nach abgebaut. Aus physikalischer Sicht ist das ein offensichtlicher Widerspruch und Grund genug, die Shannon-Entropie (siehe später in diesem Kapitel) manchmal als *unphysikalisch* zu bezeichnen (Vgl. Collier (1986), S. 18.).

möglichst kosteneffiziente und fehlerfreie Übertragung der Information zwischen der Quelle und dem Empfänger zum Zweck der Kommunikation erklärt wird. Zur Reduzierung der Übertragungskosten werden die Nachrichten komprimiert: mittels eines Kodierungsverfahrens bringt man sie in eine kompakte Form. Die Anforderung nach der Fehlerfreiheit erzielt dabei eine entgegengesetzte Wirkung: die Fehlerkorrektur bedeutet eine zusätzliche Redundanz der Daten und/oder eine steigende Datenmenge durch das Einbinden der Kontrollmechanismen. Das Vorhandensein dieser zwei gegeneinander gerichteten Anforderungen ermöglicht es, das Kommunikationsproblem auf ein Optimierungsproblem unter Nebenbedingungen zurückzuführen.

Das Ökonomische. Teil I

Eine allgemeine Interpretation der Entropie als Informationsmaß aus *finanzwirtschaftlicher* Sicht ist im Allgemeinen recht schwierig. Eine Ökonomie ist ein offenes System. Die Informationsverarbeitung auf dem Kapitalmarkt hört nie auf. Im ökonomischen Kontext scheint anstatt der Informationsmenge eine andere Quantität an Bedeutung zu gewinnen: die Informationsrate oder die Menge an Information pro Zeiteinheit. Der ökonomische Gleichgewichtszustand wird gewöhnlich mit einer effizienten Informationsverarbeitung assoziiert. Im einfachsten Fall impliziert das eine endliche Menge an preisrelevanten Informationen, die schnell genug verarbeitet wird, um keine Fehlbewertungen zuzulassen. Im Zuge der Informationsverarbeitung befinden sich die Marktteilnehmer permanent in einem Zustand mit maximaler restlicher Unsicherheit bezüglich der zukünftigen Entwicklung.

Nun hat man leider überhaupt keine Vorstellung davon, wie hoch die Gesamtunsicherheit der Zukunft ist. Dies bringt einen bei der Beurteilung, wie weit die restliche Unsicherheit von ihrem Maximum entfernt liegt, nicht viel weiter. Häufig wird die Abhilfe durch die Einschränkung der Bandbreite aller möglichen Realisationen auf einen *a-priori* festgelegten Bereich geschaffen. Bei der Normalverteilungsannahme der Renditen liegen beispielsweise die Rahmen des Möglichen *im wesentlichen* bei plus minus 3σ . Dies impliziert natürlich, dass es keine Informationen geben soll, die einen Renditeausschlag darüber hinaus bewirken können. Durch eine solche Festlegung lässt sich die Informationsverarbeitung auf dem Kapitalmarkt mit Werkzeugen der Kommunikationstheorie modellieren.

Als möglicher Nachteil der Entropie-basierten Verfahren soll das Fehlen der *semantischen Informationskomponente* (wie Informationen zu interpretieren sind?) erwähnt werden.⁴ Dies lässt sich übrigens ganz gut mit einer klassischen Annahme homogener Erwartungen der Marktteilnehmer vereinbaren, natürlich, solange eine solche Annahme nicht kritisch ist. Weitere Komplikationen können entstehen, wenn man annimmt, dass die Ankunfts- und die Verarbeitungsraten der Informationen unterschiedlich sind. Das gilt insbesondere dann, wenn die Verarbeitungsrate maximal, aber trotzdem kleiner als die Ankunftsrate ist. Im Übrigen bietet sich die Shannon-Entropie als Brücke zwischen der Informations- und der Wahrscheinlichkeitstheorie sehr natürlich an. Die Zielsetzung „Entscheidung auf Basis unvollständiger Informationen“ ist und bleibt von höchster Relevanz in vielen Gebieten der Wissenschaft. In dieser Arbeit konzentrieren wir uns auf die Anwendungen im Finance.

Einige Bits zum Begriff „Information“

Wie schon angedeutet, ist die Idee, durch Entropie die Unkenntnis über ein *physikalisches* System zu messen, relativ alt und geht auf Boltzmann zurück. Im Gegenteil dazu nimmt die *mathematische* Informationstheorie ihren Ursprung aus Modellen, die erst vor einem halben Jahrhundert zur elektronischen Kommunikation entwickelt worden sind.⁵

Um die Begriffe Information und Entropie einzuführen verwenden wir eine ganz einfache Wahrscheinlichkeitsumgebung bestehend aus einer Wahrscheinlichkeitsverteilung μ auf einer endlichen Menge \mathcal{X} und einer Zufallsvariable X mit Werten in \mathcal{X} . Damit ist $p(x) = \mathbb{P}(X = x) = \mu(\{x\})$ die Wahrscheinlichkeit, dass Ereignis $\{x\}$ eintritt.⁶ Ein klassisches Beispiel zu Shannon's fundamentalem Konzept der Information ist die Modellierung einer Folge von 2^n JA/NEIN Antworten auf n unabhängige Fragen. Dabei könnte z.B. ein Ereignis {Die Antwort ist JA}

⁴ Entropie ist rein *syntaktischer Natur*: in diesem Konzept wird nicht behauptet, dass eine Nachricht mit höherem Informationsgehalt ebenfalls aussagekräftiger oder bedeutsamer wäre. Paradox: ein weißer Fernsehschirm bei Störung ist informativer, als einer, der einwandfreie Bilder zeigt. . .

⁵ Shannon und Weaver (1949).

⁶ Da der Wertebereich \mathcal{X} endlich ist, ist die Zufallsvariable X durch die Ereignisse $\{X = x\}$ mit $x \in \mathcal{X}$ eindeutig bestimmt.

durch $\{X = 1\}$ quantifiziert werden. Jeder Endpunkt in einem binären pfadunabhängigen Baum, der $X = 2^n$ Ereignisse (=Antworten) abbildet, lässt sich in höchstens $n = \log_2 X$ Schritten (=Fragen) erreichen. Da die Fragen unabhängig sind, bietet sich ganz intuitiv eine einfache Kennzahl zur Messung des Informationsgehaltes an – nämlich die mittlere Anzahl der Fragen, die gestellt werden müssen, um die erforderliche Information zu erhalten (oder den Systemzustand mit Sicherheit zu beschreiben).

Hartley (1928) definiert die Einheit und das Maß der Information mit folgendem Satz: “*The answer to a question that can assume the two values ‘yes’ or ‘no’ (without taking into account the meaning of the question) contains one unit of information.*” Eine Menge $X = \{x_1, \dots, x_n\}$ mit binären Elementen (JA=1, NEIN=0) besitzt damit einen Informationsgehalt von

$$I(X) = -\log_2 p = \log_2 n \quad (\text{bit}). \quad (4.1.1)$$

Dieser Ausdruck für *Selbstinformation* ist als *Hartley-Formel* bekannt und gilt unter der Annahme, dass alle x_i aus X mit gleicher Wahrscheinlichkeit $p(x_i) = p = 1/n$, $i = 1, \dots, n$ auftreten. Cloud Shannon geht noch einen Schritt weiter und fragt: was ändert sich, wenn die Wahrscheinlichkeiten $p(x_i)$ nicht alle gleich sind? Es liegt nahe, den erwarteten Informationsgehalt zu berechnen.

Entropie nach Shannon

Definition 4.1.1. Sei X eine endliche oder abzählbare Menge, ε_x ein Dirakmaß (ein Punktmaß) in $x \in X$ und $\mu = \sum_x p(x) \varepsilon_x$ eine Wahrscheinlichkeitsverteilung. Die **Entropie** von μ ist definiert als

$$\mathbb{H}(\mu) \equiv -\sum_x p(x) \log p(x) \quad \text{mit} \quad 0 \log 0 \equiv 0 \quad (4.1.2)$$

oder alternativ

$$\mathbb{H}(X) = \mathbb{E}[-\log p(X)] = \mathbb{E}[I(X)]. \quad (4.1.3)$$

Bemerkung. Man spricht von der Entropie $\mathbb{H}(X)$, falls X eine X -wertige Zufallsvariable und μ ihre Verteilung ist. Mit dem Konzept der Entropie verfolgt man dasselbe Ziel, wie bei der Definition einer Zufallsvariable, nämlich die Abstraktion von den Zufallsereignissen. In Anlehnung auf

die weiterführende Literatur⁷ bezeichnen wir mit Entropie ein Funktional, das *Wahrscheinlichkeitsverteilungen* oder, im wesentlichen äquivalent, *Zufallsvariablen* auf reelle Zahlen abbildet.

In der Sprache der Informationstheorie bezeichnet man die Menge \mathcal{X} als *Alphabet* und die Elemente x_i als *Symbole*. Die Symbole sind üblicherweise Zahlen oder Buchstaben. Auf einer höheren Abstraktionsstufe werden sie zu den Realisationen der Zufallsvariablen X auf der Ereignismenge \mathcal{X} . Die Folgen aus Symbolen bilden ein *Wörterbuch* und heißen *Wörter*, *Nachrichten*, *Sequenzen* oder einfach *Daten*. Diese Nachrichten werden von einer *Quelle* oder einem *Sender* erzeugt und durch einen *Kanal* zum *Empfänger* übertragen.

Nach der Formel (4.1.3) entspricht die Entropie einem mittleren Informationsgehalt der Quelle: je höher die Entropie, desto höher ist die Unkenntnis über die Quelle bzw. desto mehr Informationen *fehlen* einem, um die Informationsquelle vollständig zu beschreiben.⁸ In dieser Hinsicht ist Entropie ein *inverses* Maß der Information. Das Rationale daran ist, dass die Entropie einer Nachricht fixierter Länge mit dem höchsten Informationsgehalt ihr Maximum annehmen muss. Aus (4.1.3) folgt:

- i) $\mathbb{H}(X)$ ist eine *nicht-negative konkave* Funktion;
- ii) $\mathbb{H}(X) \leq \log |\mathcal{X}| = \mathbb{H}_{\max}$; $\mathbb{H}(X) = \mathbb{H}_{\max}$, falls alle $p(x_i)$ gleich sind.
- iii) $\mathbb{H}(X) \geq 0 = \mathbb{H}_{\min}$; $\mathbb{H}(X) = \mathbb{H}_{\min}$, falls alle p_i bis auf ein p_k gleich Null sind. Nur in diesem letzten Fall ist die Unsicherheit über den Ausgang des Ereignisses $X = x_i$ minimal, da es fast sicher x_k sein wird.

Bei seiner Definition der Entropie als *Informationsmaß* ging Shannon von den folgenden Überlegungen aus:

- i) kleine Änderungen der Wahrscheinlichkeiten sollen vergleichsweise kleine Änderungen der Entropie verursachen (*Stetigkeit*);
- ii) sind alle Wahrscheinlichkeiten gleich, $p(x) = 1/n$, dann soll die Entropie mit der Anzahl der Zustände n monoton wachsen, wobei die Ereignisse mit $p(x) = 0$ nichts beitragen (*Monotonie*);

⁷ Vgl. Grünwald und Vitanyi (2003), S. 9.

⁸ Brillouin (1962), S. 160.

- iii) die Entropie soll nicht davon abhängen, auf welche Weise eine Auswahl in zwei aufeinanderfolgende Wahlvorgänge aufgeteilt wird. Die Zählreihenfolge soll also nicht relevant sein (*Additivität*).

Die Formel (4.1.2) kann als Definition, aber auch als Ergebnis der obengenannten Anforderungen betrachtet werden.

Satz 4.1.1 (Shannon (1948), Theorem 2) *Nur die Entropie-Formel*

$$\mathbb{H} = -k \sum_x p(x) \log p(x) \quad \text{mit } k = \text{const}$$

kann allen drei obigen Anforderungen gleichzeitig gerecht werden.

Shannon erweiterte die Hartley-Formel (4.1.1) auf die Mengen, die nicht gleichwahrscheinliche Ereignisse beschreiben. Er folgte dabei dem Grundgedanken, die Gleichverteilung als eine a-priori Verteilung von unabhängigen Zufallsvariablen zu verwenden. Sind alle Zustände gleichwahrscheinlich, erreicht der Entropiewert in (4.1.2) sein Maximum, was im Einklang mit dem Bernoulli-Prinzip ist. Jede Änderung der a-priori Wahrscheinlichkeiten würde die Existenz zusätzlicher Information über die Quelle bedeuten und folglich einen Entropierückgang verursachen.

Nun kann man vermuten, dass die Effizienz des Übertragungssystems an der Geschwindigkeit, mit der diese Informationsasymmetrie zwischen Quelle und Empfänger abgebaut wird, sowie an der Resistenz der Nachricht gegen Störungen gemessen werden kann. Diese Vermutungen hat Shannon (1948) rigoros bestätigt. Das Zusammenspiel von einzelnen Bestandteilen soll durch die Einführung einiger weiteren Kennzahlen verdeutlicht werden.

Weitere Kennzahlen und Eigenschaften der Entropie

Definition 4.1.2. *i) Gemeinsame (joint) Entropie: Die Entropie eines Paares von diskreten Zufallsvariablen (X, Y) mit der Verteilung $\mathbb{P}(X = x, Y = y) = p(x, y)$ ist*

$$\mathbb{H}(X, Y) = \sum_{x, y} p(x, y) \log \frac{1}{p(x, y)}. \quad (4.1.4)$$

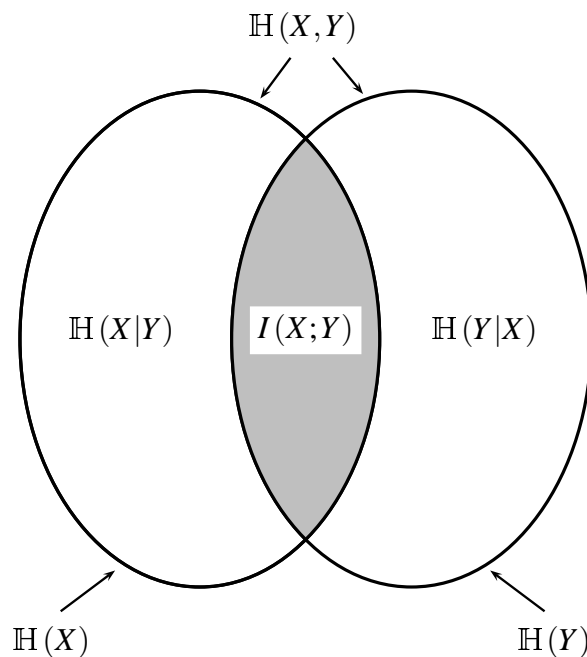


Abb. 4.1. Die Kennzahlen rund um die Entropie.

ii) **Bedingte Entropie:** zwei diskrete Zufallsvariablen (X, Y) können als Eingangs- bzw. Ausgangssignal eines Übertragungskanals betrachtet werden. Durch Beobachtung von X ändert sich $\mathbb{H}(Y)$ auf $\mathbb{H}(Y|X)$:

$$\begin{aligned} \mathbb{H}(Y|X) &= \sum_{x,y} p(x,y) \log \frac{1}{p(y|x)} = \sum_x p(x) \sum_y p(y|x) \log \frac{1}{p(y|x)} \\ &= \sum_x p(x) \mathbb{H}(Y|X=x). \end{aligned} \quad (4.1.5)$$

iii) Der Wert dieser Änderung wird durch **wechselseitige Information** zwischen X und Y erfasst:

$$I(X; Y) = \sum_x \sum_y p(x,y) \log \frac{p(x,y)}{p(x)q(y)} = D[p(x,y) \| p(x)q(y)], \quad (4.1.6)$$

iv) wobei die **relative Entropie** oder **Kullback-Leibler Distanz** $\mathbb{D}[v \| \mu]$ von $v = q(x)$ bzgl. $\mu = p(x)$ definiert ist als

$$\mathbb{D}[v|\mu] = \sum_x q(x) \log \frac{q(x)}{p(x)}. \quad (4.1.7)$$

v) Der maximale Wert wechselseitiger Information zwischen X und Y über die Verteilung $p(x,y)$ heißt **Kanalkapazität** und hat in der Kommunikationstheorie eine besondere Bezeichnung

$$C = \max_{p(x,y)} I(X; Y). \quad (4.1.8)$$

Es kann gezeigt werden⁹, dass die folgenden Aussagen gelten:

- i) $\mathbb{H}(X|X) = 0$, $\mathbb{H}(Y|X) = \mathbb{H}(Y) \iff X$ und Y sind unabhängig.
- ii) Die Kettenregel: $\mathbb{H}(X, Y) = \mathbb{H}(X) + \mathbb{H}(Y|X) \leq \mathbb{H}(X) + \mathbb{H}(Y)$.
- iii) $I(X; Y) = I(Y; X) \geq 0$ ist der mittlere Informationsgehalt über X aus Realisierung von Y und umgekehrt. Es gilt:
 - a) $I(X; Y) = \mathbb{H}(X) - \mathbb{H}(X|Y) = \mathbb{H}(Y) - \mathbb{H}(Y|X) = \mathbb{H}(X) + \mathbb{H}(Y) - \mathbb{H}(X, Y)$.
 - b) $I(X; X) = \mathbb{H}(X) - \mathbb{H}(X|X) = \mathbb{H}(X)$.
 - c) $I(X; Y)$ ist skaleninvariant: $I(cX; cY) = I(X; Y)$ mit $c = \text{const} > 0$.
- iv) $\mathbb{D}[v|\mu] \geq 0$, $\mathbb{D}[v|\mu] = 0 \iff v \equiv \mu$.

Die logarithmische Asymptotik

Betrachten wir einen Vektor $X^n = (X_1, \dots, X_n)$ von unabhängigen diskreten Zufallsvariablen X_i mit einer gemeinsamen Verteilung $\mathbb{P} := \mu^{\otimes n}$. Der Bildbereich oder die Spannweite eines solchen Vektors besteht aus $|\mathbb{A}|^n$ Elementen, wo \mathbb{A} den Bildbereich jedes der X_i bezeichnet. Im Kontext der Kommunikationstheorie ist es hilfreich, an $\Omega_n := \mathbb{A}^n$ als Menge aller Wörter der Länge n auf einem endlichen Alphabet $\mathbb{A} = \{x_1, \dots, x_r\}$ mit $|\mathbb{A}| = r$ Buchstaben zu denken. Die Sprache, die durch dieses Alphabet erzeugt wird, kann als eine Wortquelle bezeichnet werden. Zunächst soll das Auftreten eines bestimmten Buchstabens als ein Zufallsereignis modelliert werden, wobei die möglichen Korrelationen zur Vereinfachung außer Acht bleiben. Unter diesen Annahmen wird jedes Wort zu einer Zeichenfolge aus Realisationen von

⁹ Vgl. z.B. Cover und Thomas (1991).

unabhängig und identisch verteilten Zufallsvariablen. Die Wahrscheinlichkeit eines Textes $T = (x_1, x_2, \dots, x_n)$ mit jedem x_i als Buchstabe ist dann einfach $p_1^{k_1} p_2^{k_2} \dots p_r^{k_r}$, wobei k_j die Anzahl der Buchstaben x_j im Text T ist und $p_j = p(x_j)$ dessen Frequenz bezeichnet. Man kann vermuten, dass die Zahl $|A|^n$, die sozusagen für den Umfang der Sprache zuständig ist, relativ schnell mit n wachsen wird. Shannon zeigte, dass dies nicht der Fall ist.

Um das zu sehen, konzentrieren wir uns auf solche Untermengen A_ε^n , die mit einer großen Wahrscheinlichkeit $\mathbb{P}(A_\varepsilon^n) \sim 1$ vorkommen, aber im Vergleich zu $|A|^n = r^n$ für große n relativ klein sind. Es scheint sinnvoll, solche Mengen als *typische Mengen* zu bezeichnen. Die Eigenschaften typischer Mengen lassen sich aus der Asymptotic Equipartition Property (AEP)¹⁰ ableiten. Die AEP ergibt sich unmittelbar aus dem schwachen Gesetz der großen Zahlen und lautet in kompakter Form: „Fast alles ist fast gleichwahrscheinlich.“

Satz 4.1.2 (Shannon-McMillan-Breiman) *Es sei $X^n = (X_1, \dots, X_n)$ ein Vektor von i.i.d. Zufallsvariablen X_i mit Verteilung $\mu_i = \mu$ auf A . Dann gilt*

$$-\frac{1}{n} \mathbb{P}(X^n) \rightarrow \mathbb{H}(X_1) = - \sum_{x \in A} p(x) \log p(x) \quad \text{in Wahrscheinlichkeit} \quad (4.1.9)$$

Beweis: Da die X_i i.i.d. sind, ist es auch $\log \mu_i$. Das schwache Gesetz der großen Zahlen besagt, dass

$$-\frac{1}{n} \mathbb{P}(X^n) = -\frac{1}{n} \sum_{i=1}^n \log \mu_i \rightarrow -\mathbb{E}[\log \mu] = \mathbb{H}(X_1).$$

□

Bemerkung. Nach dem starken Gesetz der großen Zahlen hält (4.1.9) für $n \rightarrow \infty$ \mathbb{P} -fast sicher.¹¹

Für n groß genug haben also die beobachteten Folgen eine Wahrscheinlichkeit von der Größenordnung

$$p(x_1, \dots, x_n) \approx 2^{-n\mathbb{H}(X_1)}$$

¹⁰ Diese Eigenschaft ist auch als der Ergodische Satz der Informationstheorie bzw. als der Entropiesatz bekannt, vgl. McMillan (1953), Breiman (1957), Gray (1988), Petersen (1989), Cover und Thomas (1991).

¹¹ Breiman (1957), Shields (1998).

und sind damit im wesentlichen gleichwahrscheinlich. Mit anderen Worten, die Anzahl der Folgen der Länge n , die am wahrscheinlichsten zu beobachten sind, ist kleiner als $|\mathbb{A}|^n$ und liegt in der Größenordnung $2^{n\mathbb{H}(X_1)} \leq 2^{n\log|\mathbb{A}|}$. Deshalb spricht man von *typischen Folgen* oder *Sequenzen*.

Definition 4.1.3. Es seien X_1, \dots, X_n i.i.d. Zufallsvariablen mit einer gemeinsamen Verteilung $p(x)$. Die **typische Folge** \mathbb{A}_ε^n ist repräsentativ für eine Menge der Folgen $X^n = (X_1, \dots, X_n)$ mit der Eigenschaft:

$$\mathbb{A}_\varepsilon^n := \left\{ \left| -\frac{1}{n} \mathbb{P}(X^n) - \mathbb{H}(X_1) \right| < \varepsilon \right\} \quad (4.1.10)$$

und für n groß genug gilt:

$$\mathbb{P}(\mathbb{A}_\varepsilon^n) \geq 1 - \varepsilon, \quad (1 - \varepsilon)2^{n(\mathbb{H}(X_1) - \varepsilon)} \leq |\mathbb{A}_\varepsilon^n| \leq 2^{n(\mathbb{H}(X_1) + \varepsilon)}.$$

Zusammengefasst besagt die AEP, dass:

- i) Zeichenfolgen, die nicht zu einer Menge typischer Folgen \mathbb{A}_ε^n gehören, unwahrscheinlich sind;
- ii) Alle Wörter in \mathbb{A}_ε^n vergleichbare Wahrscheinlichkeiten haben;
- iii) Die *Entropierate* einer i.i.d. Quelle gleich der Entropie pro Zeichen ist:

$$\lim_{n \rightarrow \infty} \frac{1}{n} \mathbb{H}(X^n) = \lim_{n \rightarrow \infty} \frac{n \mathbb{H}(X_1)}{n} = \mathbb{H}(X_1). \quad (4.1.11)$$

- iv) Für eine andere Verteilung als die Gleichverteilung die Mengen \mathbb{A}_ε^n klein im Vergleich mit der Anzahl aller möglichen Wörter der Länge n sind:

$$\frac{|\mathbb{A}_\varepsilon^n|}{|\mathbb{A}^n|} \approx 2^{-n(\log r - \mathbb{H}(X_1))} \quad \text{mit } r = |\mathbb{A}|.$$

Bemerkung. Der Ausdruck $\log r - \mathbb{H}(X_1)$ kann auf die relative Entropie $\log r - \mathbb{H}(X_1) = \mathbb{H}(\mathbb{P}|\mathbb{U})$ bezüglich einer Gleichverteilung \mathbb{U} zurückgeführt werden.

Interpretation der Kennzahlen

All diese obigen Ausführungen lassen sich aus der Sicht der Kommunikationstheorie exakt und eindeutig interpretieren. Ist die Übertragung rauschfrei, kann das Maximum der Übertragungsrate (=Kanalkapazität) durch eine geschickte *Kodierung der Quelle* erreicht werden indem die Redundanz der Quelle eliminiert wird. Dies impliziert, dass sich die Quelle in einem Zustand mit Maximum Entropie befindet.

In einem gestörten Kanal ist dagegen eine fehlerfreie Übertragung nicht mehr selbstverständlich. Entzieht man der Nachricht sämtliche Redundanz, wird sie für die Störung anfällig.¹² Bleibt die Redundanz hoch, verbessert sich zwar die Resistenz gegen die Störung, dies geht aber auf Kosten der Übertragungsrate. Zum Teil kann eine hohe Redundanz durch effiziente Kodierung neutralisiert werden. Da aber jede Kodierung nachweislich nur endlich „gut“ sein kann, muss man mit einer Obergrenze der Übertragungsrate rechnen, bei der die durch die Störung verursachten Fehler noch identifiziert und in bestimmten Fällen auch korrigiert werden können. Das Verfahren, bei dem der Nachricht künstlich zusätzliche Redundanz hinzugefügt wird, heißt *Kanalkodierung* im Unterschied zur Quellenkodierung in einem rauschfreien Fall. Das bekannteste Beispiel dafür ist die Versendung von einem oder mehreren Kontrollbits als Teil der Nachricht.

Nun heißt die Obergrenze der Übertragungsrate *über alle möglichen Nachrichten* Kanalkapazität und lässt sich in vielen Fällen mit Hilfe der AEP exakt mathematisch berechnen.¹³ Liegt die tatsächliche Übertragungsrate asymptotisch unterhalb der Kanalkapazität, kann die Fehlerwahrscheinlichkeit beliebig klein gemacht werden. Übersteigt die tatsächliche Übertragungsrate dauerhaft die Kanalkapazität, ist die Fehlerwahrscheinlichkeit $\epsilon > 0$ nicht mehr vernachlässigbar. Für eine errechnete Kanalkapazität werden die Nachrichten spezifisch kodiert, um die Übertragungsrate möglichst nahe ihrem Maximum zu halten.

Eine der Kernaussagen der Kommunikationstheorie ist, dass unter bestimmten schwachen Annahmen die Quellen- und Kanalkodierung *unabhängig voneinander* vorgenommen werden kann. *Das Kommunikationsproblem wird zu einem Optimierungsproblem unter Nebenbedin-*

¹² Es fehlt in diesem Fall evtl. an Kriterien, auf deren Grundlage ein Filter zur Beseitigung einer Störung konstruiert werden kann.

¹³ Man geht davon aus, dass bei der Grenzwertbetrachtung die Menge aller möglichen Nachrichten sich auf die Menge der typischen Nachrichten reduziert (AEP).

gungen, bei dem die wechselseitige Information zwischen Quelle und Empfänger zu maximieren ist.

Das Ökonomische. Teil II

Ist eine plausible ökonomische Interpretation der oben aufgeführten Kennzahlen möglich? Wir erwägen eine eingeschränkte JA-Antwort.

Es ist hilfreich, sich die wirtschaftlichen (evtl. auch politischen und sonstigen) Ereignisse als Informationsquelle vorzustellen. In diesem Licht erscheinen die Aktienkurse bzw. Renditen als Ergebnisse der Verarbeitung der ankommenden Informationen durch die Marktteilnehmer. Unter der Annahme homogener Erwartungen der Marktteilnehmer bezüglich des Rendite/Risiko-Verhältnisses sind die Auswirkungen der ankommenden Informationen auf die Aktienkurse/Renditen *nach der Entschlüsselung* eindeutig, d.h. die semantische Komponente ist in diesem Fall irrelevant.

Warum ist die JA-Antwort eingeschränkt? Weil es ein kleines Problem mit der Quelle gibt. Wie schon vorher erwähnt, es ist zu wenig bekannt über die Informations- bzw. Unsicherheitsquelle selbst, um sich ein klares Bild darüber zu schaffen. Was lässt sich unter dem „Generator der Unsicherheit“ vorstellen? Welche Eigenschaften hat er? Und, vor allem, von welcher Größenordnung sind die *Informationsschocks*, die von den Marktteilnehmern zu verarbeiten sind? Es scheint, dass auch für die Zukunft relativ wenig Hoffnung besteht, mehr Wissen über die Unsicherheitsquelle selbst zu bekommen. *Um den Modellbildungsprozess überhaupt zu ermöglichen, müssen deshalb Annahmen über den „Generator der Unsicherheit“ getroffen werden!*

Dies impliziert, dass die Kanalkapazität, die maximal mögliche Übertragungsrates der Information, nicht berechnet werden kann, sondern ad-hoc festgelegt werden muss und zur effizienten Informationsverarbeitung (man denke an die EMH) unter Umständen nicht ausreicht. Die Normalverteilung beschränkt beispielsweise die Bandbreite für die Reaktion auf die möglichen Informationsschocks im wesentlichen auf den Bereich $\pm 3\sigma$. Liegt die tatsächliche Übertragungsrates höher, dann ist eine fehlerfreie (sprich: eindeutige) Entschlüsselung der Nachricht nicht mehr möglich, und zwar unabhängig davon, wie gut die Ressourcenausstattung (Rechnerleistung etc.) ist.

Was wären die möglichen Auswirkungen unzureichender Kanalkapazität? Erstens, die größeren Schocks würden nicht als solche empfunden und verarbeitet, sondern als eine Folge kleinerer Signale, die eine positive zeitliche Abhängigkeit aufweisen. Man würde autoregressive Effekte da sehen, wo es sie überhaupt nicht gibt (da die Schocks nicht schnell genug absorbiert werden könnten). Im Extremfall sind die Long-Memory Effekte vorstellbar. Zweitens, es würde an der notwendigen Redundanz in einem gestörten Kanal fehlen. Dies führte zu Schwierigkeiten, eine Störung als solche zu identifizieren. Die Störung könnte fälschlicherweise als Information interpretiert werden.

Weiterhin ist es vorstellbar, dass das Problem unzureichender Kanalkapazität von den Marktteilnehmern erkannt wird. Eine daraus resultierende Veränderung des Verhaltens der Marktteilnehmer würde zu Übertreibungen bei der Interpretation der Informationen (overreaction) führen. Die Unsicherheit darüber, ob die pro Zeiteinheit ankommende Informationsmenge vollständig ist (mit anderen Worten, eine lineare vs. eine konvexe Entropierate), würde eine positive Rückkopplung auslösen. Der Grund dafür wäre die Angst seitens der Marktteilnehmer, ein mögliches Ausmaß der aus der Informationsverarbeitung resultierenden Kurs/Renditeschwankungen zu unterschätzen. Die möglichen Folgen sind natürlich Fehlbewertungen der Wertpapiere aufgrund einer verlangsamten Informationsauswertung.

Bei so vielen möglichen Problemen, was spricht dagegen, die (ad-hoc festgelegte) Kanalkapazität höher zu setzen? Eigentlich nichts, außer, man würde in diesem Fall den Normalverteilungsbereich verlassen.¹⁴ Diese Tatsache ist nicht problematisch per se, abgesehen davon, dass dadurch die Grundannahmen der klassischen Finance-Theorie ernsthaft verletzt wären. Verfolgt man den Rückweg, der zu einer Annahme der Normalverteilung bei den Aktienrenditen führt, dann stellt sich heraus, dass weder ein Random Walk Modell noch eine Brownsche Bewegung als Grundlage zur Modellierung der finanzwirtschaftlichen Zeitreihen geeignet wäre, ohne die Normalverteilung als Ergebnis der Herleitungen akzeptieren zu müssen.

Bemerkenswert ist, dass die oben beschriebenen möglichen Auswirkungen der unzureichenden Kanalkapazität auf dem Aktienmarkt tat-

¹⁴ Der Grund dafür ist, dass für die gegebenen Mittelwert und Varianz die Normalverteilung eine Maximum-Entropie-Verteilung ist, d.h. sie weist die höchste Kanalkapazität auf.

BEISPIEL 4.1

Es sei $X \sim N(\mu, \sigma^2)$, $f(x) = \sqrt{\frac{1}{2\pi\sigma^2}} \exp\left\{-\frac{(x-\mu)^2}{2\sigma^2}\right\}$.

Dann

$$-\log_2 f(x) = \frac{1}{2} \log_2 (2\pi\sigma^2) + (\log_2 e) (x-\mu)^2 / (2\sigma^2).$$

Da $E[(X-\mu)^2] = \sigma^2$ gilt, ergibt sich

$$\mathbb{H}(X) = E[-\log_2 f(x)] = \frac{1}{2} \log_2 (2\pi\sigma^2) + \frac{1}{2} \log_2 e = \frac{1}{2} \log_2 (2\pi e \sigma^2).$$

Damit steht folgendes fest:

- Für kleinere σ -Werte kann Entropie beliebig klein sein
- Die Endlichkeit der Entropie ist unmittelbar an die endliche Varianz der Normalverteilung gebunden.

sächlich empirisch beobachtet werden. Wäre das nicht ein zusätzlicher Grund dafür, sich von der Normalverteilung als Modell der Aktienrenditen zu distanzieren?

Entropie: der stetige Fall

Um eine Erweiterung des Entropiebegriffes auf kontinuierliche Zufallsvariablen bzw. auf beliebige Zufallsexperimente vorzunehmen, betrachtet man eine kontinuierliche Zufallsvariable $X (\Omega, \mathcal{F}, \mathbb{P})$ mit Verteilung $F(x) \equiv \mathbb{P}(X \leq x)$. Unter der Annahme, dass die Verteilungsfunktion $F(x)$ absolutstetig ist, existiert eine Dichtefunktion $f(x) \equiv F'(x)$.

Definition 4.1.4 (Entropie, kontinuierlich). Die Entropie (oder die Differentialentropie) einer kontinuierlichen Zufallsvariable X ist definiert als

$$\mathbb{H}(X) = - \int f(x) \log_2 f(x) dx = E[-\log_2 f(x)], \quad (4.1.12)$$

wenn das Integral existiert.¹⁵ Die Integration erfolgt über den Träger der Dichte $f(x)$.

Die Erweiterung auf kontinuierliche Zufallsvariablen hat zur Folge, dass der Ausdruck (4.1.12) im Vergleich mit (4.1.2) nicht mehr endlich zu sein braucht. Eine sinnvolle Definition beschränkt sich daher auf den Bereich der endlichen Werte für $E[\max\{\ln f(x)\}]$ bzw.

¹⁵ Vgl. Shannon (1948), S. 414 bzw. Cover und Thomas (1991), S. 224.

$E[\max\{-\ln f(x)\}]$ (Vgl. Beispiel 4.1). Dazu kommt, dass die Differentialentropie einer von X nach αX ($\alpha > 0$) reskalierten Zufallsvariable um den Faktor $\log_2 \alpha$ steigt; Differentialentropie ist also *nicht skaleninvariant*. Angesichts dieser Komplikationen schlugen Kullback und Leibler (1951) (auch Kullback (1959)) drei Jahre nach Shannon ein Maß der *Divergenz* zwischen zwei Verteilungen vor.

Definition 4.1.5 (Relative Entropie). *Das Funktional*

$$\mathbb{D}(f\|g) = \int f(x) \log\left(\frac{f(x)}{g(x)}\right) dx \tag{4.1.13}$$

heißt *relative Entropie*, wobei f und g Wahrscheinlichkeitsdichtefunktionen sind. Ist der Träger von $g(x)$ nicht im Träger von $f(x)$ enthalten, ist relative Entropie unendlich. Wir vereinbaren $0 \log \frac{0}{0} = 0, 0 \cdot \infty = 0$.

Die relative Entropie ist auch als *Kullback-Leibler Informationskriterium*, *Cross-Entropie*, *Divergenz* oder als *Kullback-Leibler Distanz* bekannt.¹⁶ Denkt man in der obigen Definition an die Funktionen $f(x)$ und $g(x)$ als Wahrscheinlichkeitsdichtefunktionen zweier Wahrscheinlichkeitsmaße \mathbb{Q} bzw. \mathbb{P} , lässt sich sehr anschaulich eine Brücke zur Wahrscheinlichkeitstheorie aufbauen. Im folgenden führen wir einige Grundeigenschaften der relativen Entropie auf. Es sei \mathbb{P} ein σ -endliches¹⁷ Wahrscheinlichkeitsmaß auf einem Messraum (Ω, \mathcal{F}) und \mathbb{Q} ein für alle messbaren Mengen $A \in \mathcal{F}$ bezüglich \mathbb{P} absolutstetiges Maß mit der Radon-Nikodym-Ableitung $X \equiv d\mathbb{Q}/d\mathbb{P}$. Zur Vereinfachung setzen wir $A = \Omega$ fest. Es gilt

$$\begin{aligned} \mathbb{D}(\mathbb{Q}\|\mathbb{P}) &= \int_{\Omega} d\mathbb{Q} \ln \frac{d\mathbb{Q}}{d\mathbb{P}} = \int_{\Omega} \frac{d\mathbb{Q}}{d\mathbb{P}} \ln \frac{d\mathbb{Q}}{d\mathbb{P}} d\mathbb{P} \\ &= E^{\mathbb{Q}}[\ln X] = E^{\mathbb{P}}[X \ln X], \end{aligned} \tag{4.1.14}$$

$$\mathbb{D}(\mathbb{Q}\|\mathbb{P}) < \infty \iff \mathbb{Q} \ll \mathbb{P}. \tag{4.1.15}$$

¹⁶ Obwohl die relative Entropie für zwei Maße ν und μ nichtnegativ und ungleich Null außer $\nu = \mu$ ist, stellt die Kullback-Leibler Distanz keine „echte“ Metrik dar: sie ist weder symmetrisch, d.h. $\mathbb{D}(\nu\|\mu) \neq \mathbb{D}(\mu\|\nu)$, noch genügt sie der Dreiecksungleichung $\rho(\nu, \mu) \leq \rho(\nu, \eta) + \rho(\eta, \mu)$ für ein drittes Maß η . Insoweit ist eine weitere Bezeichnung als *gerichtete Distanz* gerechtfertigt.

¹⁷ Ein Maß μ auf einer σ -Algebra A heißt σ -endlich, wenn eine Aufteilung $\{F_n\}_{n \geq 1} \subset A$ existiert, so dass $\Omega = \sum_1^{\infty} F_n$ und $\mu(F_n) < \infty$ für alle $n \geq 1$ gilt.

Unter Berücksichtigung der Identität $\ln(x) \leq x - 1$ lässt sich leicht zeigen, dass die relative Entropie immer nichtnegativ ist:

$$\begin{aligned} \mathbb{D}(\mathbb{Q} \parallel \mathbb{P}) &= - \int_{\Omega} d\mathbb{Q} \ln \frac{d\mathbb{P}}{d\mathbb{Q}} \geq - \int_{\Omega} d\mathbb{Q} \left(\frac{d\mathbb{P}}{d\mathbb{Q}} - 1 \right) \\ &= - \int_{\Omega} (d\mathbb{P} - d\mathbb{Q}) = 0. \end{aligned} \quad (4.1.16)$$

Sie ist gleich Null genau dann, wenn $\mathbb{P} \equiv \mathbb{Q}$. Für beliebige drei absolutstetige Wahrscheinlichkeitsmaße $\mathbb{Q} \ll \mathbb{M} \ll \mathbb{P}$ gilt:

$$\begin{aligned} \mathbb{D}(\mathbb{Q} \parallel \mathbb{P}) &= \int_{\Omega} d\mathbb{Q} \ln \frac{d\mathbb{Q}}{d\mathbb{P}} = \int_{\Omega} d\mathbb{Q} \ln \left(\frac{d\mathbb{Q}}{d\mathbb{M}} \frac{d\mathbb{M}}{d\mathbb{P}} \right) \\ &= \int_{\Omega} d\mathbb{Q} \ln \frac{d\mathbb{Q}}{d\mathbb{M}} + \int_{\Omega} d\mathbb{Q} \ln \frac{d\mathbb{M}}{d\mathbb{P}}, \end{aligned} \quad (4.1.17)$$

woraus resultiert

$$\mathbb{D}(\mathbb{Q} \parallel \mathbb{P}) = \mathbb{D}(\mathbb{Q} \parallel \mathbb{M}) + \mathbb{E}^{\mathbb{Q}} \left(\ln \frac{d\mathbb{M}}{d\mathbb{P}} \right). \quad (4.1.18)$$

Für $\mathbb{Q} \ll \mathbb{P}$ gilt:

$$\mathbb{D}(\mathbb{Q} \parallel \mathbb{P}) = \sup_{\Phi} \left(\mathbb{E}^{\mathbb{Q}}[\Phi] - \ln \mathbb{E}^{\mathbb{P}}[\exp(\Phi)] \right), \quad (4.1.19)$$

wobei das Supremum über alle beschränkten Zufallsvariablen Φ berechnet wird, für die $\exp(\Phi)$ \mathbb{P} -integrierbar ist bzw. alle Momente besitzt, d.h. $\mathbb{E}^{\mathbb{P}}[\exp(\Phi)] < \infty$. Um das zu zeigen, legen wir $\Phi = \ln \frac{d\mathbb{Q}}{d\mathbb{P}} = \ln f$ fest. Es ist leicht zu sehen, dass

$$\begin{aligned} \mathbb{E}^{\mathbb{Q}}[\Phi] - \ln \mathbb{E}^{\mathbb{P}}[\exp(\Phi)] &= \int_{\Omega} d\mathbb{Q} \ln f - \ln \left(\int_{\Omega} d\mathbb{P} f \right) \\ &= \mathbb{D}(\mathbb{Q} \parallel \mathbb{P}) - \ln \left(\int_{\Omega} d\mathbb{Q} \right) = \mathbb{D}(\mathbb{Q} \parallel \mathbb{P}). \end{aligned} \quad (4.1.20)$$

Damit ist das Supremum über alle Φ nicht kleiner als die Divergenz. Weiterhin gilt für jedes Φ :

$$\begin{aligned} \mathbb{E}^{\mathbb{Q}}[\Phi] - \ln \mathbb{E}^{\mathbb{P}}[\exp(\Phi)] &= \int_{\Omega} d\mathbb{Q} \ln \left(\frac{\exp(\Phi)}{\int_{\Omega} d\mathbb{P} \exp(\Phi)} \right) = \int_{\Omega} d\mathbb{Q} \ln \left(\frac{d\mathbb{Q}/d\mathbb{P}}{d\mathbb{Q}/d\mathbb{P}^*} \right), \end{aligned} \quad (4.1.21)$$

wobei durch $\mathbb{P}^* \ll \mathbb{P}$ ein neues Wahrscheinlichkeitsmaß

$$\mathbb{P}^* = \int_{\Omega} \frac{\exp(\Phi)}{\mathbb{E}^{\mathbb{P}}[\exp(\Phi)]} d\mathbb{P} \quad \text{bzw.} \quad \frac{d\mathbb{P}^*}{d\mathbb{P}} = \frac{\exp(\Phi)}{\mathbb{E}^{\mathbb{P}}[\exp(\Phi)]} \neq 0 \quad (4.1.22)$$

definiert wird. Man nennt es auch *tilted (twisted) Distribution*. Der Übergang zu einem nach (4.1.22) definierten Wahrscheinlichkeitsmaß heißt *Cramér (Esscher) Transformation* und kann dazu verwendet werden, um ein äquivalentes Martingalmaß für die Bewertung derivativer Finanztitel zu finden.¹⁸ Da $\mathbb{Q} \ll \mathbb{P}^* \ll \mathbb{P}$, folgt aus (4.1.21)

$$\begin{aligned} \mathbb{D}(\mathbb{Q} \parallel \mathbb{P}) - (\mathbb{E}^{\mathbb{Q}}[\Phi] - \ln \mathbb{E}^{\mathbb{P}}[\exp(\Phi)]) \\ = \ln \mathbb{E}^{\mathbb{P}}[\exp(\Phi)] = \mathbb{D}(\mathbb{Q} \parallel \mathbb{P}^*) \geq 0 \end{aligned} \quad (4.1.23)$$

wegen der Eigenschaft (ii). Da die Formel (4.1.23) keiner Einschränkung bezüglich der Zufallsvariable Φ unterliegt, gilt sie für alle Φ und folglich auch für das Supremum über alle Φ . Die Ungleichung (4.1.23) wird nach der Eigenschaft (ii) genau dann zur Gleichung, wenn $\mathbb{Q} = \mathbb{P}^*$ ist.

Das Maximum Entropie Prinzip

Der Maximum Entropie Prinzip wurde erstmals von Jaynes (1957) dargestellt.¹⁹

*“It is now evident how to solve our problem; in making inferences on the basis of partial information we must use that probability distribution which has maximum entropy subject to whatever is known. This is the only unbiased assignment we can make; to use any other would amount to arbitrary assumption of information which by hypothesis we do not have.”*²⁰

Shore und Jonson (1980) geben der ersten intuitiven Idee eine formale axiomatische Form und später²¹ auch eine theoretische Begründung. Maximum Entropie (ME) ist ein Verfahren zur Schätzung einer

¹⁸ Gerber und Shiu (1994), Jensen (1995), Bühlmann u.a. (1996).

¹⁹ Zur ausführlichen Diskussion vgl. Cover und Thomas (1991), Cherny und Maslov (2004).

²⁰ Jaynes (1957), S. 623.

²¹ Jonson und Shore (1983).

BEISPIEL 4.2

Ist \mathbb{P} das Maß einer Gleichverteilung, dann gilt

$$\mathbb{P} = (1/n, 1/n, \dots, 1/n),$$

$$\mathbb{D}(\mathbb{Q} \parallel \mathbb{P}) = \sum_{i=1}^n p_i \ln \frac{p_i}{1/n} = \ln n - \left(- \sum_{i=1}^n p_i \ln p_i \right) = \ln n - \mathbb{H}(\mathbb{Q}). \quad (4.1.24)$$

Die Minimierung von $\mathbb{D}(\mathbb{Q} \parallel \mathbb{P})$ wird in diesem Fall zur Maximierung von $\mathbb{H}(\mathbb{Q})$. Im allgemeinen sind aber diese zwei Prinzipien unterschiedlich. ME ist ein Prinzip, bei dem die absolute Unsicherheit maximiert wird; beim Minimum Divergenz Prinzip minimiert man die Divergenz zwischen \mathbb{Q} und \mathbb{P} .

Wahrscheinlichkeitsverteilung auf Grundlage von bekannten Momenten (Nebenbedingungen). Dabei wird nur die in den Nebenbedingungen enthaltene Information als relevant eingestuft; der Rest unterliegt maximaler Unsicherheit. Intuitiv gesehen, haben die Verteilungen mit höherer Entropie entsprechend höhere Konzentration um den Mittelwert; sie sind in der Natur auf verschiedenen Wegen konsequent realisierbar und folglich auch öfters anzutreffen.²² Nehmen wir an, die Information über eine Wahrscheinlichkeitsverteilung wäre in der Form von r Momenten verfügbar:

$$\mathbb{E}[g_k] = \sum_{i=1}^N p_i g_{ki} = a_k, \quad k = 1, \dots, r. \quad (4.1.25)$$

Dann nimmt die ME-Lösung die folgende exponentielle Form an²³:

$$p_i = \exp \left(-\lambda_0 - \sum_{k=1}^r \lambda_k g_{ki} \right). \quad (4.1.26)$$

λ_0 und λ_k 's sind dabei Konstanten, die sich aus den Nebenbedingungen für $\sum p_i = 1$ und für $\mathbb{E}[g_k]$ ergeben. Existiert eine solche Parametermenge, die den spezifizierten Nebenbedingungen genügt, dann ist die ME-Lösung eindeutig²⁴. Die Maximierung des Informationsmaßes impliziert, dass die resultierende Verteilung genau diejenige Lösung darstellt, die alle in den Nebenbedingungen enthaltene Informationen

²² Jaynes (1982).

²³ Tribus (1969).

²⁴ Cover und Thomas (1991).

ausnutzt; andererseits wird auch die maximale Irrelevanz bezüglich unbekannter Information garantiert²⁵.

Sehr ähnlich zum ME-Prinzip lässt sich das *Minimum Divergenz Prinzip* von Kullback (1959) formulieren. Gegeben die a-priori Wahrscheinlichkeitsverteilung \mathbb{P} , sollte man aus allen zulässigen²⁶ Wahrscheinlichkeitsverteilungen diejenige auswählen, die $\mathbb{D}(\mathbb{Q}||\mathbb{P})$ minimiert. Im Finance entspricht das Minimum Divergenz Prinzip unter linearen Nebenbedingungen der Suche nach einem *Minimum-Entropie Martingalmaß* in einem einperiodigen Bewertungsmodell für derivative Finanztitel.²⁷ Für den Fall, dass alle exponentiellen Momente der Wahrscheinlichkeitsverteilung, die die Preisentwicklung der Wertpapiere beschreibt, endlich sind, ist die Lösung bekannt.²⁸ Das Minimum-Entropie Martingalmaß ist die Esscher-Transformierte des Originalmaßes \mathbb{P} .

Bei mehrperiodigen Modellen und bei Modellen in stetiger Zeit ist die Form des Minimum-Entropie Martingalmaßes nur in einigen Spezialfällen bekannt.²⁹ Die Suche nach dem Minimum-Entropie Martingalmaß kann als *duales Problem zur Maximierung einer exponentiellen Nutzenfunktion* betrachtet werden.³⁰ Cherny und Maslov (2004) geben eine explizite Lösung des Problems der Maximierung einer exponentiellen Nutzenfunktion in einem allgemeinen Modell in diskreter Zeit an. Die Dualität zweier Problemen wird aus der Lösung offensichtlich.

Einige klassische Beispiele der ME-Lösung

Teilweise in Anlehnung an Landsberg (1971) zeigen wir, wie einige Wahrscheinlichkeitsverteilungen als Lösungen eines ME-Problems dargestellt werden können.

Die Gleichverteilung. Es sei $\mathbb{P}(X = x_j) = p_j$, $j = 1, 2, \dots, N$. Hat eine Verteilung eine einzige Nebenbedingung zu erfüllen, nämlich $\sum p_j = 1$, dann ist die entsprechende ME-Verteilung eine Gleichverteilung. Mit dem Lagrange-Ansatz ergibt sich:

²⁵ Jaynes (1957).

²⁶ D.h., für die die Nebenbedingungen an die Momente nicht verletzt sind.

²⁷ Vgl. Branger (2002) und Referenzen dort.

²⁸ Vgl. zum Beispiel Föllmer und Schied (2002), Korollar 3.27.

²⁹ Fujiwara und Miyahara (2003).

³⁰ Frittelli (2000), Goll und Rüschemeyer (2001), Schachermayer (2001).

$$\mathcal{L} = - \sum_{j=1}^N p_j \ln p_j + \lambda \left(\sum_{j=1}^N p_j - 1 \right) = - \sum_{j=1}^N p_j (\ln p_j - \alpha) + C, \quad C = \text{const} \quad (4.1.27)$$

$$\frac{\partial \mathcal{L}}{\partial p_j} = - \left(\ln p_j - \alpha + p_j \frac{1}{p_j} \right) \stackrel{!}{=} 0 \quad \Rightarrow \quad \ln p_j = \alpha - 1 \stackrel{!}{=} \text{const} \quad (4.1.28)$$

Folglich sind alle p_j gleich. Durch eine Normierung erhalten wir $p_j = p = 1/N$. Der ME-Wert ist also

$$\mathbb{H}_{\max} = - \sum_{j=1}^N \frac{1}{N} \ln \frac{1}{N} = \ln N. \quad (4.1.29)$$

Kommt als weitere Nebenbedingung ein bekannter Mittelwert μ dazu, bekommt die Lagrange-Funktion einen zusätzlichen Summanden:

$$\begin{aligned} \mathcal{L} &= - \sum_{j=1}^N p_j \log p_j + \lambda_0 \left(\sum_{j=1}^N p_j - 1 \right) + \lambda_1 \left(\sum_{j=1}^N p_j x_j - \mu \right) \\ &= - \sum_{j=1}^N p_j (\ln p_j - \alpha + \beta x_j) + C \end{aligned} \quad (4.1.30)$$

$$\frac{\partial \mathcal{L}}{\partial p_j} = - (\ln p_j - \alpha + 1 + \beta x_j) \quad (4.1.31)$$

Diese Gleichung hat zwei Parameter, α und β . Die Lösung lässt sich deshalb nur in parametrischer Form angeben. Mit dem Ansatz

$$p_j = \frac{1}{Z} \exp(-\beta x_j), \quad Z \equiv \sum_{j=1}^N \exp(-\beta x_j) \quad (4.1.32)$$

erfüllt sich die Nebenbedingung an die Summe der Wahrscheinlichkeiten automatisch. Der Parameter α kommt also nicht mehr vor und der ME-Wert ist

$$\mathbb{H}_{\max} = - \sum_{j=1}^N p_j [-\beta x_j - \ln Z] = \beta \mu + \ln Z. \quad (4.1.33)$$

Die Normalverteilung. Das nächste Beispiel zeigt, dass aus allen Verteilungen mit gegebener Varianz σ^2 die Normalverteilung eine ME-Verteilung ist. Es sei $p(x)$, $x \in \mathbb{R}$ die Wahrscheinlichkeitsdichtefunktion. Die Lagrange-Funktion ist:

$$\mathcal{L} = - \int_{-\infty}^{\infty} p(x) \ln p(x) dx - \alpha \int_{-\infty}^{\infty} p(x) dx - \beta \int_{-\infty}^{\infty} x^2 p(x) dx. \quad (4.1.34)$$

Die Maximierung ist nach beliebiger Variationen von $p(x)$ vorzunehmen. Es gilt:

$$\ln p(x) - \alpha - \beta x^2 \stackrel{!}{=} 0. \quad (4.1.35)$$

Wir nehmen den Ansatz

$$p(x) = \alpha \exp(-\beta x^2), \quad \beta > 0. \quad (4.1.36)$$

Dies kann die Normierung erleichtern, da das Integral von $p(x)$ zu einer Klasse von Gausschen Integralen gehört³¹. Es ergibt sich damit

$$\alpha \int_{-\infty}^{\infty} \exp(-\beta x^2) dx = \alpha \cdot 2 \cdot \frac{1}{2} \sqrt{\frac{\pi}{\beta}} = 1, \quad (4.1.37)$$

$$\alpha \int_{-\infty}^{\infty} x^2 \exp(-\beta x^2) dx = \sigma^2 = \frac{2\alpha}{4\beta} \sqrt{\frac{\pi}{\beta}} = \frac{1}{2\beta} \quad (4.1.38)$$

und schließlich durch das Einsetzen in (4.1.36)

$$p(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{x^2}{2\sigma^2}\right).$$

Eine Erweiterung für den Erwartungswert $\mu \neq 0$ ist sehr einfach und erfolgt durch eine leichte Modifizierung der Nebenbedingungen:

$$\ln p(x) - \alpha - \beta(x - \mu)^2 - \gamma(x - \mu) \stackrel{!}{=} 0, \quad (4.1.39)$$

$$\alpha \int_{-\infty}^{\infty} x^2 \exp[-\beta(x - \mu)^2] dx = \sigma^2, \quad (4.1.40)$$

$$\int_{-\infty}^{\infty} x \exp[-\beta(x - \mu)] dx = 0. \quad (4.1.41)$$

³¹ Auf der Website <http://mathworld.wolfram.com/GaussianIntegral.html> lassen sich für das Integral $I_n(\beta) = \int_0^{\infty} x^n \exp(-\beta x^2) dx$ die folgenden Einzelwerte entnehmen: $I_0(\beta) = \frac{1}{2} \sqrt{\frac{\pi}{\beta}}$, $I_2(\beta) = \frac{1}{4\beta} \sqrt{\frac{\pi}{\beta}}$.

Wegen (4.1.41) bleiben die Parameter α und β in (4.1.39) unverändert. Die resultierende Dichte ist

$$p(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{(x-\mu)^2}{2\sigma^2}\right]. \quad (4.1.42)$$

4.2 Die „Nicht-Shannon“-Entropien

Die Modellierung der Tail-Ereignisse

Betrachtet man die Entropie als eine gewichtete Summe der Wahrscheinlichkeiten, dann führen sowohl sehr geringe als auch sehr große Wahrscheinlichkeiten zu relativ kleinen Werten der Entropie. Solche Ereignisse sind untypisch, sie verursachen nur kleine Änderungen der Entropie und bleiben nach dem Satz von Shannon-McMillan-Breiman (4.1.2) bei der Grenzwertbetrachtung außer Acht. Bei den Verteilungen vom Pareto-Typ spielen aber gerade solche Tail-Ereignisse eine wichtige Rolle. Je dicker die Tails einer Verteilung sind, desto höher sind die entsprechenden Wahrscheinlichkeiten im Vergleich zur Normalverteilung. Die Erkenntnis, dass die Tail-Ereignisse im Pareto-Fall nicht mehr vernachlässigt werden dürfen, scheint damit unmittelbar im Widerspruch mit dem Satz von Shannon-McMillan-Breiman zu stehen. Hat dieses Rätsel eine Lösung? Diese Frage kann sogar noch allgemeiner gestellt werden: *lassen sich die Verteilungen vom Pareto-Typ mittels des Maximum-Entropie Verfahrens herleiten?*

Die Antwort ist „Ja“. Allerdings wird es mit der Shannon-Entropie nicht mehr möglich sein. Die Shannon-Entropie lässt die untypischen Ereignisse bei der Grenzwertbetrachtung „aussterben“. Um das zu verhindern bzw. das Verhältnis zwischen den typischen und den untypischen Ereignissen zu Gunsten der letzteren zu verändern, ist es notwendig, die sehr kleinen Wahrscheinlichkeiten relativ zu den restlichen Wahrscheinlichkeiten zu vergrößern. Dies kann durch eine *nicht-lineare* Skalierung und eine entsprechende Normierung erfolgen. Dabei wird die Verteilungsfunktion derart verformt, dass sie an den Flanken *flacher* wird. Mathematisch lässt sich dieser Vorgang durch die Einführung eines zusätzlichen Parameters q beschreiben, der für den Grad der Nichtlinearität bei der Verformung der Wahrscheinlichkeiten

zuständig ist. Da die Wahrscheinlichkeitsverteilung durch die Maximierung der Entropie hergeleitet wird, geht der Parameter q in die Entropie-Formel ebenfalls mit ein. Die Idee liegt also in der *Parametrisierung der Entropie* bzw. der Zielfunktion im Maximum-Entropie Verfahren. Im Folgenden betrachten wir zwei Möglichkeiten, dies zu erreichen.

Die Blindheit³² der Entropie für die Feinstruktur einer Verteilung, für ihre Tails, kann durch die Bildung sogenannter *Begleitverteilungen* (*escort distributions*) korrigiert werden. Man bildet aus der ursprünglichen Verteilung $p \equiv (p_1, \dots, p_k)$ mit Hilfe eines Parameters q eine neue Verteilung

$$p^q \equiv (Z^{-1}p_1^q, \dots, Z^{-1}p_k^q)$$

mit einer Normierungskonstanten

$$Z \equiv \sum_{i=1}^k p_i^q.$$

Nach dieser Transformation lassen sich die Eigenschaften der neuen Verteilung im Vergleich mit der ursprünglichen Verteilung in Abhängigkeit vom Parameter q wie folgt charakterisieren. Der Wert $q = 1$ entspricht der ursprünglichen Verteilung. Für $q > 1$ werden alle Wahrscheinlichkeiten durch die Potenz q zunächst verkleinert. Diese Veränderung erfolgt aber nicht gleichmäßig, sondern umso stärker, je kleiner die ursprüngliche Wahrscheinlichkeiten p_i gewesen sind. Nach der Normierung um $Z < 1$ werden alle Wahrscheinlichkeiten wieder größer. Die Korrektur erfolgt wiederum nicht gleichmäßig, sondern verstärkt im Rand- und zentralen Bereich. Die Mitte wird relativ schwach angehoben, weil viel Masse ins Zentrum und in die Flanken fließt. Der zentrale und der Randbereich profitieren sozusagen auf Kosten der Mitte. Die Tails werden dicker und klingen langsamer ab. Der zentrale Bereich der Dichtefunktion wird spitzer und der Abfall beim Übergang zu den Flanken wird steiler.

Mit $0 < q < 1$ erreicht man den Gegeneffekt: es werden zwar alle Wahrscheinlichkeiten angehoben, allerdings profitieren davon die ursprünglich kleineren Wahrscheinlichkeiten verhältnismäßig mehr.³³

³² Dieser Ausdruck stammt von Eberling u.a. (1998).

³³ Die $q < 1$ -Potenz einer Zahl $0 < p < 1$ ist größer als p ; allerdings ist die relative Steigerung $p^q/p = p^{q-1}$ eine fallende Funktion von p .

Nach der Normierung um $Z > 1$ bekommt man als Ergebnis eine Verteilung, deren Rand- und Mittelbereich auf Kosten der immer flacher werdenden Spitze höhere Wahrscheinlichkeiten als ursprünglich aufweist. Dabei wird der Randbereich selbst zunehmend schmaler und die Verteilung bekommt u.U. einen kompakten Träger. Die Werte $q < 0$ führen zu den mehrgipfligen Verteilungen, im Extremfall $q \rightarrow -\infty$ ergibt sich ein Rechteck.

Grundsätzlich kann das Verfahren der Begleitverteilungen auf unterschiedliche Arten der Entropie angewendet werden. Die zwei bekanntesten sind die Rényi-Entropie³⁴ und die Tsallis-Entropie.³⁵

Es sei X eine Zufallsvariable mit Werten in einer endlichen Menge \mathcal{X} und $p(x) = \mathbb{P}(X = x)$ eine Wahrscheinlichkeitsverteilung auf \mathcal{X} .

Definition 4.2.1. Die **q-Order Rényi-Entropie** von $p(x)$ (oder von X) ist definiert für $0 < q < \infty$, $q \neq 1$ als

$$\mathbb{H}_q^R(p) \equiv \frac{1}{1-q} \ln \sum_x (p(x))^q = \frac{1}{1-q} \ln \mathbb{E} \left[p(x)^{(q-1)} \right]. \quad (4.2.1)$$

Die Rényi-Entropien der Ordnungen $q = 0$ und $q = 1$ sind als Grenzwerte von \mathbb{H}_q bei $q \rightarrow 0$ bzw. $q \rightarrow 1$ definiert und resultieren in

$$\mathbb{H}_0^R(p) = \lim_{q \rightarrow 0} \mathbb{H}_q^R(p) = \ln \mu(\{x : p(x) > 0\}), \quad \text{und} \quad (4.2.2)$$

$$\mathbb{H}_1^R(p) = \lim_{q \rightarrow 1} \mathbb{H}_q^R(p) = \mathbb{H}(p). \quad (4.2.3)$$

Der Logarithmus des Trägers der Dichtefunktion entspricht also der Rényi-Entropie der Ordnung $q = 0$ und die Shannon-Entropie kann als ein Spezialfall der Rényi-Entropie mit dem Index $q = 1$ betrachtet werden.

Die Rényi-Entropie stellt nicht die einzige Möglichkeit dar, eine (zu Shannon-) alternative Entropie zu definieren. Unter Beachtung der Identität

$$\lim_{y \rightarrow 0} \frac{\ln(1+y)}{y} = 1$$

mit $\sum_x (p(x))^q - 1$ an Stelle von y wird die hinter der Tsallis-Entropie liegende Idee schnell deutlich.

³⁴ Rényi (1961).

³⁵ Tsallis (1988). In der Literatur ist eine alternative Bezeichnung Havrda-Charvat-Entropie nach Havrda und Charvat (1967) ebenfalls geläufig.

Definition 4.2.2. Die **q-Order Tsallis-Entropie** von $p(x)$ ist definiert für $q \in \mathbb{R}$, $q \neq 1$ als

$$\mathbb{S}_q(p) \equiv \frac{1}{1-q} \left(\sum_x (p(x))^q - 1 \right). \quad (4.2.4)$$

Der funktionale Zusammenhang zwischen der Rényi- und Tsallis-Entropie ist leicht ersichtlich:

$$\mathbb{H}_q^R(p) = \frac{1}{1-q} \ln (1 + (1-q)\mathbb{S}_q(p))$$

und bei der Grenzwertbetrachtung $q \rightarrow 1$ geht die Tsallis-Entropie ebenfalls in die Shannon-Entropie über.

Die Tsallis-Entropie: wozu ist sie gut?

Sowohl die Tsallis-Entropie als auch die Rényi-Entropie kann zur Modellierung von skaleninvarianten Systemen verwendet werden. Allerdings hat die Tsallis-Entropie im Gegensatz zur Rényi-Entropie eine sehr merkwürdige Eigenschaft, die sie für die Modellierung besonders interessant macht. Sie ist nämlich von *nicht extensiver* Natur. Was ist darunter zu verstehen?

Der Hauptmerkmal der Extensivität ist nach Touchette (2002) eine asymptotisch konstante Entropierate. Dies gilt sowohl für die Shannon-Entropie³⁶

$$\mathbb{H}(X) = \lim_{n \rightarrow \infty} \frac{1}{n} \mathbb{H}(X^n) = \lim_{n \rightarrow \infty} \frac{n \mathbb{H}(X_1)}{n} = \mathbb{H}(X_1). \quad (4.2.5)$$

als auch für die Rényi-Entropie und *gilt nicht* für die Tsallis-Entropie. Die zeitlichen Abhängigkeiten bzw. die Korrelationen sind bei den extensiven Systemen *kurzfristiger* Natur und werden bei der Grenzwertbetrachtung absorbiert. Mit anderen Worten sind solche Abhängigkeiten auf die Dauer vernachlässigbar. Bei den Langzeitabhängigkeiten ist das nicht der Fall. Die Entropierate eines Systems mit den Langzeitabhängigkeiten hat u.U. keinen Grenzwert für $n \rightarrow \infty$ und steigt folglich überproportional mit n . Die Tsallis-Entropie scheint ein geeignetes Werkzeug zur Modellierung von solchen Systemen darzustellen.

³⁶ Vgl. die Eigenschaft iii) auf der Seite 64.

Eine andere Eigenschaft, die der Nicht-Extensivität sehr ähnlich, aber nicht äquivalent ist, heißt *Nicht-Additivität*. Sie kommt dadurch zustande, dass die Entropie einer Summe von zwei unabhängigen Subsystemen nicht mehr zwangsläufig gleich der Summe von entsprechenden Entropien sein muss. Stattdessen bestimmt sich die Entropie der Summe nach der Formel³⁷ (hier die Tsallis-Entropie)

$$\mathbb{S}_q(A+B) = \mathbb{S}_q(A) + \mathbb{S}_q(B) + k(1-q) \mathbb{S}_q(A) \mathbb{S}_q(B). \quad (4.2.6)$$

Dabei ist k eine Konstante und A und B sind zwei unabhängige Systeme, d.h. $p_{ij}(A+B) = p_i(A) + p_j(B)$. Für q größer bzw. kleiner Eins spricht man entsprechend von der *Sub-* bzw. *Super-Additivität*. Ein nichtextensives System ist gewöhnlich auch nichtadditiv. Es kann aber Ausnahmen geben.³⁸ Im Allgemeinen gehören die Nicht-Extensivität bzw. die Nicht-Additivität (oder die beiden Eigenschaften zusammen) zu den charakteristischen Merkmalen eines nicht homogenen Systems. Die *Homogenität* ist ein Zustand, in dem die Entropie eines Systems als Ganzes durch eine Aufteilung in mehrere Einzelteile nicht beeinflusst wird. Es gilt also:

$$\mathbb{H}(aX) = a\mathbb{H}(X), \quad a = \text{const.} \quad (4.2.7)$$

Aus (4.2.6) bzw. (4.2.4) folgt, dass diese Eigenschaft durch die Tsallis-Entropie ebenfalls verletzt ist, da die Konstante zusammen mit der Wahrscheinlichkeitsdichte nichtlinear skaliert wird. Weiterhin ist die Nicht-Homogenität ein typisches Merkmal von Systemen, die überwiegend durch die Langzeitabhängigkeiten charakterisiert werden bzw. ein (multi)fraktal-ähnliches Verhalten aufweisen.³⁹ Ein homogenes System ist *faktorisierbar*, d.h. die n -Schritt-Wahrscheinlichkeit lässt sich als Produkt der Ein-Schritt-Wahrscheinlichkeiten darstellen:

$$p(x^n) = p(x_1) p(x_2) \cdots p(x_n). \quad (4.2.8)$$

Ein nicht homogenes System kann lediglich durch das Produkt der bedingten Wahrscheinlichkeiten charakterisiert werden:

$$p(x^n) = p(x_1) p(x_2|x_1) p(x_3|x_1x_2) \cdots p(x_n|x_1x_2 \dots x_{n-1}). \quad (4.2.9)$$

³⁷ Tsallis (1988).

³⁸ Zu den Einzelheiten vgl. Touchette (2002).

³⁹ Die Einzelheiten und Definitionen solcher Systeme folgen in weiteren Kapiteln dieser Arbeit.

Die fehlende Homogenität kann mit den divergenten Momenten wie z.B. der unendlichen Varianz im Zusammenhang stehen. Dies führt wiederum zu Verteilungen vom Pareto-Typ, die mit der Shannon-Entropie nicht adäquat abgebildet werden können, für die Tsallis-Entropie aber eine ideale Anwendung sind.⁴⁰

Tsallis-Entropie: die resultierende Verteilung.

Im stetigen Fall schreibt sich die Tsallis-Entropie als

$$\mathbb{S}_q(p) = -\frac{1}{1-q} \left(1 - \int_{-\infty}^{\infty} p(x)^q dx \right). \quad (4.2.10)$$

Die Verteilung von Tsallis-Typ lässt für bestimmte Werte des Parameters q fat-tails zu und kann mittels Maximum-Entropie Verfahrens hergeleitet werden. Die Nebenbedingungen an Momente sind durch eine Art Gewichtsfunktion $p(x)^q$ explizit gegeben und erzwingen die Konvergenz der Momente.

Die Lagrange-Funktion $F_q(p)$ kann wie folgt angegeben werden:⁴¹

$$F_q[p(x)] \equiv -\frac{1}{1-q} \left(1 - \int_{-\infty}^{\infty} p(x)^q dx \right) - \beta \int_{-\infty}^{\infty} |x|^v p(x)^q dx + \alpha \int_{-\infty}^{\infty} p(x) dx. \quad (4.2.11)$$

Die Entropie (4.2.10) wird unter der Nebenbedingung maximiert, dass

- i. das v -te „ q -Moment“ (der zweite Summand in der Gleichung) konstant und endlich bleibt, und
- ii. $p(x)$ eine Wahrscheinlichkeitsdichte ist.

In einer Serie von Arbeiten aus der Physik⁴² wurde $v = 2$ gewählt. Dies würde bei den Anwendungen im Finance der Varianz einer *standardisierten Begleitverteilung* entsprechen, da die ursprüngliche Varianz wegen der (angenommenen) fat-tails nicht endlich zu sein braucht.

⁴⁰ In einer Serie von Arbeiten wie z.B. Brody u.a. (2004), Bashkurov (2004) wird der Zusammenhang zwischen der Rényi-Entropie und den Potenzgesetzen untersucht.

⁴¹ Zanette und Alemany (1995).

⁴² Alemany und Zanette (1994), Zanette und Alemany (1995), Tsallis u.a. (1995).

Als Lösung des Maximierungsproblems (4.2.11) resultiert

$$p(x) = A \left(1 - (1 - q) \beta x^2 \right)^{1/(1-q)}, \quad (4.2.12)$$

wobei A eine Normierungsfunktion ist und die beiden Lagrange-Multiplikatoren sich aus den entsprechenden (bekannten) Momenten ergeben.⁴³

Im Einzelnen lässt sich aus der Optimierung⁴⁴ (Maximierung für $q > 1$ und Minimierung für $q < 1$) des Ausdrucks

$$S_q[p] = \frac{1}{1-q} \left\{ \int_{-\infty}^{\infty} \frac{dx}{\sigma} [\sigma p(x)]^q - 1 \right\} \quad (4.2.13)$$

unter der Nebenbedingungen

$$\int_{-\infty}^{\infty} p(x) dx = 1, \quad (4.2.14)$$

$$\int_{-\infty}^{\infty} x^2 P_q(x) dx = \sigma^2, \quad (4.2.15)$$

wobei σ eine positive endliche Konstante und $P_q(x)$ eine Begleitverteilung

$$P_q(x) = \frac{[p(x)]^q}{\int_{-\infty}^{\infty} [p(x)]^q dx} \quad (4.2.16)$$

ist, die Maximum Entropie Tsallis-Verteilung $p_q(x)$ herleiten. Es gilt:⁴⁵

i) für $1 < q \leq 3$

$$p(x) = \frac{1}{\sigma} \left(\frac{q-1}{\pi(3-q)} \right)^{1/2} \frac{\Gamma\left(\frac{1}{q-1}\right)}{\Gamma\left(\frac{3-q}{2(q-1)}\right)} \left(1 - \frac{1-q}{3-q} \frac{x^2}{\sigma^2} \right)^{1/(1-q)}, \quad (4.2.17)$$

ii) für $q = 1$

$$p(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{x^2}{2\sigma^2}\right), \quad (4.2.18)$$

⁴³ Buiatti u.a. (1999), Tsallis (2001).

⁴⁴ Vgl. Prato und Tsallis (1999)

⁴⁵ Prato und Tsallis (1999).

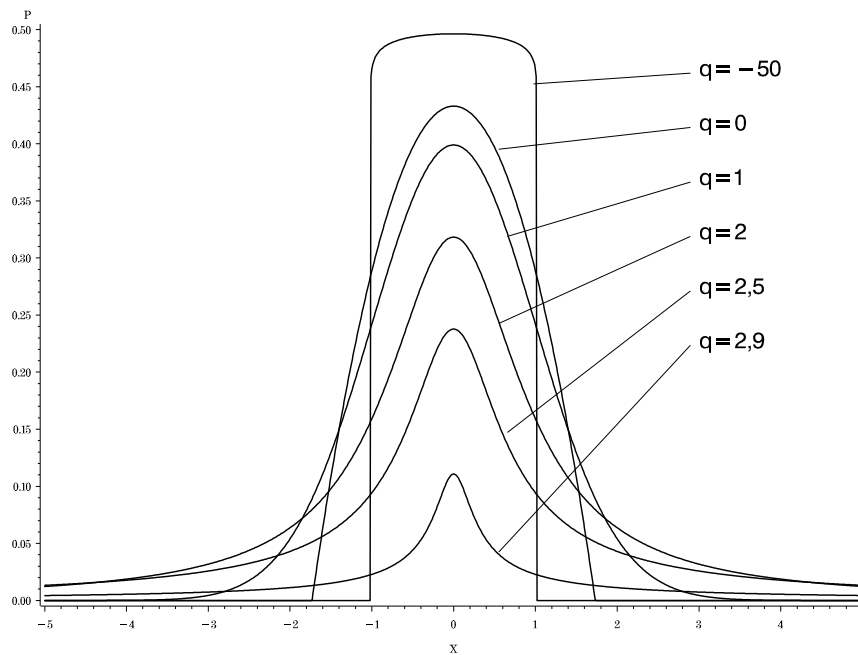


Abb. 4.2. Die Wahrscheinlichkeitsverteilungen $p_q(x)$ für einige Werte des q .

iii) für $q < 1$

$$p(x) = \frac{1}{\sigma} \left(\frac{1-q}{\pi(3-q)} \right)^{1/2} \frac{\Gamma\left(\frac{5-3q}{2(1-q)}\right)}{\Gamma\left(\frac{2-q}{1-q}\right)} \left(1 - \frac{1-q}{3-q} \frac{x^2}{\sigma^2} \right)^{1/(1-q)}, \tag{4.2.19}$$

wenn $x^2 < \sigma^2 [(3-q)/(1-q)]$ und Null sonst.

Prato und Tsallis (1999) zeigten, dass der Träger von $p_q(x)$ für $q \in (-\infty, 1)$ kompakt ist. Die Tails sind exponentiell für $q = 1$ (dieser Spezialfall entspricht der Normalverteilung) und weisen ein Potenzgesetzähnliches Verhalten für $q > 1$ auf (mit $p_q(x) \sim |x|^{2/(1-q)}$ für $|x| \gg \sigma$). Bei einer näheren Betrachtung stellt sich heraus, dass die Tsallis-Verteilung für $1 < q \leq 3$ der Student t -Verteilung mit λ Freiheitsgraden entspricht,⁴⁶ wobei $\lambda \equiv (3-q)/(q-1) > 0$ festgelegt wird:⁴⁷

⁴⁶ Souza und Tsallis (1997).

⁴⁷ μ und σ sind hier entsprechend der Lage- bzw. Skalenparameter.

$$p(x) = \frac{1}{\sigma\sqrt{\pi\lambda}} \frac{\Gamma\left(\frac{\lambda+1}{2}\right)}{\Gamma\left(\frac{\lambda}{2}\right)} \left[1 + \frac{1}{\lambda} \left(\frac{x-\mu}{\sigma}\right)^2\right]^{-\frac{\lambda+1}{2}}. \quad (4.2.20)$$

Es ist bekannt, dass bei der t -Verteilung die

$$\text{Varianz} = \left[\frac{\lambda}{\lambda-2}\right] \sigma^2 = \left[\frac{3-q}{5-3q}\right] \sigma^2 \quad (1 < q < 5/3) \quad (4.2.21)$$

ist und Momente der Ordnung k nur für $k < \lambda$ existieren. Folglich divergiert das zweite Moment für alle $0 < \lambda \leq 2$ bzw. $5/3 \leq q < 3$.⁴⁸

Der Spezialfall $q = 2$ entspricht der Cauchy-Verteilung

$$p(x) = \frac{1}{\pi\sigma \left[1 + \left(\frac{x-\mu}{\sigma}\right)^2\right]}. \quad (4.2.22)$$

Zusammenfassung

Die Entropie als inverses Informationsmaß kann zur Modellierung der Informationsverarbeitung auf dem Kapitalmarkt eingeschränkt verwendet werden. Die Einschränkungen ergeben sich vor allem bei den Annahmen über den Generator der Unsicherheit. Geht man davon aus, dass die Informationsschocks mit einer konstanten Rate die Marktteilnehmer erreichen, lässt sich mittels Entropie eine scharfe Grenze zwischen den typischen und den untypischen Ereignissen festlegen. Asymptotisch werden sämtliche preisrelevanten Informationen durch typische Ereignisse übertragen.

Eine variierende (deterministische oder stochastische) Ankunftsrate der Informationen bringt einen zusätzlichen Freiheitsgrad mit sich und macht damit eine perfekte Informationsverarbeitung entweder aufwendiger oder überhaupt nicht möglich. Es müssen nämlich zu einem Zeitpunkt zwei Aufgaben gelöst werden: die Bestimmung der Ankunftsrate und die Entschlüsselung der damit übertragenen Informationen. Als einfachste Lösung bietet sich das hypothetische Setzen der Ankunftsrate bei den Marktteilnehmern auf den maximalen Wert

⁴⁸ Die Wahrscheinlichkeitsverteilung existiert nicht für $q > 3$, da die Nebenbedingung (4.2.14) für $q > 3$ verletzt wird.

(unter der Annahme, dass dieser Wert existiert, ist er endlich und bekannt), wodurch die Unsicherheit darüber eindeutig eliminiert wird. Dies kann unter Berücksichtigung der damit verbundenen Kosten unrentabel sein.

Alternativ kann eine nichtlineare Transformation des Informationsflusses mittels Begleitverteilungen mit $q < 5/3$ vorgenommen werden mit dem Ziel, entweder eine endliche Varianz der resultierenden Verteilung zu erreichen oder die Budgetbeschränkung vollständig auszunutzen. Im letzten Fall ist die Lösung suboptimal, d.h. die ankommenden Informationen werden entweder nicht vollständig oder teilweise fehlerhaft verarbeitet. Eine scharfe Grenze zwischen den typischen und den untypischen Ereignissen lässt sich nicht mehr festlegen.

Im nächsten Kapitel werden mit Hilfe der Theorie der großen Abweichungen Kriterien aufgezeigt, die zur quantitativen Charakterisierung der untypischen Ereignisse mittels der Raten-Funktion verwendet werden können.



5. Regeln und Ausnahmen II: große Abweichungen

“So what’s new? Everyone knows: Financial markets are risky. But in the careful study of that concept, risk, lies knowledge of our world and hope of a quantitative control over it.”

– Mandelbrot und Hudson (2004), S. 4.

5.1 Large Deviations

Motivation

Die Theorie der großen Abweichungen (Large Deviations, LD) beschäftigt sich mit Konvergenzraten stochastischer Systeme und basiert auf Arbeiten von Cramér (1938) und Chernoff (1952). Nach dem zentralen Grenzwertsatz gehen bestimmte in der Folge dargestellten Wahrscheinlichkeiten gegen Null. Die LD-Theorie geht der Frage nach, wie schnell eine solche Konvergenz erfolgt. Genau genommen lautet die Fragestellung: wie schnell *die Flanken* oder *Tails* einer gegebenen Verteilung gegen die der Normalverteilung konvergieren. Der Schwerpunkt wird darauf gelegt, dass das Konvergenzverhalten an den Flanken sich radikal von dem in der mittleren Region unterscheiden kann, so dass der zentrale Grenzwertsatz nicht mehr zur Hilfe steht. Wir geben hier einige Grundprinzipien dieser Theorie an, die u.a. bei der Betrachtung der Eigenschaften von Multifraktalen sehr hilfreich sein werden. Zur ausführlichen Referenz verweisen wir auf die inzwischen sehr umfangreiche weiterführende Literatur: Ellis (1985, 1995), Deuschel und Strock (1989), Bucklew (1990), Dembo und Zeitouni (1993) sowie Gulinsky und Veretennikov (1993).

Grundlagen

Sei $\{X_k\}$ eine i.i.d. Folge mit $\mathbb{E}X_1 = m < \infty$. Nach dem schwachen Gesetz der großen Zahlen konvergiert die Teilsumme

$$\frac{1}{n} \sum_{k=1}^n X_k := \frac{S_n}{n}$$

gegen den Mittelwert m \mathbb{P} -f.s. für $n \rightarrow \infty$. Ein Ereignis

$$\sum_{k=1}^n X_k > na, \quad a > m \quad \text{oder} \quad \sum_{k=1}^n X_k < na, \quad a < m$$

wird damit zu einem *seltenen Ereignis*. Nehmen wir an, $\{X_k\}$ hat Mittelwert Null und Varianz Eins. Nach dem zentralen Grenzwertsatz konvergiert für jedes x und $n \rightarrow \infty$ die Verteilung

$$F_n(x) = \mathbb{P} \left(\frac{1}{\sqrt{n}} \sum_{i=1}^n X_k < x \right)$$

gegen die Standardnormalverteilung $N(0, 1)$:

$$\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x e^{-u^2/2} du.$$

In der Theorie der großen Abweichungen ist das grobe asymptotische Verhalten der Wahrscheinlichkeit

$$\mathbb{P} \left(\sum_{i=1}^n X_k > na \right)$$

von Interesse, wobei $a > 0$ unabhängig von n ist. Genau genommen ist es nicht die Wahrscheinlichkeit selbst, sondern ihr Logarithmus, der hier untersucht wird. Zu den klassischen Ergebnissen von Cramér (1938) und Chernoff (1952) gehört die Existenz des folgenden Grenzwertes:¹

$$\lim_{n \rightarrow \infty} n^{-1} \ln \mathbb{P} \left(\sum_{i=1}^n X_k > na \right) = -I(a), \quad (5.1.1)$$

wobei für $I(a)$ eine spezielle Bezeichnung vorgesehen ist: die *Ratenfunktion*. Sie ist mittels der Legendre Transformation mit der kumulantenzeugenden Funktion $\varphi(t) = \ln \mathbb{E} \exp(tX)$ verbunden:

¹ Siehe auch Cramér-Bedingung (5.1.4).

BEISPIEL 5.1

Raten-Funktionen einiger Verteilungen:

Bernoulli (p)

$$P(X = 1) = p, \quad P(X = 0) = 1 - p$$

$$\varphi(t) = \ln(pe^t + 1 - p)$$

$$I(a) = \begin{cases} a \ln \frac{a}{p} + (1-a) \ln \frac{1-a}{1-p} & \text{für } 0 < a < 1 \\ -\ln(p) & \text{für } a = 1 \\ -\ln(1-p) & \text{für } a = 0 \\ \infty & \text{sonst,} \end{cases}$$

Exp (λ)

$$p(x) = \lambda e^{-\lambda x} \quad \text{für } x \geq 0$$

$$\varphi(t) = \frac{\lambda}{\lambda - t} \quad \text{für } t < \lambda$$

$$I(a) = \begin{cases} \lambda a - 1 - \ln(\lambda a) & \text{für } a > 0 \\ \infty & \text{für } a \leq 0, \end{cases}$$

Poisson (λ)

$$p(X = x) = \frac{\lambda^x e^{-\lambda}}{x!} \quad \text{für } \lambda > 0$$

$$\varphi(t) = \lambda(e^t - 1)$$

$$I(a) = \begin{cases} a(\ln \frac{a}{\lambda} - 1) + \lambda & \text{für } a \geq 0 \\ \infty & \text{für } a < 0, \end{cases}$$

Cauchy

$$p(x) = \frac{1}{\pi(1+x^2)}$$

$$M(t) = \infty \quad \text{für } t \neq 0$$

$$I(a) = 0 \quad \text{für } a \in \mathbb{R}$$

$$I(a) = \sup_{t \in \mathbb{R}} [at - \varphi(t)]. \quad (5.1.2)$$

Die Raten-Funktion $I(a) : \mathbb{R} \rightarrow \mathbb{R}_+$ charakterisiert die exponentielle Seltenheit der großen Abweichungen vom Mittelwert: die *höheren* Werte dieser Funktion entsprechen asymptotisch den *weniger* wahrscheinlichen Ereignissen. Im Einklang mit dem zentralen Grenzwertsatz nimmt $I(a)$ das Minimum bei Mittelwert $a = m$ an. Aus (5.1.2) folgt, dass $I(m) = 0$. Ein sehr wichtiger Punkt ist, dass der zentrale Grenzwertsatz nur für den zentralen Bereich der Verteilung gute Ap-

proximationen liefert. Er kennt sozusagen keine Flanken außer Gauschen: unabhängig davon, wie die einzelnen „Bestandteile“ verteilt sind, ist der „Korb“ immer Normalverteilt. Daraus resultieren immer exponentielle Abschwingraten an den Flanken. Diese Näherung kann im Einzelfall sehr grob sein.

Nehmen wir als Beispiel einen Random Walk. Es seien $\{X_k\}$ i.i.d. mit $\mathbb{P}(X_k = 1) = \mathbb{P}(X_k = -1) = 1/2$. Wir definieren

$$S_n := \sum_{k=1}^n X_k, \quad x_t^{(n)} := n^{-1} S_{\lfloor nt \rfloor}. \quad (5.1.3)$$

Nach dem zentralen Grenzwertsatz konvergiert die n -indexierte Familie der Prozesse $\{x_t^{(n)}, t \geq 0\}$ gegen ihren Mittelwert (Null) und bildet eine Familie der skalierten Wiener Prozesse $n^{-1/2} \{W_t, t \geq 0\}$ nach. Allerdings ist die Gültigkeit dieser Aussage auf den Bereich von wenigen Standardabweichungen vom Erwartungswert beschränkt. Diesen Bereich nennt man *kleine Abweichungen (small deviations)* mit $S_n = O(n^{1/2})$. Für ein fixes $t > 0$ wird das Ereignis $S_{\lfloor nt \rfloor} \geq an$ für $n \rightarrow \infty$ zu einem *exponentiell unwahrscheinlichen Ereignis*. Das heißt, $\mathbb{P}(S_{\lfloor nt \rfloor} \geq an)$ verhält sich asymptotisch wie $\exp(-nI(a))$. Die Rolle der Raten-Funktion $I(a)$ wird hier besonders anschaulich: sie bestimmt die Geschwindigkeit, mit der diese Wahrscheinlichkeit gegen Null geht. Im Bezug auf die reskalierten Prozesse $x_t^{(n)}$ heißt das Ereignis $x_t^{(n)} \geq a$ ein *a-überschreitendes Ereignis* oder ein *tail-Ereignis*.

Außerhalb vom zentralen Bereich unterscheiden sich für ein fixes $t > 0$ die Wahrscheinlichkeiten $\mathbb{P}(x_t^{(n)} \geq a)$ und $\mathbb{P}(n^{-1/2}W_t \geq a)$ sehr wohl. Woran lässt sich das erkennen? Die *exponentiellen Abschwingraten* dieser Ereignisse bei $n \rightarrow \infty$ sind unterschiedlich! Ein Random Walk kann sich nicht *zu sehr* ausweiten. Die Wahrscheinlichkeiten der großen Abweichungen sind Null (f.s.). Um dies zu zeigen, berechnen wir die kumulanterzeugende Funktion $\phi(t)$:

$$\phi(t) = \ln \int_{-\infty}^{\infty} e^{tX} dF(x) = \ln \left(\frac{1}{2} (e^t + e^{-t}) \right) = \ln [\cosh(t)].$$

Damit ergibt sich nach (5.1.2) die Raten-Funktion

$$I(a)_{\text{RW}} = \begin{cases} \frac{1-a}{2} \ln(1-a) + \frac{1+a}{2} \ln(1+a) & \text{falls } |a| < 1, \\ +\infty & \text{falls } |a| \geq 1 \end{cases}$$

für den Random Walk. Sind $\{X_k\}$ unabhängig normalverteilt $N(\mu, \sigma^2)$, dann gilt

$$\varphi(t) = -t\mu + \frac{1}{2}t^2\sigma^2 \quad \text{und} \quad I(a)_{WP} = \frac{(a-\mu)^2}{2\sigma^2}$$

für einen Wiener Prozess. Es wird deutlich, dass die Wahrscheinlichkeiten der großen Abweichungen beim Random Walk *unendlich schnell* gegen Null gehen. Beim Wiener Prozess gehen sie nur *exponentiell schnell* gegen Null, wenn auch quadratisch im Exponenten.

Cramér

Der Cramérsche Satz stellt eine der grundlegenden Aussagen der LD-Theorie dar indem er Schranken für die Raten-Funktion vorgibt. Damit lässt sich u.a. das obige Beispiel formalisieren.

Es sei $\{X = (X_k), k \geq 1\}$ eine i.i.d. Folge mit Wahrscheinlichkeitsverteilung $F(x)$, definiert auf einem Wahrscheinlichkeitsraum $(\Omega, \mathcal{F}, \mathbb{P})$ mit Werten auf \mathbb{R} . Von Interesse ist das asymptotische Verhalten dieser Folge sowie das der Summe $S_n = \sum_{k=1}^n X_k$. Unter der Annahme, dass ein Mittelwert $m = \int_{\mathbb{R}} x d\mathbb{P}$ existiert, konvergiert die durch n normalisierte Summe nach dem starken Gesetz der großen Zahlen

$$S_n/n \longrightarrow m, \quad \mathbb{P}\text{-fast sicher.}$$

Dies impliziert, dass die Wahrscheinlichkeit für ein Ereignis A , den Stichprobendurchschnitt $\bar{X} = S_n/n$ für $n \rightarrow \infty$ *nicht* innerhalb einer ε -Umgebung von m zu finden, gegen Null geht. Tritt ein solches Ereignis trotzdem ein, spricht man von einer *großen Abweichung*. Intuitiv würde man einen Zusammenhang zwischen der Abweichungsgröße ε und n vermuten. Die LD-Theorie macht es möglich, solche Zusammenhänge mathematisch zu erfassen, zu begründen und zu quantifizieren.

Als erstes werden die Eigenschaften der Raten-Funktion näher spezifiziert. Eine Funktion $I(x)$ wird zur Raten-Funktion, wenn sie nichtnegativ und semistetig von unten (lower semicontinuous) ist. Sind außerdem die Mengen $\{x : I(x) \leq c\}$ für jedes c kompakt, heißt I eine *gute* Raten-Funktion.

Im zweiten Schritt wird mittels Raten-Funktion der eigentliche *Grundsatz der großen Abweichungen (large deviations principle, LDP)* eingeführt. Danach existieren unter der *Cramér-Bedingung*

$$M(t) = \mathbb{E} \exp(tX_1) < \infty \quad \text{für alle } t \in \mathbb{R} \quad (5.1.4)$$

die folgenden Grenzwerte:

$$\limsup_{n \rightarrow \infty} \frac{1}{n} \log \mathbb{P} \left(\frac{S_n}{n} \in F \right) \leq - \inf_{x \in F} I(x) \quad (5.1.5)$$

für alle *abgeschlossenen* Mengen F und

$$\liminf_{n \rightarrow \infty} \frac{1}{n} \log \mathbb{P} \left(\frac{S_n}{n} \in G \right) \geq - \inf_{x \in G} I(x). \quad (5.1.6)$$

für alle *offenen* Mengen G .

Die *Cramér-Bedingung* verlangt, dass die momenterzeugende Funktion endlich ist. Dies sichert automatisch die Existenz der kumulanterzeugenden Funktion

$$\varphi(t) \equiv \ln M(t) < \infty \quad \text{für alle } t \in \mathbb{R}. \quad (5.1.7)$$

Die Funktion $\varphi(t)$ ist eine *konvexe* Funktion. Außerdem gilt nach (A.2.29) $\varphi_n(t) = n\varphi(t)$. Damit ist $\varphi(t)$ unter der Cramér-Bedingung auch eine *glatte* Funktion. Die Legendre-Transformierte der kumulanterzeugenden Funktion $L(x) = \sup_t [xt - \varphi(t)]$ ist nichtnegativ und konvex.

Satz 5.1.1 (Cramér) *Unter der Cramér-Bedingung (5.1.4) und einer zusätzlichen Regularitätsbedingung*

$$\lim_{|t| \rightarrow \infty} \frac{\varphi(t)}{|t|} = \infty \quad (5.1.8)$$

genügt eine Folge S_n/n auf \mathbb{R} dem Grundsatz der großen Abweichungen mit einer guten Raten-Funktion

$$I(x) = L(x) = \sup_{t \in \mathbb{R}} [xt - \varphi(t)].$$

Beweis: Gulinsky und Veretennikov (1993), S. 11. □

Im ersten Teil des Satzes wird der Wachstumsbereich für die kumulanterzeugende Funktion festgelegt: sie soll langsamer steigen als $\exp(t)$, aber schneller als jede lineare Funktion. Der zweite Teil weist die Übereinstimmung der oberen (5.1.5) und der unteren (5.1.6) Schranken nach.

Aus dem Satz von Cramér folgt, dass die asymptotische Wahrscheinlichkeit eines innerhalb eines Intervalls $[a, b]$ liegenden Stichprobendurchschnittes \bar{X} an einem Punkt konzentriert ist. Liegt m ebenfalls in diesem Intervall, dann gilt für die Raten-Funktion $I_{[a,b]} = 0$ und die Wahrscheinlichkeit geht nach dem Gesetz der großen Zahlen gegen Ein. Liegt m dagegen außerhalb von $[a, b]$, handelt es sich um eine große Abweichung. Die Wahrscheinlichkeit dafür geht gegen Null mit der Rate $0 < I_{[a,b]} = \inf_{x \in [a,b]} I(x) = I(x^*)$. Die Raten-Funktion stellt also ein geeignetes Werkzeug dar, um die Schwankungen um den Mittelwert zu charakterisieren. Lassen sich die asymptotischen Werte nicht ermitteln, kann alternativ die Chernoff-Schranke verwendet werden:

$$\mathbb{P}(S_n \geq na) \leq \exp\left(-n \sup_{t \geq 0} [at - \varphi(t)]\right). \quad (5.1.9)$$

Der Satz von Cramér (5.1.1) lässt sich weit über die Summen von i.i.d. Zufallsvariablen verallgemeinern. Unter der Annahme, dass der Grenzwert

$$\varphi(t) = \lim_{n \rightarrow \infty} \frac{1}{n} \varphi_n(t) \quad \text{mit} \quad \varphi_n(t) = \ln E[\exp(tS_n)] \quad (5.1.10)$$

existiert und wohldefiniert ist, können die Schranken für eine beliebige (zum Beispiel autokorrelierte) Folge von Zufallsvariablen weiterhin durch die kumulanterzeugende Funktion $\varphi(t)$ angegeben werden.

Gärtner-Ellis

Sei $\{S_n\}$ eine Folge der Zufallsvariablen, die nicht unbedingt unabhängig sind, mit entsprechenden kumulanterzeugenden Funktionen (5.1.10). Sind X_K i.i.d., dann hängt $\varphi_n(t)$ nicht von n ab, es ist kein Übergang zum Grenzwert notwendig und $\varphi(t)$ ist stetig und differenzierbar. Im Allgemeinen ist aber die Stetigkeit nicht gegeben bzw. der Grenzwert braucht nicht zu existieren. Unter der Annahme, dass der Grenzwert existiert und die sich daraus ergebende Funktion differenzierbar ist, können die Schranken für die großen Abweichungen auch ohne die i.i.d. Bedingung hergeleitet werden. Dies ist der Gegenstand des Satzes von Gärtner-Ellis. Die obere Schranke ist allein durch die Existenz der kumulanterzeugenden Funktion $\varphi(t)$ gegeben. Unter milden technischen Bedingungen² hält auch die untere Schranke.

² Die Funktion $\varphi(t)$ soll unterhalbstetig, nicht gleich minus unendlich, glatt genug und endlich in einer Umgebung von Null sein (Gulinsky und Veretennikov (1993), S. 14.).

Satz 5.1.2 (Gärtner-Ellis) *Unter obigen Annahmen genügt $\{S_n\}$ dem Grundsatz der großen Abweichungen mit der Raten-Funktion*

$$I(x) = L(x) = \sup_t [xt - \phi(t)]. \quad (5.1.11)$$

Beweis: Dembo und Zeitouni (1993), Theorem 2.3.6. □

Mit anderen Worten, für große n und kleine $\varepsilon > 0$ gilt die Abschätzung

$$\mathbb{P}[S_n \in (na - n\varepsilon, na + n\varepsilon)] \sim \exp[-nI(a)]. \quad (5.1.12)$$

Ausgehend von (5.1.12) wird die Funktion $L(x)$ in (5.1.11) oft als *Entropie der großen Abweichungen* bezeichnet.

Sanov

Sei $\{X_k\}$ eine i.i.d. Folge. Die Wahrscheinlichkeitsverteilung dieser Folge sei nicht bekannt und soll auf Grundlage von ersten n Beobachtungen geschätzt werden. Betrachten wir ein *empirisches Maß*

$$\mu_n(A, \omega) \equiv \sum_{k=1}^n \mathbb{1}(X_k \in A).$$

Dieses Maß ist auf einem Messraum $\mathcal{M}(\mathbb{R})$ der Borel-Untermengen $\mathcal{B}(\mathbb{R})$ definiert und A ist eine davon: $A \subset \mathcal{B}(\mathbb{R})$. Das Maß $\mu_n(A, \omega)$ gibt Auskunft über die relative Frequenz, mit der die Folge $\{X_k\}$ in A vorkommt. Für jedes ω ist $\mu_n(\cdot, \omega)$ ein Element aus $\mathcal{M}(\mathbb{R})$, d.h. μ_n ist ein *Zufallsmaß*. Der Erwartungswert von $\mu_n(\cdot, \omega)$ ist durch die Teilsumme S_n/n der ersten n Beobachtungen gegeben. Wenn n gegen unendlich geht, erwarten wir nach dem starken Gesetz der großen Zahlen, dass

$$\mu_n(\cdot, \omega) \longrightarrow \mu, \quad \mathbb{P} - \text{fast sicher.} \quad (5.1.13)$$

Womit kann man aber die Verteilung \mathbb{P}_μ schätzen? Sei \mathcal{P}_n die Verteilung von $\mu_n(\cdot, \omega)$. Der Satz von Glivenko-Cantelli³ besagt, dass die Konvergenz in (5.1.13) gleichmäßig auf $\mathcal{M}(\mathbb{R})$ erfolgt, d.h.

$$D_n(\omega) \equiv \sup_A |\mu_n(A, \omega) - \mu(A)| \rightarrow 0 \quad \mathbb{P} - \text{fast sicher.}$$

³ Vgl. Billingsley (1986), S. 269.

BEISPIEL 5.2 Das Zufallsmaß (Vgl. Bryc (2000), S. 60.)

Es sei $\mathcal{M} = \{0, 1\}$ und $\{X_k\}$ eine i.i.d. Folge derart, dass $\mathbb{P}(X = 1) = p$. Dann ist μ_n ein Zufallsmaß beschrieben durch die Häufigkeit der Einsen in der Stichprobe: $\mu_n(\{1\}) = \hat{p}_n$. Man kann sich μ_n als ein Zufallsmaß $\hat{q}_n \delta_0 + \hat{p}_n \delta_1$ vorstellen.

Eine andere Möglichkeit zur Beschreibung des Maßes μ_n wäre eine \mathbb{R}^2 -wertige Zufallsvariable $(\hat{p}_n, \hat{q}_n) \in \mathbb{R}^2$, wobei $\hat{q}_n = 1 - \hat{p}_n$ die Häufigkeit der Nullen in der Stichprobe ist.

Bezeichnet man mit $\mathcal{P}(\mathcal{M})$ eine Menge aller Wahrscheinlichkeitsmaße auf \mathcal{M} , dann kann μ_n schließlich als eine $\mathcal{P}(\mathcal{M})$ -wertige Zufallsvariable interpretiert werden. Man kann sogar seine Verteilung finden. Falls $\mathbf{p} \in \mathcal{P}$, dann gilt:

$$\mathbb{P}(\mu_n = \mathbf{p}) = \begin{cases} \binom{n}{k} p^k (1-p)^{n-k}, & \text{falls } \mathbf{p}(\{1\}) = k/n, \\ 0 & \text{sonst.} \end{cases} \quad (5.1.14)$$

Dies impliziert, dass D_n nicht von einem bestimmten A abhängig ist. Daraus folgt

$$\mathcal{P}_n \longrightarrow \delta_\mu \quad \text{für } n \rightarrow \infty, \quad (5.1.15)$$

wobei δ_μ ein Punkt-Maß (Dirac Delta-Funktion) bezeichnet, das auf μ konzentriert ist. Etwas ungewöhnlich ist hier, dass es sich bei \mathcal{P}_n um die „Verteilung der Verteilungen“ handelt.⁴ Der Grundsatz der großen Abweichungen besteht in diesem Fall darin, eine Wahrscheinlichkeit zu finden, so dass das Maß $\mu_n(\cdot, \omega)$ einem anderen Maß $\nu \in \mathcal{M}(\mathbb{R})$ bei $n \rightarrow \infty$ nahe liegt. Mit anderen Worten: von Interesse ist die logarithmische Asymptotik der großen Abweichungen von \mathcal{P}_n .

Der Satz von Sanov gibt die Raten-Funktion für den Grundsatz der großen Abweichungen in Bezug auf ein empirisches Maß an. Es stellt sich heraus, dass diese Raten-Funktion mit der *relativen Entropie* oder *Kullback-Leibler Distanz* zu einem anderen Maß ν übereinstimmt. Es sei \mathcal{A} eine Menge von Wahrscheinlichkeitsmaßen auf einem Messraum $\mathcal{M}(\mathbb{R})$. Die Wahrscheinlichkeit, dass sich das empirische Maß μ_n für große n in einer Menge \mathcal{A} der Wahrscheinlichkeitsmaße ν befindet, verhält sich asymptotisch wie $\exp(-n\mathbb{D}(\mathcal{A}|\mu))$ mit

$$\exp(-n\mathbb{D}(\mathcal{A}|\mu)) = \inf_{\nu \in \mathcal{A}} \exp(-n\mathbb{D}(\nu|\mu))$$

und μ aus (5.1.13). Die Bezeichnung $\exp(-n\mathbb{D}(\nu|\mu))$ steht wie üblich für die relative Entropie zwischen ν und μ . Die Raten-Funktion

⁴ Die Fragestellungen aus diesem Bereich gehören nach Ellis (1985) zum „Level II Grundsatz der großen Abweichungen.“

$I(\nu)$ lässt sich also durch die relative Entropie $\mathbb{D}(\nu\|\mu)$ in Bezug auf den Grenzwert (5.1.13) charakterisieren, wobei ν ein anderes Wahrscheinlichkeitsmaß ist.

Es sei (Ω, \mathcal{F}) ein Messraum. Für zwei Wahrscheinlichkeitsmaße ν und μ und eine beschränkte messbare Funktion Φ auf (Ω, \mathcal{F}) definieren wir⁵

$$L(\nu) = \sup_{\Phi} \left[\int \Phi \, d\nu - \ln \int \exp(\Phi) \, d\mu \right]. \quad (5.1.16)$$

Satz 5.1.3 (Sanov) *Es sei $\{X_k\}$ eine Folge von unabhängigen Zufallsvariablen mit der Wahrscheinlichkeitsverteilung μ und μ_n ein empirisches Maß von $\{X_k\}$. Dann genügt μ_n dem Grundsatz der großen Abweichungen mit der Raten-Funktion $I(\nu) = L(\nu)$ nach (5.1.16) und es gilt*

$$I(\nu) = \mathbb{D}(\nu\|\mu).$$

Beweis: Vgl. Gulinsky und Veretennikov (1993), S. 18 bzw. Bryc (2000), S. 81 und Referenzen dort. \square

Die Erkenntnis, dass durch den Satz von Sanov die Raten-Funktion einer empirischen Verteilung angegeben werden kann, ist ohne Zweifel von großer Bedeutung. Andererseits kommt ein praktisches Problem auf: wie bestimmt man das Supremum (5.1.16) bzw. (5.1.11)? Eine Möglichkeit dafür ist die Maßtransformation.

Maßtransformation und große Abweichungen

Sei $X = (X_k)$ eine stationäre ergodische⁶ Folge mit $\mathbb{E}X_1 = 0$. Wir definieren eine Filtration $\mathbb{F} = (\mathcal{F}_k)$, $\mathcal{F}_k = \sigma(X_i, -\infty < i < k)$, $k \geq 0$ sowie eine weitere Folge der Teilsummen $S = (S_n)$, $n \geq 1$ mit $S_n = \sum_{k=1}^n X_k$ und nehmen an, dass X der Cramér-Bedingung (5.1.4) genügt. Weiterhin definieren wir einen *stochastischen Exponent*

$$\mathcal{E}_n(t) = \prod_{k=1}^n \mathbb{E}(\exp(tX_k) \mid \mathcal{F}_{k-1})$$

⁵ Vgl. Formel (4.1.19) auf S. 70.

⁶ Unter einer ergodischen Folge verstehen wir eine Folge der *typischen* Zufallsvariablen im Sinne des Satzes (4.1.2) von Shannon-McMillan-Breiman, vgl. S. 63 bzw. eine Folge, für die das schwache Gesetz der großen Zahlen gilt. Eine ausführliche Diskussion zur Ergozität findet man in Gray (1988) bzw. Petersen (1989).

und einen *stochastischen Kumulant*

$$\mathcal{K}_n(t) \equiv \ln \mathcal{E}_n(t) = \sum_{k=1}^n \ln \mathbb{E}(\exp(tX_k) | \mathcal{F}_{k-1}).$$

Sind X_k i.i.d., dann gilt auch

$$\mathcal{K}_n(t) = n \ln \mathbb{E} \exp(tX_k) = n\varphi(t). \quad (5.1.17)$$

Es ist bekannt, dass der Kumulant einer stationären ergodischen Folge selbst eine stationäre Folge darstellt. Damit existiert nach dem Ergodensatz von Birkhoff⁷ eine Funktion $\varphi(t)$ derart, dass der Grenzwert

$$\varphi(t) = \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{k=1}^n \ln \mathbb{E}(\exp(tX_k) | \mathcal{F}_{k-1}) = \mathbb{E}(\ln \mathbb{E}(\exp(tX_1) | \mathcal{F}_0)) \quad (5.1.18)$$

gegeben ist. Wir nehmen an, dass die Regularitätsbedingung (5.1.8) hält und beachten, dass der stochastische Kumulant und $\varphi(t)$ konvexe Funktionen sind. Mit der Festlegung

$$L(x) = \sup_{t \in \mathbb{R}} [tx - \varphi(t)] \quad (5.1.19)$$

bleibt nur noch festzustellen, unter welchen Bedingungen $L(x)$ die Raten-Funktion für die großen Abweichungen der Folge (S_n/n) , $n \geq 1$ ist. Gulinsky und Veretennikov (1993) zeigen, dass $L(x)$ genau dann zur Raten-Funktion wird, wenn die Folge (5.1.18) *exponentiell schnell* konvergiert, d.h. für jedes $\varepsilon > 0$ gilt:

$$\lim_{n \rightarrow \infty} \frac{1}{n} \ln \mathbb{P} \left[\left| \frac{1}{n} \sum_{k=1}^n \ln \mathbb{E}(\exp(tX_k) | \mathcal{F}_{k-1}) - \varphi(t) \right| > \varepsilon \right] = -\infty. \quad (5.1.20)$$

Wie sind diese Ergebnisse zu interpretieren? Es wurde gezeigt, dass bei oben spezifizierten Annahmen eine *große* Abweichung des stochastischen Kumulanten von seinem Erwartungswert $\varphi(t)$ äußerst unwahrscheinlich ist. Es gilt noch mehr: die Ergebnisse bleiben für eine *nicht-stationäre Folge* X_k ebenfalls gültig, solange die Bedingung (5.1.20) nicht verletzt ist.⁸

Unter der Annahme, dass (5.1.20) hält, kann aus (5.1.19) eine eindeutige Lösung für die Raten-Funktion bestimmt werden. Wegen der

⁷ Gray (1988).

⁸ Gulinsky und Veretennikov (1993), S. 88.

Regularitätsbedingung (5.1.8) wird das Supremum in einem Punkt, sagen wir τ , für ein beliebiges x erreicht. Dann gilt $L(x) = \tau x - \varphi(\tau)$ und

$$\left. \frac{d\varphi(t)}{dt} \right|_{t=\tau} = \mathbb{E}^{\mathbb{P}} \left[X_1 \frac{e^{\tau X_1}}{\mathbb{E} e^{\tau X_1}} \right] = x. \quad (5.1.21)$$

Cramér's ursprüngliche Idee, die untere Schranke der Raten-Funktion zu beweisen, bestand darin, eine Maßtransformation durchzuführen. Unter einem neuen Maß sollte ein *seltenes* Ereignis zu einem *typischen* Ereignis werden. Es sei

$$Z^n(t) = \prod_{k=1}^n \frac{e^{tX_k}}{\mathbb{E}(e^{tX_k} | \mathcal{F}_{k-1})} = \exp(tS_n - \mathcal{K}_n(t)). \quad (5.1.22)$$

Dieser Ausdruck ist ein geeigneter Kandidat für die Radon-Nikodym-Ableitung. Tatsächlich, $Z^n(t)$ ist ein Martingal:

$$\mathbb{E}(Z^n(t) | \mathcal{F}_{n-1}) = \mathbb{E} \left(\prod_{k=1}^n \frac{e^{tX_k}}{\mathbb{E}(e^{tX_k} | \mathcal{F}_{k-1})} \middle| \mathcal{F}_{n-1} \right) \quad (5.1.23)$$

$$= Z^{n-1}(t) \mathbb{E} \left(\frac{e^{tX_n}}{\mathbb{E}(e^{tX_n} | \mathcal{F}_{n-1})} \middle| \mathcal{F}_{n-1} \right) = Z^{n-1}(t), \quad (5.1.24)$$

wobei der Übergang von (5.1.23) zu (5.1.24) im einfachsten Fall unter der Annahme geschieht, die X_k seien i.i.d. Diese Annahme ist offensichtlich sehr stark. Sie kann abgeschwächt werden, indem man für die Ursprungsfolge X_k nur noch gewisse ergodische Eigenschaften unterstellt. Sogar unter diesen schwächeren Bedingungen bleibt die Gültigkeit der Ergebnisse beispielsweise für die autoregressiven Folgen oder Summen der Semimartingale erhalten.⁹ Ausgehend von (5.1.24) definieren wir ein neues Wahrscheinlichkeitsmaß \mathbb{Q} durch

$$d\mathbb{Q}^{n,t} = Z^n(t) d\mathbb{P} \quad (5.1.25)$$

und stellen fest, dass sich aus (5.1.21) nach dem schwachen Gesetz der großen Zahlen die folgenden Beziehungen ergeben:

⁹ Gulinsky und Veretennikov (1993), Kapitel 4.

$$\lim_{n \rightarrow \infty} \mathbb{Q}^{n, \tau} \left[\left| \frac{S_n}{n} - x \right| > \delta \right] = 0 \quad \text{bzw.} \quad (5.1.26)$$

$$\lim_{n \rightarrow \infty} \mathbb{Q}^{n, \tau} \left[\left| \frac{S_n}{n} - x \right| < \delta \right] = 1, \quad \delta > 0. \quad (5.1.27)$$

Die normierte Summe S_n/n genügt also dem Grundsatz der großen Abweichungen mit der Raten-Funktion $L(x)$, die für $E^{\mathbb{Q}}[S_n/n] = x$ ihren minimalen Wert $L(x) = 0$ annimmt. Bei jedem anderen Maß \mathbb{P} anstelle \mathbb{Q} entfernt man sich von den *typischen* Ereignissen, die nahe ihrem empirischen Mittelwert liegen, in Richtung *seltener* Ereignisse, die die *großen* Abweichungen von dem Mittelwert aufweisen, dafür aber mit einem exponentiellen Korrekturfaktor $L(x) > 0$ versehen werden.

Zusammenfassend liegt also eine zweifache Parametrisierung vor. Im ersten Schritt wird der erste Freiheitsgrad aufgelöst, indem die Raten-Funktion $L(x)$ unter Berücksichtigung einer der Beziehungen (5.1.17) oder (5.1.18) explizit angegeben wird. Danach ist man nicht mehr an einen Grenzwert gebunden und sucht im zweiten Schritt nach einem Wahrscheinlichkeitsmaß, das x zu einem typischen oder sogar zu dem wahrscheinlichsten Ereignis macht. Dieser Vorgang ist äquivalent zur Minimierung der Raten-Funktion $L(x)$ über t und kann sehr anschaulich mittels einer Maßtransformation dargestellt werden. Da die Funktion $\varphi(t)$ stetig und differenzierbar ist, existiert ein endliches τ derart, dass

$$\varphi(\tau) = \sup_{t \in \mathbb{R}} [tx - \varphi(t)] \quad \text{und} \quad \varphi'(\tau) = x.$$

Wird ein neues Wahrscheinlichkeitsmaß \mathbb{Q} durch

$$\frac{d\mathbb{Q}}{d\mathbb{P}}(x) = \exp(\tau x - \varphi(\tau))$$

definiert, dann ist leicht zu sehen, dass \mathbb{Q} ein Martingalmaß ist, weil unter der Cramér-Bedingung (5.1.4)

$$\int_{\mathbb{R}} d\mathbb{Q} = \frac{1}{M(\tau)} \int_{\mathbb{R}} e^{\tau x} d\mathbb{P} = 1$$

gilt. Daraus ergibt sich unmittelbar

$$E^{\mathbb{Q}}[X_1] = \frac{1}{M(\tau)} \int_{\mathbb{R}} x e^{\tau x} d\mathbb{P} = \varphi'(\tau) = x.$$

Zusammenfassung

Unter der Cramér-Bedingung (5.1.4) lassen sich die typischen und die untypischen Ereignisse (exponentiell) scharf trennen. In der Regel nimmt die Raten-Funktion kleine positive Werte an, was ein Merkmal typischer Ereignisse ist. Weiterhin impliziert die Gültigkeit der Cramér-Bedingung, dass die Flanken der Verteilung sowohl für die einzelne Zufallsvariable als auch für eine endliche Summe exponentiell schnell abklingen. Dies hat zur Folge, dass die großen Abweichungen vom Mittelwert eindeutig als Ausnahmen klassifiziert werden können und in ihrer Wirkung als Einzelereignisse *im Vergleich zur Summe* vernachlässigbar klein sind. In Bezug auf die Informationsverarbeitung auf dem Kapitalmarkt sind die durch die großen Abweichungen verursachten Renditeausschläge unter der Cramér-Bedingung nicht informativ, da der wahrscheinlichste Pfad eines stochastischen Prozesses ein Martingalpfad ist.

6. Modellierung der Langzeitabhängigkeiten

“... sheer luck made me interpret it as a symptom of scaling — that is, of fractality. For this homeless truth I created a proper home in the concept of long dependence, and extended it to the study of financial prices.”

– Mandelbrot und Hudson (2004), S. 179.

6.1 Das Langzeitgedächtnis

Grundlagen

Es ist wirklich erstaunlich, dass Versuche, statistische Abhängigkeiten in den Renditezeitreihen zu finden, nach vielen Jahren immer noch unternommen werden. Es werden auch welche gefunden, obwohl die Annahme der Unabhängigkeit oder einer schwachen (zeitlich begrenzten) Abhängigkeit immer wieder bestritten wird.

Wir wollen uns etwas tiefer mit einer bestimmten Art der Abhängigkeit, der sogenannten *Langzeitabhängigkeit* (*long memory, long-range dependence, LRD*) beschäftigen. Darunter versteht man gewöhnlich eine Skalierungseigenschaft, die sich über mehrere Zeitskalen in der Größenordnung von Intraday-Daten bis zu einigen Jahrzehnten auf täglicher Basis erstreckt. Die Langzeitabhängigkeit soll nicht mit einem globalen langfristigen Trend verwechselt werden; dies wäre nur die halbe Wahrheit. Die Abhängigkeit besteht wirklich auf *mehreren* Zeitskalen, von kleinen bis sehr großen. Im einfachsten Fall ist die Skalierungsart auf allen Skalen gleich und lässt sich mit einem einzigen Parameter beschreiben. Sollten einige Skalen das Verhalten der

Zeitreihe stärker beeinflussen als die anderen, reicht ein Parameter nicht mehr aus. In diesem Fall ist es sinnvoll, einen funktionalen Zusammenhang einzuführen, um die gewünschte Parametermenge zu beschreiben bzw. die unbekannt Parameter eines vorher festgelegten Modells zu schätzen.

Die Langzeitabhängigkeit als Forschungsgebiet hat eine relativ kurze Geschichte. Es hat aber einige Zeit gedauert, bevor ein wichtiger Schritt von der einfachen Beobachtung der Phänomene bis zur Entwicklung erster statistischer Modelle zu deren Beschreibung gemacht worden ist. Heute sind die Worte *long memory* und *Hurst* zu Synonymen geworden. Dies hat einen historischen Hintergrund. *“In 1906, a young englishman named Harold Edwin Hurst arrived in Cairo. It was to have been a short stay. But it lasted sixty-two years and ended with his solving one of the great mysteries of the pharaohs — and, inadvertently, providing a clue to the way financial markets work.”* Mit diesen Worten beginnt das Kapitel IX: *“Long Memory, from the Nile to the Marketplace”* in Mandelbrot und Hudson (2004). Der weitere Verlauf der Geschichte nimmt einige Seiten in Anspruch und kann bei Interesse in oben zitiertem Buch nachgeschlagen werden. Aus heutiger Sicht sind lediglich zwei Tatsachen von Bedeutung. Erstens, es wurde eine fachübergreifende Entdeckung gemacht, die nicht nur in der Hydrologie (Hurst war ein Hydrologe) für Aufregung sorgte. Zweitens spielt diese Entdeckung, dieses in einem anderen Umfeld erfundene Rad, eine sehr bedeutende Rolle in einem hochkomplexen Mechanismus des Finanzmarktes.

Zwischen 1951 und 1956 publizierte H. Hurst drei längere Aufsätze¹ über seine Entdeckungen, die er bei Untersuchung von Wasserstandsschwankungen des Nils und anderer Flüsse gemacht hat. Skalierungseigenschaften der hydrologischen Zeitreihen, die beispielsweise bei Analyse der Aufzeichnungen des Nil-Pegels aufgefallen sind, wurden nach und nach in vielen anderen Gebieten der Wissenschaft wiederentdeckt. So geben Lawrence und Kategoda (1977) eine Übersicht über langfristige Abhängigkeiten in geophysischen und hydrologischen Daten, Jeffreys (1961) für astronomische Daten. Motiviert durch die Beobachtungen von Hurst, führten Mandelbrot und van Ness (1958) die *fraktionale* oder *gebrochene Brownsche Bewegung* (*fBm*)²

¹ Zitiert wird gewöhnlich Hurst (1951).

² Fractional Brownian motion.

und das *fraktionale weiße Rauschen*³ ein und zeigten, dass diese Verfahren zur Modellierung des Hurst' Effektes in der Biologie, Geophysik, Meteorologie und Hydrologie geeignet sind. Seitdem sind mehrere wahrscheinlichkeitstheoretische und statistische Verfahren in diesem Bereich entwickelt worden [Beran 1994]. Das Anwendungsspektrum hat sich inzwischen auf die Ingenieurwissenschaften, insbesondere auf den Bereich Telekommunikationen [Park und Willinger 2000], die Physik [Beran 1994] und die Wirtschaftswissenschaften [Henry und Zaffaroni 2003] ausgeweitet.

Da sich die "long memory"-Effekte grundsätzlich von jenen mit "short-memory" (AR, MA, ARMA, ARCH-GARCH-Familie⁴...) unterscheiden und immer wieder beobachtet werden, handelt es sich dabei um einen selbstständigen Zweig im Bereich der Zeitreihenanalyse, der nicht mehr als etwas Exotisches betrachtet wird.

Die klassische Zeitreihenanalyse beschäftigt sich mit *stationären* Zeitreihen. Die meisten Definitionen und statistischen Verfahren nehmen Stationarität stillschweigend als gegeben an. Man spricht von einer (schwach) stationären Zeitreihe X_t , wenn ihre zwei ersten Momente zeitinvariant unter Translation sind. Dies impliziert einen konstanten Erwartungswert $EX_t = \mu$ und eine Autokovarianzfunktion $\gamma_\tau = \text{Cov}(X_t, X_{t-\tau})$, $\tau = 0, 1, \dots$, die eine Funktion von τ und nicht von t ist. Die Autokorrelationsfunktion einer stationären Zeitreihe entsteht durch Normierung der Autokovarianzfunktion: $\rho_\tau = \gamma_\tau/\gamma_0$. Im Fokus einer linearen Zeitreihenanalyse liegt die Untersuchung der Eigenschaften von ρ_τ . Flandrin u.a. (2002) geben folgende

Definition 6.1.1 (long-memory). *Ein stationärer Prozess $\{Y_t, t \in \mathbb{R}\}$ ist langzeitabhängig, wenn seine Autokorrelationsfunktion $\rho_Y(\tau)$ nicht integrierbar ist:*

$$\int_0^\infty |\rho_Y(\tau)| d\tau = \infty. \quad (6.1.1)$$

Bei einer Zeitreihe handelt es sich um eine diskrete Approximation eines stochastischen Prozesses. Es sei $\{Y_k\}$ eine stationäre Folge der Zufallsvariablen mit endlichen zweiten Moment. Wir bezeichnen

$$\mu = EX_k \quad \text{und} \quad \gamma_k = \text{Cov}(X_k, X_0)$$

³ Fractional Gaussian noise, fGn.

⁴ Bollerslev, Chou und Kroner (1992), Engle (1982).

und sagen, dass eine typische stationäre Folge $\{X_k\}$ *short memory* hat, wenn

$$\sum_{j=-\infty}^{\infty} |\gamma_j| < \infty \quad (6.1.2)$$

gilt und *long memory* sonst, d.h.

$$\sum_{j=-\infty}^{\infty} |\gamma_j| = \infty. \quad (6.1.3)$$

Solange γ_0 konstant und endlich bleibt, spielt die Normierung offensichtlich nur eine untergeordnete Rolle.

Die Beobachtung, dass die Stichprobenautokorrelation einiger Zeitreihen viel langsamer als solche von ARMA-Modellen absinkt, führte Granger und Joyeux (1980) und Hosking (1981) zur Idee der *fraktionalen Differenzen*. Daraus entstand das *fraktional Integrierte (ARFI-MA)*⁵ (p, d, q) Modell mit einem positiven nicht-ganzen Parameter d . In Bezug auf finanzwirtschaftliche Zeitreihen zeigten Ding, Granger und Engle (1993), dass für quadrierte oder absolute Renditen Langzeitabhängigkeiten typisch sind.

Ökonomische Auswirkungen

Eine mögliche Präsenz des long-memory-Effektes in Renditen führt zu interessanten Überlegungen bzw. Schlussfolgerungen sowohl theoretischer als auch empirischer Natur. Sei X_n eine stationäre Sequenz der Zufallsvariablen und $S_n = X_1 + \dots + X_n$. Sind X_n unabhängig, dann gilt: $\text{Var} S_n = n \text{Var} X_1$. Die Varianz der Summe S_n steigt *linear* mit n an. Diese Eigenschaft gilt im allgemeinen nicht nur für i.i.d. Zufallsvariablen, sondern auch für eine ganze Klasse von Markov-Prozessen. Zu einem charakteristischen Merkmal der Prozesse, die das Langzeitgedächtnis (LRD)⁶ beschreiben, zählt deren Klassifizierung außerhalb von Markov-Prozessen. Die Varianz der Summe steigt im LRD-Fall *überproportional* mit n an: $\text{Var} S_n \approx cn^{2H}$, wobei $H \in (1/2, 1)$ den *Hurst-Parameter* bezeichnet. Die Existenz der LRD in Renditen stellt damit eine lineare Modellierung von Renditen in Frage und fordert zur Entwicklung nichtlinearer Preis/Rendite-Modelle auf.

⁵ AutoRegressive Fractional Integrierte Moving Average.

⁶ Long-range dependence.

Ein weiteres wichtiges Argument für das Langzeitgedächtnis: es lässt sich recht gut mit dem empirisch beobachteten Phänomen der *volatility clustering* vereinbaren. Benoît Mandelbrot notierte schon vor vielen Jahren, dass Amplituden zeitlich aufeinanderfolgenden Renditen sich nicht zu schnell ändern. Sowohl die größeren als auch die kleineren (im absoluten Wert) Renditen kommen eher gemeinsam vor.

Es gibt aber auch einige... nein, eigentlich keine echten Argumente *dagegen*, sondern eher problematischen Aspekte. Ist long-memory tatsächlich vorhanden, stellt sich die Frage: ob und gegebenenfalls wie kann diese signifikante Autokorrelation zwischen den Beobachtungen auf mehreren Zeitskalen gewinnbringend ausgenutzt werden? Hier beginnen die Schwierigkeiten. Eine JA-Antwort würde natürlich der schwachen Form der EMH widersprechen, die besagt, dass unter Berücksichtigung historischer Kursdaten die zukünftigen Renditen unvorhersagbar sind. Weiterhin, unter EMH sind die risikoangepassten Renditen Martingale. Long-memory-Prozessen fehlt die Martingale-Eigenschaft grundsätzlich.⁷

Mandelbrot (1971) untersuchte zum ersten mal Auswirkungen der LRD auf die Wertpapierbewertung. Er zeigte, dass in Präsenz des long-memory-Effektes die neu ankommenden Informationen nicht vollständig verarbeitet werden können. Dies führt aber nicht zwangsläufig zur Arbitrage; lediglich die Anwendung der Martingale-Methoden wäre in diesem Fall inkorrekt. Mandelbrot nahm an, dass die erste Differenz des Preises einem linearen Prozess in der Form

$$P_t - P_{t-1} = \sum_{i=0}^{\infty} \alpha_i N_{t-i}$$

folgt, wobei $\{N_t\}$ eine i.i.d. Folge des weißen Rauschens ist und $\sum_{i=0}^{\infty} \alpha_i < \infty$. Weiterhin sollte der Wertpapierpreis arbitragebedingt eine lineare Funktion des Fundamentalwertes sein und die Martingaleigenschaft aufweisen. Daraus resultiert

$$P_t - P_{t-1} = N_t \sum_{i=0}^{\infty} \alpha_i$$

unter Risikoneutralität und einem Zins von Null.⁸ Sollten die α_i nicht aufsummierbar sein (=LRD im Fundamentalwert), würde ein Arbitra-

⁷ Der Zusammenhang zwischen der fBm und kontinuierlichen Martingalen wird u.a. in Grippenbergh und Norros (1996), Norros, Valkeila und Virtamo (1999), Nuzman und Poor (2000) diskutiert.

⁸ Henry und Zaffaroni (2003).

gevorgang nicht zu arbitragefreien Preisen führen und bliebe damit wirkungslos. Arbitrage, die in Finance unter gewöhnlichen Umständen wie ein Zauberstab wirkt, *versagt unter LRD!* Die Markteffizienz in Präsenz der LRD kann allein durch perfekte Arbitrage *nicht* erreicht werden. Entsprechend kann auch Derivatebewertung unter LRD mit Martingale-Methoden nicht adäquat sein.⁹

Unter der Normalverteilungsannahme ist die temporale Risikokomponente irrelevant, da das Risiko jedes Zeitintervalls durch eine lineare Reskalierung aus einem anderen Intervall ermittelt werden kann. Dies impliziert zum Beispiel, dass der jährliche Mittelwert der Renditeverteilung gleich dem 12-fachen des monatlichen bzw. dem 52-fachen des wöchentlichen Mittelwertes sein soll. Das gleiche gilt für die Varianz. Signifikante Auswirkungen für die Wertpapierbewertung (zum Beispiel im CAPM oder im Black/Scholes-Modell) entstehen, wenn LRD vorhanden ist. Eine lineare Reskalierung des Risikos würde das tatsächliche Investmentrisiko entweder unter- oder überschätzen, was auf jeden Fall zu einer inkorrekten Bewertung führte.

Eine weitere Schwierigkeit würde im Bereich des Portfolio-Managements entstehen. Optimale Handelsstrategien wären ohne Berücksichtigung der LRD nicht mehr optimal und spiegelten entsprechend Präferenzen der Investoren nicht korrekt wieder. Dies führte zu Abweichungen von der Optimalität im Rendite-Risiko-Verhältnis und möglicherweise zum Nachteil für Investoren. Unter LRD wird der Investmenthorizont zu einer zusätzlicher Komponente des Wertpapierrisikos.¹⁰ Ein *überlinearer* Anstieg des Risikos in der Zeit steht im direkten Widerspruch zur Idee der sogenannten Zeitdiversifikation. Allerdings berichtet Mandelbrot (1971), dass statistische LRD nur für Investoren mit einem unendlichen Planungshorizont auch ökonomisch signifikant sein kann.

Die *R/S*-Analyse

Die *R/S*-Analyse zählt zu den am häufigsten verwendeten Verfahren zur Aufdeckung der LRD. Im wesentlichen versucht man damit, die Variabilität der Daten über unterschiedliche Zeitintervalle auf einen

⁹ Yajima (1985), Barkoulas u.a. (2000).

¹⁰ Holton (1992).

gemeinsamen Nenner zu bringen (deshalb auch alternative Bezeichnung: *rescaled range*). Dieser gemeinsame Nenner spiegelt das Ergebnis wieder, welches man unter Unabhängigkeitsannahme erwarten würde. Der Zähler ist von der Differenz zwischen maximalen und minimalen Wert abhängig. Die Besonderheit des Verfahrens liegt darin, dass diese Differenz (range) nicht unmittelbar über den Gesamtzeitraum berechnet wird. Stattdessen wird eine Art kumulierte Trendbereinigung durchgeführt. Im Unterschied zu einer reinen Zufallssequenz verzerrt eine Renditefolge mit dem gleichen Vorzeichen (das ist natürlich ein lokaler Trend) die Extremwerte in die eine oder andere Richtung. Dies führt dazu, dass die Reihenfolge der Renditen nicht vernachlässigt werden kann.

Die R/S -Analyse hat zum Ziel, den Selbstähnlichkeitsparameter H zu schätzen. Der Parameter H gibt seinerseits Hinweise darauf, ob und gegebenenfalls wie intensiv die LRD in der Zeitreihe ausgeprägt ist. Beginnend mit Hurst (1951) wurde das Verfahren in Mandelbrot und Wallis (1969) weiterentwickelt und verfeinert. In Mandelbrot (1972) (1975), Mandelbrot und Taqqu (1979), Bhattacharia u.a. (1983) folgen einige theoretische Begründungen. Im Übrigen ist eine Liste der Publikationen, die sich direkt oder indirekt mit der R/S -Statistik beschäftigen, sehr lang. Wir halten uns in folgender Beschreibung an Weron (2002). Die R/S -Analyse beginnt mit der Aufteilung einer Renditezeitreihe Z der Länge L in d Blocks der Länge je n . Bei jedem Block $m = 1, \dots, d$ geht man wie folgt vor:¹¹

- (i) Berechnung des Mittelwertes E_m und der Standardabweichung S_m ;
- (ii) Trendbereinigung um den Mittelwert: $X_{i,m} = Z_{i,m} - E_m$;
- (iii) Erzeugung einer kumulativen Zeitreihe $Y_{i,m} = \sum_{k=1}^i X_{k,m}$ für $i = 1, \dots, n$.
- (iv) Range $R_m = \max\{Y_{1,m}, \dots, Y_{n,m}\} - \min\{Y_{1,m}, \dots, Y_{n,m}\}$.

Die R/S -Statistik ergibt sich als Mittelwert über alle Blocks der Länge n :

$$(R/S)_n = \frac{1}{d} \sum_{m=1}^d R_m/S_m.$$

¹¹ Der zweifache Index $Z_{i,m}$ bezeichnet hier i -ten Wert im m -ten Block.

Etwas weiter verbreitet ist eine andere Version, in der die ganze Stichprobe einen einzigen Block bildet.¹² Es gilt:

$$(R/S)_n = \frac{1}{S_n} \left[\max_{0 < k \leq n} (Z_k - k\bar{Z}_n) - \min_{0 < k \leq n} (Z_k - k\bar{Z}_n) \right], \quad (6.1.4)$$

mit

$$\bar{Z}_n = \frac{1}{n} \sum_{k=1}^n Z_k, \quad \text{und} \quad S_n^2 = \frac{1}{n} \sum_{k=1}^n (Z_k - \bar{Z}_n)^2, \quad n \geq 1.$$

In dieser Formel lässt sich die Idee von Hurst (*der Hurst-Effekt*) leicht erkennen und verdeutlichen. Es ging darum, einen idealen Damm zu konstruieren, der das Wasserniveau im Nil regulieren sollte. Hurst ging von einem konstanten jährlichen Abfluss \bar{Z}_n und einem variablen jährlichen Zufluss Z_k des Wassers über $k = 1, \dots, n$ Jahre aus. Wären die jährlichen Messungen unabhängig, würde man einen *linearen* Anstieg der Statistik $(R/S)_n^2$ mit n erwarten. Mit anderen Worten,

$$(R/S)_n^2 \sim cn \quad \text{oder} \quad (R/S)_n \sim cn^{1/2}$$

mit einer positiven, endlichen Konstante c , die unabhängig von n ist. Das Ratio $\log(R/S)_n / \log n$ würde dann für einen unabhängigen Prozess einen Grenzwert von $H = 1/2$ erreichen.¹³ Nach Mandelbrot und Wallis (1969) konvergiert der Hurst Exponent H bei $n \rightarrow \infty$ für jeden kurzzeitabhängigen stochastischen Prozess gegen $1/2$. Für einen Prozess mit LRD erwartet man dagegen¹⁴

$$(R/S)_n \sim cn^H \quad (6.1.5)$$

mit dem Hurst-Koeffizient $1/2 < H < 1$. *“For time intervals of varying length and starting point, he [Hurst] began by adjusting the data to remove the trend during that interval. Only then did he compute the range. Statisticians would have expected the “detrended” range to increase as the square root of the length of the time interval, as it does in Brownian motion. In fact, Hurst found the range grew faster than that.*¹⁵ Damit haben wir eine Gleichung, die zur Schätzung des Parameters H aus einer Stichprobe mit steigendem Zeithorizont durch eine einfache lineare Regression verwendet werden kann:

¹² Vgl. beispielsweise Willinger u.a. (1999).

¹³ Aus Feller (1951) ist bekannt, dass der Range R_n für einen unabhängigen Gaußschen Prozess asymptotisch proportional zu \sqrt{n} ist.

¹⁴ Mandelbrot (1975).

¹⁵ Mandelbrot und Hudson (2004), S. 179.

$$\log (R/S)_n = \log c + H \log n. \quad (6.1.6)$$

In graphischer Darstellung kann die $(R/S)_n$ -Statistik gegen n in einer doppelt-logarithmischer Auflösung eingetragen werden. Es ergibt sich ungefähr eine Gerade mit der Neigung gleich 0,5 zur X-Achse. Ist der Prozess persistent, wird die Neigung größer als 0,5 sein; im Falle eines antipersistenten Prozesses ist sie kleiner als 0,5. Für einen i.i.d. kurzzeitabhängigen Gausschen Prozess ist $H = 0,5$.

Mandelbrot hat als erster in seinen zahlreichen empirischen Untersuchungen auf die Möglichkeit der LRD in den Aktienrenditen hingewiesen. Er verwendete die klassische Form der R/S -Statistik und berechnete einen repräsentativen Wert von $H = 0,55$.

Empirie

Eine klassische R/S -Analyse hat zum Ziel, das Ratio $\log (R/S)_n / \log N$ zu schätzen. Dies hat jedoch einige Schwachstellen. Dazu gehören vor allem der Bias des Verfahrens und die fehlende Teststatistik. Dies veranlasste Lo (1991) zur Einführung einer *modifizierten* (R/S)-Analyse. Mit Lo-Modifikation kann eine statistische Hypothese

$$H_0 = \{\text{keine LRD, } H = 0,5\}$$

gegen die Alternative

$$H_1 = \{\text{es gibt LRD, } 1/2 < H < 1\}$$

getestet werden. Wir verzichten an dieser Stelle auf die technischen Details und verweisen auf Lo (1991) bzw. Teverovsky u.a. (1998). Im wesentlichen besteht die Modifikation im Austausch des Nenners S in (6.1.4) durch einen konsistenten Schätzer. Im Falle von abhängigen Zufallsvariablen ergibt sich Varianz der Teilsumme nicht als Summe der einzelnen Varianzen; die Kovarianzterme bis zu einem bestimmten *lag* müssen berücksichtigt werden. Gerade diese Anpassung für *short memory* hat Lo in seiner Statistik berücksichtigt. Mit Monte Carlo Simulation hat er einen akzeptablen p -Wert von H_0 gegenüber LRD Alternativen nachgewiesen. Eine empirische Untersuchung mit dem modifizierten Verfahren auf CRSP-Daten¹⁶ lieferte *keine* Hinweise auf

¹⁶ Center for Research in Security Prices.

LRD. So kam Lo zum Schluss, dass die Dynamik der Wertpapierrenditen auch durch gewöhnliche Verfahren adäquat beschrieben werden kann.

Allerdings ist die Lo-Modifikation in der Literatur umstritten. In der Praxis ergeben sich Schwierigkeiten, weil a) die Lo-Statistik asymptotisch ist und b) der Wert der Statistik von einem unbekanntem Parameter q abhängig ist, der ebenfalls geschätzt werden soll.¹⁷ Teverovsky u.a. (1998) zeigen weitere Probleme auf: *“... while Lo’s modified R/S-statistic represents a theoretical improvement over the classical rescaled adjusted range statistic, its practical application requires care and has a number of problematic features. . . Lo’s method has a strong preference for accepting the null hypothesis of no long-range dependence, irrespective of whether long-range dependence is present in the data or not.”*

Auf der empirischen Seite gibt es sowohl mit dem R/S -Verfahren selbst als auch mit den darauf basierenden Befunden eine ganze Menge widersprüchlicher Ergebnisse. Greene und Fielitz (1977) finden mit dem R/S -Verfahren Indizien für die Persistenz in täglichen Renditen auf dem US Markt. Peters (1991, 1994) führte zahlreiche empirische Untersuchungen durch und bekräftigte damit seine theoretischen Aussagen zugunsten der LRD. Corazza u.a. (1997) berichten über LRD in Renditen von Index- und Waren-Futures. Barkoulas u.a. (2000) finden signifikante Merkmale fraktionaler Dynamik sowie LRD am griechischen Aktienmarkt (Athens Stock Exchange). Die Preisbewegungen scheinen durch Realisationen sowohl aus unmittelbarer Vergangenheit als auch von längerer Fristigkeit beeinflusst zu sein.

Nachdem Lo (1991) feststellte, dass es nach einer Korrektur für kurzfristige Abhängigkeit keine Evidenz für LRD im US Aktienmarkt gibt, folgte eine Serie von Untersuchungen aus dem „Gegenlager.“ Mit dem nach Lo (1991) modifizierten R/S -Verfahren und spektraler Regression finden Cheung und Lai (1995) keine Anzeichen für die Persistenz in Aktienrenditen auf einigen internationalen Märkten. Crato (1994) kommt bei den G-7 Ländern mit dem Maximum Likelihood Verfahren zu ähnlichen Ergebnissen. Cheung u.a. (1993) testeten Aktienindizes aus Deutschland, Italien, Japan und UK. Das Ergebnis: *“While the conventional analysis seems to indicate the presence of long cycles in stock return, no significant evidence of long cycles can be found using the modified R/S analysis once short-term dependence and conditional heteroske-*

¹⁷ Der Parameter q dient zur Modellierung der “lag-Länge.”

*dasticity in the data are adjusted for.*¹⁸ Howe u.a. (1999) beobachten einen ähnlichen Effekt bei den Aktienindizes im Pazificraum (Japan, Australien, Hong Kong, Singapur, Korea und Taiwan). *“... These findings also raise the general specter of significant sensitivity of empirical findings to the choice of method of analysis.”*¹⁹

Es kommt immer wieder zur Kritik auf *methodischer Seite*. Bei Willinger u.a. (1999) lesen wir beispielsweise: *“Lo’s modified R/S method does not appear to provide the “ultimate” test for long-range dependence.²⁰ ... Lo’s claim of no evidence of long-run memory in the data may well be due to the fact that the modified R/S-statistic cannot, in general, guarantee a low Type II error probability.”*²¹ Dies ruft einem die endlose Diskussion rund um die EMH in Erinnerung: wie viele Anomalien müssen aufgedeckt werden, um die EMH verwerfen zu können?

Krämer u.a. (2002) berichten auf Grundlage täglicher Daten für den DAX und einiger Einzelaktien, dass sowohl die „einfachen“ als auch die quadrierten Renditen wahrscheinlich am besten durch Kurzzeitabhängigkeiten mit veränderlichen Trendkomponenten modelliert werden können. Dargestellt wurden Ergebnisse für fünf Einzelaktien und den DAX. Davon wiesen „nur“ drei (von fünf!) einen signifikanten long-memory Effekt auf. Die Schlussfolgerung: *“No matter which type of returns we take, the evidence for long memory is statistically insignificant; given a sample size of more than 9 000, this means that long memory in returns does not exist. . . The significance of the long-memory effects can be further reduced by taking still shorter periods (one year, two years).“* Keine Frage, je mehr unbequeme Daten ausgefiltert werden, desto einfacher ist es, das angestrebte Ergebnis zu erreichen. Nur welches Ergebnis war denn angestrebt?

Als Alternativverfahren zur *R/S*-Analyse bieten sich beispielweise die *Detrended Fluctuation Analysis*²² oder die *spektrale Regression* nach Geweke und Porter-Hudak (1983), der sogenannte *GPH-Test*,²³ an. Mansfield u.a. (2001) entwickelten eine Statistik, die im Unterschied zur *Lo*-Statistik keine Normalverteilungsannahme der Renditen unterstellt. Sie ist unabhängig vom Tail-Index α und eignet sich darum auch

¹⁸ Cheung u.a. (1993), S. 45.

¹⁹ Howe u.a. (1999), S. 139.

²⁰ Willinger u.a. (1999), S. 3.

²¹ Willinger u.a. (1999), S. 8.

²² Weron (2002).

²³ Mansfield u.a. (2001).

für Zeitreihen, deren Glieder stabilverteilt sind. So (2000) wendet die modifizierte Lo-Statistik und den GHP-Test auf den S&P 500 Index, auf den DJIA Index und auf die 30 DJ Aktien an. Das Ergebnis: “. . . *strong evidence of long-term dependence in volatility is found in nearly all cases. This suggests that it is important to incorporate the long memory feature in the modelling of volatility in order to produce good volatility forecasts and derivative pricing formulas.*”

Angesichts der vielen Widersprüche scheinen statistische Verfahren, die auf geeignete Weise mit long memory umgehen können, (noch) nicht ausgereift zu sein. Denken wir daran, dass die Annahme der Stationarität nicht etwas ist, was überprüft werden kann: wir müssen es glauben oder nicht. “. . . *statistical analyses cannot be expected to provide a definitive answer concerning the presence or absence of long-range dependence in asset price returns, . . . a much more challenging approach . . . will require a deeper understanding of the nature of the micro/macro-economic market forces that determine the price movements.*”²⁴

6.2 Selbstähnliche Zufallsprozesse

Skalierung pur

Die R/S -Analyse kann wohl kaum als ein anspruchsvolles theoretisches Verfahren zur Untersuchung der Langzeitabhängigkeiten angesehen werden. Letztendlich lässt sie sich auf eine einfache Regressionsgleichung zurückführen, die ihrerseits eine sehr grobe logarithmische Asymptotik zum Ausdruck bringt. Die Einführung selbstähnlicher Zufallsprozesse²⁵ markierte einen weiteren Schritt zum Verständnis einer *fraktalen* oder selbstähnlichen Welt und ihrer Verbindung zu den für uns vertrauten Begriffen wie beispielsweise Stationarität. Es stellte sich heraus, dass die Worte *fraktal*, *selbstähnlich* und *langzeitabhängig* als Synonyme verwendet werden können. Außerdem haben sie alle einen „gemeinsamen Nenner“: die *Skala*. Warum?

Ein *stationärer Prozess* $\{X_t, t \in \mathbb{R}\}$ kann dadurch charakterisiert werden, dass seine endlich-dimensionalen Verteilungen gegen Verschiebung der Zeitachse invariant sind:

²⁴ Willinger u.a. (1999), S. 12.

²⁵ Flandrin (1992), Abry u.a. (2000), Flandrin u.a. (2002).

$$\{X(t-a), t \in \mathbb{R}\} \stackrel{d}{=} \{X(t), t \in \mathbb{R}\} \quad \text{für alle } a \in \mathbb{R}, \quad (6.2.1)$$

wobei d für *Distribution* steht (die Verteilungen sind gleich). Ganz ähnlich heißt ein Prozess $\{Y_t, t \in \mathbb{R}\}$ *skalen-stationär*, falls seine endlich-dimensionalen Verteilungen bei der *Skalierung* der Zeitachse unverändert bleiben:

$$\{Y(at), t \in \mathbb{R}\} \stackrel{d}{=} \{Y(t), t \in \mathbb{R}\}, \quad a > 0. \quad (6.2.2)$$

Schließlich verhält sich die Verteilung eines *H-selbstähnlichen* (*H-self-similar*) oder *H-ss*-Prozesses auf eine ganz besondere Weise. Eine Skalierung um den Faktor $a > 0$ bewirkt das Gleiche wie eine Multiplikation mit a^H :

$$\{Y(at), t \in \mathbb{R}\} \stackrel{d}{=} a^H \{Y(t), t \in \mathbb{R}\}, \quad a > 0. \quad (6.2.3)$$

Das H steht für den *Hurst-Exponent* oder den *Selbstähnlichkeitsparameter* des Prozesses. Für $H = 0$ ist ein skalen-stationärer Prozess selbstähnlich. Das entscheidende Merkmal eines selbstähnlichen Prozesses ist die Nichtlinearität seiner Varianz /Kovarianz im allgemeinen Fall. Das zweite Moment (falls existent) skaliert in der Zeit mit dem Faktor $2H$:

$$\mathbb{E} [Y(t)^2] = t^{2H} \mathbb{E} [Y(1)^2], \quad Y(0) \equiv 0. \quad (6.2.4)$$

Es ist leicht zu sehen, dass für alle $H \neq 1/2$ diese Skalierung *nichtlinear* ist.

Ein homogener H-ss-Prozess ist ein Prozess mit stationären und unabhängigen Zuwächsen, kurz *H-ssi*. Die Klasse von *H-ssi*-Prozessen stellt eine natürliche Erweiterung gewöhnlicher stationärer Prozesse dar. Die folgende Gleichung beschreibt einen *H-ssi*-Prozess $\{Y_t, t \in \mathbb{R}\}$ im Sinne der endlich-dimensionalen Verteilungen:

$$\begin{aligned} \{Y(s+t) - Y(s), s, t \in \mathbb{R}\} &\stackrel{d}{=} a^H \{Y(s/a+t) - Y(s/a), s, t \in \mathbb{R}\} \\ &= a^H \{Y(s+t) - Y(s), s, t \in \mathbb{R}\} \end{aligned} \quad (6.2.5)$$

Die Kovarianz eines *H-ssi*-Prozesses $\{Y_t, t \in \mathbb{R}\}$ mit $Y(0) = 0$ und $Y(1) \neq 0$ lässt sich schreiben als²⁶

²⁶ Samorodnitsky und Taqqu (1994).

$$\begin{aligned}
\text{Cov}[Y(t)Y(s)] &= \frac{1}{2} \left(\mathbb{E}Y^2(t) + \mathbb{E}Y^2(s) - \mathbb{E}[Y(t) - Y(s)]^2 \right) \\
&= \frac{1}{2} \left(\mathbb{E}Y^2(t) + \mathbb{E}Y^2(s) - \mathbb{E}[Y(t-s) - Y(0)]^2 \right) \\
&= \frac{1}{2} \text{Var}[Y(1)] (|t|^{2H} + |s|^{2H} - |t-s|^{2H}) \\
&= \frac{1}{2} \mathbb{E}[Y^2(1)] (|t|^{2H} + |s|^{2H} - |t-s|^{2H}). \quad (6.2.6)
\end{aligned}$$

Eine Brücke zwischen selbstähnlichen und stationären Prozessen wird durch die *Lamperti Transformation*²⁷ hergestellt. Der *lampertisierte* stationäre Prozess ist selbstähnlich.

Definition 6.2.1. Gegeben eine Zahl $H > 0$ ist die **Lamperti Transformation** \mathcal{L}_H eines Prozesses $\{Y(t)\}_{t \in \mathbb{R}}$ definiert als

$$X(t) = (\mathcal{L}_H Y)(t) = t^H Y(\ln t), \quad t > 0. \quad (6.2.7)$$

Die entsprechende inverse Transformation \mathcal{L}_H^{-1} des Prozesses $\{X(t)\}_{t > 0}$ ist definiert als

$$Y(t) = (\mathcal{L}_H^{-1} X)(t) = e^{-Ht} X(e^t), \quad t \in \mathbb{R}. \quad (6.2.8)$$

Satz 6.2.1 Ein Prozess $Y(t)$ ist stationär genau dann, wenn seine Lamperti Transformation $X(t)$ ein H -ss-Prozess ist.

Beweis: Samorodnitsky und Taqqu (1994). □

Flandrin u.a. (2002) geben folgenden Ausdruck für die Korrelationsfunktion eines Lamperti-transformierten Prozesses Y_t an:

$$\mathbb{E}[X(t)X(s)] = \text{Cov}_{\mathcal{L}_H Y}(t, s) = (ts)^H \text{Cov}_Y(\ln t, \ln s), \quad (6.2.9)$$

und, falls Y_t stationär ist,

$$\text{Cov}_{\mathcal{L}_H Y}(t, s) = (ts)^H \gamma_Y(\ln(t/s)) \quad (6.2.10)$$

mit $\text{Cov}_Y(ts) = \gamma_Y(t-s)$. Auf Grundlage dieser Erkenntnisse scheinen die für die stationären Prozesse bekannten Analyseverfahren bei der Anwendung auf *delampertisierte* H -ss-Prozesse ihre Gültigkeit zu behalten. Diese Idee wurde in Nuzman und Poor (2000) bei der Untersuchung von Vorhersage- und Interpolationseigenschaften einer gebrochenen Brownschen Bewegung aufgegriffen und weiterentwickelt.

²⁷ Lamperti (1962).

6.3 Gebrochene Brownsche Bewegung

Das klassische Beispiel eines H -ssi-Prozesses ist die gewöhnliche Brownsche Bewegung. Sie hat unabhängige stationäre Zuwächse, stetige Pfade und ist selbstähnlich mit Parameter $H = 1/2$. Zur selben Klasse gehört die „gebrochene“ Version.

Definition 6.3.1 (Taqqu (2003)). Ein H -ssi Gaussprozess B_t^H mit $0 < H < 1$ heißt **gebrochene Brownsche Bewegung (fBm)**. In Standardform gilt

$$\sigma^2 = \text{Var} B^H(1) = 1.$$

Die fBm ist ein stetiger, zentrierter Gaussprozess $\{B_t^H, t \in \mathbb{R}\}$ mit abhängigen Zuwächsen. Sie ist durch die Kovarianzfunktion

$$\text{Cov}[B_t^H B_s^H] = \frac{1}{2} (|s|^{2H} + |t|^{2H} - |t-s|^{2H}) \quad s, t \in \mathbb{R} \quad (6.3.1)$$

bis auf den Skalierungsfaktor H vollständig spezifiziert.²⁸ Die Selbstähnlichkeit mit Parameter H impliziert, dass

$$\left\{ a^H B_{t/a}^H, t \in \mathbb{R} \right\} \text{ dieselben Verteilungen wie } \left\{ B_t^H, t \in \mathbb{R} \right\} \text{ hat.}$$

Wir fassen die wichtigsten Eigenschaften der fBm zusammen.

Satz 6.3.1 Falls

- (i) Y_t ein Gaussprozess mit $EY_t = 0$ ist,
- (ii) $EY_t^2 = \sigma^2 |t|^{2H}$ für $\sigma > 0$ und $0 < H < 1$ gilt,
- (iii) Y_t stationäre Zuwächse besitzt,

dann ist Y_t eine fBm.

Beweis: Wir folgen Taqqu (2003). Da Y ein zentrierter Gaussprozess ist, sind seine endlich-dimensionale Verteilungen durch die Kovarianzfunktion eindeutig charakterisiert. Weiterhin folgt aus (ii) und (iii), dass diese Kovarianzfunktion in der Form (6.3.1) sein muss. \square

²⁸ Vervaat (1987).

Charakterisierung der gebrochenen Brownschen Bewegung: der Hurst-Parameter H

Mit der gebrochenen Brownschen Bewegung scheint ein Modell entstanden zu sein, das die long-memory-Eigenschaft mathematisch zum Ausdruck bringt. Im Mathematical Finance macht die fBm als long-memory stochastisches Modell der gewöhnlichen Brownschen Bewegung immer mehr Konkurrenz.²⁹

Die fBm ist erstmals von Mandelbrot und van Ness (1958) als eine Familie von Zufallsfunktionen eingeführt worden, “*that could in some way be expected to be ‘typical’ of what happens in the absence of asymptotic independence*”. A. N. Kolmogorov modellierte bereits in den frühen vierziger Jahren³⁰ die hochentwickelte Turbulenz mittels einer fBm mit $H = 1/3$ und hat unter anderem gezeigt, dass (6.3.1) genau dann eine Kovarianz-Funktion ist, wenn H im Intervall $(0, 1]$ liegt.

Bezogen auf den Parameter H lässt sich die fBm als parametrische Erweiterung oder Verallgemeinerung einer gewöhnlichen Brownschen Bewegung interpretieren, die hier als Spezialfall $H = 1/2$ auftritt. Dabei wird das Intervall zwischen 0 und 1 in zwei Hälften aufgeteilt. Für $H < 1/2$ spricht man von einer *negativen Korrelation* der Zuwächse oder *Antipersistenz*, bei $H > 1/2$ sind dagegen die Zuwächse *positiv korreliert* oder *persistent*. Die *Unabhängigkeit* der Zuwächse ist ausschließlich für $H = 1/2$ möglich und gegeben (ein Wiener Prozess). Für $H > 1/2$ nimmt die Korrelation zwischen den Zuwächsen mit der Zeit *hyperbolisch* (sprich: langsamer als exponentiell) ab

$$\text{Cov}(t) = \mathbb{E}[Y(1)(Y(t+1) - Y(t))] \sim 2H(2H - 1)t^{2H-2} \quad (6.3.2)$$

und divergiert asymptotisch für große t :

$$\sum_{t=1}^{\infty} \text{Cov}(t) = \infty. \quad (6.3.3)$$

Dies deutet laut Definition (6.1.1) auf die Langzeitabhängigkeit hin. Für $H = 1$ degeneriert die fBm zu einer Linie, die mit der Zeitachse einen zufälligen Winkel $Y(1)$ bildet: $Y(t) = tY(1)$.

²⁹ Cutland u.a. (1995), Salopek (1998), Sottinen (2001).

³⁰ Kolmogorov (1940).

Darstellungsformen der gebrochenen Brownschen Bewegung

In der Literatur gibt es inzwischen mehrere sehr unterschiedliche Darstellungsformen für die fBm. Eine der bekanntesten Formeln stammt aus Mandelbrot und van Ness (1958). Dieser Arbeit wird gewöhnlich auch die Herkunft des Begriffes *gebrochene Brownsche Bewegung* zugeschrieben. Wir ziehen hier eine modernere Variante vor. Hu (2001), S. 225, gibt die folgende *Konstruktionsformel* für die fBm $\{B_t, t \in \mathbb{R}\}$ mit $0 < H < 1$ an:

$$B_t^H = c_H \left[\int_{-\infty}^t (t-s)^{H-1/2} dW_s - \int_{-\infty}^0 (-s)^{H-1/2} dW_s \right], \quad (6.3.4)$$

mit einer Normierungskonstante c_H . Durch das zweite Integral wird die Anfangsbedingung $B_0^H = 0$ gewährleistet. Nach (6.3.4) stellt die fBm das Ergebnis einer Konvolution zwischen einem Wiener Prozess W_t und einem Kern $K_H(t, s)$ dar. Mit Festlegung

$$K_H(t, s) := \begin{cases} (t-s)^{H-1/2} - (-s)^{H-1/2}, & \text{falls } s < 0, \\ (t-s)^{H-1/2}, & \text{falls } 0 < s < t, \\ 0, & \text{falls } s > t \end{cases}$$

ergibt sich eine kompakte Darstellung

$$B_t^H = \int_{-\infty}^{\infty} K_H(t, s) dB_s. \quad (6.3.5)$$

Mit Hilfe eines Normierungsfaktors

$$\sigma_H^2 = \mathbb{E} \left[(B^H(1))^2 \right] = \frac{\Gamma(1-2H) \cos(\pi H)}{\pi H}$$

kann die fBm in die Standardform mit $\mathbb{E} \left[(B^H(1))^2 \right] = 1$ gebracht werden. Durch einige Umformungen des *Memory-Kerns* $K_H(t, s)$ entsteht eine weitere Darstellung nach Decreusefond und Üstünel (1999)³¹:

$$B_k(t) = \int_0^t K_k(t, s) dW(s), \quad (6.3.6)$$

$$K_k(t, s) = \frac{(t-s)_+^k}{\Gamma(k+1)} F(-k, k, k+1, 1-t/s), \quad k = H - 1/2.$$

³¹ Siehe auch Pipiras und Taquu (2001).

$F(\alpha, \beta, \gamma, z)$ bezeichnet die Gaussche hypergeometrische Funktion.³² Noch eine Integraldarstellung der fBm, die wir hier nicht aufführen, geben Nuzman und Poor (2000) an. Mathematisch sehr fortgeschritten ist schließlich eine Darstellung der fBm als ein *fraktionales* oder *gebrochenes* Integral der Ordnung $H + 1/2$ eines Wiener Prozesses dW_t .³³

6.4 Problematik stochastischer Integration in Bezug auf die gebrochene Brownsche Bewegung

Eine recht komplizierte Darstellung der fBm lässt sich auf die ebenso komplizierte Form der Langzeitabhängigkeit zurückführen, die mittels fBm zum Ausdruck gebracht wird. Damit beginnen aber erst die Probleme *qualitativer* Natur, die im Zusammenhang mit der fBm als Modellierungswerkzeug im Finance auftauchen.

Zu den bedeutendsten Erkenntnissen über die fBm gehört die Tatsache, dass sie weder ein Markov-Prozess noch ein Semimartingal ist. Dies bedeutet unter anderem, dass die sehr gut entwickelte Theorie stochastischer Integration in Bezug auf die fBm ungeeignet ist. Einen möglichen Ausweg bietet eine erweiterte Definition des stochastischen Integrals unter Berücksichtigung spezifischer Eigenschaften der fBm.

Eine gute Übersicht zur geschilderten Problematik geben Pipiras und Taqqu (2000). Die ausführliche Quelle zur *fraktionalen* Integration ist das Buch von Samko, Kilbas und Marichev (1993). Speziell mit der fBm beschäftigen sich Decreusefond und Üstünel (1999), Carmona u.a. (1999), Alos u.a. (2000). Eine neuere Quelle zu diesem Themengebiet ist das Buch von Dounkan, Oppenheim und Taqqu (2003).

Abgesehen vom mathematischen Interesse per se, interessiert uns im finanzwirtschaftlichen Kontext die entscheidende Frage: *Sind Modelle (u.a. zur Derivatbewertung) mit der fBm als Underlying arbitragefrei?*

³² Anhang A.7.

³³ Das Riemann-Liouville fraktionale Integral der Ordnung α einer Funktion $\varphi(t)$ ist definiert als

$$I_+^\alpha \varphi(t) = \frac{1}{\Gamma(\alpha)} \int_{-\infty}^t (t-s)^{\alpha-1} d\varphi(s), \quad \Gamma(\alpha) = \int_0^\infty s^{\alpha-1} e^{-s} ds$$

wobei $\Gamma(\alpha)$ die Gammafunktion ist. Wir gehen im Rahmen dieser Arbeit auf die Einzelheiten fraktionaler Integrale nicht ein und verweisen auf Samko u.a. (1993).

Diese Frage kann noch breiter aufgefasst werden: *lassen sich die durch die fBm modellierte Langzeitabhängigkeiten zu einer Vorhersage des zukünftigen Verlaufs eines stochastischen Prozesses verwenden und, falls ja, wie gut ist eine solche Vorhersage?*

Quadratische Variation der gebrochenen Brownschen Bewegung

Ein gemeinsames Problem, das bei allen Ansätzen zu stochastischer Integration zu lösen ist, lässt sich wie folgt formulieren. Man steht vor der Aufgabe, für einen *geeigneten* Integrand $f(t, \omega)$ einen Ausdruck der Art

$$\int_0^T f(t, \omega) dB_t^H \quad (6.4.1)$$

sinnvoll zu definieren. Die Idee dabei ist, in einem Itô-ähnlichem Verfahren $f(t, \omega)$ so zu konstruieren, dass

$$S_n = \sum_{k=1}^{n-1} f(\xi_k) (B^H(t_{k+1}) - B^H(t_k)) \quad (6.4.2)$$

konvergiert, wenn die Intervalllänge

$$|\Delta| = \max_k (t_{k+1} - t_k)$$

des größten Teilintervalls

$$0 = t_0 < t_1 < \dots < t_k < \dots < t_n = T$$

gegen Null geht. Im allgemeinen Fall ist die Kontinuität des Integrals in Bezug auf den Integrand f nicht gewährleistet. Dies lässt sich aber für *deterministische* Integranden umgehen.³⁴ Im nächsten Schritt folgt die Erweiterung auf nicht-deterministische Integranden, natürlich unter Berücksichtigung einiger Besonderheiten der fBm. Die bedeutendste davon ist, dass die fBm „nicht stochastisch genug“ für L^2 ist. Formal gilt der folgende Satz.

Satz 6.4.1 *Die Quadratische Variation der fBm für $1/2 < H < 1$ ist ein Null-Prozess. Es gilt für jede Teilung Δ_k des Intervalls $[0, T]$ mit $|\Delta| \rightarrow 0$ bei $n \rightarrow \infty$,*

$$\sum_{k=1}^{n-1} (B^H(t_{k+1}) - B^H(t_k))^2 \rightarrow 0 \quad \text{in Wahrscheinlichkeit.} \quad (6.4.3)$$

³⁴ Grippenberg und Norros (1996), Pipiras und Taqqu (2001).

Beweis: Vgl. Theorem 2.1 Lin (1995); Korollar 1.12 Cheridito (2001).
Wegen Stationarität der Zuwächse der fBm gilt für alle $s, t \geq 0$

$$\mathbb{E} \left[(B^H(s+t) - B^H(s))^2 \right] = \mathbb{E} \left[(B^H(t))^2 \right] = t^{2H}. \quad (6.4.4)$$

Für $1/2 < H < 1$ lässt sich für jedes Teilintervall

$$0 = t_0 < t_1 < \dots < t_k < \dots < t_n = T$$

die folgende Abschätzung treffen:

$$\begin{aligned} \mathbb{E} \left[\sum_{k=1}^{n-1} (B^H(t_{k+1}) - B^H(t_k))^2 \right] &= \sum_{k=1}^{n-1} (t_{k+1} - t_k)^{2H} \\ &= |\Delta|^{2H-1} \sum_{k=1}^{n-1} (t_{k+1} - t_k) \cdot \left[\frac{(t_{k+1} - t_k)}{|\Delta|} \right]^{2H-1} \\ &\leq |\Delta|^{2H-1} \rightarrow 0 \quad \text{bei } n \rightarrow \infty. \end{aligned} \quad (6.4.5)$$

□

Bemerkung. Mit ähnlicher Argumentation kann gezeigt werden, dass die quadratische Variation der fBm mit $0 < H < 1/2$ unendlich groß ist.

Zusammengefasst kommen wir zum Ergebnis, dass die quadratische Variation der fBm im ganzen Bereich $0 < H < 1$ nur drei Werte annehmen kann: unendlich, eine endliche Konstante und Null. Aus diesem Grund scheint die quadratische Variation als Maß der Variabilität für Prozesse außerhalb der gewöhnlichen Brownschen Bewegung nicht aussagekräftig zu sein. Die *p-Variation* ist besser dafür geeignet.

p-Variation

Eine ausführliche Beschreibung der *p-Variation* (und weitere interessante Überlegungen rund um das stochastische Integral) bieten Dudley und Norvaiša (1998) und Norvaiša (2000). Wir versuchen die Darstellung möglichst kompakt zu halten und konzentrieren uns deshalb auf ausgewählte Aspekte.

Es sei f eine reellwertige Funktion auf einem Intervall $[0, T]$ und $0 < p < \infty$. Wir teilen das Intervall $[0, T]$ in Teilintervalle

$$0 = t_0 < t_1 < \cdots < t_n = T$$

mehrmals auf unterschiedliche Weise auf und bilden die Summen

$$s_p = \sum_{i=0}^{n-1} |f(t_{i+1}) - f(t_i)|^p \quad (6.4.6)$$

über alle solche Aufteilungen $k = \{t_i : i = 0, 1, \dots, n\}$. Die Aufteilungen k werden zu einer Menge $J \subset \mathbb{R}$ der *Punktteilungen* $\text{PP}(J)$ zusammengefasst.

Definition 6.4.1. Die kleinste obere Schranke der Summen (6.4.6) über alle Aufteilungen k aus $\text{PP}(J)$ heißt **p-Variation**:

$$v_p(f) = v_p(f; \cdot) := \sup \{s_p(f, k) : k \in \text{PP}(J)\}. \quad (6.4.7)$$

Für jede Funktion f ist ein **Index der p-Variation** $v(f)$ wie folgt definiert:

$$v(f; [0, T]) := \inf \{p > 0 : v_p(f; [0, T]) < \infty\} \quad (6.4.8)$$

$$= \sup \{p > 0 : v_p(f; [0, T]) = \infty\} \quad (6.4.9)$$

Mit $\mathcal{W}_p(J)$ und $\mathcal{W}_q(J)$ bezeichnen wir die Mengen aller Funktionen $f : J \rightarrow \mathbb{R}$, für die $v_p(J)$ bzw. $v_q(J)$ endlich ist.

Die obige Definition macht es möglich, die feinen Unterschiede zwischen gewöhnlichen und gebrochenen Versionen der Brownschen Bewegung zu erkennen. Als erstes gilt, dass die 2-Variation (also, eine p -Variation mit $p = 2$) und die früher als *Erwartungswert* der Summen quadratischer Differenzen definierte quadratische Variation unterschiedliche Kennzahlen sind. Quadratische Variation $\langle W_t \rangle$ ist gleich t im „klassischen“ Fall bzw. $\langle B_t^H \rangle$ ist gleich Null nach dem Satz (6.4.1). Die 2-Variation eines Wiener Prozesses W_t ist fast sicher unendlich. Allerdings ist v_p für alle $p > 2$ nach oben beschränkt.³⁵ Daraus ergibt sich $v(W_t) = 2$ fast sicher für einen gewöhnlichen Wiener Prozess W_t .

Bei einer fBm könnte man einen Zusammenhang zwischen p -Variation und H vermuten. Er scheint tatsächlich gegeben zu sein. Dudley und Norvaiša³⁶ zeigten, dass für die fBm B^H mit $H \in (1/2, 1)$ der Index der p -Variation gleich dem Kehrwert von H ist:

³⁵ Dudley und Norvaiša (1998), S. 47.

³⁶ Dudley und Norvaiša (1998), S. 47-48.

$$v(B^H) = 1/H \quad \text{fast sicher.} \quad (6.4.10)$$

Decreusefond und Üstünel (1999) definieren

$$\text{die } 1/H\text{-Variation der fBm} := \begin{cases} \sigma_H^2 \cdot t & \text{für } pH = 1; \\ 0 & \text{für } pH > 1; \\ +\infty & \text{sonst} \end{cases} \quad (6.4.11)$$

Cheridito (2001, 2003) zeigt,³⁷ dass für $T, p, q > 0$

$$\begin{aligned} a) \quad & n^{pH-1+q} \sum_{j=0}^{n-1} \left| B_{\frac{j+1}{n}T}^H - B_{\frac{j}{n}T}^H \right|^p \xrightarrow{(n \rightarrow \infty)} 0 \quad \text{in } L^1 \\ b) \quad & n^{pH-1+q} \sum_{j=0}^{n-1} \left| B_{\frac{j+1}{n}T}^H - B_{\frac{j}{n}T}^H \right|^p \xrightarrow{(n \rightarrow \infty)} \infty \quad \text{in Wahrscheinlichkeit,} \end{aligned}$$

d.h. es existiert für alle $L > 0$ ein n_0 derart, dass für alle $n \geq n_0$

$$\mathbb{P} \left[n^{pH-1+q} \sum_{j=0}^{n-1} \left| B_{\frac{j+1}{n}T}^H - B_{\frac{j}{n}T}^H \right|^p < L \right] < \frac{1}{L}.$$

All diese Ausführungen in Zusammenhang mit fehlender Semimartingaleigenschaft³⁸ führen dazu, dass die Itô-Kalküle in ihrer ursprünglichen Form für die fBm nicht gelten. Ein Semimartingal M erfordert $v(M) \in [0, 1] \cup \{2\}$.³⁹ Diese Bedingung ist für alle $H \neq 1/2$ nicht erfüllt. Da die Anwendung der Itô-Formel auf Semimartingale beschränkt ist, wird die Lage zunächst unbestimmt: es ist nicht klar, wie ein Integral in Bezug auf die fBm zu definieren ist bzw. wie Funktionen der fBm berechnet werden können.

Diese Fragestellung scheint von rein mathematischem Interesse zu sein. Zugleich ist aber *die ökonomische Bedeutung des Verfahrens, mit dem das stochastische Integral über die fBm definiert wird, kaum zu unterschätzen. Von den zwei im folgenden dargestellten Verfahren ermöglicht nur eines eine arbitragefreie Modellierung des Marktes.*

³⁷ Cheridito (2003), Lemma 3.5.

³⁸ Zum ausführlichen Beweis siehe zum Beispiel Lipster und Shiryaev (1989), S. 300, Lin (1995), Korollar 2.2.

³⁹ Sottinen und Valkeila (2001), Dudley und Norvaiša (1998), S. 46–48.

Riemann-Stieltjes vs. Skorohod

Satz 6.4.2 *Es sei $f \in \mathcal{W}_p[a, b]$ und $h \in \mathcal{W}_q[a, b]$ für $p, q > 0$, $p^{-1} + q^{-1} > 1$. Das Integral $\int_a^b f dh$ existiert im Riemann-Stieltjes-Sinne (pfadweise), falls f und h keine Unstetigkeiten am selben Punkt haben.*

Beweis: Dudley und Norvaiša (1998), S. 33. □

Als erstes lassen sich damit zwei Fälle unterscheiden, in dem Integrale $\int Y_t dX_t$ pfadweise definiert werden können:

- (i) die Funktionen $t \mapsto Y_t$ sind von beschränkter p -Variation der Ordnung $p < 2$ oder
- (ii) die Prozesse X_t sind von beschränkter q -Variation mit $q < 2$.

Die beiden Anforderungen sind bei einer fBm mit $1/2 < H < 1$ erfüllt. Problematisch wird es lediglich für zwei unabhängige fBm's B_1^H, B_2^H mit jeweils $1/4 < H < 1/2$. Da hier die Bedingung $p^{-1} + q^{-1} > 1$ evtl. verletzt ist, kann das Integral $\int_0^1 B_1^H dB_2^H$ nicht pfadweise definiert werden.⁴⁰ Das gleiche gilt natürlich für eine gewöhnliche Brownsche Bewegung mit $H = 1/2$. Dies begründet indirekt die Anwendung des Itô-Integrals in diesem Fall.

Satz 6.4.3 *Für fast alle Pfade W_t eines gewöhnlichen Wiener Prozesses ist das Integral $\int_0^1 W_t dW_t$ im Riemann-Stieltjes-Sinne nicht definiert.*

Beweisidee. Die Ursache der Nichtexistenz liegt an der Bedingung $p^{-1} + q^{-1} > 1$, die hier verletzt ist. Die Partialsummen werden unbeschränkt groß und divergieren. Vollständiger Beweis: Dudley und Norvaiša (1998), S. 8. □

Fassen wir zusammen. Nach (6.4.2) und (6.4.3) ist es möglich, sowohl für einen Wiener Prozess W_t als auch für eine fBm B_t^H ein stochastisches Integral pfadweise zu definieren, solange der Integrand nicht stochastisch ist. Für einen stochastischen Integrand besteht diese Möglichkeit weiterhin, solange die Regularitätsbedingung $p^{-1} + q^{-1} > 1$ nicht verletzt ist. Handelt es sich um ein Integral vom Typ $\int_0^1 W_t dW_t$, dann ist diese Bedingung verletzt. Da die fBm davon nicht betroffen ist, besteht von mathematischer Seite aus kein Handlungsbedarf. Von ökonomischer Seite aber sehr wohl!

⁴⁰ Carmona, Coutin und Montseny (1999), S. 4.

Beide, Wiener Prozess und fBm, sind *Gaussprozesse*. Es bietet sich damit ein Standardverfahren an, mit dem das Integral zumindest für deterministische Funktionen durch die L^2 -Isometrie definiert wird.⁴¹ Tatsächlich weisen Duncan u.a. (2000) darauf hin, dass in einem Itô-ähnlichen Verfahren die Riemannschen Summen

$$S_n = \sum_{k=1}^{n-1} f(\xi_k) (B^H(t_{k+1}) - B^H(t_k)) \quad (6.4.12)$$

mit den Auswertungspunkten $\xi_k \in [t_k, t_{k+1}]$ für die Partitionen

$$0 = t_1 < t_2 < \dots < t_k < \dots < t_n = T,$$

für eine beschränkte Funktion f und $\Delta = \max_k (t_{k+1} - t_k) \rightarrow 0$ mit $n \rightarrow \infty$ zu

$$\int_0^T f_s \delta B_s^H \quad (6.4.13)$$

in L^2 konvergieren. Überraschend ist die Unabhängigkeit des Grenzwertes (6.4.13) von der Wahl des Auswertungspunktes ξ_k in (6.4.12). Sowohl für $\xi_k = t_{k+1}$ als auch für $\xi_k = t_k$ konvergiert S_n in L^2 gegen die gleiche Zufallsvariable.⁴² Gleichung (6.4.13) definiert damit ein stochastisches Integral vom Stratonovich-Typ für die fBm mit $1/2 < H < 1$. Der Erwartungswert dieses Integrals ist im allgemeinen ungleich Null, und genau das führt bekanntlich in finanzwirtschaftlichen Anwendungen zu Problemen.

In der gleichen Arbeit definieren Duncan u.a. (2000) ein stochastisches Integral vom Itô-Typ.⁴³ Dieses sogenannte *Skorohod-Integral* erweitert das Wiener-Itô-Integral auf eine besondere Weise. Es ist durch das *Wick-Produkt* \diamond definiert und kann mit Zufallsintegranden ebenso gut umgehen wie mit deterministischen Funktionen. Die Wick-Kalküle unterscheiden nicht zwischen antizipierenden und nicht antizipierenden Quantitäten. Es gibt auch keine Punktwerte $F(\omega)$ für spezifische Realisationen ω ; vielmehr sind das *Mittelwerte bezüglich geglätteten Zufallsvariablen* $\eta = \eta(\omega)$.⁴⁴ Es gilt

$$\int_0^T f_s dB_s^H = \lim_{|\Delta| \rightarrow 0} \sum_{k=0}^{n-1} f(\xi_k) \diamond (B^H(t_{k+1}) - B^H(t_k)) \quad \text{in } L^2 \quad (6.4.14)$$

⁴¹ Holden u.a. (1996), Norros u.a. (1999), Pipiras und Taqqu (2000).

⁴² Duncan u.a. (2000), S. 595.

⁴³ Siehe auch Hu u.a. (2003).

⁴⁴ Holden u.a. (1996), S. 6. Ausführliche Details zum Wick-Produkt sind in Lindstrøm u.a. (1992) und Holden u.a. (1996) zu finden.

mit Auswertungspunkten $\xi_k = t_k$ und den sonst üblichen Bezeichnungen für t_k und Δ . Dieses Integral hat den Erwartungswert Null und unterscheidet sich von dem Integral (6.4.13) um einen zusätzlichen Term namens *Malliavin-Derivative* $D_t^\phi f$:⁴⁵

$$\int_0^T f_s dB_s^H = \int_0^T f_s \delta B_s^H - \int_0^T D_s^\phi f_s ds \quad \text{fast sicher.} \quad (6.4.15)$$

Wir verfolgen diesen Zweig der Forschung nicht mehr weiter und verweisen auf Dunkan u.a. (2000)⁴⁶ für Definition und weitere Eigenschaften von Malliavin-Derivativen.

Itô-Formel für die gebrochene Brownsche Bewegung

In der Literatur sind mehrere Versionen der Itô-Formel für die fBm zu finden. Viele davon gelten nur sehr eingeschränkt für einen bestimmten Bereich des Parameters H . Eine umfassende Übersicht sowie eine eigene Version der Formel bietet Bender (2003). Wir folgen Dunkan u.a. (2000).

Satz 6.4.4 *Es sei $f : \mathbb{R} \rightarrow \mathbb{R}$ eine zweimal stetig differenzierbare Funktion mit beschränkter erster und zweiter Ableitung. Dann gilt*

$$f(B_T^H) - f(B_0^H) = \int_0^T f'(B_s^H) dB_s^H + H \int_0^T s^{2H-1} f''(B_s^H) ds \quad (6.4.16)$$

fast sicher.

Beweis: Dunkan u.a. (2000), Theorem 4.1; Bender (2003) Theorem 4.4. □

Die gute Nachricht ist, dass diese Formel für $H = 1/2$ wie erwartet mit der gewöhnlichen Itô-Formel übereinstimmt. Die scheinbare Einfachheit ist aber trügerisch: das zweite Integral ist nur in den seltensten Fällen leicht zu berechnen. Für praktische Zwecke kann es sich lohnen, auf die Wick-Kalküle zurückzugreifen, wie das folgende Beispiel aus Hu und Øksendal (2003) zeigt:

⁴⁵ Theorem 3.12 in Dunkan u.a. (2000)

⁴⁶ Siehe auch Nualart (1995), Øksendal (1997), Malliavin (1997) und Fournié u.a. (2001).

$$\begin{aligned}\int_0^t B_s^H dB_s^H &= \int_0^t B_s^H \diamond W_s^H ds \\ &= \int_0^t B_s^H \diamond \frac{d}{ds} B_s^H ds = \frac{1}{2} (B_t^H)^{\diamond 2} = \frac{1}{2} (B_t^H)^2 - \frac{1}{2} t^{2H},\end{aligned}$$

wobei W^H für *fractional white noise* steht.

6.5 Überlegungen zur Arbitragefreiheit des Marktes

Ein sehr spannendes Thema in Zusammenhang mit der Anwendung der fBm als Underlying in Finance ist die Frage nach der Arbitragefreiheit der darauf basierenden Modelle. Rogers (1997) war die erste Arbeit mit einer klaren Nachricht: *die fBm-basierten Modelle sind nicht arbitragefrei!* Es folgen Shiryaev (1998) und Salopek (1998), die die entsprechenden Arbitragestrategien aufzeigen. Nach Salopek (1998) wird Arbitrage durch Pfadregularität verursacht. Die Tatsache, dass die quadratische Variation der fBm Null ist, spielt dabei eine entscheidende Rolle. Sottinen (2001) zeigt die Existenz eines Binärbaums in diskreter Zeit, also, eines Semimartingals. Die Konvergenz dieser Struktur in kontinuierlicher Zeit lässt Arbitrage zu. Cheridito (2001) und (2003) schlägt eine Restriktion der Handelsstrategien vor, um die Arbitragefreiheit wieder zu gewinnen.

Rogers (1997) leitet eine Art von *Halbierungsstrategien* (im Gegensatz zu Bernoulli's *Verdoppelungsstrategien* im St. Petersburgen Paradox) ab. Der Handel erfolgt in diskreter Zeit. Der Investor sucht die sogenannten „vielversprechenden“ (im Original: „promising“) Perioden⁴⁷ und investiert einen Einheitsbetrag. Desweiteren legt er eine Bandbreite für die Renditeentwicklung fest. Wird dieses Band in die eine oder andere Richtung überschritten, erfolgt eine sofortige Auflösung der Position. Hat sich ein Gewinn ergeben, bleibt der Einsatz unverändert und beträgt eine Geldeinheit; beim Verlust wird in der folgenden Periode nur eine halbe Geldeinheit investiert. Das Spiel ist beendet, wenn der kumulierte Gewinn zum ersten Mal ein vorher spezifiziertes Niveau erreicht hat. Dies erfolgt im Rogers-Modell fast sicher. Dadurch, dass sich die Zwischenverluste im Gegenteil zu einer gewöhnlichen Brownschen Bewegung in Grenzen halten, erhält die Arbitragestrategie eine endgültige Bestätigung.

⁴⁷ Es wird gezeigt, dass unendlich viele solche Perioden fast sicher existieren.

Shiryaev (1998)⁴⁸ führt ein einfaches Beispiel einer Arbitragestrategie mit der fBm an. Sei $r > 0$ ein kontinuierlicher risikoloser Zins, $A_t = e^{rt}$ und $S_t = e^{rt+B_t^H}$ entsprechend Bond- und Aktienpreis zum Zeitpunkt t . Die Unsicherheit wird durch die fBm B_t^H modelliert. Ausgehend davon, dass die quadratische Variation der fBm Null ist, verwendet Shiryaev (1998) eine Itô-Formel ohne zweiten Term

$$F(B_t^H) - F(B_0^H) = \int_0^t f(B_u^H) dB_u^H, \quad (6.5.1)$$

die eine *pfadweise Integration* in Bezug auf B_t^H rechtfertigt. Es wird gezeigt, dass ein fraktionaler Black/Scholes-Markt (A_t, S_t)

$$\begin{cases} dA_t = rA_t dt = re^{rt} dt \\ dS_t = S_t (r dt + dB_t^H) = re^{rt+B_t^H} dt + e^{rt+B_t^H} dB_t^H \end{cases}$$

nicht arbitragefrei ist; d.h., es existiert ein Arbitrageportfolio θ mit

$$V_0^\theta = 0, \quad V_T^\theta \geq 0 \text{ (P-f.s)} \quad \text{und} \quad \mathbb{P}(V_T^\theta > 0) > 0.$$

Man legt wir die Gewichte $\beta_t = 1 - e^{2B_t^H}$ und $\gamma_t = 2(e^{B_t^H} - 1)$ für das Portfolio $\theta = (\beta, \gamma)$ fest. Nach einer Vereinfachung ergibt sich

$$\begin{aligned} V_t^\theta &= \beta_t A_t + \gamma_t S_t = (1 - e^{2B_t^H}) e^{rt} + 2(e^{B_t^H} - 1) e^{rt+B_t^H} \\ &= e^{rt} (1 + e^{2B_t^H} - 2e^{B_t^H}) = e^{rt} (e^{B_t^H} - 1)^2. \end{aligned}$$

Es ist leicht zu sehen,⁴⁹ dass das Portfolio θ selbstfinanzierend ist:

$$dV_t^\theta = \beta_t dA_t + \gamma_t dS_t.$$

Schließlich gilt

$$V_0^\theta = 0, \quad V_T^\theta \geq 0 \text{ für } t > 0 \quad \text{und} \quad \mathbb{E}[V_t^\theta] > 0 \text{ für } t > 0.$$

⁴⁸ Siehe auch Dasgupta (1997), Hu und Øksendal (2003) und Norvaiša (2000).

⁴⁹
$$\begin{aligned} dV_t^\theta &= re^{rt} (e^{B_t^H} - 1)^2 dt + 2e^{rt+B_t^H} (e^{B_t^H} - 1) dB_t^H = re^{rt} (e^{2B_t^H} - 2e^{B_t^H} + 1) dt \\ &+ 2(e^{B_t^H} - 1) e^{rt+B_t^H} dB_t^H = re^{rt} (1 - e^{2B_t^H} + 2e^{2B_t^H} - 2e^{B_t^H}) dt + \gamma_t S_t dB_t^H = \beta_t dA_t \\ &+ 2(e^{B_t^H} - 1) e^{B_t^H} re^{rt} dt + \gamma_t S_t dB_t^H = \beta_t dA_t + \gamma_t S_t (r dt + dB_t^H) = \beta_t dA_t + \gamma_t dS_t. \end{aligned}$$

Das Portfolio θ ist also ein Arbitrageportfolio. Wie lässt sich dieses Puzzle lösen?

Hu und Øksendal (2003) [HØ] weisen darauf hin, dass allein die Wahl des Integraltypes für die Arbitragefreiheit bei der Definition eines selbstfinanzierenden Portfolios entscheidend sein kann. Es spielt natürlich auch die Spezifik der fBm eine große Rolle. Wir müssen folgendes beachten. Erstens, eine solche Strategie wäre mit einer gewöhnlichen Brownschen Bewegung nicht zulässig. Da weder β_t noch γ_t deterministische Funktionen sind, müsste man über $W_t dW_t$ bzw. eine Funktion davon integrieren. Wir wissen aber, dass nach dem Satz (6.4.3) das Riemann-Stieltjes-Integral über $W_t dW_t$ nicht existiert. Selbstverständlich kann auch die Anwendung der Formel (6.5.1) auf die gewöhnliche Brownsche Bewegung nicht gerechtfertigt werden. Dies ist der klassische Fall, in dem man auf die pfadweise Integration im Riemann-Stieltjes-Sinne verzichten muss, um korrekte Ergebnisse zu erhalten. Durch Anwendung der Itô-Formel entsteht bekanntlich ein zusätzlicher Term zweiter Ordnung, der sonst mögliche Arbitragegewinne neutralisiert.

Etwas anders ist die Lage bei der gebrochenen Version der Brownschen Bewegung. Es existiert zwar ein Riemann-Stieltjes-Integral über $B_t^H dB_t^H$, *seine Anwendung im Finance führt aber zu Modellen, die nicht mehr arbitragefrei sind*. Formel (6.5.1) berücksichtigt allein die Tatsache, dass die quadratische Variation der fBm ein Null-Prozess ist. Sie lässt aber eine andere wichtige Feststellung außer Betracht: die $1/H$ -Variation der fBm ist nämlich für alle $pH < 1$ unendlich! Mit anderen Worten kann (und soll) die fBm *nicht* als deterministische Funktion behandelt werden.

Aus diesem Grund folgen wir [HØ] und ziehen eine Sichtweise vor, dass ausschließlich ein stochastisches Integral vom Itô-Typ zur Modellierung der Märkte geeignet ist. Nach [HØ] liegt das Problem nicht an der fehlenden Markov- bzw. Semimartingaleigenschaft der fBm. Es ist vielmehr durch einen ungeeigneten Integraltyp verursacht. Diese Sichtweise ist insoweit gerechtfertigt, da in den letzten Jahren von der „norwegischen Schule“ (B. Øksendal und Kollegen) zahlreiche Anwendungen der fBm entwickelt worden sind. Das Spektrum reicht von Girsanov-Satz, Clark- und Itô-Formel⁵⁰ bis fraktionaler Black/Scholes-Formel bzw. Erweiterung und Lösung des Merton-

⁵⁰ Aase u.a. (2000).

Problems der Portfolio- und Konsumoptimierung in einem fraktionalem Markt.⁵¹ Alle diese Modelle schließen Arbitragemöglichkeiten aus.

6.6 Vorhersage der gebrochenen Brownschen Bewegung

Aus dem Blickwinkel der Langzeitabhängigkeit ist der bedingte Erwartungswert $E[B_t^H | B_s^H]$, $0 < s < t < \infty$ einer fBm von besonderem Interesse. Die Markov-Eigenschaft eines gewöhnlichen Wiener Prozesses W_t besagt, dass sämtliche Informationen über die vergangene Entwicklung in seinem aktuellen Wert enthalten sind. Da die fBm kein Markov-Prozess ist, stellt sich die Frage: lässt sich eine ähnliche Aussage bezüglich der fBm gegebenenfalls in einer abgeschwächten Form treffen?

Die Langzeitabhängigkeit impliziert, dass der Speicherbedarf zur Abbildung der Vergangenheit bei der fBm ohne Grenzen wächst. Bleibt dabei eine Möglichkeit zu effizienter Informationsverarbeitung erhalten? Offensichtlich ergibt sich ein unlösbares Dilemma: entweder können die Informationen nur im beschränkten (wenn auch sehr großen) Umfang bei der Preisbildung berücksichtigt werden oder man approximiert eine fBm durch einen Prozess, dessen Abhängigkeiten sehr weit (aber nur endlich weit) in die Vergangenheit zurückreichen. In beiden Fällen geht ein Stück Information verloren. Welche Auswirkungen ergeben sich daraus? Wie groß ist der Approximationsfehler?

Unerwartet hilfreich erweist sich der Gedanke, zur Lösung dieses Problems eine spezifische Eigenschaft der fBm, die *Selbstähnlichkeit*, heranzuziehen. Bei einem selbstähnlichen Prozess können Verteilungen über die längeren Zeitintervalle durch das Skalieren ihrer Gegenstücke, die sich auf die kürzeren Zeitintervalle beziehen, zurückgewonnen werden. Dafür variiert man die Länge der Historie bei einem bedingten Erwartungswert $E[B_t^H | B_s^H]$, $0 < s < t < \infty$ und versucht auf dieser Basis, die Güte der Vorhersagbarkeit zu vergleichen. Ist die Vergangenheit als Informationsquelle tatsächlich etwas Wert, dann sollte sich dies im Hinblick auf die zukünftige Entwicklung bemerkbar machen. Man würde erwarten, dass sich die Güte der Vorhersagbarkeit mit abnehmender Periodenlänge der Vergangenheitswerte reduziert.

⁵¹ Hu u.a. (2003).

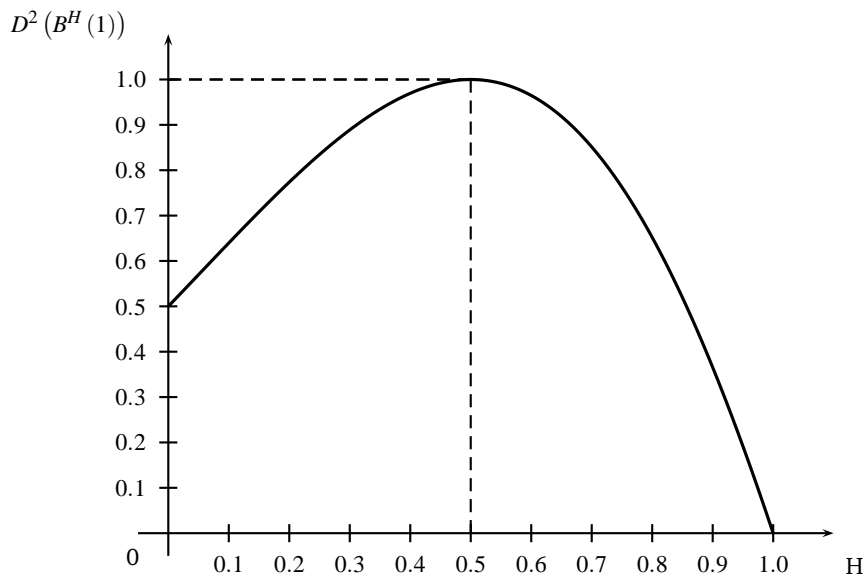


Abb. 6.1. Relative Varianz des Schätzers $D^2(B^H(1))$ als Funktion von H . Quelle: Nuzman und Poor (2000).

Ein Mathematischer Ausdruck für den bedingten Erwartungswert als Lösung des sogenannten *prediction problems* der fBm ist von mehreren Autoren gefunden worden.⁵² Die Länge des Beobachtungszeitraums aus der Vergangenheit spielt dabei eine wichtige Rolle. Im Idealfall ist sämtliche Historie $I = (-\infty, 0)$ zugänglich. Im allgemeinen sind wir nach Beobachtung eines H -ss Prozesses $\{Y_t, t \in I\}$ an einer linearen Schätzung einer Zufallsvariable Z interessiert. Nuzman und Poor (2000) untersuchen anhand der mittleren quadratischen Abweichung

$$D^2(Z) \equiv \mathbb{E} \left[(Z - \hat{Z})^2 \right]$$

die Varianz des Schätzers und zeigen, dass $D^2(Y(a))$ proportional zu a^{2H} ist, wobei a die Intervalllänge der Historie bezeichnet. Intuitiv würde man erwarten, dass die unmittelbare Zukunft leichter als die weit entfernte Zukunft vorherzusagen ist. Dies impliziert einen besseren Schätzer für kleinere a 's. Dies ist hier tatsächlich der Fall: der

⁵² Gripenberg und Norros (1996), Pipiras und Taqqu (2001) und Nuzman und Poor (2000).

Schätzfehler für sehr kleine Zeitintervalle $a \rightarrow 0$ kann beliebig klein gemacht werden. Allerdings bleibt die *normierte Varianz des Schätzers*

$$\begin{aligned} d^2(Y(a)) &\equiv \frac{D^2(Y(a))}{\mathbb{E}[Y(a)^2]} = \frac{a^{2H} D^2(Y(1))}{a^{2H} \mathbb{E}[Y(1)^2]} = d^2(Y(1)) \\ &= \frac{\sin(\pi(H - 1/2))}{\pi(H - 1/2)} \cdot \frac{\Gamma(3/2 - H)^2}{\Gamma(2 - 2H)} = \frac{\Gamma(3/2 - H)}{\Gamma(2 - 2H)\Gamma(H + 1/2)} \end{aligned}$$

unabhängig von a . In diesem Sinne “... it is just as difficult to predict the next millisecond of a self-similar process, based on the infinite past, as it is to predict the next millenium.”⁵³

Abbildung 6.1 zeigt die relative Varianz des Schätzers. Es ist erkennbar, dass die prediktive Kraft der Vergangenheit im mittleren H -Bereich nicht allzu hoch ist. Die Historie vor dem Zeitpunkt 0 erklärt die Hälfte der Varianz von $B^H(a)$ für $H \sim 0,85$. Für $H = 1/2$ ist die Vorhersage natürlich unmöglich. Im rechten Bereich des Bildes, in dem die fBm immer mehr persistent wird, fällt der relative Fehler langsam (im Extremfall $H = 1$ bis auf Null) ab. Andererseits, für $H < 1/2$ hilft die Kenntnis der gesamten Vergangenheit nur die Hälfte der Varianz zu erklären. Grippenberg und Norros (1996) kommen zum Schluss, dass es relativ wenig Unterschied macht, ob uns die Information über die Zufallsvariable Z aus einem Intervall $(-a, 0)$ oder $(-\infty, 0)$ vorliegt. Trotzdem bleibt das Intervall $(-\infty, 0)$ als Idealfall erhalten.

6.7 Überlegungen zur Arbitragefreiheit des Marktes. P.S.

Nachdem diese Arbeit fast fertiggestellt war, sind einige neue Arbeiten zum Thema Arbitragefreiheit der fBm veröffentlicht worden. Diese Tatsache sollte nicht unberücksichtigt bleiben. So entstand dieses Kapitel als Stellungnahme zu diesen Arbeiten.

Der Hauptvorteil der fBm als mathematisches Modell ist die Möglichkeit, eine globale zeitliche Abhängigkeit mittels eines stochastischen Prozesses zu modellieren. Das Hauptproblem bei ihrer Anwendung als Underlying Prozess im Finance ist die theoretische Unzulässigkeit der Vorhersagbarkeit der Renditen als Geschenk für die Investoren. Jede

⁵³ Nuzman und Poor (2000), S. 435.

Veränderung des Risikos zu Gunsten bzw. zu Lasten der Investoren hat ihren Preis bzw. wird mit einer Risikoprämie belohnt. Abgesehen vom Diversifikationseffekt ist dieser Preis positiv und endlich.

Die Modellierung der Aktienkursverläufe durch Semimartingale impliziert, dass der wahrscheinlichste (typische) Pfad immer ein Martingalpfad ist. Dieser Pfad repräsentiert gleichzeitig den Verlauf der gesamten Unsicherheit über die Zeit. Die andere, zweite Komponente eines Semimartingals ist ein vorhersagbarer Prozess, der völlig risikolos ist und im Finance gewöhnlich als Handelsstrategie interpretiert wird. Wählt ein Investor durch die Festlegung der Portfoliogewichte den Martingalpfad, ist er dem Maximum des Risikos (gemessen durch die Varianz der Trajektorie) ausgesetzt. Seine Gewinne bzw. Verluste bei der Glattstellung der Position sind zufällig, da der Martingalpfad ein typischer Pfad ist. Eine Reduzierung bzw. eine Erhöhung des Risikos ist ausschließlich durch die Änderung der Portfoliogewichte möglich. Das Portfolio besteht dabei lediglich aus zwei Wertpapieren: einem risikobehafteten Wertpapier, das dem Martingalpfad folgt, und einem risikolosen Wertpapier, das eine sichere Auszahlung leistet.

Was bedeutet ökonomisch die Tatsache, dass mit der fBm als Underlying Prozess die Semimartingalumgebung verlassen wird? Ist die fehlende Semimartingaleigenschaft eine hinreichende Bedingung für die Arbitrage? Nach unserem Kenntnisstand bleibt diese Frage derzeit immer noch offen. Offensichtlich existiert in diesem Fall kein äquivalentes Martingalmaß. Dies ist aber nicht hinreichend für die Existenz einer Arbitragestrategie. Weiterhin haben Delbaen und Schachermayer (1994, 2000) gezeigt, dass ein „Nicht-Semimartingal“ die NFLVR⁵⁴-Eigenschaft nicht erfüllt. Diese Eigenschaft ist strenger als die Arbitragefreiheit (NA, No Arbitrage) und kann in erster Näherung auf die Unzulässigkeit von Verdoppelungsstrategien zurückgeführt werden: sind die Verluste aus einer Handelsstrategie (gleichmäßig) beschränkt, dann sollen auch die Gewinne (in Wahrscheinlichkeit) beschränkt sein.⁵⁵

Rogers (1997) hat demonstriert, wie die Verletzung der NFLVR-Eigenschaft gewinnbringend ausgenutzt werden kann. Dabei hat er sich auf Annahmen gestützt, die aus unserer Sicht fragwürdig erscheinen. Erstens, es wird angenommen, dass die gesamte Entwicklung des Pro-

⁵⁴ No Free Lunch with Vanishing Risk.

⁵⁵ Delbaen und Schachermayer (2000).

zesses bis auf minus Unendlich bekannt ist:⁵⁶

$$\mathbb{E}[X_t | \mathcal{G}_0] = \int_{-\infty}^0 \left\{ (t-s)^{H-1/2} - (-s)^{H-1/2} \right\} dW_s,$$

wobei $\mathcal{G}_t \equiv \sigma(\{W_u : u \leq t\})$. Es wird nicht erläutert, wie dieser Erwartungswert zu bestimmen ist (im Folgenden wird klar, dass die technischen Details von entscheidender Bedeutung sind). Nehmen wir an, der Erwartungswert ist korrekt bestimmt. Dann entsteht ein direkter Widerspruch zur EMH: was ist der Grund dafür, dass ein *bekannt* Erwartungswert $\mathbb{E}[X_t | \mathcal{F}_0] > 0$ von den restlichen Marktteilnehmern in ihren Handelsstrategien unberücksichtigt bleibt? Die fBm ist schließlich ein *zentrierter* Gausscher Prozess und die Feststellung, dass seine Varianz für $H > 1/2$ mit der Zeit überlinear ($\sim t^{2H}$) ansteigt, kann für einen rationalen Investor nur eines bedeuten: das Risiko einer langfristigen Handelsstrategie kann durch eine laufende Umschichtung des Portfolios (mindestens zum Teil) eliminiert werden. Ist die Risikoreduktion kostenlos möglich, handelt es sich um eine Art von „Anti-Zeitdiversifikation“, weil eine langfristige *buy and hold* Strategie aus der Sicht des Rendite/Risiko-Verhältnisses ganz klar suboptimal ist. Andernfalls ist eine Option auf das Langzeitgedächtnis vorstellbar, die den Wert der vorgeschlagenen Arbitragestrategie quantifizieren soll.

Zweitens, Rogers (1997) verwendet kein risikoloses Wertpapier bei der Konstruktion der Arbitragestrategie. Abgesehen davon, dass der Markt dadurch unvollständig wird, ist dies unter gewöhnlichen Umständen irrelevant, da die Zinsverluste in einem endlichen Zeitintervall selbstverständlich beschränkt sind. Dies impliziert aber eine versteckte Annahme über die Absorptionszeit oder die First Passage Time (FPT) für den Portfolio-Prozess X_t bei einem festgelegten Level b , definiert als

$$\tau_b = \inf\{t \geq 0 : X_t \geq b\}, \quad X_0 = x < b. \quad (6.7.1)$$

Die Arbitragestrategie in Rogers (1997) basiert auf der Glättstellung der Position sobald sich ein Gewinn $X_t - b$ ergibt. Mit einem risikolosen Wertpapier ist dies nur dann sinnvoll, wenn die FPT τ_b endlich ist. Es ist bekannt, dass

$$\mathbb{P}(\tau_b \leq T) = \mathbb{P}\left(\sup_{t \leq T} X_t \geq b\right). \quad (6.7.2)$$

⁵⁶ Rogers (1997), S. 2.

Die Verteilung von τ_b steht damit in einem unmittelbaren Zusammenhang mit der Verteilung von $\sup_{t \leq T} X_t$. Nun bleibt zu zeigen, dass die fBm mit $H > 1/2$ über ein beliebiges stochastisches Intervall sehr schwach variieren kann.

Satz 6.7.1 Sei $H \in (1/2, 1)$. Für alle $\varepsilon, T > 0$ und alle Stoppzeiten $\tau \leq T$ f.s. gilt:

$$\mathbb{P} \left(\sup_{t \in [\tau, T]} |B_t^H - B_\tau^H| < \varepsilon \right) > 0. \quad (6.7.3)$$

Beweis: Guasoni (2004), S. 7. □

Daraus folgt $\mathbb{P}(\tau_b \geq T) > 0$ für jedes $B_1^H < b$ und alle T : das Warten auf die Gewinne kann sehr lange dauern. Die sicheren Zinsverluste sind dann nicht mehr vernachlässigbar, da sie linear mit der Zeit steigen und zum Zeitpunkt T_b ebenfalls den Level b erreichen. Gilt $\mathbb{P}(\tau_b \geq T_b) > 0$, ist die Strategie nicht mehr risikolos. Zum Vergleich: für eine gewöhnliche Brownsche Bewegung als Spezialfall der fBm mit $H = 1/2$ gilt⁵⁷ $\mathbb{P}(\tau_b < \infty) = 1$. Dies ist ein klassisches Resultat: die Brownsche Bewegung bricht jede Schranke $b < \infty$ f.s. (wobei $E[\tau_b] = \infty$ ist). Die Dichtefunktion der FPT für $H = 1/2$ ist

$$f_{\tau_b}(t) = \frac{1}{\sqrt{\pi}} \frac{|b|}{t^{3/2}} \exp\left(-\frac{b^2}{2t}\right). \quad (6.7.4)$$

Für $H > 1/2$ ist die entsprechende Verteilung nicht bekannt. Es konnte aber gezeigt werden,⁵⁸ dass die Verteilung der FPT für $H > 1/2$ und große b ein Pareto-ähnliches Verhalten aufweist:

$$\mathbb{P}(\tau_b > T) \sim T^{-(1-H)}, \quad T \rightarrow \infty. \quad (6.7.5)$$

Molchan (1999) zeigte, dass

$$\mathbb{P} \left(\max_{t \in \Delta} B_t^H < x \right) \rightarrow 0 \quad \text{für } x = T^{-H}, \quad (6.7.6)$$

wobei Δ ein Intervall $(0, 1)$ oder eine beschränkte Umgebung in der Nähe von Null ist. Er gab auch eine logarithmische Abschätzung für das Maximum der fBm an.

⁵⁷ Siehe z.B. Steele (2001), Theorem 4.6 S. 56.

⁵⁸ Sinai (1997), Molchan (1999), Ding und Yang (1995) (Vermutung), Rangarajan und Ding (2000).

Satz 6.7.2 Für das Maximum M_t der fBm B_t^H auf dem Intervall $[0, t]$ gilt

$$\ln \mathbb{P}(M_t < 1) = D \ln t^{-1} \left(1 + O\left((\ln t)^{-1/2}\right) \right), \quad t \rightarrow \infty, \quad (6.7.7)$$

mit $D = 1 - H$. Mit anderen Worten, $\lim_{t \rightarrow 0} \mathbb{P}(M_t < 1) = 0$.

Beweis: Molchan (1999), Theorem 1, S. 101. □

Dies spricht dafür, dass *lange Exkursionen* der fBm B_t^H abseits von B_0^H für höhere Werte $H > 1/2$ immer wahrscheinlicher werden, und lässt die Aussichten auf die sicheren Arbitragegewinne in weite Ferne rücken. Als Nebeneffekt entsteht ein wichtiger Ansatzpunkt für die fBm als mathematisches Werkzeug zur Modellierung empirischer Phänomene wie *Bubbles*, die im Rahmen der EMH als Anomalien bezeichnet werden.

Die obigen Ausführungen zeigen, dass eine Beweisführung auf Grundlage von versteckten Annahmen zu falschen Schlussfolgerungen führen kann. Allein die Verletzung der NFLVR-Eigenschaft reicht noch nicht aus, um Arbitragestrategien konstruieren zu können. Übrigens kommt es auch allgemein in der Wissenschaft nicht selten vor, dass die versteckten Annahmen unbemerkt bleiben (vgl. Beispiel 6.1).

In einer aktuellen Veröffentlichung lehnen Björk und Hult (2005) im Gegensatz zu Elliott und van der Hoek (2003)⁵⁹ sowie Hu und Øksendal (2003) die Verwendung von Wick-Kalkülen als Grundlage für ein stochastisches Integral für die fBm mit dem Argument ab, dass *“the definition of the self-financing trading strategies and/or the definition of the value of a portfolio used in the above papers does not have a reasonable economic interpretation, and thus that the results in these papers are not economically meaningful.”* Folglich bleibe die Definition des stochastischen Integrals, die pfadweise erfolgt, alternativlos. In dieser Umgebung könnten selbstfinanzierende Strategien angegeben werden, die zur Arbitragegewinnen führen. Aus diesem Grund sei die fBm als Underlying für ein risikobehaftetes Wertpapier ungeeignet.

Diese Argumentation lässt sich ebenfalls widerlegen, wenn man berücksichtigt, dass bei den diskreten Handelsstrategien das Langzeitgedächtnis der fBm berücksichtigt werden muss. Dies impliziert u.a.,

⁵⁹ Diese Arbeit ist uns leider erst im Zusammenhang mit Björk und Hult (2005) bekannt geworden.

BEISPIEL 6.1 Entdeckung des Neptuns

Im 19-ten Jahrhundert berechneten Astronomen die Laufbahn des Urans mit Hilfe der Newton'schen Mechanik und stellten fest, dass die berechneten Werte nicht mit den Beobachtungen übereinstimmen. Wären sie streng nach Popper vorgegangen, hätte die Newton'sche Mechanik als wissenschaftliche Disziplin für ungültig erklärt werden müssen. Stattdessen postulierten die Astronomen die Existenz eines anderen, noch unentdeckten Planeten. Sie berechneten seine erwartete Laufbahn und fanden den Planet tatsächlich in der geschätzten Richtung. Es war der Planet Neptun.

Die Lösung des Rätsels lag in einer versteckten Annahme (auxiliary hypothesis), dass es keinen anderen unbekannteten Planeten gibt, der die Urans-Laufbahn signifikant beeinflussen kann.

dass wegen der fehlenden Semimartingaleigenschaft der zu einem Arbitragegewinn führende Erwartungswert eigentlich kein korrekter Erwartungswert ist, weil er nicht an die korrekte Filtration adaptiert ist. Verwendet man dagegen einen *quasi-bedingten* Erwartungswert $\hat{E}_t X := \hat{E}_t [X | \mathcal{F}_t^H]$ mit $\mathcal{F}_t^H = \mathcal{B}(B_s^H, s \leq t)$, bleibt das selbstfinanzierende Portfolio arbitragefrei und ökonomisch wohldefiniert.⁶⁰

Zusammenfassend steht aus ökonomischer Sicht fest: solange die EMH nicht endgültig widerlegt ist, kann man vermuten, dass die long-memory-Abhängigkeiten in den Aktienrenditen, Volatilitäten usw. vorhanden sind. Man muss aber begründen, warum diese Abhängigkeiten nicht von allen Investoren gewinnbringend ausgenutzt werden können. Es kommt nicht auf die mathematischen Verfahren an, sondern auf die viel allgemeinere Hypothese, z.B. die EMH, die durch diese Verfahren widerlegt oder bekräftigt werden soll. In Bezug auf die schwache Form der Markteffizienz ist es ökonomisch sinnvoll, über die Reduzierung von Risiken zu sprechen unter Berücksichtigung sämtlicher preisrelevanten Informationen. Solange diese Informationen kostenlos erhältlich sind, gibt es keinen Grund, eine zusätzliche Rendite dafür zu verlangen. *“It is a particular property of most long-memory processes that seeming patterns arise and fall, appear and disappear. They could vanish at any instant. They have no real permanence. They cannot be predicted. ... Human nature yearns to see order and hierarchy in the world. It will invent it where it cannot find it.”*⁶¹

⁶⁰ Für eine detaillierte Darstellung vgl. Necula (2002), Hu und Øksendal (2003) und Benth (2003).

⁶¹ Mandelbrot und Hudson (2004), S. 189.

7. Skalierung anschaulich: Fraktale und Multifraktale

“Until Mandelbrot, we never had a proper theory of the irregular, the rough – all the annoying imperfections that we normally try to ignore in life.”

– Richard L. Hudson,

Einleitung zu Mandelbrot und Hudson (2004)

7.1 Fraktale

Die Motivation

Man kann sich vorstellen, dass manche Wirtschaftswissenschaftler beim Wort „Fraktal“ schnell misstrauisch werden. Dieses Misstrauen ist aber absolut nicht gerechtfertigt. Die fraktale Geometrie¹ ermöglicht eine anschauliche Darstellung der Irregularität in ihrer einfachsten Form, der Selbstähnlichkeit. *“Recall the definition of a fractal: a pattern or object whose parts echo the whole, only scaled down.”*²

Die Anwendung der Fraktale im Finance ist leicht zu motivieren.³ Natürlich sind die Aktienkurse/Renditen keine Schneeflocken und ginge es allein darum, dass *“all charts look the same”*, wäre es kaum notwendig, darüber Bücher zu schreiben. Der Grundgedanke besteht einfach darin, das Risiko eines Wertpapiers an die Irregularität der entsprechenden Zeitreihe zu knüpfen. Unter dieser Annahme kann die Risikomessung durch Quantifizierung der Irregularität des Aktienkursver-

¹ Mandelbrot (1987), Falconer (1993).

² Mandelbrot und Hudson (2004), S. 208.

³ Mandelbrot (1997).

laufs erfolgen. Da die Quantifizierung der Irregularität ein Schwerpunkt der fraktalen Geometrie ist, können auf dieser Grundlage die anspruchsvollen Modelle im Finance entwickelt werden.

*“What have we here? A new tool to measure, not how long, heavy, hot, or loud something is, but how convoluted and irregular it is. It provides science with its first yardstick for roughness.”*⁴ Der Vorteil einer fraktalen Analyse ist offensichtlich: die Irregularität ist ein allgemeinerer Begriff als beispielsweise Volatilität. Damit lässt sich über die Annahme der Pfadstetigkeit hinaus auch die Intensität der Sprünge in den Renditezeitreihen erfassen. Die fraktale Analyse ermöglicht eine Modellierung der Schwankungsbandbreite auf unterschiedlichen Zeit- und Raumskalen. Es lassen sich zeitliche Abhängigkeiten nachbilden, die beliebig lange in die Vergangenheit zurückreichen. Die befürchteten Volatilitätscluster stellen modellmäßig kein Problem mehr dar. Schließlich können die wildesten Preisbewegungen nicht als Ausreißer oder Ausnahme, sondern als Regelfall modelliert werden. Ist man bereit zuzugeben, dass die Märkte um ein vielfaches riskanter sind, als man gewöhnlich annimmt, erhält man mit der fraktalen Analyse ein sehr wirkungsvolles Tool zur Modellierung der Finanzrisiken.

Die Ideen, die Benoît Mandelbrot in den letzten vierzig Jahren hervorgebracht hat, erscheinen auf den ersten Blick sehr gewagt. Es ist auch nicht überraschend, dass die neueren Erkenntnisse, die das gewohnte Weltbild auf den Kopf stellen, zunächst mit Skepsis empfunden werden. Manchmal müssen einige Jahrzehnte vergehen, bevor eine wissenschaftliche Idee eine breite Unterstützung findet. Es hat beispielsweise ein halbes Jahrhundert gedauert, bis die Arbeit von Louis Bachelier (1900) durch Paul Samuelson um 1954 wiederentdeckt wurde.

Die auf der Grundlage der fraktalen Geometrie basierende Theorie der Fraktale ist viel mehr als nur ein *„äußerst effizientes und wirkungsvolles Instrument zur Beschreibung komplexer Phänomene und nichtlinearer dynamischer Systeme“*.⁵ Sie formalisiert eine neue Denkweise, die ihre Wurzeln aus der Beobachtung der Natur nimmt und die zu helfen versucht, die Gesetze der Natur (*“The Fractal Geometry of Nature”*) und der Finanzwirtschaft (*“Fractals and Scaling in Finance”*) besser zu ver-

⁴ Mandelbrot und Hudson (2004), S. 131.

⁵ Nordmeier (1999), S. 10.

stehen. Mandelbrot selbst schreibt darüber:⁶ *“My aim: To change the way people think, so that reform may go forward.”*

Fraktale sind geometrische Objekte: Punkte, Linien, Ebenen. All diese Objekte werden in der Mathematik unter dem Sammelbegriff „Mengen“ erfasst. Im Unterschied zu den herkömmlichen Objekten können Fraktale gerade wegen ihrer spezifischen Eigenschaften nicht mit konventionellen Methoden erfasst und beschrieben werden; die Werkzeuge der fraktalen Geometrie sind anders als die der euklidischen Geometrie. Diese Unmöglichkeit einer adäquaten Beschreibung mittels der euklidischen Geometrie ist ein charakteristisches Merkmal der Fraktale.

Mandelbrot hat erkannt, dass viele irreguläre Flächen in der Natur nicht nur durch Zufall, sondern auch durch sehr einfache Formeln generiert werden können. Beschreibt die klassische Geometrie die von Menschen gemachten Objekte, so stehen im Mittelpunkt der fraktalen Geometrie die von der Natur erzeugten Strukturen. *„Warum wird die Geometrie oft als „nüchtern“ und „trocken“ bezeichnet? Nun, einer der Gründe besteht in ihrer Unfähigkeit, solche Formen zu beschreiben, wie etwa eine Wolke, einen Berg, eine Küstenlinie oder einen Baum. Wolken sind keine Kugeln, Berge keine Kegel, Küstenlinien keine Kreise. Die Rinde ist nicht glatt – und auch der Blitz bahnt sich seinen Weg nicht gerade.“*⁷

Damit entwickelte sich die „fraktale Idee“ weit über die Beschreibung der geometrischen Objekte hinaus. Zu den weiteren Einsatzgebieten der Fraktale zählt die Modellierung komplexer nichtlinearer dynamischer Systeme, die Chaosforschung, die DNA-Analyse und viele mehr. Mehrere Jahrhunderte wird der Mensch damit konfrontiert, dass sein Auge manchmal mehr sieht, als viele mathematische Formeln zum Ausdruck bringen können. Auf der Abb. 7.2 sind einige Beispiele fraktaler Objekte aufgezeichnet. Trotz eines einfachen Musters ist es lange Zeit keinem Wissenschaftler gelungen, die geeigneten Beschreibungsmethoden für solche Objekte zu finden, obwohl die Muster selbst keineswegs neu sind. Ein globales *Umdenken* war erforderlich, um eine scheinbare Komplexität auf die einfacheren Konzepte reduzieren zu können.

⁶ Mandelbrot und Hudson (2004), S. 27.

⁷ Mandelbrot (1987).

Das Optische

Man spricht über die *Selbstähnlichkeit*, wenn die geeignet gewählte Ausschnittvergrößerungen eines Objektes dem Ausgangsgegenstand ähneln. Eine selbstähnliche Menge $S \subset \mathbb{R}^N$ kann als eine Vereinigung von einigen disjunkten Untermengen S_1, \dots, S_k dargestellt werden. Diese Untermengen entstehen ihrerseits durch Skalierung, Verschiebung und Rotation der ursprünglichen Menge.

Fraktale Strukturen sind selbstähnlich. Das zeigen Peitgen u.a. (1992) am Beispiel eines Blumenkohls. Der Blumenkohl ist aus kleinen Rosen aufgebaut. Bricht man eine solche Rose heraus und vergrößert sie, so ist das Bild schwer von einem ganzen Blumenkohl zu unterscheiden. Die Rose selbst ist wieder aus noch kleineren Röschen zusammengesetzt, deren vergrößerte Darstellung wiederum das Bild des ganzen Blumenkohls liefert. Entscheidend dabei ist, dass auf jeder Skala immer das gleiche Muster auftritt. Der Blumenkohl hat eine *skaleninvariante* oder *selbstähnliche* Struktur. In der fraktalen Geometrie untersucht man den Aufbau und die Eigenschaften solcher Strukturen.

Rein äußerlich sind Fraktale keine geometrischen Formen, sondern eher neue Arten, bestehende Formen aneinanderzufügen. Man nennt sie *Anordnungsvorschriften*. Eine Anordnungsvorschrift beschreibt, wie ein neues geometrisches Objekt aus einem vorhandenen durch Skalierung, Verschiebung, Rotation, Aufteilung oder Vervielfältigung erzeugt wird. Grundsätzlich besteht die Konstruktion fraktaler Objekte aus einem Basisobjekt (*Initiator*), auf den die Anordnungsvorschrift (*Generator*) mehrmals angewendet wird. Dabei werden die verkleinerten Kopien des Initiators auf die vom Generator vorgeschriebene Weise zusammengefügt. Das gesuchte fraktale Objekt oder einfach das Fraktal entsteht nach mehreren Iterationen, wobei das Abbruchkriterium individuell festgelegt wird.

Zu den klassischen fraktalen Mengen zählen die schon in dem einleitenden Kapitel erwähnten *Cantor-Mengen* (Abb. 7.1). Die Anordnungsvorschrift zum Aufbau einer Cantor-Menge ist sehr einfach: beginnend mit dem Einheitsintervall $[0, 1]$ entfernt man Schritt für Schritt das mittlere Drittel. Dann bleiben nach der ersten Iteration die beiden Intervalle $[0, 1/3]$ und $[2/3, 1]$ der Länge $1/3$ zurück. In der zweiten Iteration wiederholt sich das Vorgehen mit den zwei neu erzeugten Intervallen, woraus jetzt vier Intervallen der Länge $1/9$ entstehen. Nach

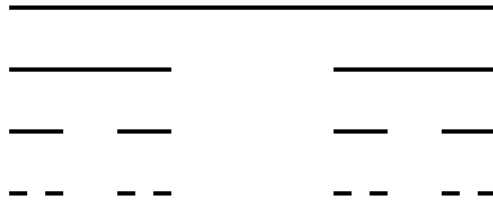


Abb. 7.1. Die Aufbau einer gleichmäßigen Cantor-Menge. In jedem Iterationsschritt wird ein mittleres Drittel von jedem Segment entfernt.

dem n -ten Schritt verbleiben 2^n Intervalle der Länge $1/3^n$ und der Rest wird entfernt. Nach einer solchen sukzessiven Entfernung der Teile bleiben nach der n -ten Iteration $2^n + 2$ Elementen der Cantor-Menge zurück, nämlich die Endpunkte aller geschlossenen Intervalle. Setzt man das iterative Verfahren unendlich fort, so entsteht die Cantor-Menge. In der Grenzwertbetrachtung besteht sie nur aus den isolierten Punkten, allerdings aus unendlich vielen.

Ein weiteres, optisch sehr ansprechendes Beispiel einer effizienten Naturbeschreibung mit Hilfe der fraktalen Geometrie ist die Koch-Kurve (Abb. 7.2), erzeugt von der schwedischen Mathematikerin Helge von Koch in 1904.

Man beginnt mit einer geraden Linie (Initiator). Das mittlere Drittel wird dann durch ein gleichseitiges Dreieck ohne dessen Grundlinie ersetzt. Durch das Zusammenfügen mehrerer auf diese Weise gedrehter Abschnitte entsteht eine Schneeflocken-Kurve.

Die Anwendung solcher Anordnungsvorschriften führt zweifellos zur Entstehung selbstähnlicher oder skaleninvarianter Figuren: nimmt man ein Mikroskop und spielt mit dem Okular, dann ist auf den unterschiedlichen Skalen immer dieselbe Struktur zu beobachten. Alternativ bietet sich zur Charakterisierung der Fraktale das Konzept *charakteristischer Länge* an. Dieser Begriff ist für das Verständnis der Fraktale von besonderer Bedeutung: sie sind nämlich durch das Fehlen dieser

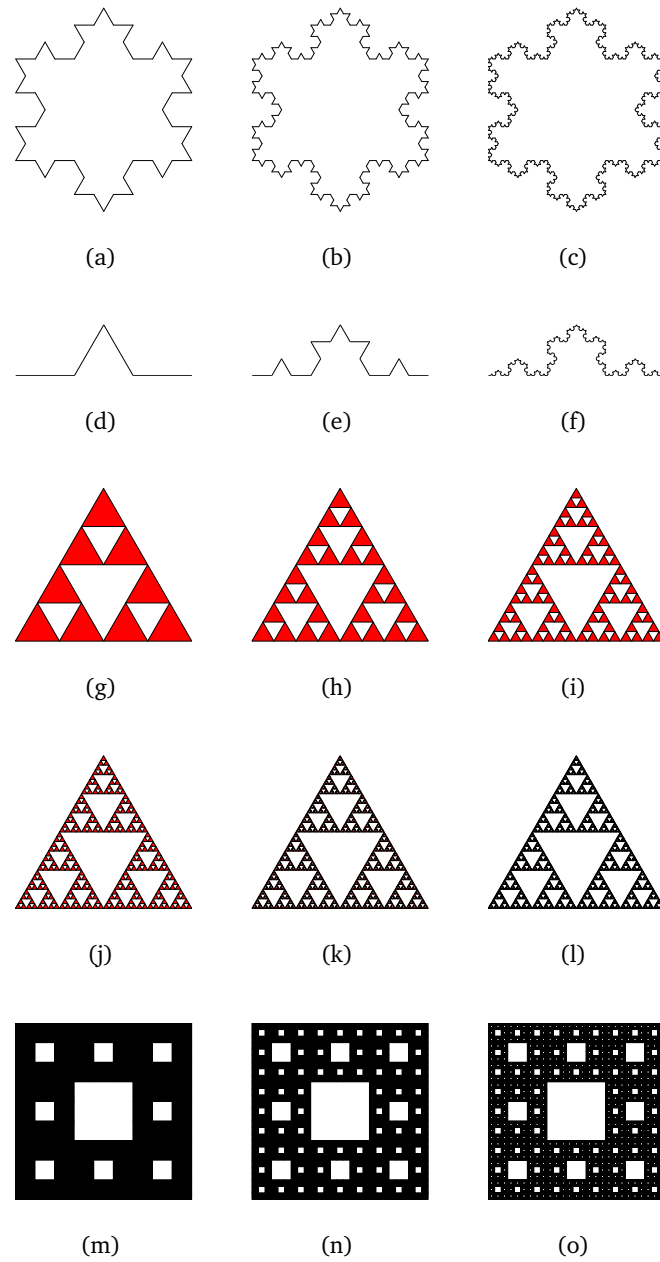


Abb. 7.2. Die iterative Konstruktion verschiedener Fraktale: Eine *Koch-Schneeflocke* (a)-(c) setzt sich aus drei *Koch-Kurven* (d)-(f) zusammen; Sierpinski-Dreieck (g)-(l); Sierpinski-Teppich (m)-(o).

Eigenschaft gekennzeichnet⁸. Einer Cantor-Menge wird mit jeder Iteration immer mehr entnommen, so dass eine anfangs durchgezogene Linie im Endeffekt zu einer Punktmenge degeneriert. Für die Koch-Kurve ist dagegen ein anderes Extremum charakteristisch: sie „bläht sich immer mehr auf“ und die Perimeter-Länge geht gegen unendlich. Dabei zeigt sich ein weiteres typisches Merkmal der Fraktale besonders deutlich: die *fehlende Glattheit (smoothness)*. An keinem Punkt der Koch-Kurve lässt sich eine Tangente anlegen. Das Gebilde besteht ausschließlich aus Ecken! Die mathematischen Begriffe dafür sind die Stetigkeit und die Nichtdifferenzierbarkeit. Dies erklärt auch die Herkunft des Wortes Fraktal: ‘fractus’ auf Latein heißt „gebrochen“.

Ganz unten in der Abb. 7.2 sind schließlich noch zwei sehr verbreitete Arten der Fraktale zu sehen: das Sierpinski-Dreieck und der Sierpinski-Teppich. Das Verfahren zur Erzeugung dieser Strukturen ähnelt dem der Cantor-Menge. Man beginnt im ersten Fall mit einem Dreieck auf einer Ebene und entfernt aus der Mitte eine verkleinerte auf 180 Grad gedrehte Kopie. Beim „Teppich“ ist die Ursprungsfigur kein Dreieck, sondern ein Quadrat. Es wird zunächst in neun kleinere Quadrate unterteilt; dann entfernt man dasjenige, das im Zentrum liegt. Damit stellen die beiden Figuren eine Verallgemeinerung der Cantor-Menge dar.

Die fraktale Dimension

Wenn auch eine rein qualitative und intuitive Beschreibung der Fraktale bei der Modellierung der Vorgänge in den Natur- und Sozialwissenschaften hilfreich und sinnvoll ist, hätte sich die fraktale Idee ohne Entwicklung geeigneter quantitativer Kennzahlen wahrscheinlich kaum weiter verbreitet. Es stellte sich heraus, dass die Grundeigenschaft der Fraktale, die Irregularität, mit einer einzigen Kennzahl beschrieben werden kann. Sie heißt *fraktale Dimension*⁹ und stellt eine Verallgemeinerung des herkömmlichen Dimensionsbegriffs dar, indem sie eine Aussage darüber ermöglicht, bis zu welchem Grad das Fraktal seinen Trägerraum (also zum Beispiel \mathbb{R}^2 oder \mathbb{R}^3) ausfüllt. Allein die Tatsache, dass diese Kennzahl nicht ganzzahlig sein braucht, macht die Besonderheit der fraktalen Mengen noch einmal recht deutlich. Kann man sich etwa eine Dimension von 1,7 vorstellen?

⁸ Takayasu (1990).

⁹ Mandelbrot (1974).

Die herkömmliche Dimension bezeichnet man zur Klarheit als *topologische Dimension*. Sie ist gut dafür geeignet, die *glatten* geometrischen Objekte zu charakterisieren: die topologische Dimension eines Punktes ist Null, einer Linie Eins, einer Ebene Zwei, einer Kugel Drei. Durch eine Dimension wird die minimale Anzahl von Parametern festgelegt, die zur vollständigen Bestimmung der Lage eines Punktes im Raum notwendig ist. Man kann dabei an linear unabhängige Vektoren in einem Vektorraum denken. Bei einem alternativen Ansatz überführt eine Dimension D eine Länge ε (oder eine Messlatte) durch das Potenzgesetz

$$\mu \sim \varepsilon^{-D}$$

in ein Maß μ für den „Inhalt“ (also eine Länge, eine Fläche oder ein Volumen). Wie kommt ein solcher Zusammenhang zustande?

Die glatten, euklidischen Kurven haben eine endliche Länge, die mit einer geeigneten Messlatte beliebig genau gemessen werden kann. Im Grenzfall lassen sich solche Kurven durch eine Gerade approximieren, so dass zumindest stückweise eine Ableitung (oder eine Tangente) existiert. Die Skaleninvarianz der Fraktale bedeutet dagegen immer dieselbe funktionale Form auf jeder (auch beliebig kleinen) Skala. Die Irregularität bleibt trotz abnehmender Intervalllänge bzw. zunehmender Präzision der Messlatte unverändert! Dies hat zur Folge, dass die herkömmliche Länge zur Charakterisierung fraktaler Strukturen ungeeignet ist. Die Skaleninvarianz impliziert, dass nicht die Länge der Kurve, sondern ihre Irregularität, gemessen durch die fraktale Dimension, unverändert bleibt. Die Messlatte muss dagegen mit Änderung der Skala selbst entsprechend angepasst werden. Die Grundeigenschaft der Fraktale besteht also in einer gleichmäßigen Aufteilung des Maßes (=Irregularität der Form) zwischen den verschiedenen Skalen, wengleich dafür die Ganzzahligkeit der Dimension (=die herkömmliche Messlatte) geopfert werden muss.

Die Länge der Küstenlinie von Grossbritannien sollte beispielsweise nach den herkömmlichen Methoden, gemessen in Metern und, sagen wir, in Dezimetern, nach der Umrechnung exakt übereinstimmen. Das tut sie aber nicht! *“But now try measuring that jagged coastline mentioned earlier. Something unusual develops as you use ever-smaller rulers: The length you measure is growing faster than the rulers shrinking. And that phenomenon is measured by a quantity called fractal dimension.”*¹⁰

¹⁰ Mandelbrot und Hudson (2004), S. 130.

BEISPIEL 7.1

Wir skizzieren hier nur den Gedankengang, der zur Definition der Hausdorff-Dimension führt, und verweisen auf Falconer (1993) für eine exakte Definition und technische Details.

Es sei U eine geschlossene und beschränkte Untermenge des N -dimensionalen Euklidischen Raums \mathbb{R}^N . Eine Überdeckung von U der Größe ε mit einer Menge von Hyperwürfeln \mathcal{H}_ε wird so definiert, dass

- i) jeder Hyperwürfel die Seitenlänge ε hat und
- ii) die Menge U in der Vereinigung $\bigcup \mathcal{H}_\varepsilon$ enthalten ist.

Ist $N(\varepsilon)$ die minimale Anzahl der zur Überdeckung von U notwendigen Hyperwürfel für jedes ε , dann kann die Hausdorff-Dimension von U anhand der Skalierungseigenschaften von Überdeckungen $N(\varepsilon)$ definiert werden, wenn ε gegen Null geht. Nach Falconer (1993), S. 38, existiert ein solcher Grenzwert für jede Teilmenge U von \mathbb{R}^n .

Es gibt mehrere zum Teil unterschiedliche Möglichkeiten, diesen Ansatz zu formalisieren. Deshalb spricht man von einer *fraktalen Dimensionen* als Sammelbegriff zur quantitativen Charakterisierung der fraktalen Mengen. Stimmen Ergebnisse unterschiedlicher Definitionen wenigstens überein? Es gibt keine Universalantwort auf diese Frage. Manchmal ja, manchmal nein. Eine sehr detaillierte Darstellung geben Falconer (1993) und Peitgen u.a. (1992), sehr einfach liest sich Schroeder (1991). Die wohl bekannteste Dimension heißt *Hausdorff-Dimension*. Motiviert durch die Beobachtung, dass einige Mengen, gemessen auf unterschiedlichen Skalen, unterschiedliche Größen aufweisen, hat der deutsche Mathematiker Felix Hausdorff am Anfang des zwanzigsten Jahrhunderts nach Möglichkeiten einer geeigneten Skalennormierung gesucht. In seinem Artikel „Dimension und äußeres Maß“ (1918) leistete er nach eigenen Worten einen „kleinen Beitrag“ zur Maßtheorie. Sein Beitrag erlangte aber später eine so große Bedeutung, dass es zum Entstehen einer ganzen Theorie nichtganzzahliger Dimensionen gekommen ist.

Die Hausdorff-Dimension ist eine theoretische Größe, die für praktische Berechnungen wenig geeignet ist. Für rechnerische oder experimentelle Zwecke verwendet man gewöhnlich eine *Kapazitätsdimension*.¹¹ Man kann sie als eine vereinfachte Version der Hausdorff-Dimension beschreiben. Zur Berechnung der Kapazitätsdimension wird ein Fraktal (oder eine fraktale Menge S) in N ähnliche Stücke

¹¹ Hentchel und Procaccia (1983).

der Größe l^D (wobei l die Skala und N die Anzahl der Teile bezeichnet) derart aufgeteilt, dass die fraktale Struktur erhalten bleibt. Dieses Verfahren genügt der Konstruktionsgleichung

$$N \times l^D = 1. \quad (7.1.1)$$

Dabei ist zu beachten, dass die Anforderung nach Strukturhaltung eines Fraktals nur für eine einzige Konstante Zahl $0 < D_C < \infty$ erfüllt werden kann, wenn die Anzahl der Teile gegen unendlich (oder die Skalengröße gegen Null) geht. Für alle anderen $D \neq D_C$ divergiert der Ausdruck in (7.1.1) oder er verschwindet; beides ist einem Strukturbruch gleichzusetzen. Das Produkt $N \times l^D$ bleibt nur für $D = D_C$ konstant und endlich. Die Berechnungsvorschrift für die Kapazitätsdimension lautet:

$$D_C := \lim_{l \rightarrow 0} \frac{\ln N}{\ln(1/l)}. \quad (7.1.2)$$

Relevant für diese Berechnung ist also die minimale Anzahl $N(l)$ der Hyperwürfel (oder einfach Kugeln mit Radius l), die zur Überdeckung des Fraktals S notwendig ist. Ist S eine gewöhnliche geometrische Figur mit der endlichen Länge ($D = 1$), Fläche ($D = 2$) oder mit dem endlichen Umfang ($D = 3$), dann gilt $N(l) \sim l^{-D}$ und

$$D = -\lim_{l \rightarrow 0} \frac{\ln N(l)}{\ln(l)}. \quad (7.1.3)$$

Die Kapazitätsdimension D_C von S verallgemeinert diesen Zusammenhang auf fraktale Mengen und ist definiert als

$$D := -\liminf_{l \rightarrow 0} \frac{\ln N(l)}{\ln(l)}. \quad (7.1.4)$$

Das Maß von S ist dann¹²

$$\mu = \limsup_{l \rightarrow 0} N(l) l^D. \quad (7.1.5)$$

Es kann endlich oder unendlich sein. Um die Eindeutigkeit zu bewahren, verstehen wir im folgenden unter fraktaler Dimension die Kapazitätsdimension. Entstehen beispielsweise bei einer Koch-Kurve $_/__$ aus einem Teil in jedem Iterationsschritt $N = 4$ neue Teile der Länge $l = 1/3$, dann braucht man $N(l) = 4^n$ Kugeln der Größe $l = 3^{-n}$, um

¹² Mallat (1999), S. 202.

die ganze Kurve zu überdecken. Für die minimale Anzahl der Kugeln gilt

$$D = -\liminf_{l \rightarrow 0} \frac{\ln N(l)}{\ln(l)} = \frac{\ln 4}{\ln 3} \approx 1,26 > 1.$$

Die Cantor-Menge besteht in n -ter Iteration aus $N = 2^n$ Intervallen der Länge $l = 3^{-n}$:

$$D = -\liminf_{l \rightarrow 0} \frac{\ln N(l)}{\ln(l)} = \frac{\ln 2}{\ln 3} \approx 0,63.$$

Der Weg zur Singularität

Es liegt nur ein kleiner Gedankenschritt zwischen den deterministischen Fraktalen und den singulären Wahrscheinlichkeitsverteilungen. Dies lässt sich anhand eines alternativen Verfahrens zur Konstruktion der Cantor-Menge zeigen. Dieses Verfahren ist in der Literatur wegen seiner spezifischen Form als *Zeltabbildung* (*Tent-Map*) bekannt (Abb. 7.3). Wir bezeichnen das Einheitsintervall mit A und legen eine Iterationsvorschrift fest:

$$x_{n+1} := \begin{cases} 3x_n & \text{für } x \leq 1/2, \\ 3(1-x_n) & \text{für } x > 1/2. \end{cases}$$

Alle Punkte mit $x < 0$ werden dabei monoton auf $-\infty$ abgebildet. Jeder Punkt mit $x > 1$ wird zunächst auf den Punkt $3(1-x) < 0$ abgebildet und dann nach $-\infty$. Es wird insbesondere kein Punkt von außerhalb von A auf A abgebildet. Diese wichtige Beobachtung impliziert, dass ein einmal herausgenommener Punkt für immer außerhalb von A bleiben wird; man spricht von einer "kommie-zurück"-Eigenschaft. Für die Gesamtlänge der nach n -ter Iteration verbleibenden Intervalle gilt:

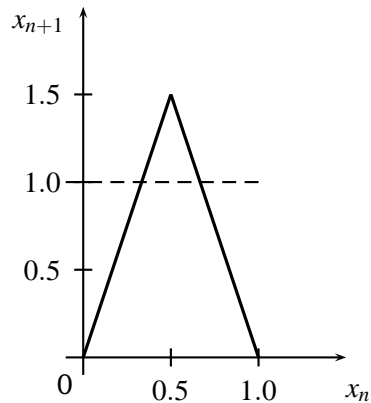


Abb. 7.3. Die Zeltabbildung.

$$L(n) = \left(\frac{2}{3}\right)^n = \exp(-\gamma n), \quad (7.1.6)$$

mit $\gamma = -\ln(2/3)$. Die Menge der verbleibenden Punkte heißt *Gefangenen-Menge*: gehört ein Punkt einmal dazu, wird er für immer darin bleiben. Bei der Grenzwertbetrachtung $n \rightarrow \infty$ erscheint diese Gefangenen-Menge sehr seltsam. Sie hat 2^∞ Punkte, also überabzählbar viele. Sie ist *intermittent*: es bleibt so viel Platz zwischen den Punkten, dass die Menge selbst überhaupt keinen Platz belegt. Für $n \rightarrow \infty$ bzw. $L(n) \rightarrow 0$ ist es eine Lebesgue-Null-Menge!

Ähnlich ist die Situation bei anderen Fraktalen. Ein Sierpinski-Dreieck besteht ebenfalls aus überabzählbar vielen Punkten, obwohl die Gesamtfläche entfernter Dreiecke sich zu eins aufsummiert:

$$\frac{1}{4} + \frac{3}{4^2} + \frac{3^2}{4^3} + \dots = \frac{1}{3} \sum_{i=1}^{\infty} \left(\frac{3}{4}\right)^i = \frac{1}{3} \left(\frac{1}{1-\frac{3}{4}} - 1\right) = 1.$$

Dasselbe gilt für die Gesamtfläche der aus dem Sierpinski-Teppich entfernten Quadrate:

$$\frac{1}{3^2} + \frac{8}{3^4} + \frac{8^2}{3^6} + \dots = \frac{1}{8} \sum_{i=1}^{\infty} \left(\frac{8}{9}\right)^i = \frac{1}{8} \left(\frac{1}{1-\frac{8}{9}} - 1\right) = 1.$$

Einige sehr interessante Bilder findet man in einer klassischen Arbeit über die algorithmische Informationstheorie.¹³ Zunächst geht es um die Koeffizienten des Pascal-Dreiecks. Von Relevanz sind eigentlich nicht die Koeffizienten selbst, sondern das Ergebnis einer arithmetischen Operation (mod 2) darüber. Die Zahlenpyramide in der Abb. 7.4 wurde nach der Regel

$$\binom{n+1}{k+1} \equiv \binom{n}{k+1} + \binom{n}{k} \pmod{2} \quad (7.1.7)$$

erzeugt. Die Anwendung dieser Regel bewirkte, dass alle geraden Glieder durch Nullen und alle ungeraden Glieder durch Einsen ersetzt worden sind. Mit anderen Worten, jeder Eintrag in der Pyramide ergibt sich aus einer logischen Operation EXCLUSIVE OR aus zwei Einträgen in der Zeile darüber: einem Eintrag in derselben Spalte und einem Eintrag in der nächstgelegenen Spalte links. Im zweiten Schritt werden die Nullen durch Leerzeichen ersetzt (Abb. 7.5). Auf diese Weise entsteht eine fraktale Fläche. Die Ähnlichkeit mit dem Sierpinski-Dreieck ist einfach nicht zu übersehen!

¹³ Chaitin (1987), S.27–29.

```

0  1
1  11
2  101
3  1111
4  10001
5  110011
6  1010101
7  11111111
8  100000001
9  1100000011
10 10100000101
11 111100001111
12 1000100010001
13 11001100110011
14 101010101010101
15 1111111111111111
16 10000000000000001
17 110000000000000011
18 1010000000000000101
19 11110000000000001111
20 100010000000000010001
21 1100110000000000110011
22 10101010000000001010101
23 111111110000000011111111
24 1000000010000000100000001
25 11000000110000001100000011
26 101000001010000010100000101
27 1111000011110000111100001111
28 10001000100010001000100010001
29 110011001100110011001100110011
30 10101010101010101010101010101
31 11111111111111111111111111111111
32 100000000000000000000000000000001
33 1100000000000000000000000000000011
34 10100000000000000000000000000000101
35 111100000000000000000000000000001111
36 1000100000000000000000000000000010001
37 11001100000000000000000000000000110011
38 101010100000000000000000000000001010101
39 1111111100000000000000000000000011111111
40 10000000100000000000000000000000100000001
41 110000001100000000000000000000001100000011
42 1010000010100000000000000000000010100000101
43 11110000111100000000000000000000111100001111
44 100010001000100000000000000000001000100010001
45 1100110011001100000000000000000011001100110011
46 10101010101010100000000000000000101010101010101
47 111111111111111100000000000000001111111111111111
48 100000000000000010000000000000001000000000000001
49 1100000000000000110000000000000011000000000000011
50 10100000000000001010000000000000101000000000000101

```

Abb. 7.4. Das Pascal-Dreieck Mod 2.

```

0  1
1  11
2  1 1
3  1111
4  1  1
5  11 11
6  1 1 1 1
7  11111111
8  1      1
9  11     11
10 1 1    1 1
11 1111   1111
12 1  1   1  1
13 11 11  11 11
14 1 1 1 1 1 1 1 1
15 1111111111111111
16 1              1
17 11             11
18 1 1            1 1
19 1111           1111
20 1  1           1  1
21 11 11          11 11
22 1 1 1 1        1 1 1 1
23 11111111       11111111
24 1      1       1      1
25 11     11      11     11
26 1 1    1 1    1 1    1 1
27 1111   1111   1111   1111
28 1  1   1  1   1  1   1  1
29 11 11 11 11 11 11 11 11
30 1 1 1 1 1 1 1 1 1 1 1 1
31 1111111111111111111111111111
32 1                                1
33 11                               11
34 1 1                              1 1
35 1111                             1111
36 1  1                             1  1
37 11 11                            11 11
38 1 1 1 1                           1 1 1 1
39 11111111                          11111111
40 1      1                           1      1
41 11     11                          11     11
42 1 1    1 1                         1 1    1 1
43 1111   1111                        1111   1111
44 1  1   1  1                         1  1   1  1
45 11 11 11 11                         11 11 11 11
46 1 1 1 1 1 1 1 1                     1 1 1 1 1 1 1 1
47 1111111111111111                    1111111111111111
48 1              1                     1              1
49 11             11                     11             11

```

Abb. 7.5. Das Pascal-Dreieck Mod 2, in dem die Nullen durch die Leerzeichen ersetzt sind.

l	Muster	$N(l)$	$D_n = \frac{\log N(l)}{\log(l)}$
1	1	1	.
2	11	3	1,5 849 625 007
3	101	5	1,4 649 735 207
4	1111	9	1,5 849 625 007
5	10001	11	1,4 898 961 024
6	110011	15	1,5 113 915 945
7	1010101	19	1,5 131 423 106
8	11111111	27	1,5 849 625 007
9	100000001	29	1,5 325 223 761
10	1100000011	33	1,5 185 139 399
11	1010000101	37	1,5 058 697 323
12	111100001111	45	1,5 319 136 798
13	1000100010001	49	1,5 173 088 259
14	11001100110011	57	1,5 320 058 501
15	101010101010101	65	1,5 414 733 701
16	1111111111111111	81	1,5 849 625 007
17	10000000000000001	83	1,5 596 568 529
18	110000000000000011	87	1,5 450 981 717
19	1010000000000000101	91	1,5 319 928 647
20	11110000000000001111	99	1,5 338 886 892
21	100010000000000010001	103	1,5 223 172 379
22	1100110000000000110011	111	1,5 236 057 972
23	10101010000000001010101	119	1,5 242 010 235
24	111111110000000011111111	135	1,5 434 838 553
25	1000000010000000100000001	139	1,5 329 805 192
26	11000000110000001100000011	147	1,5 317 018 782
27	101000001010000010100000101	155	1,5 302 411 260
28	1111000011110000111100001111	171	1,5 430 216 065
29	10001000100010001000100010001	179	1,5 405 197 725
30	110011001100110011001100110011	195	1,5 503 362 395
31	1010101010101010101010101010101	211	1,5 584 968 187
32	1111111111111111111111111111111111	243	1,5 849 625 007

Tabelle 7.1. Die Berechnung der fraktalen Dimension des Pascal-Dreiecks (mod2).

Die Berechnung der fraktalen Dimension gestaltet sich sehr einfach. Man summiert in der Pyramide die Einsen von oben nach unten auf und normiert diese Summe durch die Zeilennummer (beides auf der Logskala). Die Ergebnisse sind in Tabelle 7.1 zusammengefasst. Ein exakter Wert ist $D = \ln 3 / \ln 2 = 1,5849625$.

Eine Brücke zwischen Fraktalen, Skalierung und Potenzgesetzen entsteht durch den Gedanken, dass die Skaleninvarianz nicht exakt, sondern

nur *statistisch* gegeben ist. Wird ein Fraktal als Träger für eine Wahrscheinlichkeitsverteilung verwendet, ist das Wahrscheinlichkeitsmaß vollständig auf einer Lebesgue-Null-Menge konzentriert. Eine solche Verteilung ist singulär.¹⁴ Die Präsenz eines fraktalen Trägers ist dabei hinreichend, aber nicht notwendig. Zur Verdeutlichung bietet sich wiederum die Zeltabbildung bzw. die Cantor-Funktion an, in dem sie selbst als eine Wahrscheinlichkeitsverteilung auf einem (nicht fraktalen!) Intervall $[0, 1]$ betrachtet wird (Beispiel 7.2). Das Wahrscheinlichkeitsmaß verteilt sich rekursiv auf die einzelnen Skalen um. Im folgenden Kapitel geben wir weitere Beispiele singulärer Verteilungen, deren Träger kein Fraktal ist.

Eine rekursive Umverteilung des Maßes ist ein typisches Merkmal einer ganzen Klasse der *skaleninvarianten* Verteilungen. Die Änderung der Skala entspricht einer Transformation des Argumentes um einen konstanten Faktor: $x \rightarrow \lambda x$. Skaleninvarianz einer Verteilung liegt vor, falls für den Flankenbereich $\tilde{F}(x) = \mathbb{P}(X > x)$ der folgende Zusammenhang gegeben ist: $\tilde{F}(x) \sim \tilde{F}(\lambda x)$. Dieser Eigenschaft genügen in erster Linie die Potenz- oder ParetoGesetze vom Typ

$$\tilde{F}(x) \sim x^{-D},$$

weil das Verhältnis $\tilde{F}(\lambda x)/\tilde{F}(x) = \lambda^D$ unabhängig von x ist. Sie werden durch *fat tails* und unendliche Momente (spätestens ab dem zweiten Moment) gekennzeichnet. Im Gegensatz dazu gilt zum Beispiel für eine Standardnormalverteilung¹⁵

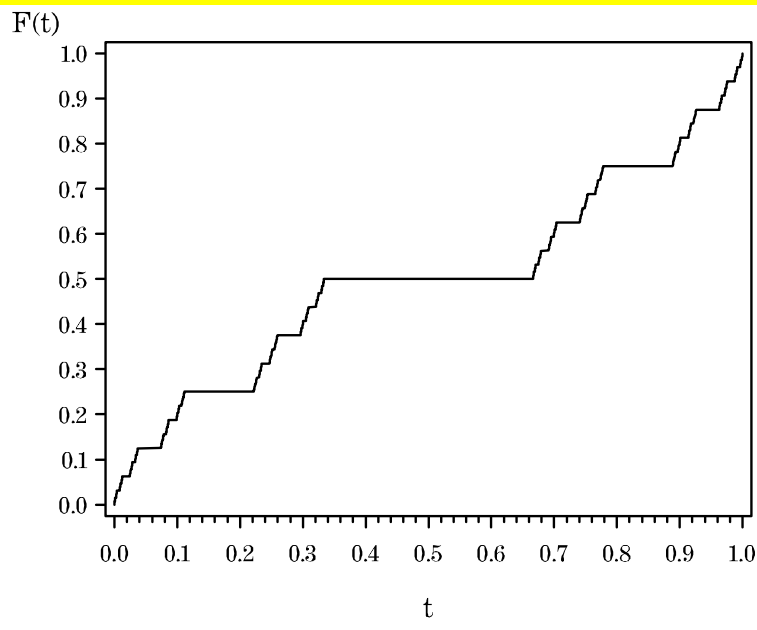
$$\tilde{F}(x) \sim \frac{1}{x\sqrt{2\pi}}e^{-x^2/2} \quad \text{und} \quad \tilde{F}(\lambda x)/\tilde{F}(x) \sim e^{-(\lambda^2-1)x^2/2}.$$

Die Tails sind im Vergleich zu einer Paretoverteilung nicht mehr so dick und klingen exponentiell ab. Die Standardnormalverteilung ist nicht skaleninvariant.

Im folgenden Kapitel soll eine weitere Möglichkeit zur Modellierung der Skaleninvarianz aufgezeigt werden: die Multifraktale. Sie lassen sich als eine (möglicherweise zufällige) Mischung von mehreren singulären Verteilungen interpretieren, jede auf dem eigenen fraktalen Träger. Mandelbrot (1997) zeigte, dass multifraktale Verteilungen ebenfalls zu einem Potenzgesetz führen und *fat tails* haben, ihre Momente bleiben aber endlich.

¹⁴ Vgl. Seite 204.

¹⁵ Feller (1957).



BEISPIEL 7.2 Die „Teufelstreppe.“

Das wohl berühmteste Exemplar einer fraktalen oder singulären Verteilung ist eine auf Grundlage der Cantor-Menge aufgebaute Verteilung, die wegen sehr spezifischer Eigenschaften ihrer Verteilungsfunktion oft als „Teufelstreppe“ bezeichnet wird. Sie ist stetig, aber nicht absolut stetig und hat keine Dichte, da für jede Borel-Menge $B \in \mathcal{B}(\mathbb{R})$ gilt: $\mathbb{P}(X \in B) = 0$. Im k -ten Schritt der Konstruktion wird das Maß \mathbb{P} zwischen 2^{k-1} Säulen *gleichverteilt*. Mathematisch lässt sich die „Teufelstreppe“ als Integral über das Cantor-Maß darstellen [Mallat (1999), S. 203.]:

$$F(t) = \int_0^t d\mu_\infty(x),$$

mit $\mu_\infty(x)$ als Grenzwert der Cantor-Iterationen. Diese Funktion ist rekursiv definiert:

$$F(t) = \begin{cases} p_1 F(3t) & \text{falls } t \in [0, 1/3], \\ p_1 & \text{falls } t \in [1/3, 2/3], \\ p_1 + p_2 F(3t - 2) & \text{falls } t \in [2/3, 1] \end{cases} \quad (7.1.8)$$

mit den Gewichtungsfaktoren $p_1 + p_2 = 1$. Funktionen mit einem kompakten Träger S , die auf disjunkten Untermengen S_i eine affine Transformation von sich selbst darstellen, heißen *selbstähnlich*. Im einfachsten Fall genügen sie der Iterationsregel $F(t) = p_i F(a_i t)$ für alle $t \in S_i$ auf S_i mit dem Skalen-Faktor a_i und dem Gewichtungsfaktor p_i .

7.2 Multifraktale

Grundlagen

Bei dem Begriff *Multifraktal* handelt es sich, wie man auch intuitiv vermuten kann, um eine Erweiterung des Begriffes *Fraktal*. Der Zusatz *Multi* deutet auf die Existenz einer ganzen *Familie von Skalierungsexponenten* hin, die untereinander hierarchisch geordnet sind. Einige sind stärker, die anderen schwächer ausgeprägt. *“If the fractals scale in many different ways at different points, they are **multifractal** – and their mathematical properties become intricate and powerful.”*¹⁶

Wurde die fraktale Geometrie eingeführt, um den Zusammenhang zwischen den Mustern (Mengen) und der Messskala (die sogenannte *Skalierungsbeziehung*) zu beschreiben, bilden die Multifraktale eine (womöglich auch unendliche) *Hierarchie* solcher Mengen. Zweifellos können Multifraktale als eine qualitative Erweiterung der Fraktale betrachtet werden. Eine (exakt oder statistisch) selbstähnliche fraktale Menge kann durch eine fraktale Dimension charakterisiert werden. Multifraktale sind keine Mengen, sondern singuläre Maße. Durch das Maß μ bekommt jedes Mitglied einer Menge eine Zahl oder eine Quantität zugewiesen. Diese Quantität kann eine Zufallsvariable sein. In diesem Fall wird ein Multifraktal zu einem *Zufallsmaß*. Im Allgemeinen wird durch ein Multifraktal eine singuläre Verteilung über der zugrundeliegenden Menge definiert. Die Singularitäten dieser Verteilung können einerseits mittels einer verallgemeinerten Dimension D_q bzw. einer Funktion $\tau(q)$ (die beiden Größen werden im Folgenden definiert) beschrieben werden. Eine alternative Beschreibung bietet ein *multifraktales Spektrum*, in dem die Hierarchie von lokalen Skalierungsexponenten α mittels einer Funktion $f(\alpha)$ analytisch und graphisch dargestellt wird. In der *multifraktalen Analyse* wird gezeigt, dass die beiden Alternativen über eine Legendre Transformation verbunden sind. Dies erklärt auch die unterschiedlichen Definitionen der Multifraktale in der Literatur.¹⁷

¹⁶ Mandelbrot und Hudson (2004), S. 127.

¹⁷ Die Literatur zum Thema Multifraktale ist sehr umfangreich, beispielsweise Mandelbrot (1974, 1997, 1999a, 1999b, 2001a, 2001b, 2001c, 2001d, 2004), Grassberger (1983), Hentchel und Procaccia (1983), Schertzer und Lovejoy (1983), Parisi und Frisch (1985), Meneveau und Sreenivasan (1987, 1991), Arneodo u.a. (1995), Jaffard (1999), Seuront u.a. (1999), Riedi (2003) u.v.m.

Definition 7.2.1 (Parisi und Frisch (1985), Mandelbrot (1990)). Ein Maß ist ein **selbstähnliches Multifraktal**, wenn es über die folgenden geometrischen Skalierungseigenschaften verfügt:

A. Es existiert ein Exponent α auf einem Intervall $[t, t + dt]$ derart, dass

$$\mu(dt) \text{ in der Größenordnung von } (dt)^\alpha \text{ ist.} \quad (7.2.1)$$

B. Die Menge der t -Werte mit gegebenem α ist ein Fraktal. D.h., die Anzahl der Intervalle mit einem solchen α bestimmt sich nach der Formel

$$N(dt) = (dt)^{-f(\alpha)}. \quad (7.2.2)$$

Die fraktale Dimension dieser Menge ist also keine Funktion von t , sondern nur eine Funktion $f(\alpha)$ von α .

Definition 7.2.2 (Holley und Waymire (1992)). Die **multifraktale Struktur** eines (möglicherweise Zufalls-) Maßes μ beschreibt Dimensionen der Mengen

$$E(\alpha) := \{x \in \text{supp}(\mu) : \mu B_\varepsilon(x) \sim \varepsilon^\alpha \text{ für } \varepsilon \rightarrow 0\} \quad (7.2.3)$$

als Funktion des Parameters α . $B_\varepsilon(x)$ bezeichnet eine abgeschlossene Kugel mit Radius $\varepsilon > 0$ mit Mittelpunkt x , $\text{supp}(\mu)$ ist der Träger des Maßes μ , und mit $B_\varepsilon(x) \sim \varepsilon^\alpha$ ist gemeint, dass $\lim_{\varepsilon \rightarrow 0} (\ln B_\varepsilon(x) / \ln \varepsilon) = \alpha$ ist.

Mit anderen Worten, der Träger des Maßes μ wird so auf die einzelnen Mengen $E(\alpha)$ zerlegt, dass das Maß μ auf $E(\alpha)$ immer dieselbe fraktale Dimension α hat. Enthält das ursprüngliche Maß mehrere unterschiedliche α 's, bedeutet eine solche Zerlegung bezogen auf den Träger des Maßes μ eine Dekomposition dieses Trägers in Abhängigkeit von α . Dabei wird jede Menge $E(\alpha)$ selbst zu einem Fraktal mit der fraktalen Dimension $f(\alpha)$. Im Unterschied zu einer globalen fraktalen Dimension beschreibt die Funktion $f(\alpha)$ nur die lokalen Eigenschaften des Maßes μ auf dem Träger $E(\alpha)$. Auf diese Weise kann jedem „Teilmaß“ $\mu_r = \mu(\alpha) |_{\alpha=\alpha_r}$ durch die Funktion $f(\alpha) |_{\alpha=\alpha_r}$ ein eigener Träger zugeordnet werden.

Die Vorteile einer multifraktalen Analyse werden besonders dann ersichtlich, wenn die in Betracht kommenden Signale oder Zeitreihen eine hohe Variabilität auf mehreren Zeitskalen aufweisen bzw. eine

sehr irreguläre lokale Struktur besitzen. In solchen Fällen ermöglicht die multifraktale Analyse eine kompakte Beschreibung komplexer Formen der Skalierung. Im folgenden gehen wir auf die Einzelheiten der multifraktalen Modellierung und die entsprechenden Werkzeuge detailliert ein.

Die verallgemeinerte Dimension

Eine der wichtigsten Kennzahlen der multifraktalen Analyse heißt verallgemeinerte Dimension. Der allgemeine Dimensionsbegriff für ein beliebiges Objekt wurde erstmals von Hentchel und Procaccia (1983) aufgestellt.

Sei $M_l \in \mathbb{N}$ die minimale Anzahl der Hyperwürfel i der Kantenlänge l , die zu einer vollständigen Überdeckung eines gegebenen Objektes notwendig sind. Weiterhin soll eine invariante Wahrscheinlichkeitsverteilung $\mathbb{P}(l)$ über diese Hyperwürfel i existieren. Die verallgemeinerte Dimension D der Ordnung $q \in \mathbb{N}$ ist definiert als

$$D_q = \frac{1}{q-1} \lim_{l \rightarrow 0} \frac{1}{\ln l} \ln \left[\sum_{i=1}^{M_l} p_i^q \right]. \quad (7.2.4)$$

Für unterschiedliche q ergeben sich entsprechend Dimensionen verschiedener Ordnungen. Das D_q ist dabei diejenige reelle Zahl, für die $M_l l^{D_q}$ im Limes $l \rightarrow 0$ endlich bleibt. Es gilt:

$$D_i \leq D_j \quad \forall i > j.$$

Die klassischen Beispiele der verallgemeinerten Dimensionen beziehen sich auf die ersten drei Ordnungen. Für $q = 0$ reduziert sich (7.2.4) auf

$$D_0 = - \lim_{l \rightarrow 0} \frac{\ln M_l}{\ln l} \quad (7.2.5)$$

und die resultierende Dimension D_0 heißt einfach *fraktale Dimension*. Es ist natürlich dieselbe Kennzahl, die im vorherigen Kapitel schon ausführlich diskutiert wurde. Wird D_0 für Objekte der euklidischen Geometrie berechnet, stimmt ihr Wert mit der bekannten topologischen Dimension (z.B. $D_0 = 1$ für eine Linie, $D_0 = 2$ für eine Fläche usw.) überein. Von der fraktalen Dimension spricht man gewöhnlich, wenn $D_0 \notin \mathbb{N}$ ist. Um das D_1 herzuleiten, muss auf (7.2.4) die

L'Hospital-Regel angewendet werden. Nach einigen Berechnungen ergibt sich

$$D_1 = \lim_{l \rightarrow 0} \frac{1}{\ln l} \left[\sum_{i=1}^{M_l} p_i \ln p_i \right]. \quad (7.2.6)$$

Die Kennzahl D_1 heißt *Informationsdimension* und erfasst die Menge an Informationen, die zur Beschreibung des gegebenen Objektes durch Hyperwürfel der Kantenlänge l benötigt werden. Für $q = 2$ wird aus (7.2.4) schließlich

$$D_2 = \lim_{l \rightarrow 0} \frac{1}{\ln l} \ln \left[\sum_{i=1}^{M_l} p_i^2 \right]. \quad (7.2.7)$$

Die resultierende Kennzahl D_2 heißt *Korrelationsdimension* und ist umso höher, je höher die Wahrscheinlichkeit ist, zwei zufällig ausgewählten Punkte des gegebenen Objektes unter dem gleichen Hyperwürfel zu finden.

Der Hölder-Exponent

Eine weitere Kennzahl, die in der multifraktalen Analyse von Bedeutung ist, heißt *Hölder-Exponent*.¹⁸ Der Hölder-Exponent wird gewöhnlich für die Charakterisierung der Regularität bzw. der Irregularität einer mathematischen Funktion verwendet. Dabei unterscheidet man zwischen einer *lokalen Regularität* an einem Punkt $x = x_0$ und einer *globalen Regularität* über ein geschlossenes Intervall $x \in [a, b]$.

Intuitiv¹⁹ versteht man unter einer solchen Quantifizierung die Möglichkeit, mit Hilfe einer reellen Zahl α zu zeigen, dass der Graph einer Funktion $f(x)$ im Punkt x_0 eine bestimmte „Glattheit“ besitzt. Die kleinstmögliche Stufe der Regularität ist die Stetigkeit: eine Funktion f ist stetig in x_0 , wenn im Limit $x \rightarrow x_0$ der Ausdruck $|f(x) - f(x_0)|$ gegen Null geht. Nach diesem Prinzip hat eine beschränkte, aber nicht stetige Funktion $\alpha = 0$, da in diesem Fall die Distanz zwischen $f(x)$ und $f(x_0)$, wenn sie größer Null ist, auch mit $x \rightarrow x_0$ nicht mehr verkürzt werden kann. Für $\alpha > 0$ ist $f(x)$ stetig; für $\alpha \geq 1$ ist sie auch differenzierbar. Je höher das α , desto „glatter“ ist die Funktion f in x_0 .

Daraus folgt, dass eine bestimmte Klasse der Funktionen, die stetig, aber nicht differenzierbar sind, ein α zwischen Null und Eins haben

¹⁸ Die andere verbreitete Bezeichnung ist der *Lipschitz-Exponent*.

¹⁹ In Anlehnung an Jaffard (1999).

muss. Mit anderen Worten, hat eine Funktion den Hölder-Exponenten α mit $\alpha > 0$ und $n \leq \alpha < n+1$, $1 \leq n \in \mathbb{N}$, dann ist sie notwendig n Mal, aber nicht $n+1$ Mal differenzierbar. In diesem Fall ist die Funktion f *singulär* in ihrer n -ten Ableitung. Durch den Hölder-Exponenten α wird dieses singuläre Verhalten charakterisiert bzw. quantifiziert.

Wir geben eine Definition des Hölder-Exponenten in Anlehnung an Mallat und Hwang (1992).

Definition 7.2.3. Sei n eine positive ganze Zahl und $n \leq \alpha \leq n+1$.

- Eine Funktion $f(x)$ hat einen Hölder-Exponent α in x_0 genau dann, wenn es zwei Konstanten C und $h_0 > 0$ sowie ein Polynom des Grades n , $P_n(x)$, existieren, so dass für $h < h_0$ gilt:

$$|f(x_0 + h) - P_n(h)| \leq C|h|^\alpha. \quad (7.2.8)$$

- Die Funktion $f(x)$ hat einen globalen Hölder-Exponenten α auf einem Intervall (a, b) genau dann, wenn (7.2.8) für $(x_0 + h) \in (a, b)$ erfüllt ist.
- Die Hölder-Regularität von $f(x)$ in x_0 ist das Supremum über alle α , so dass $f(x)$ den Hölder-Exponenten α in x_0 hat.
- Die Funktion $f(x)$ ist *singulär* in x_0 , wenn sie nicht den Hölder-Exponenten 1 in x_0 hat.

Nach dieser Vorgehensweise wird die Funktion $f(x)$ in einer kleinen Umgebung von x_0 durch das Polynom $P_n(x)$ *lokal* approximiert. Dies entspricht für die ganzzahligen α natürlich einer Taylor-Reihen Zerlegung. Doch geht die Hölder-Regularität über diese Zerlegung hinaus und verfeinert die obere Schranke für den Restterm durch die Einführung eines nichtganzzahligen Exponenten. Der Vorteil dieser verfeinerten Approximation ist offensichtlich: dadurch wird es überhaupt erst möglich, die nichtdifferenzierbaren Funktionen in einer Punktumgebung von x_0 zu beschreiben.

Bei einem schwach stationären stochastischen Prozess $\{X_t, t \in \mathbb{R}\}$ impliziert ein Hölder-Exponent $\alpha > 0$ die Stetigkeit der Pfade in L^2 . Die Differenzierbarkeit der Pfade kann allerdings nur bei $\alpha > 1$ erreicht werden. Mit anderen Worten, für $0 < \alpha < 1$ sind die Pfade von X stetig, aber nirgendwo differenzierbar. Das klassische Beispiel dazu ist die (gebrochene) Brownsche Bewegung.

Die Besonderheit der multifraktalen Maße liegt in ihrer Irregularität, die, gemessen am Hölder-Exponenten, von Punkt zu Punkt stark variiert. Der Hölder-Exponent eines multifraktalen Maßes bleibt nicht einmal innerhalb einer Skala konstant: die Punkte mit einem fixierten Wert des Hölder-Exponenten bilden selbst fraktale Mengen und breiten sich über mehrere Skalen aus.

Daher ist es nicht von Bedeutung, einen exakten Wert des Hölder-Exponenten in jedem einzelnen Punkt oder Intervall²⁰ zu ermitteln. Stattdessen werden die relevanten Informationen über die Größe und die Geometrie der Singularitäten extrahiert und in Form eines multifraktalen Spektrums dargestellt.

Die turbulenten Zeiten...

Das klassische Beispiel einer stark variierender Irregularität stammt übrigens aus der Physik. Es ist die hochentwickelte Turbulenz.

Die Forschung der turbulenten Strömungen geht auf die sogenannte K-41 Theorie von Kolmogorov (1941) zurück.²¹ Die ursprüngliche Annahme, dass der turbulente Fluss einen homogenen Prozess darstellt, wurde empirisch widerlegt. Es stellte sich heraus, dass nicht nur die skalaren Größen wie die Temperatur und die Geschwindigkeit der turbulenten Strömung, sondern auch die Verteilungen dieser Größen einer erheblichen Schwankung auf allen Zeitskalen unterliegen. Spätestens mit den Arbeiten von Kolmogorov (1962) und Obukhov (1962) stand fest: der turbulente Fluss ist äußerst inhomogen. Die Irregularität der Geschwindigkeit unterscheidet sich von Punkt zu Punkt. Dieses Phänomen (*Intermittenz* genannt) lässt vermuten, dass sich durch die Bestimmung des Singularitätsspektrums der Flussgeschwindigkeit (es stellt vermutlich eine nichttriviale Funktion dar) wichtige Informationen über die Natur der Turbulenz gewinnen lassen.²²

Ist es nur ein Zufall, dass die Worte „Aktienmarkt“ und „turbulent“ in Börsenberichten oft zusammen vorkommen? Mandelbrot (1997, 2001d, 2004) und Shiryaev (1999a) (Kapitel III und IV) diskutieren ausführlich die Anwendungen auf Kapitalmarktdaten. In 1999 wird

²⁰ Bei Intervallen spricht man von einem *groben (coarse)* Hölder-Exponenten.

²¹ Vgl. auch Obukhov (1941, 1949).

²² Jaffard (1999).

bei MaPhySto²³ sogar ein ganzer Workshop zum Thema “Turbulence and Finance” organisiert.²⁴

“Turbulence is dangerous. Its output – the pressure or velocity of water, the average or change in price – can swing wildly, suddenly. It is hard to predict, harder to protect against, hardest to engineer and profit from. Conventional finance ignores this, of course.” (Mandelbrot und Hudson (2004), S. 230).

Ghashghaie u.a. (1996) geben Hinweise auf die Ähnlichkeit in den statistischen Eigenschaften der Daten im Finance und in der hochentwickelten Turbulenz: “Guided by this similarity, we claim that there is an information cascade in FX market dynamics that corresponds to the energy cascade in the hydrodynamic turbulence ... The analogy gives a conceptual framework for understanding the short-term dynamics of speculative markets.” Die Schlussfolgerungen dieser Arbeit sind optimistisch: “We have reason to believe that the qualitative picture of turbulence that has been developed during the past 70 years, will help our understanding of the apparently remote field of financial markets.”

Das Binomialmaß

Ein multifraktales Maß kann durch ein elementares iteratives Verfahren, die sogenannte *multiplikative Kaskade*, erzeugt werden. In der einfachsten Variante einer solchen Kaskade wird das Maß in jedem Schritt des Verfahrens auf eine festgelegte Art zweigeteilt und heißt darum *Binomialmaß*.

Die Konstruktion eines Binomialmaßes besteht, wie es auch bei den Fraktalen üblich ist, aus einem Initiator und einem Generator. Der Initiator ist ein Maß μ_0 , das auf dem Einheitsintervall $[0, 1]$ definiert ist. Gewöhnlich ist das Maß μ_0 ein Wahrscheinlichkeitsmaß, d.h. $\mu_0[0, 1] = 1$. Der Generator oder die Anordnungsvorschrift für die Erzeugung der Kaskade kann wie folgt beschrieben werden.

Es werden zwei Zahlen $m_0, m_1 > 0$ so gewählt, dass $m_0 \neq m_1$ und $m_0 + m_1 = 1$. Diese Zahlen heißen *Multiplikatoren* und bleiben im Laufe des Verfahrens konstant. Im ersten Schritt wird das ursprüngliche Maß μ_0

²³ Centre for Mathematical Physics and Stochastics – MaPhySto, University of Aarhus, Denmark.

²⁴ <http://www.maphysto.dk/oldpages/events/Turbulence99/>, vgl. auch Shiryaev (1999b).

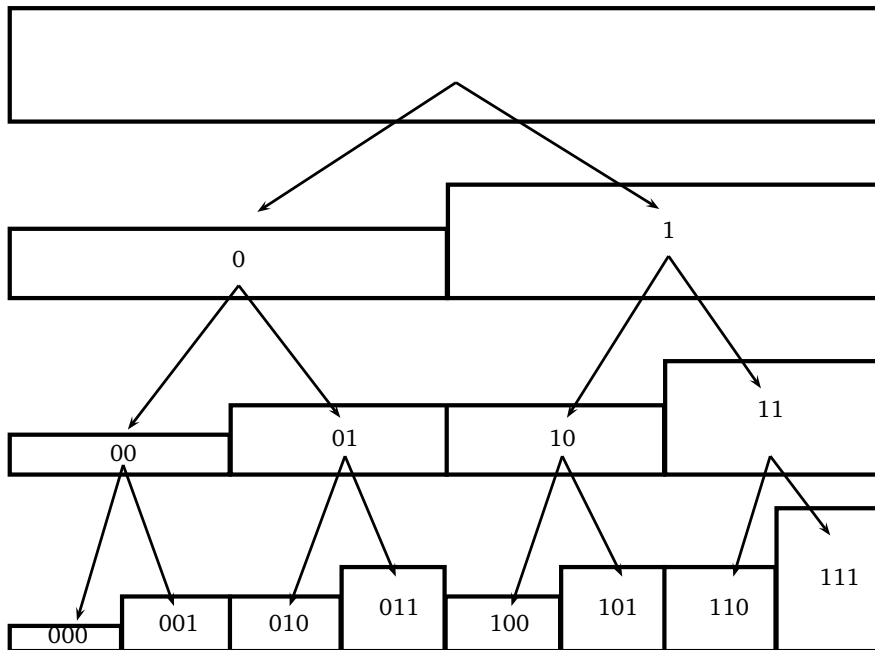


Abb. 7.6. Eine multiplikative Kaskade.

in das linke $[0, 1/2]$ und das rechte $[1/2, 1]$ Halbointervall im Verhältnis m_0/m_1 aufgeteilt. Daraus ergibt sich $\mu_1[0, 1/2] = m_0$ und $\mu_1[1/2, 1] = m_1$. Im zweiten Schritt wiederholt man diese Vorgehensweise bei jedem Teilintervall. Das Maß $\mu_1[0, 1/2]$ wird jetzt in $[0, 1/4]$ und $[1/4, 1/2]$ aufgeteilt, das Maß $\mu_1[1/2, 1]$ entsprechend in $[1/2, 3/4]$ und $[3/4, 1]$. Nach $n = 2$ Schritten sind $2^n = 4$ Intervalle der Länge $2^{-n} = 1/4$ mit den Maßen

$$\begin{aligned} \mu_2[0, 1/4] &= m_0 m_0, & \mu_2[1/4, 1/2] &= m_0 m_1, \\ \mu_2[1/2, 3/4] &= m_1 m_0, & \mu_2[3/4, 1] &= m_1 m_1 \end{aligned}$$

erzeugt. Bei jeder weiteren Iteration verdoppelt sich die Anzahl der Intervalle bzw. halbiert sich die Intervalllänge (vgl. Abb. 7.6). Im Weiteren entstehen nach dem n -ten Schritt $N_n = 2^n$ Intervalle der Länge $dt = 2^{-n}$ mit den Maßen

$$\mu(dt) = m_0^k m_1^{n-k}, \quad k = 0, 1, 2, \dots, n. \quad (7.2.9)$$

Davon bekommen

$$N_n(k) = \binom{n}{k} = \frac{n!}{k!(n-k)!} \quad (7.2.10)$$

Intervalle das gleiche Maß $m_0^k m_1^{n-k}$ zugewiesen. Dies ist leicht zu sehen, da die „Adresse“ eines beliebigen dyadischen Intervalls $[dt] = [t, t + 2^{-n}]$ durch eine Zahl zur Basis zwei

$$t = 0, \eta_1 \eta_2 \dots \eta_n = \sum_{i=1}^n \eta_i 2^{-i}, \quad \eta_i \in \{0, 1\} \quad (7.2.11)$$

eindeutig darstellbar ist. Für die Zuteilung des Maßes ist also ausschließlich die relative Frequenz der Nullen bzw. Einsen in der binären Zerlegung der Intervalladresse zuständig. Mit den Bezeichnungen φ_0 und φ_1 für die obigen relativen Frequenzen ergibt sich das Maß eines dyadischen Intervalls im n -ten Schritt:

$$\mu(dt) = m_0^{\sum_{i=1}^n \eta_i} m_1^{n - \sum_{i=1}^n \eta_i} = m_0^{n\varphi_0} m_1^{n\varphi_1} = (2^{-n})^\alpha. \quad (7.2.12)$$

Nach Eggleston (1949) hat für ein fixiertes $0 \leq \varphi_0 \leq 1$ und $\varphi_1 = 1 - \varphi_0$ die Untermenge $E(\alpha)$ der Zahlen t in $[0, 1]$, für die

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=1}^n \eta_i \rightarrow \varphi_0$$

gilt, die Hausdorff Dimension $-\varphi_0(\alpha) \log_2 \varphi_0(\alpha) - \varphi_1(\alpha) \log_2 \varphi_1(\alpha)$, wobei sich das $\varphi_0 \equiv \varphi_0(\alpha)$ aus der Lösung der Gleichung

$$-\varphi_0 \log_2 m_0 - \varphi_1 \log_2 m_1 = \alpha \quad (7.2.13)$$

ergibt.²⁵ Die Selbstähnlichkeit des Binomialmaßes ist aus der Konstruktion unmittelbar ersichtlich. Durch eine Aggregation des Maßes über die Intervalle mit unterschiedlichen φ_0 bzw. α entsteht ein multifraktales Maß auf $[0, 1]$.²⁶

Das Binomialmaß ist wie viele andere Multifraktale ein stetiges, aber (für $m_1 \neq 1/2$) nicht ein absolutstetiges Wahrscheinlichkeitsmaß. Damit ist die entsprechende Verteilungsfunktion singulär und hat keine Dichte (vgl. Abb. 7.7). Nach dem Binomialsatz bleibt das Originalmaß

$$\mu_0 = \sum_{k=0}^n \binom{n}{k} m_0^{n\varphi_0} m_1^{n\varphi_1}$$

in jedem Schritt der Konstruktion erhalten.

²⁵ Holley und Waymire (1992), S. 820

²⁶ Arneodo (1996), Riedi (2003).

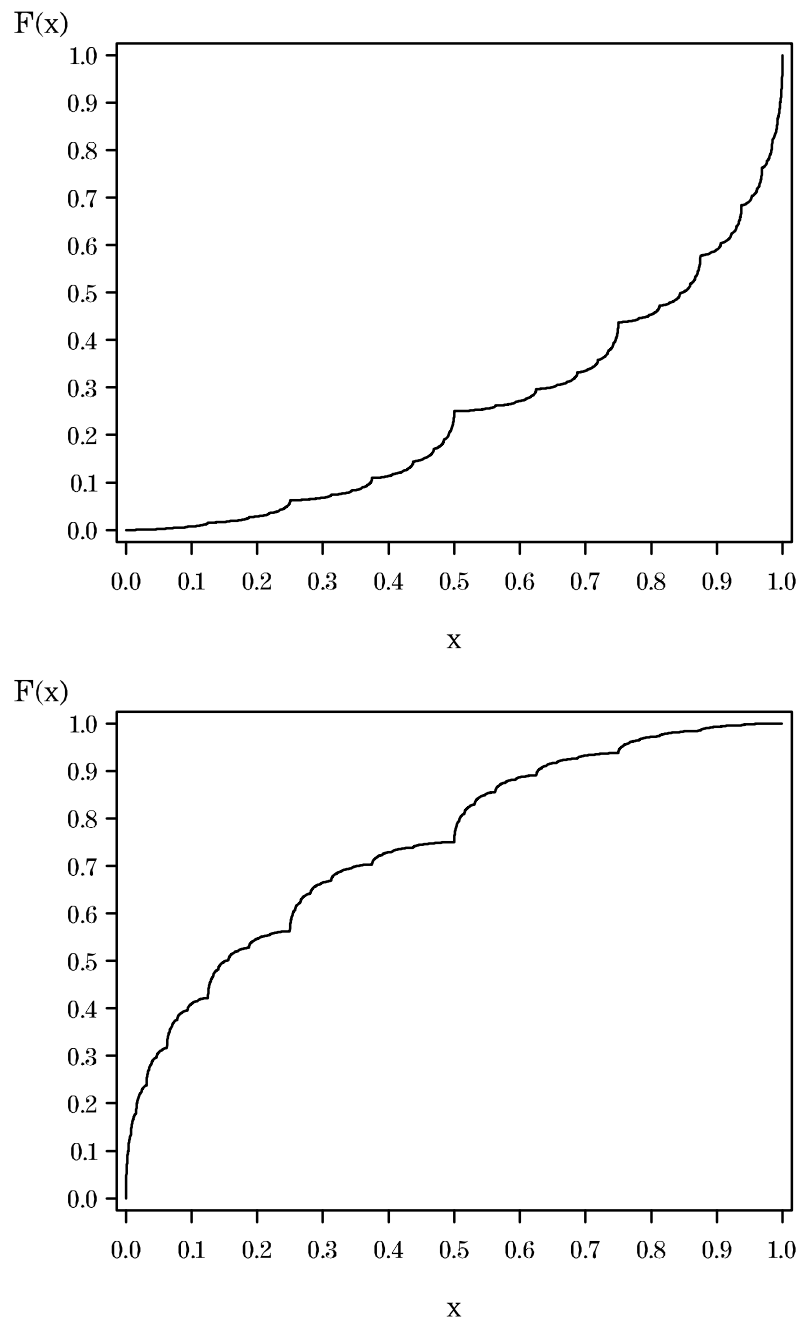


Abb. 7.7. Die mittels einer deterministischen multiplikativen Kaskade nach 7 Iterationen erzeugte singuläre Verteilung mit $m_0 = 0,25$ (oben) bzw. $m_0 = 0,75$ (unten).

Multifraktale auf einem fraktalen Träger

Das Verfahren zur Erzeugung einer multiplikativen Kaskade kann auf unterschiedliche Weise erweitert werden. Als Erstes ist es denkbar, eine Basis $b > 2$ zu verwenden. Dadurch entsteht ein *multinomiales* Maß. Diese Änderung ist unwesentlich und wird hier nicht weiter verfolgt.

Die nächste Erweiterung besteht darin, ein multifraktales Maß auf einem *fraktalen Träger* zu definieren. Zur Verdeutlichung greifen wir noch einmal auf die Cantor-Menge zurück, die als ein Spezialfall eines Multifraktals auf dem fraktalen Träger mit dem Multiplikator $m_1 = 1/2$ interpretiert werden kann (Abb. 7.8).

Der Initiator einer Cantor-Menge ist ein Maß μ_0 , das auf dem Einheitsintervall $[0, 1]$ definiert ist. Wir legen $\mu_0[0, 1] = 1$ fest. Der Generator teilt dieses Maß in jedem Schritt je zur Hälfte in zwei Intervalle der Länge $l_{n+1} = (1/3)l_n$ auf. Damit sind lediglich zwei Kennzahlen von Bedeutung: $\mu_G = 1/2$ und $l_G = 1/3$.

Im ersten Schritt wird das ursprüngliche Maß μ_0 in das linke $[0, 1/3]$ und das rechte $[2/3, 1]$ Drittel des Intervalls aufgeteilt. Daraus ergibt sich $\mu_1[0, 1/3] = 1/2$ und $\mu_1[2/3, 1] = 1/2$. In jedem weiteren Schritt wiederholt sich diese Vorgehensweise und im Schritt n hat man schließlich $N = 2^n$ Intervalle jeweils mit dem Maß $\mu_n = 2^{-n}$ und der Länge $l_n = 3^{-n}$. Die Hausdorff Dimension D_0 einer Cantor-Menge ergibt sich aus der Anforderung der Maßerhaltung auf jeder Stufe des rekursiven Verfahrens

$$\lim_{n \rightarrow \infty} N l_n^{D_0} \equiv \mu_0 = 1 \quad (7.2.14)$$

und folglich gilt $D_0 = \ln 2 / \ln 3$ wegen der Identität $3^{\ln 2 / \ln 3} = 2$. Die Gleichung (7.2.14) lässt sich auch mit μ_G und l_G beschreiben:

$$\lim_{n \rightarrow \infty} \left(\mu_G l_G^{-D_0} \right)^n = \mu_0 \quad \text{oder} \quad \mu_G = l_G^{D_0}. \quad (7.2.15)$$

Die Skalierungsbeziehung (7.2.15) kann auch mittels einer verallgemeinerten Dimension D_q (7.2.4) bzw. einer Funktion

$$\tau(q) = (q - 1)D_q \quad (7.2.16)$$

beschrieben werden. Für $q = 0$ ist die Funktion $\tau(q)$ identisch mit D_0 . Für $q \neq 0$ übernimmt der Ausdruck $\tau(q) / (q - 1)$ die Rolle der verallgemeinerten Dimension D_q . Ein *verallgemeinerter Generator* hat zwei Skalierungsexponenten: τ für den Träger des Maßes μ und q für das Maß

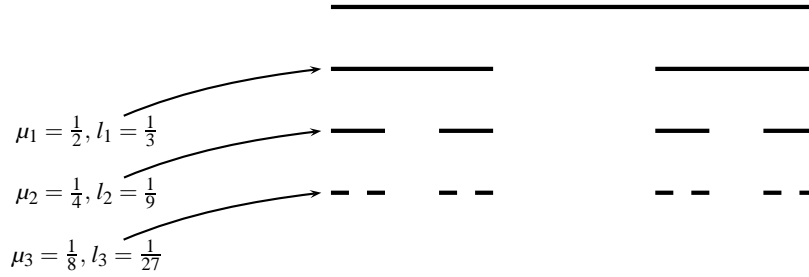


Abb. 7.8. Der Aufbau einer Cantor-Menge. Jedes Segment hat ein Maß $\mu_n = (\frac{1}{2})^n$ und eine Skala $l_n = (\frac{1}{3})^n$.

μ selbst. In dieser Umgebung erweitert sich die Gleichung (7.2.14) zu

$$\lim_{n \rightarrow \infty} N l_n^{D_q} = \lim_{n \rightarrow \infty} N^{(1-q)} l_n^{-\tau} \equiv \mu_0 = 1. \tag{7.2.17}$$

Da das Maß gleichmäßig in die Intervalle auf derselben Skala aufgeteilt wird, d.h. $\mu_n = N^{-1}$ und $N = 2^n$, gilt wegen (7.2.15) und (7.2.17):

$$\lim_{n \rightarrow \infty} N (\mu_G^q l_G^{-\tau})^n = \lim_{n \rightarrow \infty} (2 \mu_G^q l_G^{-\tau})^n = \mu_0. \tag{7.2.18}$$

Im Falle einer ungleichmäßigen Aufteilung bzw. einer multinomialen Kaskade zur Basis b schreibt sich (7.2.18) als

$$\lim_{n \rightarrow \infty} \left(\sum_{i=1}^b \mu_{G_i}^q l_{G_i}^{-\tau} \right)^n = \mu_0. \tag{7.2.19}$$

Damit (7.2.19) endlich bleibt, suchen wir nach q und τ , so dass

$$\sum_{i=1}^2 \mu_{G_i}^q l_{G_i}^{-\tau} = 2 \left(\frac{1}{2} \right)^q \left(\frac{1}{3} \right)^{-\tau} = 1 \tag{7.2.20}$$

gilt. Da die Cantor-Menge kein Multifraktal ist, erwarten wir, dass $D_q = D_0 = \ln 2 / \ln 3$ für alle q gilt und folglich $\tau = (q - 1) \ln 2 / \ln 3$. Nach dem Einsetzen in (7.2.20) bestätigt sich diese Erwartung:

$$2 \left(\frac{1}{2} \right)^q \left[\left(\frac{1}{3} \right)^{\frac{\ln 2}{\ln 3}} \right]^{(1-q)} = 2 \left(\frac{1}{2} \right)^q \left(\frac{1}{2} \right)^{(1-q)} = 1. \tag{7.2.21}$$

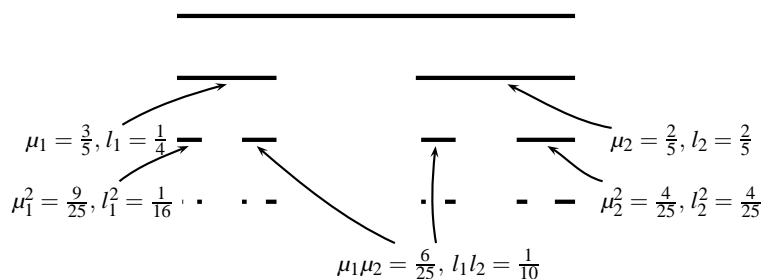


Abb. 7.9. Der Aufbau eines Multifraktals auf einem fraktalen Träger.

Die Herleitung des multifraktalen Spektrums ist in diesem Fall recht einfach. Für große n verhält sich das Maß μ in einem kleinen Intervall dt als dt^α . Die Anzahl der Intervalle dt mit dem gegebenen Wert α zwischen α und $\alpha + d\alpha$ ist $dt^{-f(\alpha)}$. Da die Cantor-Menge einen einzigen konstanten Hölder Exponenten $\alpha = \ln 2 / \ln 3$ hat, folgt $f(\alpha) = \alpha = \ln 2 / \ln 3$. Das multifraktale Spektrum ist in diesem Fall trivial und besteht aus einem einzigen Punkt. Weiterhin gilt $f(\alpha) = \alpha q - \tau(q)$.

Die Abb. 7.9 zeigt ein weniger triviales Beispiel. Das Maß $\mu_0 = 1$ wird hier ungleichmäßig zwischen zwei Intervallen unterschiedlicher Länge aufgeteilt. Die Anzahl der Intervalle spielt dabei eine untergeordnete Rolle und kann natürlich anders gewählt werden. Wir verwenden die Zahlen aus Hasley u.a. (1986) und belassen es bei zwei Intervallen zur Vermeidung unnötiger Komplexität.²⁷ Essentiell ist bei dieser Erweiterung eine *ungleichmäßige Aufteilung des Maßes* in die Intervalle. Dies führt zur Entstehung von mehreren unterschiedlichen Werten des Hölder-Exponenten α und einem nichttrivialen multifraktalen Spektrum.

Wie aus der Abb. 7.9 zu entnehmen ist, wird das Maß in jedem Schritt vom Generator $\Gamma(q, \tau, l)$ im Verhältnis 3 : 2 in zwei Intervalle der Länge $l_{1,n+1} = \frac{1}{4} l_{1,n}$ bzw. $l_{2,n+1} = \frac{2}{5} l_{2,n}$ aufgeteilt. Die Bedingung (7.2.19) verlangt für jedes n

$$\Gamma_n(q, \tau, l) \equiv (\mu_1^q l_1^{-\tau} + \mu_2^q l_2^{-\tau})^n = \mu_0 = 1, \tag{7.2.22}$$

wobei die Funktion $\tau(q)$ nach (7.2.16) von n unabhängig ist. Die Entwicklung von (7.2.22) nach dem Binomialsatz ergibt

²⁷ Argoul u.a. (1990) führen ein ähnliches Beispiel mit fünf Intervallen auf.

$$\Gamma_n(q, \tau, l) = \sum_{m=0}^n \binom{n}{m} \mu_1^{mq} \mu_2^{(n-m)q} (l_1^m l_2^{n-m})^{-\tau} = 1. \quad (7.2.23)$$

Für große n lässt sich der Ausdruck (7.2.23) als eine Summe über $\binom{n}{m}$ Intervalle der Länge $\Delta l = l_1^m l_2^{n-m}$ interpretieren, dessen Hölder-Exponent zwischen α und $\alpha + \Delta\alpha$ liegt. Der Träger dieser Intervalle ist ein Fraktal mit der fraktalen Dimension $f(\alpha)$. Jedes dieser Fraktale kann als eine Projektion des ursprünglichen Multifraktals betrachtet werden, bei der ein bestimmter Hölder-Exponent α für ein gegebenes q die Summe dominiert und die restlichen Summanden vernachlässigbar klein sind. Diese Überlegungen führen zur Annahme, dass die Summe (7.2.23) durch ihren größten Term approximiert werden kann, d.h.

$$\binom{n}{m} \mu_1^{mq} \mu_2^{(n-m)q} (l_1^m l_2^{n-m})^{-\tau} = 1. \quad (7.2.24)$$

Der *globale* Skalierungsexponent τ ergibt sich aus der Bedingung²⁸

$$\frac{d \ln \Gamma_n(q, \tau, l)}{dm} \stackrel{!}{=} 0 \quad (7.2.25)$$

und gilt für einen *durchschnittlichen* Summanden in (7.2.23). Die Formel (7.2.25) folgt aus der Überlegung, dass der relative Beitrag eines durchschnittlichen Gliedes zur Summe (7.2.23) konstant bleibt. Die Lösung zu (7.2.25) ist:

$$\tau(q) = \frac{\ln[(1-\xi)/\xi] + q \ln(\mu_1/\mu_2)}{\ln(l_1/l_2)}, \quad \xi \equiv m/n. \quad (7.2.26)$$

Die Werte des *lokalen* Hölder-Exponenten α und der Funktion $f(\alpha)$ werden durch die folgenden Gleichungen festgelegt

$$\mu_1^{mq} \mu_2^{(n-m)q} = (l_1^m l_2^{n-m})^{\alpha q}, \quad \binom{n}{m} = (l_1^m l_2^{n-m})^{-f(\alpha)} \quad (7.2.27)$$

und charakterisieren den relativen Beitrag des größten Terms zur Summe. Aus (7.2.24) und (7.2.27) folgt

$$(l_1^m l_2^{n-m})^{\alpha q - f(\alpha) - \tau} = 1. \quad (7.2.28)$$

Andererseits gilt wegen (7.2.23)

²⁸ Hasley u.a. (1986).

$$\sum_{m=0}^n (l_1^m l_2^{n-m})^{\alpha q - f(\alpha) - \tau} = 1. \quad (7.2.29)$$

Um zu sehen, dass (7.2.28) und (7.2.29) keineswegs im Widerspruch stehen, erläutern wir kurz die Vorgehensweise bei der Approximation durch das Laplace Integral.²⁹ Alternativ zur Intervalllänge ε kann die Integration angesichts der vorherigen Überlegungen über den Hölder-Exponenten α erfolgen. Zu berechnen ist der Ausdruck

$$Z_q(\varepsilon) \sim \int l^{\alpha q - f(\alpha)} \rho(\alpha) d\alpha. \quad (7.2.30)$$

Für *große* Werte von $(\alpha q - f(\alpha))$ wird $l^{\alpha q - f(\alpha)}$ *klein*³⁰ sein und das Verhalten des Integrals wird im wesentlichen durch die Funktion $\rho(\alpha)$ bestimmt, die hier die Rolle einer Dichtefunktion übernimmt. Für *kleine* Werte von $(\alpha q - f(\alpha))$ wird dagegen der Beitrag von $l^{\alpha q - f(\alpha)}$ im Vergleich zu dem von $\rho(\alpha)$ sehr groß werden und das Integral dominieren. In diesem Fall erfolgt eine Approximation durch den größten Term. Die Anwendung dieses Verfahrens auf (7.2.28) und (7.2.29) wird durch eine zusätzliche Bedingung ermöglicht. Sie lautet

$$\tau(q) = \inf_{\alpha} (\alpha q - f(\alpha)). \quad (7.2.31)$$

Je kleiner die Differenz $(\alpha q - f(\alpha))$, desto besser ist die lokale Approximation. Für große Werte von $(\alpha q - f(\alpha))$ ist dagegen der Beitrag des globalen Exponenten $\tau(q)$ relativ hoch. Die Gleichung (7.2.31) besagt, dass die Funktionen $\tau(q)$ und $f(\alpha)$ durch die Legendre Transformation miteinander verbunden sind.³¹

Als Hilfsmittel zur Bestimmung von $f(\alpha)$ aus (7.2.27) kann eine Abschätzung durch die Stirling-Formel³² verwendet werden:

$$\binom{n}{m} = \left(\frac{m}{n}\right)^{-m} \left(1 - \frac{m}{n}\right)^{m-n} + O(\ln n). \quad (7.2.32)$$

Mit der Festlegung $\xi \equiv m/n$ folgt

$$\frac{1}{n} \ln \binom{n}{m} \approx -\xi \ln \xi - (1 - \xi) \ln (1 - \xi). \quad (7.2.33)$$

²⁹ Siehe auch Anhang A.7.

³⁰ Da $l \ll 1$ ist. Zur Orientierung geben wir zwei Werte an: $0,1^{0,1} = 0,7943$; $0,1^{0,9} = 0,1259$.

³¹ Dies kann auch rigoros gezeigt werden, siehe z.B. Jaffard (1999).

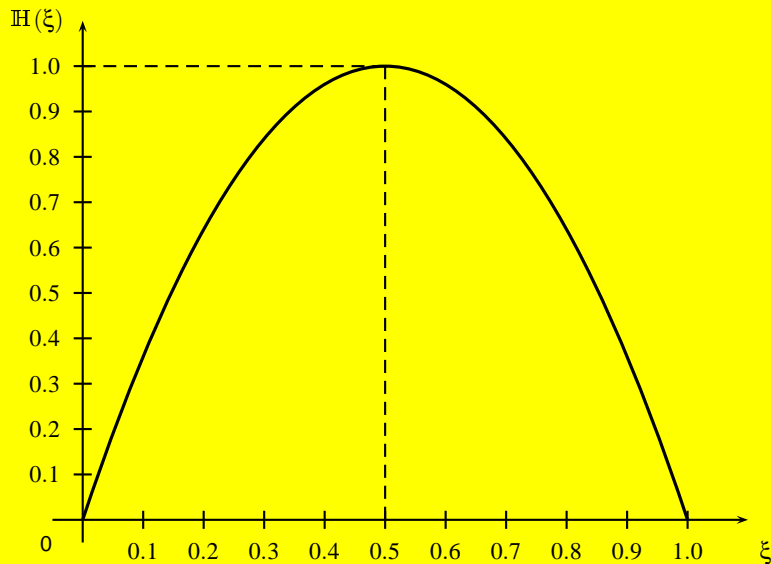
³² Siehe Anhang A.7. Wir ignorieren hier den unwesentlichen Faktor $\sqrt{2\pi n}$.

BEISPIEL 7.3

In einem Bernoulli-Fall mit nur zwei Zuständen und Wahrscheinlichkeiten ξ und $1 - \xi$ ist die (Shannon-) Entropie

$$\mathbb{H}(\xi) = -\xi \log_2 \xi - (1 - \xi) \log_2 (1 - \xi).$$

Als Funktion von ξ stellt die Entropie eine nach unten gerichtete Parabel dar. Die minimalen Werte sind $\mathbb{H}(0) = \mathbb{H}(1) = 0$, das Maximum $\mathbb{H} = 1$ wird bei $\xi = 1/2$ erreicht.



Diese Formel ähnelt einem Ausdruck für die Entropie einer Binomialverteilung (vgl. Beispiel 7.3), wobei ξ in (7.2.33) natürlich keine Zufallsvariable ist. Aus (7.2.27) und (7.2.33) resultiert

$$\alpha(\xi) = \frac{\xi \ln \mu_1 + (1 - \xi) \ln \mu_2}{\xi \ln l_1 + (1 - \xi) \ln l_2}, \quad (7.2.34)$$

$$f(\alpha(\xi)) = \frac{\xi \ln \xi + (1 - \xi) \ln (1 - \xi)}{\xi \ln l_1 + (1 - \xi) \ln l_2}. \quad (7.2.35)$$

Wegen einer nichtlinearen Abhängigkeit von ξ in den beiden Gleichungen (7.2.34) und (7.2.35) lassen sich nicht alle Kennzahlen des resultierenden multifraktalen Spektrums (Abb. 7.10) exakt berechnen. Beachtet man dennoch, dass ξ ausschließlich Werte zwischen Null und

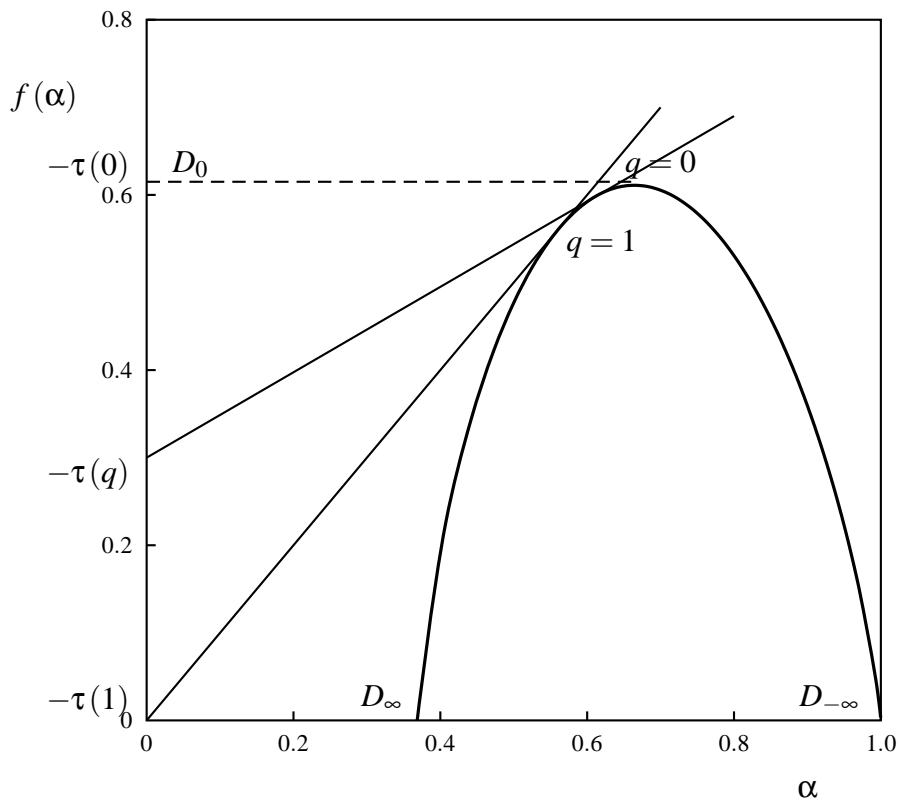


Abb. 7.10. Ein multifraktales Spektrum zur Abb. 7.9.

Eins annehmen kann, ist $\alpha(\xi)$ in (7.2.34) eine monoton fallende Funktion von ξ auf dem Intervall $[0, 1]$ mit den Randwerten

$$\alpha_{\min} = \alpha(1) = \frac{\ln \mu_1}{\ln l_1} = 0,3685$$

$$\alpha_{\max} = \alpha(0) = \frac{\ln \mu_2}{\ln l_2} = 1.$$

Die Funktion $f(\alpha(\xi))$ ist eine nach unten gerichtete Parabel. Ihr Maximum entspricht der Dimension D_0 und wird in Argoul u.a. (1990) aus der Gleichung (7.2.22) für $q = 0$ und $n = 1$ numerisch bestimmt.³³ Aus

³³ Es stellte sich heraus, dass diese Vorgehensweise nicht ganz unproblematisch ist. Obwohl diese Aussage für $n = 1$ ohne weiteres gilt, muss für $n > 1$ beachtet werden,

der Ableitung

$$\frac{\partial f(\alpha(\xi))}{\partial \alpha(\xi)} = \frac{\partial f(\alpha(\xi))/\partial \xi}{\partial \alpha(\xi)/\partial \xi} = \frac{\ln \xi \ln l_2 - \ln(1-\xi) \ln l_1}{\ln \mu_1 \ln l_2 - \ln \mu_2 \ln l_1} \quad (7.2.36)$$

ist ersichtlich, dass für $\partial f(\alpha)/\partial \alpha = 1$ der entsprechende Wert von ξ durch $\xi = \mu_1$ gegeben ist. Weiterhin gilt in diesem Punkt wegen (7.2.34) und (7.2.35) $f(\alpha) = \alpha$, die Gleichung (7.2.26) liefert $\tau(q) = 0$ und aus der Gleichung (7.2.31) folgt schließlich $q = 1$. Die Bedingung für das Maximum des multifraktalen Spektrums

$$\ln \xi \ln l_2 = \ln(1-\xi) \ln l_1$$

ergibt sich aus (7.2.36) und kann mit dem Ansatz

$$\xi^* = \frac{l_1^{D_0}}{l_1^{D_0} + l_2^{D_0}}, \quad l_1^{D_0} + l_2^{D_0} \equiv 1$$

für $D_0 \approx 0,61098$ aufgelöst werden. Die optimalen Werte sind:

$$\begin{aligned} \xi^* &= l_1^{D_0} = 0,4287 \\ f(\alpha(\xi^*)) &= -\tau(0) |_{\xi=\xi^*} = D_0 \\ \alpha(\xi^*) &= \frac{l_1^{D_0} \ln \mu_1 + l_2^{D_0} \ln \mu_2}{l_1^{D_0} \ln l_1 + l_2^{D_0} \ln l_2} = 0,66423. \end{aligned}$$

Nach der Gleichung (7.2.31) stellt die Funktion $-\tau(q)$ eine Menge der Schnittpunkte der zum multifraktalen Spektrum angelegten Tangenten mit der Ordinaten-Achse dar und es gilt $q = \partial f(\alpha)/\partial \alpha$. Auf diese Art ergibt sich eine alternative Beschreibung des Spektrums durch die Geraden $f(\alpha) = -\tau(q) + q\alpha$ mit der Steigung q . Die Menge der Schnittpunkte dieser Geraden miteinander bildet das multifraktale Spektrum. Durch diese Beobachtung bestätigt sich die Feststellung, dass die Funktionen $f(\alpha)$ und $\tau(q)$ durch die Legendre Transformation verbunden sind:

$$f(\alpha) = \inf_q [q\alpha - \tau(q)]. \quad (7.2.37)$$

dass die Funktion $\tau(q)$ nach (7.2.26) auch von $\ln \xi = \ln(m/n)$ abhängig ist. Solange kein funktionaler Zusammenhang zwischen ξ und D_0 spezifiziert ist, lässt sich das Maximum des multifraktalen Spektrums nicht eindeutig bestimmen. Weder in Hasley u.a. (1986) noch in Argoul u.a. (1990) ist ein Hinweis darauf zu finden. Nach unserer Version besteht zwischen ξ und der Intervalllänge l eine Abhängigkeit nach dem Potenzgesetz, und zwar mit dem Exponenten D_0 . Addieren sich l_1 und l_2 zu Eins auf, ist diese Abhängigkeit linear und trivial (vgl. die folgende Lösung).

Ein multifraktales Zufallsmaß

Unfortunately, those who expect the world to be simple, hence manageable by simple mathematics, will be disappointed.

– B. Mandelbrot (1990)

Die nächste qualitative Erweiterung bei der Modellierung der Multifraktale besteht im Austausch der deterministischen Multiplikatoren m_i bzw. der Maßanteile μ_i gegen Zufallsvariablen M_i .³⁴ Was ändert sich dadurch? Erstens wird das multifraktales Maß zu einem Zufallsmaß. Zweitens bedient man sich nun stochastischer Konvergenz, die in der Kombinatorik der deterministischen Kaskade nicht zugänglich war. Eine weitere Unterstützung bei der Beschreibung von Eigenschaften eines multifraktalen Zufallsmaßes erhält man von der Theorie der großen Abweichungen, u.a. ist der Satz von Sanov (5.1.3) von Bedeutung.

Ein stochastischer Multiplikator \mathcal{M} repräsentiert eine Menge $\mathcal{M} = (M_0, M_1, \dots)$ von Zufallsvariablen auf $[0, 1]$, die unabhängig, aber nicht unbedingt identisch verteilt sind. Ähnlich wie bei der deterministischen Kaskade wird das Intervall $[0, 1]$ sukzessiv in zwei Teilen gleicher Länge aufgeteilt, wobei jedem Teilintervall auf der Teilung $J([0, 1/2), [1/2, 1])$ ein Maß $M \in \mathcal{M}$ zugeordnet wird. Auf der Stufe n des rekursiven Verfahrens kann das Maß $\psi_n(t)$ eines Intervalls

$$B_{2^{-n}}(\eta) = J(\eta_1, \dots, \eta_n) = \left[\sum_{i=1}^n \eta_i 2^{-i}, \sum_{i=1}^n \eta_i 2^{-i} + 2^{-n} \right), \quad \eta_i \in \{0, 1\}$$

der Länge $\Delta t = 2^{-n}$ als Produkt

$$\psi_n(t) = M_1(\eta) M_2(\eta) \cdots M_n(\eta), \quad t \in J(\eta_1, \dots, \eta_n), \quad (7.2.38)$$

$$\eta = (\eta_1, \dots, \eta_n)$$

der Zufallsmultiplikatoren entlang des Pfades η dargestellt werden. Das Maß auf jeder Stufe wird nur im Mittel³⁵ erhalten:

$$\mathbb{E}[M_{i,0}(\eta) + M_{i,1}(\eta)] = 1. \quad (7.2.39)$$

³⁴ Everetsz und Mandelbrot (1992).

³⁵ Eine andere Variante sieht eine exakte Erhaltung des Maßes vor: $M_i(\eta) = M_{i,0}(\eta) + M_{i,1}(\eta)$.

Eine Folge von *Zufallsmaßen*, die durch eine Dichtefunktion (Radon-Nikodym Ableitung) $\mu_n(dt) = \psi_n(t) dt$, $n \geq 1$ definiert ist, konvergiert (schwach) zu einem Grenzwert μ_∞ , weil für jede beschränkte messbare stetige Funktion q auf J die Folge $\{\int_J q d\mu_n\}$ ein L_1 -beschränktes Martingal bezüglich der Folge von σ -Algebren $\mathcal{F}_n := \sigma\{M(\eta_1, \dots, \eta_n) : \eta_i \in \{0, 1\}\}$ ist.³⁶

Die Problematik der Nichttrivialität des Maßes μ_∞ , der Existenz von Momenten sowie der Berechnung der fraktalen Dimension des Trägers ist in Mandelbrot (1974) und Kahane und Peyrière (1976) abgehandelt. Der folgende Satz bietet weitere Erkenntnisse über die Eigenschaften des multifraktalen Zufallsmaßes, das durch eine stochastische Binomialkaskade erzeugt wird. Diese Eigenschaften werden mittels einer „modifizierten“ kumulanterzeugenden Funktion

$$\chi_c(h) = \log_c \mathbb{E} M^h + 1$$

beschrieben.³⁷ Für eine Binomialkaskade gilt $c = 2$.

Satz 7.2.1 (Kahane und Peyrière (1976)) *Sei M ein Generator einer stochastischen Binomialkaskade μ_∞ und $\chi_c(h) = \log_c \mathbb{E} M^h + 1$. Es gilt:*

i. (Nichttrivialität)

$\mathbb{E}[\mu_\infty([0, 1])] > 0$ genau dann, wenn $-D = \chi'_c(1^-) = c \mathbb{E}[M \log_c M] < 0$.

ii. (Existenz der Momente)

Sei $h > 1$. Dann hat $Y = \mu_\infty([0, 1])$ ein endliches Moment der Ordnung h genau dann, wenn $\mathbb{E} M^h < 1/c$. Weiterhin gilt $0 < \mathbb{E} M^h < \infty$ für alle $h > 0$ genau dann, wenn M durch c essentiell beschränkt ist und $\mathbb{P}(M = c) < 1/c$.

iii. (Träger)

Nehmen wir an, dass $\mathbb{E}[M \log_c M] < \infty$. Der Träger von μ_∞ ist eine Borel-Menge der Dimension

$$D = -c \mathbb{E}[M \log_c M] = -\chi'_c(1^-).$$

Calvet und Fisher (2002) verwenden an Stelle von $-\chi_c(h)$ eine Funktion $\tau(q) = -\log_c \mathbb{E} M^h - 1$. Die Unabhängigkeit der Multiplikatoren auf

³⁶ Für eine detaillierte Darstellung vgl. Holley und Waymire (1992), S. 823.

³⁷ Vgl. auch Gilbert u.a. (1999).

unterschiedlichen Stufen der Kaskade impliziert $E[\mu(\Delta t)^q] = [E(M^q)]^k$ mit $\Delta t = 2^{-k}$ oder

$$E[\mu(\Delta t)^q] = (\Delta t)^{\tau(q)+1}. \quad (7.2.40)$$

Im nächsten Schritt wird das Konzept der Multifraktale von einem Zufallsmaß auf einen stochastischen Prozess erweitert. Auf dieser Basis hat B. Mandelbrot ein Modell vorgeschlagen, das „wilde Schwankungen“ in den Aktienrenditen nachbilden soll.

7.3 Das Multifraktale Modell of Asset Returns

Das Multifraktale Modell of Asset Returns ist derzeit Mandelbrot's letztes Wort in seiner mehr als vierzigjährigen Forschung empirischer Eigenschaften der Wertpapierrenditen und geht auf eine Serie von Working Papers³⁸ aus dem Jahr 1997 zurück. Im Jahr 2000 folgten weitere Veröffentlichungen in *Quantitative Finance*³⁹, zwei Jahre später erschien ein Artikel von Calvet und Fisher in *The Review of Economics and Statistics*.⁴⁰ „A Multifractal Walk Down Wall Street“⁴¹ in *Scientific American* sowie ein jüngstes Buch „*The (mis)Behavior of Markets: A Fractal View of Risk, Ruin and Reward*“⁴² das aus einer gemeinsamen Arbeit mit Richard Hudson, Wissenschaftsredakteur beim „Wall Street Journal,“ hervorgegangen ist, geben eine populäre Darstellung von Multifraktalen und verwandter Themen. Unsere Darstellung ist in Anlehnung an Calvet und Fisher (2002).

Definition 7.3.1. *Es seien \mathcal{T} und Q Intervalle auf \mathbb{R} derart, dass $0 \in \mathcal{T} = [0, T]$ und $[0, 1] \subseteq Q$. Ein stochastischer Prozess $\{X(t), t \in \mathcal{T}\}$ mit stationären Zuwächsen heißt **multifraktaler Prozess**, falls die Momente von $X(t)$ folgender Relation genügen:*

$$E|X(t)|^q = c(q) t^{\tau(q)+1} \quad \text{für alle } t \in \mathcal{T}, q \in Q, \quad (7.3.1)$$

wobei $c(q)$ und $\tau(q)$ reellwertige Funktion auf Q sind.

³⁸ Calvet u.a. (1997), Fisher u.a. (1997) und Mandelbrot u.a. (1997).

³⁹ Mandelbrot (2001a, 2001b, 2001c, 2001d).

⁴⁰ Calvet und Fisher (2002).

⁴¹ Mandelbrot (1999a).

⁴² Mandelbrot und Hudson (2004).

Die Funktion $\tau(q)$ heißt *Skalierungsfunktion* des multifraktalen Prozesses und ist so konstruiert, dass sie den Einfluss der Zeit auf die Momente q berücksichtigt. Am Beispiel der gebrochenen Brownschen Bewegung (fBm) kann am einfachsten gezeigt werden, dass der multifraktale Prozess ein selbstähnlicher Prozess ist. Für die fBm gilt

$$X(t) \stackrel{d}{=} t^H X(1)$$

und folglich

$$\mathbb{E}|X(t)|^q \stackrel{d}{=} t^{Hq} \mathbb{E}|X(1)|^q = c(q) t^{\tau(q)+1} \quad (7.3.2)$$

mit $c(q) = \mathbb{E}|X(1)|^q$ und $\tau(q) = Hq - 1$. Im Allgemeinen ist die Skalierungsfunktion nichtlinear und ihre *konkave* Form ist ein Merkmal von Multifraktalität.

Das Multifraktale Modell of Asset Returns beschreibt einen Renditeprozess

$$X(t) = \ln S(t) - \ln S(0) \quad (7.3.3)$$

eines Wertpapiers (einer Aktie, eines Indexes, einer Währung) als eine gebrochene Brownsche Bewegung in multifraktaler Zeit.

Definition 7.3.2. *Der Renditeprozess $X(t)$ ist ein untergeordneter (subordinated) Prozess*

$$X(t) \equiv B^H[\theta(t)], \quad (7.3.4)$$

wobei $B^H(t)$ eine gebrochene Brownsche Bewegung und $\theta(t)$ eine stochastische Marktzeit ist. Weiterhin ist $\theta(t)$ ein multifraktaler Prozess mit stetigen nichtfallenden Pfaden und stationären, von $B^H(t)$ unabhängigen Zuwächsen.

Diese Konstruktion ermöglicht die gleichzeitige Modellierung von *fat tails* durch $\theta(t)$ und *long memory* durch $B^H(t)$ und ist somit universal.

Die Skalierungsfunktion $\tau(q)$ im MMAR genügt folgender Relation:

$$\tau_X(q) = \tau_\theta(Hq). \quad (7.3.5)$$

Durch die Gleichung (7.3.5) wird ein Zusammenhang zwischen beiden Skalierungsfunktionen – vom Renditeprozess und vom Marktzeitprozess – hergestellt. Dies impliziert, dass aus der geschätzten Skalierungsfunktion des Renditeprozesses auf die Eigenschaften der Marktzeit $\theta(t)$ geschlossen werden kann.

Calvet und Fisher (2002) betonen, dass die Langzeitabhängigkeiten nicht notwendig durch die gebrochene Brownsche Bewegung modelliert werden müssen. Für einen stochastischen Prozess Z mit stationären Zuwächsen können die Abhängigkeiten in Potenzen von absoluten Renditen zu Langzeitabhängigkeiten führen, falls die Kovarianz in den Levels

$$\delta_Z(t, q) = \text{Cov}(|Z(a, \Delta t)|^q, |Z(a+t, \Delta t)|^q) \quad (7.3.6)$$

bei $t/\Delta t \rightarrow \infty$ ein hyperbolisches Verhalten in t aufweist.⁴³ $\delta_Z(t, q)$ in (7.3.6) ist für $E|Z(a, \Delta t)|^{2q} < \infty$ wohldefiniert.

Zur Schätzung der Skalierungsfunktion aus empirischen Daten wird gewöhnlich die Zustandsumme

$$S_q(T, \Delta t) \equiv \sum_{i=0}^{N-1} |X(i\Delta t + \Delta t) - X(i\Delta t)|^q \quad (7.3.7)$$

verwendet, wobei das Intervall $[0, T]$ in $N \in \mathbb{N}$ Intervalle der Länge Δt aufgeteilt wird. Ist $X(t)$ ein multifraktaler Prozess, dann sind die Summanden identisch verteilt und aus der Gleichung (7.3.1) folgt

$$E[S_q(T, \Delta t)] = Nc(q) t^{\tau_X(q)+1},$$

falls das q -te Moment existiert. Dies führt zu einer empirisch testbaren Relation

$$\ln E[S_q(T, \Delta t)] = \tau_X(q) \ln(\Delta t) + \ln c(q) + \ln T. \quad (7.3.8)$$

Für ein Multifraktal erwartet man eine lineare Beziehung zwischen $\ln E[S_q(T, \Delta t)]$ und $\ln(\Delta t)$ für unterschiedliche Werte von q , falls das q -te Moment existiert. Oft wird diese Analyse auf einer doppellogarithmischen Skala graphisch durchgeführt.

Schließlich kann der Schätzer der Skalierungsfunktion $\tau_X(q)$ zur Berechnung des multifraktalen Spektrums $\hat{f}(\alpha) = \min_q [q\alpha - \tau_X(q)]$ mittels Legendre Transformation verwendet werden.

Derzeit existieren nur wenige empirische Untersuchungen zur Multifraktalität in den Aktienkursen, um eine repräsentative Aussage bezüglich dieses Konzeptes treffen zu können. Das Modell selbst ist relativ jung und hat noch keine breite Akzeptanz gefunden. Calvet und

⁴³ Vgl. auch Matteo u.a. (2005).

Fisher (2002) finden Evidenz für ein multifraktales Verhalten von Momenten in DM/USD-Wechselkursen. Die verwendeten Daten von Olsen and Associates beinhalten die Handelsfrequenzen von ca. zwei Stunden bis zu 180 Tagen über den Gesamtzeitraum von 23 Jahren. Fillol (2003) berichtet über Anzeichen von Multifraktalität beim französischen Aktienindex CAC40, Sun u.a. (2001) kommen zu ähnlichen Ergebnissen beim Hang Seng Index.

Zusammenfassung

Die multifraktale Analyse ist ein mathematisches Framework zur Beschreibung und Simulation komplexer Systeme mit lokalen Singularitäten und nichttrivialen Interaktionen zwischen unterschiedlichen Skalen. Immer häufiger wird der Kapitalmarkt in der Literatur als ein solches System bezeichnet. Die Erwartungen der Marktteilnehmer können sich sehr schnell oder sehr langsam ändern, weil die langfristigen preisrelevanten Informationen fast immer unvollständig sind. Dies führt zu überhöhten Volatilitäten der Renditen bzw. zu andauernden Abweichungen der Aktienkurse vom Fundamentalwert.

Ausgehend von diesen empirisch beobachteten Eigenschaften ist die Anwendung der multifraktalen Analyse zur Modellierung und Simulation der Aktienkursbewegungen gerechtfertigt und sinnvoll. Da die Momente unterschiedlicher Ordnungen eines multifraktalen stochastischen Prozesses auf eine nicht triviale Weise skalieren, liegt die Stärke dieser Modellklasse an der Fähigkeit, die Nicht-Gaussischen Prozesse nachzubilden zu können.

1

2

3

4

8. Analyse der Ergebnisse

“Economists adjust their models until they both fit the data and give “reasonable” results.”

– Leeper u.a. (1996), S. 5.

8.1 Skalierung: Modell oder Hypothese?

Es ist üblich, die Wissenschaft durch die zwei wichtigsten Bestandteile zu charakterisieren: a) empirische Beobachtung und b) Theorie. Diese zwei Komponenten werden von Stevens (1951) in seiner Definition von Wissenschaft aufgeführt: *“Science seeks to generate confirmable propositions by fitting a formal system of symbols (language, mathematics, logic) to empirical observation.”* (S. 22). Damit die Beobachtungen nicht zu einer Unmenge von “stylized facts” degenerieren, werden Theorien formuliert. Die Theorien übernehmen zwei wichtige Funktionen: a) sie ordnen empirische Beobachtungen und b) sie geben Richtlinien für die zukünftige empirische Forschung vor. Eine weitere Funktion der wissenschaftlichen Theorie ist nach Stevens die Erzeugung von bestätigungsfähigen Aussagen.¹ Sind die theoretischen Aussagen mit den empirischen Beobachtungen konsistent, gewinnt die Theorie an Erklärungskraft, andernfalls. . . Was passiert andernfalls?

Man könnte erwarten, dass durch Inkonsistenzen zwischen Theorie und Empirie erstere geschwächt wird. Erzeugt die Empirie zu viele Gegebenheiten, die im Widerspruch zu theoretischen Aussagen stehen, muss die Theorie entweder überarbeitet oder aufgegeben werden. Diese klassische Vorgehensweise nach Popper (vgl. Beispiel 8.1)

¹ Orig. *confirmable propositions*.

BEISPIEL 8.1

Nach Karl Popper brauchen Wissenschaftler keine Induktion, nur Falsifizierung. Die Idee:

- I. Eine Hypothese H besagt, dass unter Annahme A ein Effekt E beobachtet wird;
- II. A ist wahr;
- III. E ist nicht beobachtet worden;
- IV. Also ist H falsch.

Ist E nicht beobachtet worden, dann ist H falsifiziert. Sonst ist H bekräftigt (corroborated).

Popper's *scientific methodology*:

- I. Stelle falsifizierbare Hypothesen auf;
- II. Versuche durch mehrere Testversuche Hypothesen zu falsifizieren;
- III. Behalte die Hypothese bei solange sie nicht falsifiziert ist;
- IV. Verwerfe die Hypothese, wenn sie falsifiziert ist.

Popper (1949): „Durch die Falsifikation unserer Annahmen bekommen wir tatsächlich Kontakt mit der „Wirklichkeit“. Die Widerlegung unserer Irrtümer ist die positive Erfahrung, die wir aus der Wirklichkeit gewinnen.“

scheint in Bezug auf die empirische Kapitalmarktforschung und die These vom informationseffizienten Kapitalmarkt (EMH) keine Anwendung zu finden. Die EMH besagt kurzgefasst, dass alle preisrelevanten Informationen von den Marktteilnehmern laufend berücksichtigt werden, und zwar aus dem Gesichtspunkt eines rationalen Investors.

Eine Vielzahl von Ausnahmen von dieser Regel wurde empirisch festgestellt mit einem Ergebnis: die Ausnahmen wurden als Anomalien eingestuft und für nichtig erklärt, weil sie entweder von geringer Bedeutung sind oder nicht systematisch vorkommen. Ein weiteres Argument ist: „... *the alternative hypothesis is ... unacceptable.*“ (Fama (1998), S. 284). Mit anderen Worten, die EMH ist nach Fama nicht falsifizierbar. Durch das Auffinden von immer neuen Ausnahmen können sie in ihrer Gesamtheit noch nicht zu einer neuen Regel führen, die eine bessere ökonomische Interpretation des Marktgeschehens als die EMH geben könnte. Eine gängige Erklärung für dieses Phänomen verweist auf eine *zusammengesetzte Hypothese*.

Was ist eine zusammengesetzte Hypothese? Nehmen wir einen logischen Ausdruck A und B . Ist der Ausdruck *wahr*, folgt daraus notwendig, dass A *wahr ist* und B *wahr ist*. Es ist eindeutig. Zum Ergebnis *falsch* führen dagegen drei mögliche Kombinationen aus A und B , die nach außen *nicht unterscheidbar sind*: a) A *ist falsch*, B *ist irrelevant*,

b) *B ist falsch, A ist irrelevant* und c) *A ist falsch und B ist falsch*. Da die EMH nur im Zusammenhang mit einem Asset Pricing Model überprüft werden kann, lässt sich die Aussage, dass falsche Ergebnisse nicht durch die Hypothese, sondern durch das dazugehörige Modell verursacht sind, kaum widerlegen. Dabei sollte zwischen *Modellen* und *Hypothesen* unterschieden werden.

Modelle sind vereinfachte Abbildungen der Realität oder Projektionen der Theorie auf eine ideale Welt. Man verwendet Modelle, um die Theorie (oder ein Teil davon) zu veranschaulichen. *“The whole point in building an analytical model is to construct a representation that is simpler than the real thing.”* [Gordon 1991]. Modelle machen keine Behauptungen und basieren auf Annahmen. Dies bringt zwei wichtige Implikationen mit sich. Erstens, Modelle können ausschließlich durch das Vorhandensein interner Widersprüche widerlegt werden. Wissenschaftler können sich darauf beschränken, *konsistente* Modelle zu bilden. *“A valid argument is one that does not contain obvious errors of logic. By analogy, a model that does not contain known or detectable flaws and is internally consistent can be said to be valid.”* [Oreskes u.a. 1994]. Die zweite Folgerung ergibt sich aus der ersten und besagt, dass Modelle nicht getestet werden können. Das Testen der Modelle ist einfach überflüssig, da der *induktive Schritt* schon bei der Formulierung der Annahmen erfolgt. Die Schlussfolgerungen sind lediglich Verfeinerungen dieser Annahmen und werden *deduktiv* hergeleitet. *“If the conclusion of an inference cannot be false given that all the premises are true, then the inference is logical/deductive. The conclusion does not add any new empirical content to the premises.”* [Suppe 1998]. Die Annahmen können zutreffend sein oder nicht, führen aber höchstens dazu, dass ein Modell als „korrekt“ oder „inkorrekt“ eingestuft wird, anstatt Kategorien „richtig“ oder „falsch“ zu verwenden.² Modelle werden nicht getestet, sondern *kalibriert*: man verändert die unabhängigen Variablen solange, bis ein voraus festgelegtes Gütekriterium an die abhängigen Variablen seinen maximalen Wert erreicht.

Eine deduktive Natur der Modelle schmälert keinesfalls ihre Bedeutung. Je nach Modell oder Darstellung können nicht nur unterschiedliche Aspekte einer Theorie besser betont werden, sondern es ändert sich möglicherweise unser Verständnis der Wirklichkeit. Morgan (2001) argumentiert, dass Modelle zum Verständnis der Wirklichkeit

² Vgl. Dohmen (2002), S. 193.

beitragen, indem sie eine *story-telling*-Funktion übernehmen (vgl. Beispiel 3.1, S. 36). Eine Story ist keineswegs eindeutig. Es ist selbst ein theoretisches Konzept, das mehr oder weniger plausibel ist. *“Not at least because the purpose of building a model determines its form, there cannot exist a representative model in economics so that different categories of models might accentuate different aspects raised in theoretical model accounts, rendering models mainly deductive, mainly metaphoric, mainly instrumentalist etc.”* [Dohmen 2002], S. 208.

Was erwartet man von einer Hypothese? *“If an inference makes the conclusion plausible or probable (but not necessarily true) given that all premises are true, this is an inductive inference. The inference is ampliative in the sense that the conclusion adds some empirical content which was not contained in the premises.”* [Suppe 1998]. Dies ist ein sehr interessanter Punkt. Eine Hypothese braucht realistische Annahmen. Sie kann und muss getestet werden. Aber: beim Testen erwartet man von einer Hypothese in erster Linie *nicht* die induktive Inferenz, sondern die Möglichkeit, die Hypothese selbst *falsifizieren zu können*. Diese Möglichkeit scheint sich, wenn überhaupt, aus dem oben zitierten *empirical content* zu ergeben. Interne Widersprüche reichen nicht mehr aus, um eine Hypothese zu widerlegen. Ausnahmen in Form sogenannter *Anomalien* sind zugelassen. Induktive Inferenz entsteht lediglich als „Beiprodukt“ aus dem Teil der Hypothese, der noch nicht falsifiziert ist. Von *Bestätigung* einer Hypothese kann eigentlich überhaupt keine Rede sein: *“Under most conditions, an hypothesis cannot be proved true or valid. It remains on the table so long as it cannot be rejected.”* [Frankfurter und McGoun 2000]. Bei der Formulierung der Hypothese müssen Aussagen über die möglichen empirischen Erkenntnisse getroffen werden, die zu ihrer Falsifizierung führen würden.

Welche Stellung nehmen dann die Skalierungs- oder die Selbstähnlichkeitseigenschaften der Aktienkursbewegungen in der obigen Klassifikation ein? In erster Linie stellen sie zweifellos eine Story dar: *all charts look the same*. Ausgehend davon hat B. Mandelbrot Modelle entwickelt, die eine quantitative Darstellung dieser Beobachtung im Lichte der fraktalen Geometrie bzw. multifraktaler Maße ermöglichen. Man nimmt die empirisch beobachteten extremen Kursschwankungen und die Langzeitabhängigkeiten als Ausgangspunkt für die Modellentwicklung und zeigt, wie gut sich die zu modellierenden *stylized facts* in den Ergebnissen niederschlagen. Die Annahmen über die Selbstähnlichkeit sind verfeinert, die Kennzahlen entwickelt, die statistische

Anpassung ist im Vergleich zu den ähnlichen Modellklassen (GARCH, FIGARCH) recht gut. Die Schlussfolgerung: „Die tatsächlichen Risiken – sowohl nach oben als auch nach unten – sind wesentlich größer, als wir annehmen.“³

Diese Aussage ist von deskriptiver Natur und führt lediglich auf die Annahmen zurück, die den Modellen zu Grunde gelegt wurden. Welche Handlungen soll ein rationaler Investor vornehmen, wenn er mit diesen Erkenntnissen konfrontiert wird? Was sind die “reasonable results”? Mandelbrot gibt keine Antwort auf diese Frage. Er bleibt bei Modellen. Die Aufstellung von Hypothesen überlässt er seinen „ambitionierteren Kollegen.“

8.2 Die Skalenrisiken

In der vorliegenden Arbeit gehen wir auf die Beschreibung von typischen und untypischen Ereignissen aus dem Gesichtspunkt der Informationsverarbeitung auf dem Aktienmarkt ausführlich ein. Da die großen Risiken gewöhnlich durch die untypischen Ereignisse verursacht werden, ist diese Unterscheidung essentiell.

Die extremen Kursschwankungen, die in den letzten Jahren immer häufiger beobachtet werden und zu verheerenden Folgen führen können bzw. in der Vergangenheit auch geführt haben, sind zwar für das Risiko- und Portfoliomanagement von großer Bedeutung, werden aber unter der EMH als Anomalien eingestuft und statistisch als Ausreißer behandelt. Dies impliziert, dass diese als Informationsquelle für die Investoren nutzlos sind: man kann daraus nichts lernen und soll sie bei den Investmententscheidungen vernachlässigen. Wie ist das zu begründen?

Da die Aktienbewertung letztendlich auf die Ermittlung des Fundamentalwertes zurückgeführt wird, tritt ein Argument in den Vordergrund, dass sich der Fundamentalwert nicht so schnell ändern könnte, um die sprunghaften Veränderungen des Aktienkurses hervorzurufen. Die Sprünge können entweder durch eine ungleichmäßige Ankunft der Informationen auf den Markt erfolgen (Offenlegung der Quartalsergebnisse, Ad hoc Mitteilungen etc.) oder durch eine Überreaktion der Marktteilnehmer. Beides sind Ergebnisse der Anpassung von

³ Mandelbrot (2005).

Erwartungen über den Fundamentalwert und nicht der Veränderung des Fundamentalwertes selbst. Im ersten Fall werden die Erwartungen aus der Vergangenheit angepasst. Die großen Abweichungen, die dabei ausgeglichen werden, deuten unter der EMH *nicht* auf eine ineffiziente Informationsverarbeitung in der Vergangenheit hin, sondern implizieren, dass der Aktienkurs bisher einen sehr untypischen Pfad genommen hat. Im zweiten Fall werden die Erwartungen über die zukünftige Entwicklung gebildet. Die Überreaktionen der Marktteilnehmer gleichen sich aber im Mittel mit den Unterreaktionen aus und sind damit unproblematisch. *“If anomalies split randomly between underreaction and overreaction, they are consistent with market efficiency.”*⁴ Sie gehören damit ebenfalls zu Anomalien, also zu den untypischen Ereignissen, obwohl sie immer wieder zu beobachten sind.

Sollte die von der EMH postulierte Kontinuität des Informationsflusses mit den empirisch beobachteten Renditesprüngen und Langzeitabhängigkeiten in Übereinstimmung gebracht werden, dann bietet sich eine *Skalinghypothese* an. Sie besagt, dass *die preisrelevanten Informationen auf mehreren Zeitskalen ankommen, aber überwiegend auf wenigen feinsten Skalen verarbeitet werden.*

Die Überlagerung von Informationen aus unterschiedlichen Skalen kann sowohl zu einer gegenseitigen Verstärkung als auch zu einer Reduzierung von entsprechenden Auswirkungen führen. Können die langfristigen Veränderungen des Fundamentalwertes auf einer grobkörnigen Skala aus der Sicht von mehreren feineren Skalen antizipiert werden, kommt es zu Langzeitabhängigkeiten. Stellt sich heraus, dass diese Antizipation ungerechtfertigt war, erfolgt die Anpassung auf allen feineren Skalen: die Blase platzt. Im Falle eines Katastrophenereignisses oder durch politische Risiken kann es ebenfalls zu einer zwangsläufigen Überlagerung von Informationen aus unterschiedlichen Skalen kommen. Daraus folgt, dass sowohl die typischen als auch die untypischen Ereignisse nur im Zusammenhang mit der entsprechenden Zeitskala ausgewertet werden können.

Dies ist die Grundidee der vorliegenden Arbeit und ein Versuch, einen induktiven Schritt in Richtung ökonomischer Erklärung von empirisch beobachteten stylized facts zu machen. Die Abgrenzung zu den bereits existierenden Hypothesen in der Literatur wie Peters (1994) und Dacorogna u.a. (2001) besteht darin, dass sich die Aufteilung der

⁴ Fama (1998), S. 284.

Marktteilnehmer als langfristig bzw. kurzfristig orientierte Investoren endogen ergibt. Dafür gibt es zwei plausible Gründe: die Skalenunabhängigkeit der preisrelevanten Informationen und die Knappheit der zur Verfügung stehenden Ressourcen. Der Informationsfluss divergiert und kann nicht mit einer hohen, aber konstanten Verarbeitungsrate bewältigt werden. Von Zeit zu Zeit erfolgt eine sprunghafte Anpassung an die neuen Verhältnisse, d.h. es ist untypisch, dass die Informationen aus einer langfristigen Komponente kontinuierlich in die Preise eingehen.

Die Skalingshypothese in unserer Formulierung impliziert aus der Sicht von Marktteilnehmern mit einem konstanten endlichen Investmenthorizont, dass die Ankunftsrate der Informationen auf der gewählten Skala variabel ist. Die Verarbeitungsrate wird so festgelegt, dass sie dem Erwartungswert der Ankunftsrate entspricht. Da letztere skaleninvariant ist, werden die Abweichungen zwischen den beiden Größen exponentiell verstärkt.

Nehmen wir an, dass die längerfristigen Informationen, die erst aus der Sicht von mehreren hundert Jahren von Relevanz sind, innerhalb einer kurzen Zeit auf den Markt gelangen. Als Beispiel kann der 11. September 2001 betrachtet werden (vgl. Abb.8.1). Aufgrund einer *untypischen* Skala sind die Investoren aus kurzfristiger Sicht einfach überfordert, die Auswirkungen dieses Ereignisses abzuschätzen. Eine Log-Transformation der Zeit von mehreren hundert Jahren auf wenige Tage oder sogar Stunden entspricht einer Verstärkung von Risiken auf der feinsten Skala. Dies führt zu extremen Renditeauschlägen, die aus kurzfristiger Sicht nicht einmal als untypische Ereignisse eingestuft werden können, weil man keine Grundlage für die Bestimmung von Wahrscheinlichkeiten hat. Folglich kann eine fehlerfreie Verarbeitung in diesem Fall nicht garantiert werden und die *Modellrisiken* gewinnen an Bedeutung.

Heuristisch lässt sich die Idee einer explosionsartigen Risikoverstärkung mit Hilfe des Laplace-Integrals veranschaulichen. Sei

$$I_q(\Delta t) \sim \int (\Delta t)^{(V-Aq)} \rho(A) dA \quad (8.2.1)$$

die Gesamtmenge der zu verarbeitenden Informationen über das Zeitintervall Δt , V und A entsprechend eine konstante Verarbeitungsrate und eine zufällige Ankunftsrate der Informationen. Mit q wird ein Skalenparameter bezeichnet: die höheren Werte von q stehen

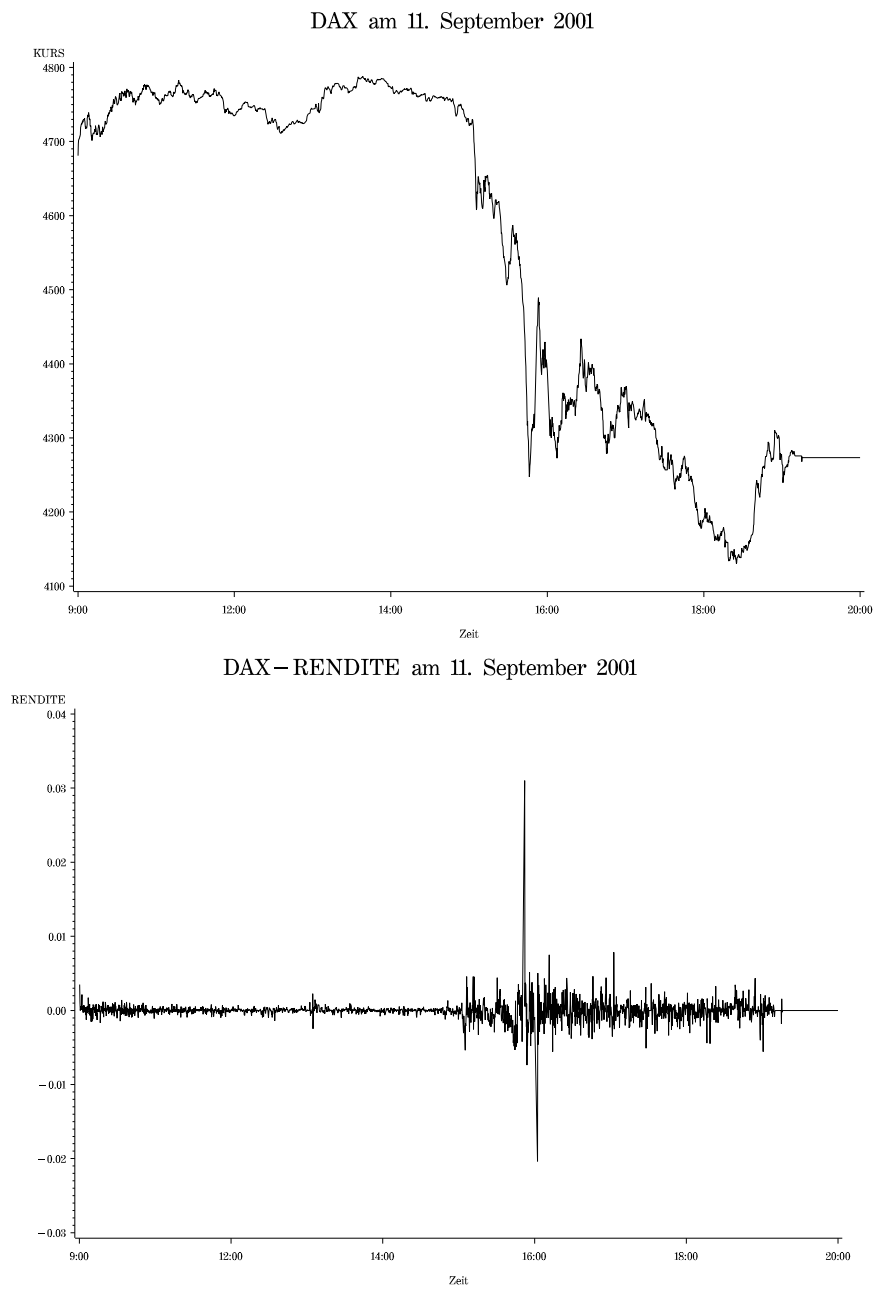


Abb. 8.1. Der DAX-Index und seine Rendite am 11. September 2001.

für eine feinere Skala, da $\Delta t \ll 1$ ist. Solange die Verarbeitungsrate um ein Vielfaches höher als die Ankunftsrate ist, bleibt der Ausdruck $(V - Aq)$ groß und $(\Delta t)^{(V-Aq)}$ entsprechend klein. Die Informationsmenge wird im wesentlichen durch eine Dichtefunktion $\rho(A)$ bestimmt. Das ist der „Normalzustand.“ Probleme beginnen, wenn $(V - Aq)$ klein und $(\Delta t)^{(V-Aq)}$ entsprechend groß wird. Dann wird der Beitrag von $(\Delta t)^{(V-Aq)}$ im Vergleich zu dem von $\rho(A)$ sehr groß werden und das Integral dominieren. Liegt die Ursache an der gestiegenen Ankunftsrate auf derselben Skala, kann das Modell durch die Wahl $V = A + \text{const}$ kalibriert werden.⁵ Wie ist aber ein höherer Wert des Parameters q zu interpretieren?

Es ist genau der Fall, in dem man von einer untypischen Skala spricht. Kommt ein seltenes Ereignis vor, ändert sich die Skalenordnung. Der kleinstmögliche Wert von q wird jetzt mit diesem Ereignis assoziiert. Entsprechend „erhöht sich der Zähler“ so, dass $q_{\text{neu}} = q_{\text{alt}} + \varepsilon$ ist und aus den vorherigen Darstellungen folgt, dass sich die Ankunftsrate A um den Faktor $\Delta = \varepsilon/q$ erhöht. Als Gegenmaßnahme ist eine *proportionale* Erhöhung von V auf $V(1 + \Delta)$ erforderlich. Aus Kostengründen kann aber diese Lösung nicht immer dauerhaft tragbar sein und das Problem lässt sich auf die Optimierung unter Nebenbedingungen zurückführen, wobei ε eine unbekannte Größe bleibt und subjektiv festgelegt werden muss.

Welche Möglichkeiten hat ein Investor, diese Art von Risiken, die wir Skalenrisiken nennen, in den Griff zu bekommen? Es bieten sich mehrere Alternativen. Die historische Erfahrung zeigt, dass die Portfolio-Versicherung durch die stufenweise gestalteten Stop-Loss-Orders nur selten eine gute Lösung darstellt, da in turbulenten Zeiten die Liquiditätsrisiken sehr hoch sind. Alternativ bietet sich ein Kauf von weit aus-dem-Geld liegenden Optionen an. Der Zeitwert einer Option ist zweifellos ein bedeutender Kostenfaktor, dafür bleiben die Risiken überschaubar und beschränkt. Sucht man nach einem optimalen Kosten/Folgen-Verhältnis,⁶ kann dieses durch die Anwendung von einem oder mehreren spezifischen Risikomaßen gefunden werden.⁷

Bisher eher von theoretischem Interesse ist das Verfahren von Begleitverteilungen (vgl. Abb. 8.2). Man sucht nach einer Skala, die den ex-

⁵ Nehmen wir an, die Verteilung von A hat einen kompakten Träger.

⁶ Wir verwenden diesen Ausdruck alternativ zum Rendite/Risiko-Verhältnis, wenn es sich um extreme Risiken handelt.

⁷ Zur ausführlichen Darstellung vgl. z.B. Ortobelli u.a. (2004) und Referenzen dort.

tremen Risiken am wenigsten ausgesetzt ist. Dabei werden durch eine nichtlineare Transformation sowohl die sehr großen als auch die sehr kleinen (im absoluten Wert) Wahrscheinlichkeiten so reskaliert, dass die resultierenden Werte gerade noch mit den herkömmlichen Verfahren wie Entropie verarbeitet werden können. In dem auf der Abb. 8.2 gezeigten Beispiel hat sich die Ungleichmäßigkeit der Maßaufteilung, gemessen an $(\sum q^2)^{1/2}$, nach einer nichtlinearen Transformation von 0,0493 (100%, $k=1$) zu 0,0279 (56,59%, $k=0,62$) sowie 0,0167 (33,87%, $k=0,32$) und 0,0115 (23,33%, $k=0,07$) reduziert.

Ein wichtiges Ergebnis dieser Arbeit ist ein Beitrag zur Klärung einer bisher offenen Frage, nämlich ob die Modelle auf Grundlage gebrochener Brownschen Bewegung (fBm) notwendig zu Arbitragemöglichkeiten führen. Aus mathematischer Sicht gibt es zwei Möglichkeiten, für diesen Prozess ein stochastisches Integral zu definieren: eine pfadweise Definition und eine Definition durch die Wick-Kalküle. Da die quadratische Variation der fBm im relevanten Bereich $1/2 < H < 1$ Null ist, spricht (im Unterschied zu einer gewöhnlichen Brownschen Bewegung) nichts gegen die pfadweise Definition. Die letztere ermöglicht aber Arbitragestrategien.⁸ Eine Definition durch die Wick-Kalküle lässt zwar keine Arbitragemöglichkeiten zu, führt aber ihrerseits ökonomisch zu unüberwindlichen Interpretationsschwierigkeiten. Dies führt in der Literatur oft zur Schlussfolgerung, dass die gebrochene Brownsche Bewegung als Underlying Prozess für ein risikobehaftetes Wertpapier keine gute Modellwahl ist.

Die aufgezeigten Arbitragestrategien sind selbstverständlich ein wichtiges ökonomisches Argument gegen die gebrochene Brownsche Bewegung als Underlying-Prozess für ein risikobehaftetes Wertpapier. Speziell in Rogers (1997) liegt aber eine Modellschwäche durch Annahme eines risikolosen Zinses von Null vor. Andererseits kann man die Arbitragefreiheit im Modell durch die Einführung eines minimalen nicht-zufälligen Zeitintervalls zwischen zwei benachbarten Handelszeitpunkten wiedergewinnen [Cheridito 2001]. Dies deutet auf eine starke Abhängigkeit der ökonomischen Implikationen von den Modellvoraussetzungen hin.

Das Thema „Langzeitabhängigkeit“ ist im Finance sehr wichtig und sehr kontrovers. Es ist bekannt, dass seinerzeit durch die Einschränkung auf „reguläre“ Handelsstrategien eine arbitragefreie Modellie-

⁸ Rogers (1997), Salopek (1998), Shiryaev (1999a), Cheridito (2003, 2001).

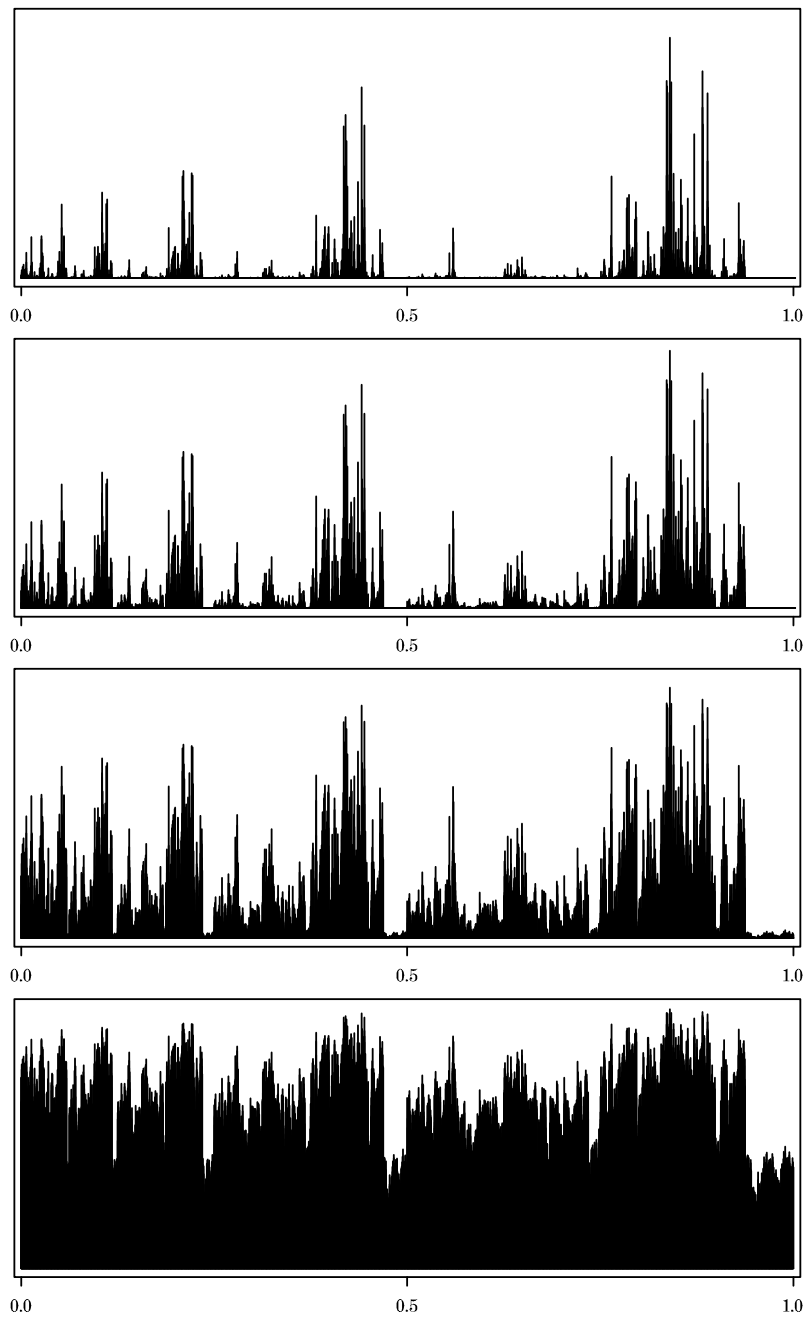


Abb. 8.2. Ein multifraktales Zufallsmaß (ganz oben) und Ergebnisse seiner Transformation mittels Begleitverteilungen $q_i(x) = p_i^k(x) [\sum_{i=1}^n p_i^k(x)]^{-1}$ mit $k = 0,62$, $k = 0,32$ und $k = 0,07$ (von oben nach unten).

rung mit der gewöhnlichen Brownschen Bewegung ermöglicht wurde. Ob eine ähnliche Vorgehensweise bei der gebrochenen Brownschen Bewegung alternativ zu Cheriditos Vorschlag zielführend ist, bleibt noch offen. Es besteht also ein weiterer Diskussionsbedarf über die verwendeten Verfahren und über die ökonomische Bedeutung der erzielten Ergebnisse.

Damit schließen wir die Arbeit ab und hoffen, Anregungen zu weiteren Untersuchungen dieses spannenden Forschungsgebiets gegeben zu haben.

Anhang

1

2

3

4

A. Hilfsmittel der Stochastik

“To use the tools of probability is not to say chance governs global commerce and finance. Sure, after the fact, with enough time and effort, we can piece together a tolerable cause-and-effect story of why a price moved the way it did. But who cares? It is too late by then. Fortunes have been gained and lost. Before the fact, in the real world of fast markets, veiled motives, and uncertain outcomes, probability is the only tool at our disposal.

First and foremost, random need not mean simple. There is more to probability than coins and dice.”

– Mandelbrot und Hudson (2004), S. 30.

Motivation

Es besteht kein Zweifel daran, dass die mathematische Erfassung der Finanzrisiken mit Mitteln der Wahrscheinlichkeitstheorie sich in Theorie und Praxis durchgesetzt hat. Paradoxe Weise lässt sich in dieser Umgebung das Wesen des Zufalls *im Allgemeinen* nicht umfassend beschreiben. Es ist beispielsweise keine definitive Aussage darüber möglich, ob eine gegebene Zeichenkette $a_1 a_2 \dots a_n$ zufällig ist oder nicht. Dies wäre ein Ausgangspunkt für die *Komplexitätstheorie*, einer sehr interessanten Weiterentwicklung, die ihre Wurzeln in den Arbeiten von A. N. Kolmogorov hat.¹

Ein weiterer Punkt liegt in Verbindung zwischen Wahrscheinlichkeitstheorie und Statistik. Eigentlich weiß niemand genau, warum die empirischen Frequenzen der Ereignisse eine gute Annäherung für die

¹ Eine sehr detaillierte Darstellung dazu bieten Li und Vitanyi (1993).

mathematischen Wahrscheinlichkeiten darstellen sollen. In den meisten Fällen denkt man an das Gesetz der großen Zahlen bzw. an den Zentralen Grenzwertsatz, ohne sich besonders viel Sorgen über deren Gültigkeitsbereiche zu machen. Einige weitere Konzepte wie Entropie beanspruchen für sich die Rolle, das Ganze anschaulich zu machen.

Doch wirklich spannend wird es, wenn die *seltenen* oder die *extremen* Ereignisse ins Spiel kommen. Diese beiden Ereignistypen bezeichnen wir vereinfacht mit dem Sammelbegriff *extreme* Ereignisse. Ökonomisch gesehen stellen sie die größten Risiken dar, gleichzeitig werden solche Ereignistypen meistens als *untypisch* eingestuft. Dies lässt die Hoffnung auf Hilfe von Seiten geeigneter statistischer Verfahren nicht all zu groß werden: Statistik spezialisiert sich vorzugsweise auf Ereignissen, die immer wieder auftreten. Es sei denn, die extremen Ereignisse häufen sich, wie beispielsweise ein schwerer Autounfall nach einem „Sechser“ im Lotto. Wie auch immer genießen die beiden Extremwertbereiche – Raum und Zeit – seit langem erhöhte Aufmerksamkeit und Interesse der Forscher unter anderem im Finance.

Die entsprechenden Verfahren nehmen eine Sonderstellung im Bereich zwischen Maß- und Wahrscheinlichkeitstheorie ein. Die dazugehörigen Wahrscheinlichkeitsverteilungen sind merkwürdig: sie haben keine Dichten und heißen *singulär*. Die Maße können einen fraktalen Träger bekommen, der weder eine „richtige“ Punktmenge noch eine glatte Linie oder eine vollständig gefüllte Fläche ist. Die zeitlichen Abhängigkeiten der Zufallsvariablen (in unserem Kontext sind das natürlich meistens Wertpapierrenditen) dürfen bis ins Unendliche reichen. Deren Schicksal teilt unter Umständen das zweite Moment einer Verteilung bzw. die Varianz: sie kann ebenfalls unendlich sein.

Ist eine solche Modellierung überhaupt ökonomisch sinnvoll? Sind Arbitragestrategien in Präsenz des Langzeitgedächtnisses wirksam? Welche Risikoprämie hat ein Investor zu verlangen, wenn die entsprechende Renditeverteilung eine unendliche Varianz hat? Um den Antworten näher zu kommen, versuchen wir zunächst darzustellen, wie solche oder ähnliche Effekte mathematisch mit Mitteln der Maß- bzw. Wahrscheinlichkeitstheorie modelliert werden können.

Der folgende Abschnitt enthält einige Begriffe und Konzepte, die für ein weiteres Verständnis der Vorgehensweise essentiell sind. Keineswegs erheben wir Anspruch auf eine umfassende Darstellung. Dies würde sicherlich den Rahmen dieser Arbeit sprengen. Manchmal las-

sen wir Beweise und technische Bedingungen aus und verweisen auf entsprechende Stellen in Lehrbüchern.

Die am häufigsten verwendeten Quellen sind Chung (1974), Williams (1991), Gnedenko und Korolev (1996), Bening und Korolev (2002). Sehr hilfreich fanden wir Petrov (1975), Cohn (1980), Kloeden und Platen (1992), Borovskikh und Korolyuk (1997), Pollard (2002) sowie Gärtner (2001). Feller I und II ist ein Muss.

A.1 Das (Wahrscheinlichkeits-) Maß

Die Axiome von Kolmogorov

Eine klassische Definition der Wahrscheinlichkeit nach Laplace geht von einer Menge unterschiedlicher Ausgänge oder Elementarergebnisse eines Zufallsexperimentes, die gleichwahrscheinlich sind, aus. Dann werden diejenigen Ergebnisse, die eine relevante Eigenschaft aufweisen, die sogenannten *günstigen Ergebnisse*, gezählt. Das Verhältnis aus der Anzahl A der günstigen Ergebnisse und der Anzahl der möglichen Ergebnisse heißt Wahrscheinlichkeit von A . Ungeachtet der weiten Verbreitung und der intuitiven Klarheit, führt diese Sichtweise leider zu technischen Schwierigkeiten bei der fortgeschrittenen mathematischen Modellierung.

Der Ausgangspunkt der modernen Wahrscheinlichkeitstheorie nach A. N. Kolmogorov (1933) ist *axiomatischer* Natur und besteht zunächst aus einer nichtleeren Menge Ω aller möglichen *Ergebnisse* $\omega \in \Omega$ eines Zufallsexperimentes. Hinzu kommt eine Kollektion \mathcal{F} von Teilmengen oder *Ereignissen* $A \subseteq \Omega$, wobei jedes Ereignis selbst eine Menge von Ergebnissen darstellt. In einem trivialen Fall besteht ein *Elementar-Ereignis* aus einem einzigen Ergebnis: $A = \{\omega\}$. Wir sagen, dass das Ereignis A *eingetreten* ist, wenn das Ergebnis ω in A liegt. Es ist sinnvoll, das Ereignissystem \mathcal{F} so zu gestalten, dass es unter Komplementenbildung und unter Vereinigung von endlichen oder abzählbaren Elementen abgeschlossen ist.

Definition A.1.1 (σ -Algebra). Ein Mengensystem \mathcal{F} heißt **σ -Algebra**, falls gilt:

- (i) sowohl ein Gesamttraum Ω als auch eine leere Menge \emptyset sind in \mathcal{F} ;

- (ii) für jedes A aus \mathcal{F} ist sein Komplement $A^c = \Omega \setminus A$ ebenfalls in \mathcal{F} ;
 (iii) ist A_1, A_2, \dots eine abzählbare Sequenz der Mengen in \mathcal{F} , dann sind beide

$$\bigcup_{i=1}^{\infty} A_i \quad \text{und} \quad \bigcap_{i=1}^{\infty} A_i \quad \text{in } \mathcal{F}.$$

Ein *sicheres Ereignis* ist hier durch den Gesamttraum Ω vertreten. Ein zu Ω komplementäres Ereignis \emptyset ist folglich ebenfalls ein Ereignis und heißt *unmögliches Ereignis*. Damit ist $\{\emptyset, \Omega\}$ eine σ -Algebra, die sogenannte *triviale σ -Algebra*.

Durch die σ -Algebra \mathcal{F} wird das Fundament für weitere wichtige Begriffe wie Wahrscheinlichkeitsmaß und Wahrscheinlichkeit gelegt. Wir definieren einen Wahrscheinlichkeitsraum als einen messbaren Raum, in dem ein Maß bestimmte Anforderungen erfüllen muss: die „Axiome der Wahrscheinlichkeitstheorie“. Mengen stellen im Wahrscheinlichkeitsraum unterschiedliche Arten von Ereignissen dar. Damit entspricht die Wahrscheinlichkeit des Eintretens eines bestimmten Ereignisses genau dem Maß der entsprechenden Menge. Alle Schlussfolgerungen der klassischen Theorie (nach Laplace) fließen in diese Umgebung mit ein, wenn sie zutreffen. Man sieht lediglich von den gleichwahrscheinlichen Elementarergebnissen ab, gewinnt aber dafür ein stabiles Fundament: die Möglichkeit, auf die Maßtheorie zurückgreifen, bleibt erhalten und wird immer wieder genutzt.

Definition A.1.2 (Maß).

- (i) Ein **Maß** (ein *endliches additives Maß*) ist eine Funktion $\mu : \mathcal{F} \rightarrow [0, \infty]$ derart, dass $\mu(\emptyset) = 0$ und

$$\mu\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} \mu(A_i) \quad (\sigma\text{-Additivität})$$

für jede abzählbare Sequenz von paarweise disjunkten² Mengen $A_1, A_2, \dots \in \mathcal{F}$.

- (ii) Das Paar (Ω, \mathcal{F}) heißt ein **Messraum**.
 (iii) Ein **Maßraum** ist ein Tripel $(\Omega, \mathcal{F}, \mu)$ mit einer σ -Algebra \mathcal{F} und einem Maß μ .

² Zwei Mengen A und B sind genau dann *disjunkt*, wenn $A \cap B = \emptyset$.

Definition A.1.3 (Wahrscheinlichkeitsmaß).

- (i) Ein Maß \mathbb{P} ist ein Wahrscheinlichkeitsmaß, falls $\mathbb{P}(\Omega) = 1$.
- (ii) Ein Wahrscheinlichkeitsraum ist ein Tripel $(\Omega, \mathcal{F}, \mathbb{P})$ mit einer σ -Algebra \mathcal{F} und einem Wahrscheinlichkeitsmaß \mathbb{P} .

Mit $\mathbb{P}(A)$ bezeichnet man die Wahrscheinlichkeit dafür, dass Ereignis A unter dem gegebenen Wahrscheinlichkeitsmaß \mathbb{P} eintritt. Wir sagen, dass ein Ereignis *mit Wahrscheinlichkeit 1* oder *fast sicher* (\mathbb{P} -f.s.) auftritt, wenn $\mathbb{P}(A) = 1$ für alle $A \subset \Omega \setminus N$ mit $\mathbb{P}(N) = 0$, das heißt, abgesehen von den Null-Mengen. Analog spricht man bei einem Maß μ von einer Eigenschaft, die *fast überall* (f.ü.) gilt.

Unendlich ist nicht gleich unendlich...

Von besonderem Interesse in diesem Gerüst aus Mengen und Sequenzen ist die Idee abzählbarer Additivität. Um sie zu verstehen, greift man gewöhnlich auf das Längenkonzept von Henry Lebesgue zurück. In seiner Dissertation aus dem Jahre 1902 weist er jedem Intervall $(a, b) \subset \mathbb{R}$ ein Maß μ_L zu, das genau der Länge $(b - a)$ entspricht. Ein solches Maß lässt sich auf viele, aber nicht auf alle Untermengen von \mathbb{R} anwenden.³ Ein Teil von Lebesgue-messbaren Mengen gehört zu einer Klasse \mathcal{B} von Borel-Mengen, oder besser, Borel-Untermengen aus \mathbb{R} . Borel-Mengen haben sehr viel mit σ -Algebren gemeinsam. Es ist bekannt, dass ein Durchschnitt beliebiger nichtleerer Mengen von σ -Algebren eine σ -Algebra ist. Im Fokus steht dabei die feinste σ -Algebra, die sich aus der Sammlung von Untermengen \mathcal{F} einer Menge B erzeugen lässt. Diese σ -Algebra ist eindeutig und heißt *durch \mathcal{F} erzeugte σ -Algebra*.

$$\mathcal{B} = \sigma(\mathcal{F}) = \bigcap_{\substack{B \supseteq \mathcal{F} \\ B \text{ ist } \sigma\text{-Algebra}}} B. \quad (\text{A.1.1})$$

Definition A.1.4 (Borel-Menge). Die feinste σ -Algebra \mathcal{B} , die durch jede Vereinigung, Durchschnitt oder Komplement von semi-unendlichen Intervallen $\{x \in \mathbb{R}; -\infty < x \leq 1\}$ erzeugt ist, heißt *Borel- σ -Algebra* $\mathcal{B}(\mathbb{R})$. Elemente aus $\mathcal{B}(\mathbb{R})$ heißen *Borel-Mengen*.

³ Die Problematik der Erweiterung des Längenkonzeptes auf *alle* Untermengen aus \mathbb{R} wird ausführlich in Pollard (2002), S. 4. ff. diskutiert.

Man kann zeigen, dass $\mathcal{B}(\mathbb{R})$ alle Intervalle, alle offenen Mengen und alle abgeschlossenen Mengen enthält. Zu solchen „guten Mengen“ gehören zum Beispiel die natürlichen und rationalen Zahlen (letztere lassen sich als Bruch aus zwei natürlichen Zahlen darstellen). Die „weniger guten“ sind die irrationalen und (als eine Obermenge davon) die reellen Zahlen.

Georg Cantor (1845–1918) hat in zwei seiner Artikel (1895 und 1897) namens „Beiträge zur Begründung der transfiniten Mengenlehre“ gezeigt, dass die Mengen natürlicher und reeller Zahlen „nicht dieselbe Anzahl der Elemente“ haben. Die reellen Zahlen lassen sich nämlich nicht „auflisten“; jeder Versuch dies zu tun würde daran scheitern, dass sich immer eine reelle Zahl finden würde, die nicht in der Liste sei. Folglich ist die Menge der reellen Zahlen überabzählbar. Auf der Gegenseite stehen die natürlichen Zahlen, die sich „zählen lassen“. Eine der berühmten merkwürdigen Mengen heißt Cantor-Menge. Sie ist überabzählbar, Lebesgue-messbar mit $\mu_L(C) = 0$ und sie ist keine Borel-Menge.⁴

Es gibt Nullmengen, denen das Maß Null nicht zugeordnet werden kann, weil sie einfach nicht messbar sind. Nach Lebesgue besitzt jede abzählbare Punktmenge das Maß Null, da jeder einzelne Punkt einen „Null-Beitrag“ leistet. Nach Cantor gibt es einige überabzählbare Mengen, die ebenfalls Lebesgue-Null aber *nicht Borel-Null* sind. Sie sind überhaupt keine Borel-Mengen! Damit belegt $\mathcal{B}(\mathbb{R})$ nicht durchgehend die reelle Achse und ist unvollständig.

Es sei $(\Omega, \mathcal{F}, \mu)$ ein Maßraum. Betrachten wir die drei Relationen:

$$B \in \mathcal{F}, \quad \mu(B) = 0 \quad \text{und} \quad A \subset B.$$

Folgt daraus $A \in \mathcal{F}$, dann ist das Maß μ (oder der Maßraum $(\Omega, \mathcal{F}, \mu)$) *vollständig*. Die Untermenge A von Ω heißt μ -Nullmenge oder einfach *Nullmenge*. Das Maß μ ist vollständig genau dann, wenn jede μ -Nullmenge von Ω in \mathcal{F} ist.⁵ Für $A \notin \mathcal{F}$ ist dagegen das Maß μ unvollständig. Mit anderen Worten, für $A \subset B$ und B gleich Borel-Lebesgue-Null gilt $\mu(A) = 0$ auch wenn A nicht Borel ist. Auch solche Mengen A sind Lebesgue-Null. Eine Menge $A \subset \mathbb{R}$ ist Lebesgue-messbar, falls gilt

⁴ Georg Cantor hat seine Überraschung nach dieser Entdeckung in einem nicht weniger berühmten Satz erfasst: *“I see it, but I don’t believe it!”*

⁵ Cohn (1980), S. 35.

$A = B \cup N$ mit B Borel und N Lebesgue-Null.⁶ Kurzgefasst: um Konsistenz im Umgang mit überabzählbaren Mengen u.a. in Bezug auf die Definition der Wahrscheinlichkeiten zu behalten, beschränkt man sich im ersten Schritt auf Borel-Mengen. Danach werden die (restlichen) Nullmengen hinzugenommen, um das Mengensystem Lebesgue-messbar zu machen. Dieses Vorgehen bezeichnet man als Vervollständigung (Augmentation) eines Maßes.

Im Allgemeinen ist die Idee abzählbarer Additivität eher theoretischer als praktischer Natur: eine strenge Annahme, dass eine Vereinigung unendlich vieler Ereignisse selbst ein Ereignis darstellt (und eine Wahrscheinlichkeit besitzt), ist eine Abstraktion mit großen theoretischen Nutzen. Wir behalten stets in Erinnerung, dass *Wahrscheinlichkeiten immer Ereignissen zugewiesen werden*. Ergebnis eines Zufallsexperimentes kann ein Ereignis sein, muss aber nicht (vgl. Seite 195)! Im ersten Fall bekommt es auch eine Wahrscheinlichkeit zugewiesen. Oft kann die Wahrscheinlichkeit eines Ereignisses aus den entsprechenden Wahrscheinlichkeiten individueller Ergebnisse *deduziert* werden; doch für *alle* Ereignisse gilt dies nicht. Auch die Wahrscheinlichkeit eines individuellen Ergebnisses lässt sich nicht immer aus der eines Ereignisses ableiten. Das kann bei technischen Beweisen sehr wichtig sein.

Gewöhnlich bringt die Beschränkung auf Borel-Mengen keine Nachteile mit sich. Eigentlich „würde man keinen großen Fehler machen ausgehend davon, dass alle einem begegnenden Mengen Borel-Mengen sind.“⁷ Leider stellen gerade fraktale Mengen, die für uns im Weiteren besonders interessant werden, eine Ausnahme dar. Ohne Augmentation wäre es beispielsweise unmöglich, ein Maß auf einem fraktalen Träger zu definieren.

Zufallsvariablen und Verteilungen

Unter *Zufallsvariable* versteht man eine Regel für die Zuweisung reeller Zahlen zu Ergebnissen eines Zufallsexperimentes.

⁶ Cohn (1980), S. 22 stellt zwei berühmte Fragen:

- (i) Ist jede Untermenge von \mathbb{R} Lebesgue-messbar?
- (ii) Ist jede Lebesgue-messbare Menge eine Borel-Menge?

Die Antwort lautet in beiden Fällen „Nein.“

⁷ Falconer (1993), S. 6.

Definition A.1.5 (Zufallsvariable). Es sei $(\Omega, \mathcal{F}, \mathbb{P})$ ein Wahrscheinlichkeitsraum. Eine Funktion $X : \Omega \rightarrow \mathbb{R}$ heißt Zufallsvariable, falls sie \mathcal{F} -messbar ist: für jede Borel-Menge $B \in \mathcal{B}(\mathbb{R})$ liegt das Urbild $X^{-1}(B)$ in \mathcal{F} .

Durch die Zufallsvariable X wird jedem Ergebnis ω in Ω ein Wert $X(\omega)$ zugewiesen. Eine Zufallsvariable $X(\omega) = 5$ könnte zum Beispiel das Ergebnis eines Zufallsexperimentes mit dem Ausgang „eine rote Kugel wurde gezogen“ quantifizieren. Das Konzept einer Zufallsgröße wird aus rein theoretischen Überlegungen eingeführt. Man ersetzt einen natürlichen, aus allen möglichen Ergebnissen ω bestehenden Zufallsraum Ω durch einen erweiterten Raum \mathbb{R} , in dem Ergebnisse reelle Zahlen sind und legt fest, welche Teilmengen davon Ereignisse sind. Die Einschränkung auf Borel-Mengen macht es möglich, einen Wahrscheinlichkeitsraum so zu konstruieren, dass sowohl die Ergebnismenge der Menge $\mathcal{B}(\mathbb{R})$ entspricht als auch jede Untermenge von $\mathcal{B}(\mathbb{R})$ ein Ereignis darstellt. Die letzte Bedingung für das Urbild $X^{-1}(B)$ garantiert, dass die Ergebnismenge ω , an die durch X ein Wert in B zugewiesen wird, tatsächlich ein Ereignis ist. Im nächsten Schritt weist man den Zufallsvariablen Wahrscheinlichkeiten zu.

Definition A.1.6. Es sei $(\Omega, \mathcal{F}, \mathbb{P})$ ein Wahrscheinlichkeitsraum und X eine reellwertige Zufallsvariable. Durch \mathcal{P}_X wird auf $\mathcal{B}(\mathbb{R})$ ein Wahrscheinlichkeitsmaß definiert, für das gilt

$$\mathcal{P}_X[B] = \mathbb{P}(X^{-1}(B)) = \mathbb{P}(\{\omega : X(\omega) \in B\}) \equiv \mathbb{P}(X \in B) \text{ für alle } B \in \mathcal{B}.$$

Das Maß \mathcal{P}_X wird **Wahrscheinlichkeitsverteilung** von X genannt und gibt die Wahrscheinlichkeit an, dass die Zufallsvariable X Werte in der Borel-Menge B annimmt. Der **Träger (Support)** einer Verteilung ist die kleinste abgeschlossene Menge B mit der Eigenschaft $\mathbb{P}(X \in B) = 1$. Die reellwertige Funktion $F_X : \mathbb{R} \rightarrow \mathbb{R}$

$$F_X(x) \equiv F(x) \equiv \mathcal{P}_X[(-\infty, x]] = \mathbb{P}(X \leq x)$$

heißt **Verteilungsfunktion** der Zufallsvariable X oder einfach **Verteilung** von X .

Jede Verteilungsfunktion ist nicht fallend, rechtsseitig-stetig und es gilt:

$$\lim_{x \rightarrow -\infty} F(x) = 0, \quad \lim_{x \rightarrow \infty} F(x) = 1.$$

BEISPIEL A.1

Das gemäß (A.1.6) definierte Wahrscheinlichkeitsmaß heißt

(i) *Binomialverteilung* (n, p) , falls

$$\mathbb{P}(X = k) = \binom{n}{k} p^k (1-p)^{n-k}, \quad k = 0, 1, \dots, n. \quad (\text{A.1.2})$$

(ii) *Poissonverteilung* (λ) , falls für alle $\lambda > 0$

$$\mathbb{P}(X = k) = e^{-\lambda} \frac{\lambda^k}{k!}, \quad k = 0, 1, \dots \quad (\text{A.1.3})$$

(iii) *Normalverteilung* $N(\mu, \sigma^2)$, falls für alle $a < b$

$$\mathbb{P}(a < X < b) = \frac{1}{\sigma\sqrt{2\pi}} \int_a^b e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx. \quad (\text{A.1.4})$$

Damit haben wir für jedes $B = (a, b] \cup \mathcal{B}$:

$$\mathbb{P}(X \in (a, b]) = F(b) - F(a), \quad \mathbb{P}(\{a\}) = F(a) - F(a-), \quad (\text{A.1.5})$$

wobei die Differenz $F(a) - F(a-)$ der Sprunghöhe in a entspricht. Die Verteilungsfunktion einer stetigen Verteilung hat keinen Sprung in B . Darum gilt für eine *stetige* Verteilung notwendigerweise $\mathbb{P}(X \in B) = 0$ für jedes $B \in \mathcal{B}$. Ist jede solche Menge auch Lebesgue-Null, heißt die Verteilung *absolutstetig* und lässt sich in Integralform mit einer *Dichtefunktion* $f(u) \geq 0$ darstellen

$$\mathbb{P}(X \leq x) = \int_{-\infty}^x f(u) du.$$

Definition A.1.7 (Absolutstetigkeit, äquivalente Maße). Es seien μ und ν Maße auf einem Messraum (Ω, \mathcal{F}) . Das Maß ν heißt *absolutstetig* zur Basis μ (in Zeichen: $\nu \ll \mu$), falls jede μ -Nullmenge auch ν -Nullmenge ist:

$$A \in \mathcal{F}, \mu(A) = 0 \quad \Rightarrow \quad \nu(A) = 0.$$

Gilt zusätzlich $\mu \ll \nu$, dann heißen die Maße μ und ν *äquivalent*.

Bemerkung. Die Nullmengen zweier äquivalenter Maße stimmen exakt überein.

Satz A.1.1 (Radon-Nikodym, 1930) Sei $(\Omega, \mathcal{F}, \mathbb{P})$ ein Wahrscheinlichkeitsraum und \mathbb{Q} ein Maß auf (Ω, \mathcal{F}) mit $\mathbb{Q} \ll \mathbb{P}$. Dann existiert eine nichtnegative messbare numerische Funktion f auf (Ω, \mathcal{F}) mit

$$\mathbb{Q}(A) = \int_A f \, d\mathbb{P} \quad \text{für alle } A \in \mathcal{F}.$$

Die Funktion f ist \mathbb{P} -f.ü. eindeutig bestimmt und kann reellwertig gewählt werden.

Bemerkung. Die Funktion f im Satz A.1.1 heißt *Radon-Nikodym-Ableitung* oder *Dichte* von \mathbb{Q} bezüglich \mathbb{P} und wird mit $d\mathbb{Q}/d\mathbb{P}$ bezeichnet.

Beweis: Billingsley (1986), S. 376. □

Beweisidee. Sei $(\Omega, \mathcal{F}, \mathbb{P})$ ein Wahrscheinlichkeitsraum und f eine nichtnegative messbare numerische Funktion. Für $A \subseteq \Omega$ setzen wir

$$\mathbb{1}_A(\omega) = \begin{cases} 1, & \omega \in A, \\ 0, & \omega \in \Omega \setminus A. \end{cases}$$

Die Funktion $\mathbb{1}_A$ heißt *Indikator-Funktion* der Menge A . Sie ist genau dann messbar, wenn A messbar ist, $A \in \mathcal{F}$. Dann ist das durch

$$\mathbb{Q}(A) := \int_A f \, d\mathbb{P}, \quad A \in \mathcal{F},$$

auf (Ω, \mathcal{F}) definierte Maß \mathbb{Q} absolutstetig bezüglich \mathbb{P} . Tatsächlich, aus $\mathbb{P}(A) = 0$ folgt $\mathbb{1}_A f = 0$ \mathbb{P} -f. ü. und hieraus

$$\mathbb{Q}(A) = \int_{\Omega} \mathbb{1}_A f \, d\mathbb{P} = 0.$$

□

Es bietet sich folgende Rechenregel für die Radon-Nikodym-Ableitungen an.

Satz A.1.2 Es seien \mathbb{P} und \mathbb{Q} zwei Wahrscheinlichkeitsmaße auf (Ω, \mathcal{F}) mit $\mathbb{Q} \ll \mathbb{P}$. Dann gilt für alle messbaren numerischen Funktionen g auf (Ω, \mathcal{F})

$$\int g \, d\mathbb{Q} = \int g \frac{d\mathbb{Q}}{d\mathbb{P}} \, d\mathbb{P},$$

d.h., existiert eines der beiden Integrale, so existiert auch das andere, und die beiden sind gleich.

*Beweisidee.*⁸ Die Behauptung gilt für die Indikatorfunktionen $g = \mathbb{1}_A$ ($A \in \mathcal{F}$):

$$\int \mathbb{1}_A d\mathbb{Q} = \mathbb{Q}(A) = \int_A \frac{d\mathbb{Q}}{d\mathbb{P}} d\mathbb{P} = \int_A \mathbb{1}_A \frac{d\mathbb{Q}}{d\mathbb{P}} d\mathbb{P}.$$

Hiervon ausgehend überlegt man sich schrittweise, dass die Behauptung auch im allgemeinen Fall gilt. \square

Der Erwartungswert einer Zufallsvariable X auf $(\Omega, \mathcal{F}, \mathbb{P})$ existiert⁹ für $\int_{\Omega} |X| d\mathbb{P} < \infty$:

$$\mathbb{E}X = \int_{\Omega} X d\mathbb{P} = \int_{-\infty}^{\infty} x dF_X(x). \quad (\text{A.1.6})$$

Gilt auch $\mathbb{E}X^2 < \infty$ und $\mathbb{E}Y < \infty$, dann existieren *Varianz* und *Kovarianz*:

$$\text{Var}X = \sigma^2 = \mathbb{E}[(X - \mathbb{E}X)^2], \quad (\text{A.1.7})$$

$$\text{Cov}(X, Y) = \mathbb{E}[(X - \mathbb{E}X)(Y - \mathbb{E}Y)]. \quad (\text{A.1.8})$$

Ausgehend von der bedingten Wahrscheinlichkeit eines Ereignisses B relativ zum Ereignis A $\mathbb{P}(B|A) = \mathbb{P}(B \cap A) / \mathbb{P}(A)$ lässt sich der *bedingte Erwartungswert* $\mathbb{E}[X|A]$ einer beliebigen integrierbaren Zufallsvariable definieren. Es bietet sich an, anstatt des Ereignisses A eine Sub- σ -Algebra $\mathcal{G} \subset \mathcal{F}$ zu verwenden. Dies hat einen technischen Vorteil, falls \mathcal{G} (abzählbar) unendlich viele Ereignisse einschließt. Dann setzen wir Integrierbarkeit von X voraus, legen $Y = \mathbb{E}[X|\mathcal{G}]$ fest und verlangen, dass Y \mathcal{G} -messbar ist.

Definition A.1.8. Sei X eine Zufallsvariable und \mathcal{G} eine Sub- σ -Algebra von \mathcal{F} . Dann heißt eine \mathcal{G} -messbare Zufallsvariable $Y = \mathbb{E}[X|\mathcal{G}]$ **bedingter Erwartungswert von X bezüglich \mathcal{G}** und es gilt

$$\mathbb{E}[\mathbb{1}_C X] = \mathbb{E}[\mathbb{1}_C \mathbb{E}[X|\mathcal{G}]] \quad \text{für alle } C \in \mathcal{G}. \quad (\text{A.1.9})$$

Bemerkung. Man kann zeigen, dass eine solche Zufallsvariable existiert und eindeutig ist. Außerdem gilt für $\mathcal{G} \subset \mathcal{F}$ die „Tower-Property“ $\mathbb{E}[\mathbb{E}[X|\mathcal{F}]|\mathcal{G}] = \mathbb{E}[X|\mathcal{G}]$, d.h. jedes Ereignis von \mathcal{G} ist auch Ereignis von \mathcal{F} . Weiterhin ermöglicht die Eigenschaft

⁸ Gärtner (2001), S. 123.

⁹ Ist die Bedingung erfüllt, heißt die Zufallsvariable X *integrierbar*.

$$\int_G X \, d\mathbb{Q} = \int_G \mathbb{E}[X | \mathcal{F}] \, d\mathbb{Q} \quad \text{für alle } G \subseteq \mathcal{F}$$

eine Maßstransformation für $\mathbb{Q} \ll \mathbb{P}$, da $\mathbb{E}[X | \mathcal{F}]$ \mathcal{G} -messbar ist:

$$\mathbb{E}^{\mathbb{Q}} X = \int_G X \, d\mathbb{Q} = \int_G \mathbb{E}[X | \mathcal{F}] \, d\mathbb{Q} = \int_G X \frac{d\mathbb{Q}}{d\mathbb{P}} \, d\mathbb{P} = \mathbb{E}^{\mathbb{P}} \left[\frac{d\mathbb{Q}}{d\mathbb{P}} X \right]. \quad (\text{A.1.10})$$

Man sagt, dass ein Maß auf einer Menge $A \in \mathcal{F}$ *konzentriert* ist, falls $\mu(A^c) = 0$ (und damit $\mu(B) = 0$ für alle messbaren $B \subseteq A^c$) gilt. Die kleinste abgeschlossene Menge A mit $\mu(\mathbb{R}^n \setminus A) = 0$ ist der Träger von μ . Wir sagen, dass μ ein Maß auf einer Menge A ist, wenn A den Träger von μ enthält.¹⁰

Definition A.1.9. Eine **singuläre** Wahrscheinlichkeitsverteilung ist eine kontinuierliche Verteilung, die vollständig auf dem Lebesgue-Maß Null ist. Eine **diskrete** Wahrscheinlichkeitsverteilung ist eine singuläre Verteilung, die höchstens abzählbar viele Werte annimmt.

Damit existiert eine endliche oder abzählbare Punktmenge $B \in \mathcal{B}$ derart, dass $\mathbb{P}(X \in B) = 1$. Die Verteilungsfunktion einer diskreten Zufallsvariable X mit $\mathbb{P}(X = x_i) = p_i$ ist eine Treppenfunktion mit Sprunghöhen p_i : $F_X(x) = \sum_{i: x_i < x} p_i$.

Satz A.1.3 (Lebesgue) Jede Verteilungsfunktion $F(x)$ ist auf eine eindeutige Weise als Summe aus drei Komponenten darstellbar:

$$F(x) = a_1 F_1(x) + a_2 F_2(x) + a_3 F_3(x). \quad (\text{A.1.11})$$

Dabei sind

- a_1, a_2 und a_3 nichtnegative Zahlen, die sich zu eins addieren; und
- $F_1(x), F_2(x)$ und $F_3(x)$ ist jeweils eine absolutstetige, eine diskrete und eine singuläre Verteilungsfunktion.

Beweis: Petrov (1975). □

Bemerkung. Eine diskrete Verteilung kann aufgrund der vorherigen Überlegungen als singulär gelten. So besteht eine Verteilung aus einem absolutstetigen und einem singulären Teil.

¹⁰ Falconer (1993), S. 12.

A.2 Charakterisierung der Wahrscheinlichkeitsverteilungen

L^p -Räume

Mit $L^p(\Omega, \mathcal{F}, \mu) = L^p(\mu)$ oder einfach L^p bezeichnen wir einen Raum aller p -fach ($p \geq 1$) μ -integrierbaren reellwertigen Funktionen auf Ω . Danach gehört eine reellwertige Funktion f genau dann zu L^p , wenn sie \mathcal{F} -messbar ist und $\int |f|^p d\mu < \infty$. Durch

$$\|f\|_p \equiv \left[\int |f|^p d\mu \right]^{1/p} \quad (\text{A.2.1})$$

ist eine Seminorm auf L^p definiert.¹¹ Ein vollständiger, normierter Raum heißt *Banachraum*. Nach dem Satz von *Riesz-Fischer*¹² ist L^p vollständig. Damit ist $L^p(\mu)$ ein Banachraum unter entsprechender Norm.¹³ Von besonderem Interesse sind oft die sogenannten *Hilberträume* L^2 . Als zusätzliche Struktur lassen sie ein *Skalarprodukt* zu:

$$\langle f, g \rangle = \int f g d\mu, \quad \langle f, f \rangle = \|f\|_2^2, \quad f, g \in L^2.$$

Momente

Bezogen auf Zufallsvariablen interpretieren wir L^p als eine Auswahl der Zufallsvariablen auf $(\Omega, \mathcal{F}, \mathbb{P})$ mit endlichem p -ten Moment. Ein *absolutes Moment* der Zufallsvariable X ist definiert als

$$\mathbb{E}|X|^k = \int_{-\infty}^{\infty} |x|^k dF(x), \quad k = 1, 2, \dots \quad (\text{A.2.2})$$

Für jedes k , für welches (A.2.2) endlich ist, hat X ein *Moment der Ordnung k* oder ein k -tes Moment:

$$\mathbb{E}X^k = \int_{-\infty}^{\infty} x^k dF(x). \quad (\text{A.2.7})$$

¹¹ $\|\cdot\|_p$ ist im allgemeinen keine Norm: $\|f\|_p = 0$ ist gleichbedeutend mit $f = 0$ μ -fast überall, woraus nicht unbedingt $f \equiv 0$ folgen muss. Für den Fall $p = \infty$ ist das wesentliche Supremum von f definiert durch $\text{ess sup}_{\omega \in \Omega} |f(\omega)| := \inf \{c : |f| \leq c \text{ } \mu\text{-fast überall}\}$.

¹² Billingsley (1986), Theorem 19.1.

¹³ Elemente aus L^p sind eigentlich keine richtigen Funktionen, sondern eher Äquivalenzklassen von Funktionen. In diesem Sinne lassen sich zwei Funktionen f und g mit $\|f - g\|_p = 0$ nicht unterscheiden und gelten als ein Element des Raumes.

BEISPIEL A.2

Für $Y = |X|$ und $g(x) = x$ ergibt sich aus A.2.8 die *Markov-Ungleichung*

$$P(|X| \geq \varepsilon) \leq \frac{E|X|}{\varepsilon} \quad \text{für alle } \varepsilon > 0, \quad (\text{A.2.3})$$

für $Y = |X - EX|$ und $g(x) = x^2$ folgt die *Chebyshev-Ungleichung*

$$P(|X - EX| \geq \varepsilon) \leq \frac{\text{Var} X}{\varepsilon^2} \quad \text{für alle } \varepsilon > 0 \quad (\text{A.2.4})$$

oder für $Y = |X|$ und $g(x) = x^p$

$$P(|X| \geq \varepsilon) \leq \frac{1}{\varepsilon^p} E|X|^p \quad \text{für alle } \varepsilon > 0. \quad (\text{A.2.5})$$

Außerdem gilt für $g(x) = e^{tx}$, $t > 0$

$$P(X \geq \varepsilon) \leq \frac{E e^{tX}}{e^{t\varepsilon}} \quad \text{für alle } \varepsilon > 0. \quad (\text{A.2.6})$$

Ungleichungen

Sei g eine nichtnegative monoton wachsende Funktion auf \mathbb{R} . Dann gilt für alle Zufallsvariablen X und alle $\varepsilon > 0$ die *verallgemeinerte Chebyshev-Ungleichung*:

$$\mathbb{P}(|X| \geq \varepsilon) \leq \frac{E[g(X)]}{g(\varepsilon)}. \quad (\text{A.2.8})$$

Unter Berücksichtigung der Identität $E[\mathbb{1}_A] = \mathbb{P}(A)$ ist leicht zu sehen, dass

$$\begin{aligned} E[g(X)] &= E[g(X) \mathbb{1}_{|X| \geq \varepsilon}] + E[g(X) \mathbb{1}_{|X| < \varepsilon}] \\ &\geq E[g(X) \mathbb{1}_{|X| \geq \varepsilon}] \geq g(\varepsilon) E[\mathbb{1}_{|X| \geq \varepsilon}] = g(\varepsilon) \mathbb{P}(|X| \geq \varepsilon). \end{aligned} \quad (\text{A.2.9})$$

Daraus ergeben sich mehrere Spezialfälle (Beispiel A.2).

Für eine Zufallsvariable X mit endlichem Erwartungswert und eine konvexe Funktion g besagt die *Jensen-Ungleichung*, dass

$$Eg(X) \geq g(EX). \quad (\text{A.2.10})$$

Für $p, q \in [1, \infty]$ und $p^{-1} + q^{-1} = 1$ heißen die Indizes p und q *konjugiert*. Die *Hölder-Ungleichung* besagt, dass für konjugierte Indizes p und q und Zufallsvariablen $X \in L^p$ und $Y \in L^q$ folgendes gilt:

BEISPIEL A.3 Spezialfälle verallgemeinerter Durchschnitte

q	$(E X ^q)^{1/q}$	Ergebnisse
$-\infty$	$\min [x : X = x]$	degeneriert zum minimalen Wert
-1	$\frac{1}{\frac{1}{x_1} + \frac{1}{x_2} + \dots + \frac{1}{x_n}}$	Harmonisches Mittel
0	$\left(\prod_{i=1}^n X_i\right)^{1/n}$	Geometrisches Mittel
1	$\frac{1}{n} \sum_{i=1}^n X_i$	Arithmetisches Mittel
2	$\sqrt{\frac{1}{n} \sum_{i=1}^n X_i^2}$	Quadratisches Mittel
$+\infty$	$\max [x : X = x]$	degeneriert zum maximalen Wert

$$E[|XY|] \leq [E|X|^p]^{1/p} \cdot [E|Y|^q]^{1/q}, \quad \frac{1}{p} + \frac{1}{q} = 1. \tag{A.2.11}$$

Mit $1/p = 1/q = 1/2$ wird daraus *Ungleichung von Schwarz*:

$$E|XY| \leq \sqrt{EX^2 \cdot EY^2}. \tag{A.2.12}$$

Für $0 < \alpha < \beta$ setzen wir in (A.2.11) $p = \beta/\alpha$, $q = \beta/(\beta - \alpha)$, $Y(\omega) = 1$ ein und ersetzen X durch $|X|^\alpha$. Das Ergebnis ist als *Lyapunov-Ungleichung* bekannt¹⁴

$$\left[E|X|^\alpha\right]^{1/\alpha} \leq \left[E|X|^\beta\right]^{1/\beta}, \quad 0 < \alpha < \beta. \tag{A.2.13}$$

Einige Spezialfälle der sogenannten *verallgemeinerten Durchschnitte* oder *q-Mittel* aus (A.2.13) sind im Beispiel A.3 aufgeführt.¹⁵

Integraltransformationen

Definition A.2.1. Die **charakteristische Funktion (c.F.)** einer Zufallsvariable X oder einer Verteilungsfunktion $F(x)$ ist eine komplexwertige

¹⁴ Billingsley (1986), S. 81.

¹⁵ Zhang (1998), S. 137.

Funktion

$$\Psi(t) \equiv \mathbb{E} e^{itX} = \int_{-\infty}^{\infty} e^{itx} dF(x). \quad (\text{A.2.14})$$

Existiert eine Dichtefunktion $f(x) = F'(x)$, dann ist die c.F. bis auf das Vorzeichen im Exponenten gerade gleich der Fourier-Transformierten dieser Dichte

$$\Psi(t) = \int_{-\infty}^{\infty} e^{itx} f(x) dx. \quad (\text{A.2.15})$$

Für eine diskrete Zufallsvariable gilt entsprechend $\Psi(t) = \sum_{k=1}^{\infty} e^{itx_k} p_k$. Die wichtigsten Eigenschaften charakteristischer Funktionen sind in Kürze:¹⁶

1. Die c.F. einer Zufallsvariable existiert immer und bestimmt mittels *Umkehrformel* eindeutig die Verteilungsfunktion:

$$F(a+h) - F(a) = \lim_{T \rightarrow \infty} \frac{1}{2\pi} \int_{-T}^T \frac{1 - e^{-ith}}{it} e^{-ita} \Psi(t) dt \quad (\text{A.2.16})$$

(Integrierbarkeit vorausgesetzt). Zwei Verteilungsfunktionen mit gleichen c.F. sind identisch.

2. Die c.F. einer transformierten Zufallsvariable $Y = g(X)$ berechnet sich als $\mathbb{E} e^{itY}$. Damit gilt insbesondere für eine affine Transformation $Y = aX + b$

$$\Psi_Y(t) = \Psi_X(at) e^{ibt} \quad (\text{A.2.17})$$

3. Die c.F. einer Summe von n unabhängigen Zufallsvariablen $Z = X_1 + X_2 + \dots + X_n$ ist gleich dem Produkt ihrer c.F.:

$$\begin{aligned} \Psi(t) &= \mathbb{E} e^{itZ} = \mathbb{E} e^{it(X_1 + X_2 + \dots + X_n)}, \\ &= \mathbb{E} e^{itX_1} \mathbb{E} e^{itX_2} \dots \mathbb{E} e^{itX_n} \quad (\text{Unabhängigkeit}) \\ &= \Psi_1(t) \Psi_2(t) \dots \Psi_n(t). \end{aligned} \quad (\text{A.2.18})$$

Zum Vergleich: die *Verteilungsfunktion* einer Summe von unabhängigen Zufallsvariablen X und Y ist gleich einer *Faltung* oder *Konvolution* der entsprechenden Verteilungsfunktionen:

$$\begin{aligned} F_{X+Y}(x) &= \int_{-\infty}^{\infty} F_X(x-z) dF_Y(z) \\ &= \int_{-\infty}^{\infty} F_Y(x-z) dF_X(z) = F_X(x) * F_Y(x). \end{aligned} \quad (\text{A.2.19})$$

¹⁶ Zur ausführlichen Darstellung siehe Petrov (1975).

4. Existiert das Moment n -ter Ordnung, so ist die c.F. n -mal differenzierbar und

$$\mathbb{E} X^k = \frac{1}{i^k} \psi^{(k)}(t) \Big|_{t=0} \quad \text{für } k \leq n \quad (\text{A.2.20})$$

5. Eine Zufallsvariable und ihre Verteilung heißen *symmetrisch*, wenn die Verteilungen von X und $(-X)$ gleich sind. Die c.F. einer symmetrischen Verteilung ist gleich ihrer komplex konjugierten Funktion; damit ist sie reellwertig:

$$\psi(t) = \mathbb{E} e^{itX} = \mathbb{E} e^{-itX} = \psi(-t) = \overline{\psi(t)}. \quad (\text{A.2.21})$$

Existieren Momente aller Ordnungen, bietet sich neben der c.F. die *momenterzeugende Funktion* an. Sie ist definiert als

$$M(t) \equiv \mathbb{E} e^{tX} = \int_{-\infty}^{\infty} e^{tX} dF(x) = \psi(-it). \quad (\text{A.2.25})$$

Aus der Entwicklung der Funktion e^{tX} in die Taylor-Reihe

$$M(t) = \mathbb{E} e^{tX} = \mathbb{E} \left[1 + \sum_{k=1}^{\infty} \frac{t^k}{k!} X^k \right] = 1 + \sum_{k=1}^{\infty} \frac{t^k}{k!} \mathbb{E} [X^k] \quad (\text{A.2.26})$$

ergeben sich Momente einer Verteilung. Konvergiert die obige Reihe, dann gilt es für alle $k \geq 1$:

$$\mathbb{E} [X^k] = M^{(k)}(0) \equiv \frac{d^k M(t)}{dt^k} \Big|_{t=0}.$$

Für eine positive Zufallsvariable X fällt die Ähnlichkeit mit der *Laplace-Stieltjes Transformation* auf:

$$\mathcal{L}(s) := \int_{-\infty}^{\infty} e^{-sX} dF(x) = \int_0^{\infty} e^{-sX} dF(x).$$

Die Taylor-Reihen-Entwicklung des Logarithmus der momenterzeugenden Funktion führt zu *Kumulanten* C_k von X bzw. zur *kumulantenerzeugenden Funktion*

$$\varphi(t) \equiv \ln M(t) = \sum_{k=1}^{\infty} \frac{t^k}{k!} C_k. \quad (\text{A.2.27})$$

BEISPIEL A.4

Die charakteristische Funktion einer Standardnormalverteilung $N(0, 1)$ lässt sich durch die Umschichtung der Terme im Exponenten berechnen

$$\begin{aligned}\Psi(t) &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-x^2/2} e^{itx} dx \\ &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-(x-it)^2/2} dx e^{-t^2/2} = 1 \cdot e^{-t^2/2} = e^{-t^2/2}. \quad (\text{A.2.22})\end{aligned}$$

Nach der Transformationsregel ergibt sich für $N(\mu, \sigma^2)$ die charakteristische Funktion

$$\Psi(t) = e^{it\mu - \frac{t^2\sigma^2}{2}}. \quad (\text{A.2.23})$$

Die charakteristische Funktion einer Poissonverteilung

$$\mathbb{P}(X = k) = \sum_{k=0}^{\infty} e^{-\lambda} \frac{\lambda^k}{k!}, \quad k = 0, 1, \dots$$

berechnet sich zu

$$\Psi(t) = \sum_{k=0}^{\infty} e^{itk} e^{-\lambda} \frac{\lambda^k}{k!} = e^{-\lambda} \sum_{k=0}^{\infty} \frac{(\lambda e^{it})^k}{k!} = e^{-\lambda} e^{\lambda it} = e^{\lambda(e^{it} - 1)}. \quad (\text{A.2.24})$$

Weitere Beispiele:

Verteilung	Dichte	Bereich	charakteristische Funktion
Gleichverteilung	$\frac{1}{b-a}$	$a \leq x \leq b$	$\frac{1}{b-a} \frac{e^{itb} - e^{ita}}{it}$
Exponentialverteilung	$\lambda e^{-\lambda x}$	$x \geq 0$	$\frac{\lambda}{\lambda - it}$
Gammaverteilung	$\frac{\lambda^\alpha x^{\alpha-1}}{\Gamma(\alpha)} e^{-\lambda x}$	$x \geq 0$	$\left(\frac{\lambda}{\lambda - it}\right)^\alpha$
Cauchyverteilung	$\frac{a}{\pi(a^2 + x^2)}$	$a > 0$	$e^{-a t }$

Die kumulanterzeugende Funktion ist immer eine konvexe Funktion und hat die sehr bequeme Eigenschaft, dass Erwartungswert und Varianz einer Verteilung gerade der ersten bzw. der zweiten Kumulanten entsprechen:

$$\varphi'(0) = C_1 = EX, \quad \varphi''(0) = C_2 = EX^2 - (EX)^2 = \text{Var } X.$$

BEISPIEL A.5 Chernoff-Schranke

Es seien die X_i Exp(1)-verteilt. Die momenterzeugende Funktion errechnet sich zu

$$M(t) = \int_0^{\infty} e^{tx} e^{-x} dx = \frac{1}{1-t}, \quad 0 \leq t < 1.$$

$M(t)$ ist stetig und differenzierbar, deshalb suchen wir nach

$$t_0 = \arg \max_{0 \leq t < 1} [-at + n\varphi(t)] = \arg \max_{0 \leq t < 1} [-at - n \ln(1-t)] = 1 - \frac{n}{a} \quad \text{für } a > n.$$

Beim Einsetzen in (A.2.30) ergibt sich die Chernoff-Schranke

$$P(S_n \geq a) \leq e^{n-a} \left(\frac{a}{n}\right)^n \quad \text{für } a > n.$$

Die Chernoff-Schranke

Mit Hilfe der oben definierten erzeugenden Funktionen lässt sich die Chernoff-Schranke¹⁷ herleiten. Es seien X_1, \dots, X_n i.i.d. Zufallsvariablen mit $M(t) = \mathbb{E} e^{tX_1}$. Wir setzen $S_n \equiv \sum_{i=1}^n X_i$ und interessieren uns für die Wahrscheinlichkeit $\mathbb{P}(S_n \geq na)$ mit $a > 0 = \text{const.}$ Anwendung der Chebyshev-Ungleichung (A.2.6) resultiert in

$$\mathbb{P}(S_n \geq na) \leq \exp(-nat) \mathbb{E} \exp(tS_n) = \exp(-nat) M_n(t). \quad (\text{A.2.28})$$

Offensichtlich gilt in diesem Fall

$$M_n(t) = (M(t))^n \quad \text{sowie} \quad \varphi_n(t) = n\varphi(t)$$

für die moment- bzw. kumulanterzeugende Funktion der Partialsumme S_n . Weiterhin, wegen der Unabhängigkeit der X_i genügt $\varphi(t)$ folgender Identität:

$$\varphi_n(t) = \ln \mathbb{E} \exp(tS_n) = \ln \prod_{k=1}^n \mathbb{E} \exp(tX_k) = \sum_{k=1}^n \ln \mathbb{E} \exp(tX_k) = n\varphi(t). \quad (\text{A.2.29})$$

Damit lässt sich (A.2.28) weiter vereinfachen zu

$$\mathbb{P}(S_n \geq na) \leq \exp(-n(at - \varphi(t))). \quad (\text{A.2.30})$$

Diese Ungleichung gilt für alle t , damit auch für ein solches t , das die Schranke minimiert. Mit $I(a) \equiv \sup_{t \geq 0} [at - \varphi(t)]$ kommen wir schließlich zum Ergebnis

¹⁷ Chernoff (1952).

$$\mathbb{P}(S_n \geq na) \leq \exp(-nI(a)). \quad (\text{A.2.31})$$

Formel (A.2.31) gibt eine exponentielle Schranke für die Summe von i.i.d. Zufallsvariablen an und findet eine breite Anwendung in der Theorie der großen Abweichungen (Kapitel 5).

Konvexe Dualität und Legendre-Fenchel Transformation

Zu den klassischen Ergebnissen nach Rockafellar (1970) zählt die Erkenntnis, dass eine konkave Funktion $f(x)$ durch eine konjugierte oder duale Funktion $f^*(\lambda)$ dargestellt werden kann:

$$f(x) = \min_{\lambda} [\lambda x - f^*(\lambda)], \quad (\text{A.2.32})$$

die ihrerseits an die Originalfunktion gebunden ist

$$f^*(\lambda) = \min_x [\lambda x - f(x)]. \quad (\text{A.2.33})$$

Entsprechend spiegelbildlich gilt diese Beziehung auch für konvexe Funktionen:

$$f(x) = \max_{\lambda} [\lambda x - f^*(\lambda)], \quad (\text{A.2.34})$$

$$f^*(\lambda) = \max_x [\lambda x - f(x)]. \quad (\text{A.2.35})$$

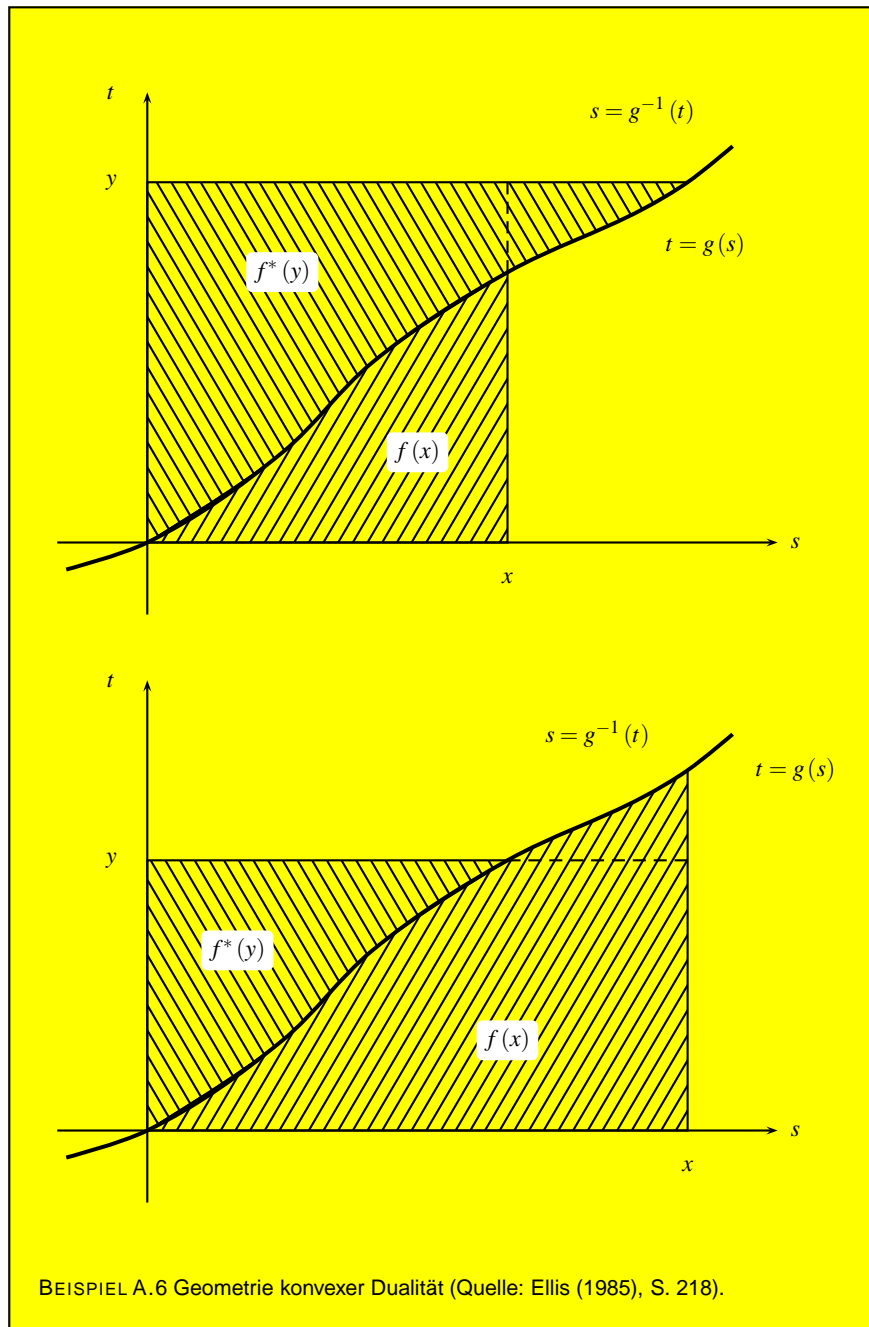
Auf diese Weise wird eine konkave bzw. eine konvexe Funktion als *parametrische Lösung eines Optimierungsproblems* dargestellt.

Definition A.2.2. Sei f eine Funktion auf \mathbb{R} . Die **Legendre-Transformierte** f^* von f ist definiert als

$$f^*(y) = \sup_{x \in \mathbb{R}^n} [xy - f(x)], \quad y \in \mathbb{R}. \quad (\text{A.2.36})$$

Die Legendre-Fenchel- oder einfach Legendre-Transformierte¹⁸ einer Funktion f auf \mathbb{R} ist eine andere Funktion f^* auf \mathbb{R} . Ist die ursprüngli-

¹⁸ Zu detaillierter Darstellung siehe Rockafellar (1970), Roberts und Varberg (1973) (eindimensionaler Fall), Ekeland und Temam (1976), Ellis (1985), Alekseev u.a. (1987).



che Funktion f konvex und unter-halbstetig,¹⁹ dann gilt für die Rücktransformation $f^{**} = f$; die Legendre Transformation ist also invertierbar.²⁰ Unabhängig von der Form der Funktion f ist ihre Legendre-Transformierte f^* eine *konvexe* Funktion. Eine Verallgemeinerung auf \mathbb{R}^n ist möglich. Um eine graphische Darstellung der Legendre Transformation (Beispiel A.6) zu verdeutlichen, bezeichnen wir mit g eine stetige Funktion auf \mathbb{R} derart, dass

$$g(0) = 0, \quad \lim_{x \rightarrow -\infty} g(x) = -\infty, \quad \lim_{x \rightarrow \infty} g(x) = \infty. \quad (\text{A.2.37})$$

Dasselbe gelte auch für die inverse Funktion g^{-1} . Wir definieren zusätzlich

$$f(x) = \int_0^x g(s) \, ds \quad \text{für } x \in \mathbb{R} \quad (\text{A.2.38})$$

$$f^*(y) = \int_0^y g^{-1}(t) \, dt \quad \text{für } y \in \mathbb{R}. \quad (\text{A.2.39})$$

Satz A.2.1 (Ellis (1985), Theorem VI.4.1.) *Es seien f und f^* Funktionen nach (A.2.37), (A.2.38) und (A.2.39). Dann sind diese beiden Funktionen streng konvex und genügen folgenden Eigenschaften:*

- i) $xy \leq f(x) + f^*(y)$ für alle x und y in \mathbb{R} (**Fenchel-Ungleichung**)
- ii) $xy = f(x) + f^*(y)$ genau dann, wenn $y = f'(x)$
- iii) $(f^*)' = (f')^{-1}$
- iv) $f^*(y) = \sup_{x \in \mathbb{R}} [xy - f(x)]$ für $y \in \mathbb{R}$
- v) $f(x) = \sup_{y \in \mathbb{R}} [xy - f^*(y)]$ für $x \in \mathbb{R}$,

wobei mit f^* die Legendre-Fenchel-Transformierte von f bezeichnet wird.

A.3 Grenzwertsätze

Wie lassen sich Wahrscheinlichkeiten, die bekanntlich axiomatischer Natur sind, mit empirisch beobachteten relativen Häufigkeiten in Einklang bringen?

¹⁹ Eine Funktion f heißt *unter-halbstetig* (lower semi-continuous) auf einer Menge $M \subset \mathbb{R}$, falls für alle $x \in M$ und alle Folgen $\{x_n\} \subset M$ mit $x_n \rightarrow x$ gilt: $\liminf_{n \rightarrow \infty} f(x_n) \geq f(x)$. Mit anderen Worten, eine Funktion $f(x)$ ist unter-halbstetig, wenn sie an ihren Sprungstellen den *unteren* Wert als Funktionswert annimmt.

²⁰ Rockafellar (1970), S. 104.

Den ganzen Komplex von Fragestellungen, bei denen aus beobachteten Ereignissen Rückschlüsse auf unbeobachtete (vornehmlich zukünftige) Zustände gezogen werden, fasst man unter dem Begriff *statistische Inferenz* zusammen. Es ist eine empirische Tatsache, dass wir aus der Erfahrung lernen können. Andererseits scheint nichts, was im Zusammenhang mit der Zukunft liegt, sich *logisch* aus der Erfahrung der Vergangenheit ableiten zu lassen. Die Brücke zwischen Wahrscheinlichkeitstheorie und Statistik wird u.a. mittels sogenannter Grenzwertsätze hergestellt. Wir verwenden dieses Wort als Sammelbegriff für mehrere Formulierungen, beginnend bei Bernoulli, dann Borel, Chebyshev, Khinchin, Kolmogorov.

Im Grunde geht es darum, wie sich die geeignet normierte Summe $S_n = X_1 + \dots + X_n$ von i.i.d. Zufallsvariablen X_k bezüglich Kennzahlen wie Erwartungswert und/oder Varianz verhält. Was ist unter geeigneter Normierung zu verstehen? In allgemeiner Form lässt sich der Normierungsfaktor als n^α schreiben (mit n : Anzahl der Summanden, α : positive Konstante).

Für $\alpha = 1$ bekommen wir *Gesetze der großen Zahlen* jeweils in schwacher (Chebyshev) oder starker (Kolmogorov) Form. Sie besagen, dass bei steigender Anzahl der Summanden die Information über die genaue Form der Wahrscheinlichkeitsverteilung verloren geht; das einzige, was übrigbleibt, ist der Mittelwert. Die gemeinsame Idee beider Versionen dieses Gesetzes besteht darin, dass der Stichprobenmittelwert nicht weit vom Mittelwert der Population liegen kann.

Satz A.3.1 (schwaches Gesetz der großen Zahlen) Seien X_1, X_2, \dots i.i.d. Zufallsvariablen mit $\mathbb{E}X_1 \equiv m$, $\text{Var}X_1 \equiv \sigma^2$ und $S_n = X_1 + \dots + X_n$. Dann gilt

$$\lim_{n \rightarrow \infty} \mathbb{P} \left[\left| \frac{S_n}{n} - m \right| > \delta \right] = 0 \quad (\text{A.3.1})$$

für jedes $\delta > 0$.

Mit steigendem Umfang der Stichprobe konvergiert deren Mittelwert in Wahrscheinlichkeit gegen den Erwartungswert. Der Beweis erfolgt durch Anwendung der Chebyshev-Ungleichung

$$\mathbb{P} \left[\left| \frac{S_n}{n} - m \right| > \delta \right] \leq \frac{1}{\delta^2} \text{Var} \left(\frac{S_n}{n} \right) = \frac{\sigma^2}{n\delta^2}.$$

In dieser schwachen Form stellt das Gesetz der großen Zahlen nur eine sehr grobe Abschätzung dar. Im Jahre 1926 hat A. N. Kolmogorov eine strengere Form bewiesen. Danach ist die Existenz des Erwartungswertes für eine fast sichere Konvergenz an Stelle der Konvergenz in Wahrscheinlichkeit notwendig und hinreichend.

Satz A.3.2 (starkes Gesetz der großen Zahlen)

1. Es seien X_1, X_2, \dots i.i.d. Zufallsvariablen mit endlichem Erwartungswert $\mathbb{E}X_1 = m$ und $S_n = X_1 + \dots + X_n$. Dann gilt

$$\mathbb{P} \left[\lim_{n \rightarrow \infty} \frac{S_n}{n} = m \right] = 1. \quad (\text{A.3.2})$$

2. Umgekehrung. Sind X_1, X_2, \dots i.i.d. derart, dass (A.3.2) für einige $m \in \mathbb{R}$ hält, dann existiert der Erwartungswert $\mathbb{E}X_1$ und er ist gleich m .

Beweis: Bening und Korolev (2002), Theorem 1.6.2. □

Im Unterschied zu den Gesetzen der großen Zahlen operiert der *zentrale Grenzwertsatz* mit dem Normierungsfaktor $n^{1/2}$ und wird in der Literatur oft als Versuch interpretiert, die Frage nach der Konvergenzrate der Summe gegen den Mittelwert zu beantworten. Warum ist das so wichtig? Die Konvergenzrate gibt an, wie genau die Approximation des Erwartungswertes durch das arithmetische Mittel der i.i.d. Zufallsvariablen ist. Die Idee hinter dem zentralen Grenzwertsatz ist sehr alt und stammt von Laplace: die Summe einer großen Anzahl zufälliger Einflüsse, von denen jeder im Vergleich zur Summe der übrigen klein ist, folgt einem allgemeinen Gesetz, und dieses Gesetz ist die Normalverteilung.

Satz A.3.3 (Zentraler Grenzwertsatz) Es seien X_1, X_2, \dots i.i.d. Zufallsvariablen mit $\mathbb{E}X_1 = m$ und $\text{Var}X_1 = \sigma^2 < \infty$. Dann konvergiert die Verteilung der normierten und skalierten Summe $S_n = X_1 + \dots + X_n$ schwach gegen die Standardnormalverteilung $N(0, 1)$

$$\lim_{n \rightarrow \infty} \mathbb{P} \left[\frac{S_n - nm}{\sigma\sqrt{n}} < x \right] = \Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x e^{-u^2/2} du, \quad x \in \mathbb{R}. \quad (\text{A.3.3})$$

Die Idee hinter dem Satz ist, dass eine große Anzahl von i.i.d. Zufallsvariablen in deren Zusammensetzung eine Standardnormalverteilung

nachbildet und zwar unabhängig davon, wie die einzelnen Summanden verteilt sind. Insbesondere wird nicht verlangt, dass sie normalverteilt sind. Die Frage nach der Konvergenzrate scheint damit auch beantwortet zu sein. Für endliche Varianz ergibt sich durch elementare Umformung die Konvergenzrate zu $n^{-1/2}$:

$$\left(\frac{S_n - nm}{\sigma\sqrt{n}} \right) = \frac{\sqrt{n}}{\sigma} \left(\frac{S_n}{n} - m \right).$$

Eine *hinreichende* Bedingung für die Gültigkeit des zentralen Grenzwertsatzes ist in der Literatur²¹ als *Lindeberg-Feller Bedingung* bekannt. Sie erfasst mathematisch die Anforderung nach der Geringfügigkeit des Beitrages von jedem einzelnen Summanden relativ zur Summe. Ist diese Bedingung verletzt, ergeben sich interessante Konsequenzen. Zu welcher Verteilung können beispielsweise normalisierte Summen von i.i.d. Zufallsvariablen konvergieren unter der Annahme, dass die Größenordnung der Summe allein von wenigen Summanden entscheidend beeinflusst werden kann? Ist die Konvergenz gegeben, dann kommt nur die Klasse *stabiler Verteilungen* in Frage.²²

A.4 Stochastische Prozesse

In diesem Abschnitt geben wir eine kompakte Übersicht der wichtigsten Eigenschaften stochastischer Prozesse. Insbesondere sind Martingale bzw. Semimartingale als mathematische Grundlage für die Modellierung des Aktienkursverlaufs von Bedeutung. In dieser Arbeit werden Prozesse betrachtet, die keine Semimartingale sind. Damit die Wichtigkeit der Semimartingal-Eigenschaft deutlicher wird, gehen wir auf die Eigenschaften der stochastischen Integrale näher ein.

Standardreferenzen sind Jacod und Shiryaev (1987), Karatzas und Shreve (1988) und Protter (1990).²³

²¹ Bening und Korolev (2002), S. 41 Theorem 1.6.8.

²² Vgl. Kapitel A.5.

²³ Sehr hilfreich sind ebenfalls Borovskikh und Korolyuk (1997) sowie umfangreiche Ressourcen auf der Homepage von Jan Kallsen www.stochastik.uni-freiburg.de/homepages/kallsen/start.html

Grundlagen

Definition A.4.1 (stochastischer Prozess). Ein stochastischer Prozess ist eine Familie von Zufallsvariablen $X(t) = X(t, \omega)$ auf einem Wahrscheinlichkeitsraum $(\Omega, \mathcal{F}, \mathbb{P})$.

Alternativ werden die Bezeichnungen $\{X_t, t \in \mathcal{T}\}$ oder X_t verwendet. Der Parameter t hat gewöhnlich die Bedeutung *Zeit* und stammt aus einer Parametermenge \mathcal{T} . Diese Zeitindexmenge \mathcal{T} kann *diskret* $\mathcal{T} := \{0, 1, 2, \dots\}$ oder *kontinuierlich* $\mathcal{T} = \mathbb{R}_+ := [0, \infty)$ gewählt werden; dementsprechend wird auch der Prozess klassifiziert.

Stochastische Prozesse können auf unterschiedliche Weise interpretiert werden. Betrachtet man den Prozess $X(t, \omega)$ als eine Zufallsfunktion X auf dem Produktraum $\mathcal{T} \times \Omega$, dann handelt es sich um eine ganze Familie von Zufallsvariablen.²⁴

Für ein fixes $\omega \in \Omega$ reduziert sich ein Prozess zu einer Zeitfunktion oder einer Familie der Zeitfunktionen zusammen mit einem Wahrscheinlichkeitsmaß. Unter X_t stellt man sich dann den „Zustand“ oder die „Position“ des Prozesses zum Zeitpunkt t vor. Desweiteren kann $X(\cdot, \omega)$ als eine Funktionsmenge auf \mathcal{T} interpretiert werden, je eine Funktion per ω . Jede Funktion $t \mapsto X(t, \omega)$ heißt *Pfad (sample path)*, *Trajektorie* oder *Realisierung* des stochastischen Prozesses.

Bei der Modellierung der Unsicherheit auf dem Aktienmarkt ist insbesondere ein bestimmter Prozesstyp, der sogenannte *càdlàg (continu à droite, limité à gauche)* oder *RCLL (right continuous, left limit)* Prozess von Interesse. Wir bezeichnen mit $X_- := \{X_{t-}\}$ einen linksseitigen Grenzwertprozess ($X_{0-} \equiv X_0$) und mit $\Delta X := \{\Delta X_t\}$ einen Sprungprozess von X :

$$\Delta X_t := X_t - X_{t-}, \quad t \geq 0.$$

Definition A.4.2 (äquivalente Prozesse). Zwei stochastische Prozesse X_t und Y_t auf dem gleichen Wahrscheinlichkeitsraum $(\Omega, \mathcal{F}, \mathbb{P})$ sind *äquivalent*, falls $\mathbb{P}(X_t = Y_t) = 1$ für jedes $t \geq 0$ gilt.

Äquivalente Prozesse heißen auch *Modifikationen* oder *Versionen* voneinander. Sie müssen keine identischen Pfade haben.

²⁴ Pollard (2002), S. 212.

BEISPIEL A.7 Brownsche Bewegung

Sei $\mathcal{T} = \mathbb{R}_+$ eine stetige Zeitindexmenge auf $(\Omega, \mathcal{F}, \mathbb{P})$.

Definition A.4.3. Ein reellwertiger stochastischer Prozess $\{W_t, t \in \mathcal{T}\}$ mit $W_0 = 0$ \mathbb{P} -f. s. und

- (i) stationären und unabhängigen Zuwächsen $[W_t - W_s]$,
- (ii) normalverteilten Zuwächsen $[W_t - W_s] \sim N(0, t - s)$,
- (iii) \mathbb{P} -f. s. stetigen Pfaden $t \mapsto B(\omega, t)$ für alle $\omega \in \Omega$

heißt **Brownsche Bewegung** oder **Wiener Prozess**.

Bemerkung. Eigenschaft (iii) kann *definiert* oder *bewiesen* werden. Der erste Beweis dafür, dass eine Brownsche Bewegung ein *stetiger* Prozess ist, wurde im Jahr 1923 von Norbert Wiener erbracht. Gewöhnlich wird dafür das *Kolmogorov'sche Kontinuitätskriterium* (Kloeden und Platen (1992), S. 39) verwendet.

Definition A.4.4 (ununterscheidbare Prozesse). Zwei Prozesse X_t und Y_t sind \mathbb{P} -ununterscheidbar, falls $X(\cdot, \omega) = Y(\cdot, \omega)$ für alle $\omega \in \Omega \setminus N$ mit $\mathbb{P}(N) = 0$. Das heißt, abgesehen von einer \mathbb{P} -Nullmenge haben X_t und Y_t identische Pfade.

Zwei ununterscheidbare Prozesse sind äquivalent. Zwei RCLL-Versionen sind ununterscheidbar. Manchmal ist der Wahrscheinlichkeitsraum $(\Omega, \mathcal{F}, \mathbb{P})$ nicht gegeben, so dass man bei der Konstruktion eines stochastischen Prozesses auf die endlich-dimensionalen Verteilungen angewiesen ist.

Definition A.4.5 (endlich-dimensionale Verteilungen). Die gemeinsamen Verteilungen der Zufallsvektoren $\{X_t, t \in S\}$, wobei S über alle endliche Untermengen von \mathcal{T} variiert, nennt man *endlich-dimensionale Verteilungen des Prozesses*.

Nach einem bekannten Satz von Kolmogorov²⁵ ist ein stochastischer Prozess unter bestimmten Konsistenzbedingungen durch ein System seiner endlich-dimensionalen Verteilungen vollständig spezifiziert. Dabei werden alle Modifikationen eines Prozesses durch dasselbe System von endlich-dimensionalen Verteilungen charakterisiert. Endlich-dimensionale Verteilungen eines *Gauss'schen Prozesses* sind mehrdimensionale Normalverteilungen. Ein stochastischer Prozess $\{X_t, t \geq 0\}$, dessen endlich-dimensionale Verteilungen ein endliches

²⁵ Bening und Korolev (2002), Theorem 1.10.1, S. 63.

BEISPIEL A.8 Poisson Prozess

Die Stärke eines Wiener Prozesses liegt bei der Modellierung zufälliger *Bewegungen*. Ein Poisson Prozess eignet sich dagegen eher zur Modellierung zufälliger *Beobachtungen* N_t bis zum Zeitpunkt t . Gewöhnlich ist N ein *Zählprozess*: Beobachtungen werden nur einmal pro Zeitpunkt registriert. Mit Ausnahme von Sprüngen der Höhe eins ist N eine Konstante. Für $s < t$ kann $N_t - N_s$ als Anzahl der Beobachtungen im Zeitintervall $(s, t]$ interpretiert werden.

Definition A.4.6. Ein Poisson Prozess $\{N_t, t \geq 0\}$ ist durch folgende Eigenschaften definiert:

- (i) $\lambda(t) := \mathbb{E} N_t$ ist eine stetige Funktion für alle $t \in \mathbb{R}_+$ und heißt Intensität von N ;
- (ii) Zuwächse $[N_t - N_s]$ sind unabhängig für alle $s < t$ und sind Poisson-verteilt:

$$\mathbb{P}\{N_t - N_s = n\} = e^{-(\lambda(t) - \lambda(s))} \frac{(\lambda(t) - \lambda(s))^n}{n!}, \quad n \geq 0.$$

Ein homogener Poisson Prozess ist definiert durch seine Intensität $\lambda(t) = \lambda t$ ($\lambda > 0$).

zweites Moment $\mathbb{E} X_t^2 < \infty$ für alle $t \in \mathcal{T}$ aufweisen, heißt *Prozess zweiter Ordnung (second order, L^2)*. Zur Spezifikation eines solchen Prozesses werden lediglich zwei Kennzahlen benötigt: der Erwartungswert und die Autokovarianz.

$$m_X(t) = \mathbb{E} X_t = \mathbb{E}[X(t)], \quad \sigma_X^2(t) = \text{Var} X_t = R_X(t, t), \quad (\text{A.4.1})$$

$$R_X(s, t) = \text{Cov}(X_s, X_t) = \mathbb{E}\left[(X_s - m_X(s))(X_t - m_X(t))\right]. \quad (\text{A.4.2})$$

Definition A.4.7. Ein stochastischer Prozess $\{X_t, t \in \mathcal{T} \in \mathbb{Z}\}$ heißt *Prozess mit unabhängigen Zuwächsen*, falls für alle $n \geq 1$ und $t_i \in \mathcal{T}$, $i = 0, \dots, n$ mit $t_0 < t_1 < \dots < t_n$ die Zufallsvariablen $X_{t_0}, X_{t_1} - X_{t_0}, \dots, X_{t_n} - X_{t_{n-1}}$ unabhängig sind.

Definition A.4.8. Ein Prozess mit unabhängigen Zuwächsen heißt **homogen** oder **stationär**, falls für alle $s < t$ die Verteilung von $X(t) - X(s)$ ausschließlich von $(t - s)$ abhängig ist. Für einen L^2 -Prozess ist dies äquivalent zu

$$m_X(t) = m, \quad R_X(s, t) = R_X(s + h, t + h)$$

für alle $s, t, h \in \mathbb{Z}$.

Durch die Festlegung $\gamma_X(h) = R_X(h, 0)$ schreibt sich die Autokovarianzfunktion $R_X(s, t) = \gamma_X(s - t)$. Die Funktion $\gamma_X(h)$ ist symmetrisch bezüglich h und es gilt $\gamma_X(0) = \text{Var}(X_t)$. Mittelwert und Varianz eines stationären Prozesses sind konstant.

Die Filtration

Definition A.4.9 (Filtration). Eine Filtration $\mathbb{F} = \{\mathcal{F}_t, t \in \mathcal{T}\}$ ist eine Familie von Sub- σ -Algebren mit $\mathcal{F}_s \subseteq \mathcal{F}_t$ für alle $s < t$. Eine Filtration \mathbb{F} erfüllt die üblichen Bedingungen, falls

- (i) \mathbb{F} rechtsseitig-stetig ist: $\mathcal{F}_t = \mathcal{F}_{t+} := \bigcap_{s>t} \mathcal{F}_s$;
- (ii) \mathcal{F}_0 alle \mathbb{P} -Nullmengen aus \mathcal{F} enthält;
- (iii) $\mathcal{F}_\infty := \mathcal{F}$.

Eine solche verschachtelte Struktur eignet sich sehr gut dazu, einen Informationsfluss abzubilden. Sie wird oft von einem stochastischen Prozess, zum Beispiel einem Aktienkursprozess, erzeugt. Zur Filtration \mathcal{F}_t gehören alle Vergangenheitswerte oder alle verfügbare Informationen bis zum Zeitpunkt t . Die Messbarkeit einer Funktion X_t auf Ω bezüglich \mathcal{F}_t bedeutet, dass der Funktionswert X_t in t bekannt ist.

Bei einer alternativen Betrachtung ist eine steigende Familie von Sub- σ -Algebren gegeben. Man verlangt dagegen für einen stochastischen Prozess $\{X_t, t \geq 0\}$ auf dem Wahrscheinlichkeitsraum $(\Omega, \mathcal{F}, \mathbb{P})$, für jedes $t \geq 0$ \mathcal{F}_t -messbar zu sein. Ein solcher Prozess X_t heißt **adaptiert** zur Familie $\{\mathcal{F}_t, t \in \mathcal{T}\}$. Dabei ist X_t zum jeden Zeitpunkt t eine Zufallsvariable auf (Ω, \mathcal{F}) .

Die durch eine gewöhnliche Brownsche Bewegung erzeugte natürliche oder kanonische Filtration $\mathcal{F}_t^B := \sigma\{W_s : s \leq t\}$ erscheint intuitiv ein plausibles Modell für die in t verfügbare Information zu sein. Aus technischen Gründen wird aber eine Standard-Filtration $\mathbb{F} = \{\mathcal{F}_t, t \in \mathcal{T}\}$ definiert, die die üblichen Bedingungen erfüllt.

Ein gefilterter Wahrscheinlichkeitsraum $(\Omega, \mathcal{F}, \mathbb{F}, \mathbb{P})$ heißt **stochastische Basis**.

Die Stoppzeit

Es ist relativ einfach, sich einen stochastischen Prozess als Funktion zweier Variablen $(\omega, t) \mapsto X(\omega, t)$, $\omega \in \Omega$, $t \in \mathcal{T}$ vorzustellen. Etwas weniger intuitiv ist die Erkenntnis, dass der Zeitpunkt, zu dem ein fixes ω realisiert wird, rein zufällig ist.

BEISPIEL A.9 Chung (1974), S. 18.

Sei $(\Omega, \mathcal{F}, \mathbb{F}, \mathbb{P})$ eine stochastische Basis, X ein RCLL-Prozess und $B \in \mathcal{B}(\mathbb{R})$ eine Borel-Menge. Dann ist die Zufallsvariable

$$T_B(\omega) = \inf \{0 \leq t < \infty : X(\omega, t) \in B\} \quad (\text{A.4.3})$$

eine Stoppzeit, die auch als *hitting time* bezeichnet wird.

Definition A.4.10. Eine Zufallsvariable $\tau = \tau(\omega) \in \mathcal{T} \cup \{\infty\}$ auf der Basis $(\Omega, \mathcal{F}, \mathbb{F}, \mathbb{P})$ heißt **Stoppzeit**, falls das Ereignis $\{\omega : \tau(\omega) \leq t\}$ für alle $t \in \mathcal{T}$ in \mathcal{F}_t ist.

Der Grundgedanke dabei ist, dass allein aus der Beobachtung eines fixen ω_0 bestimmt werden kann, ob es sich dabei um die Vergangenheit oder um die Zukunft handelt.

Es sei X ein \mathbb{F} -adaptierter Prozess, τ eine Stoppzeit und X^τ ein in τ gestoppter Prozess. Dann gilt

$$\begin{cases} X_t^\tau = X_t & \text{für } t \leq \tau, \\ X_t^\tau = X_\tau & \text{für } t > \tau \end{cases} \quad (\text{A.4.4})$$

oder $X_t^\tau = X_{t \wedge \tau} = X_t 1_{(t < \tau)} + X_\tau 1_{(t \geq \tau)}$, wobei $t \wedge \tau := \min\{t, \tau\}$. Mit anderen Worten, bei der Entscheidung, einen Prozess in τ zu stoppen, darf ausschließlich die in \mathcal{F}_τ enthaltene Information verwendet werden.

Martingale und Semimartingale

Sei $(\Omega, \mathcal{F}, \mathbb{F}, \mathbb{P})$ eine stochastische Basis.

Definition A.4.11 (Martingal). Ein reellwertiger stochastischer Prozess $\{M_t, t \in \mathcal{T}\}$ heißt **Martingal** (bzw. **Sub-** oder **Supermartingal**), falls

- (i) M_t \mathbb{F} -adaptiert ist;
- (ii) $E |M_t| < \infty$ für alle $t \in \mathcal{T}$;
- (iii) $E[M_t | \mathcal{F}_s] = M_s$ für alle $s < t$ (Martingal)
 $E[M_t | \mathcal{F}_s] \geq M_s$ (Submartingal)
 $E[M_t | \mathcal{F}_s] \leq M_s$ (Supermartingal).

BEISPIEL A.10

Es ist leicht zu sehen, dass die gewöhnliche Brownsche Bewegung ein stetiges L^2 -Martingal ist. Aus Messbarkeit von W_s bezüglich \mathcal{F}_s und Unabhängigkeit der Zuwächse folgt für $s \leq t$

$$E[W_t | \mathcal{F}_s] = E[(W_t - W_s) | \mathcal{F}_s] + E[W_s | \mathcal{F}_s] = E[W_t - W_s] + E[W_s | \mathcal{F}_s] = W_s. \quad (\text{A.4.5})$$

BEISPIEL A.11 **Waldsches Martingal** (Hesse (2003), S. 304.)

Sei $(X_n)_{n \in \mathbb{N}}$ eine i.i.d. Folge mit momenterzeugender Funktion

$$M(t) = E \exp(tX_1) < \infty$$

für ein $t \neq 0$ und Partialsummenfolge $S_n = \sum_{i=1}^n X_i$, $n \in \mathbb{N}$. Die Folge der Zufallsvariablen $(Z_n)_{n \in \mathbb{N}}$ mit $Z_n = \exp(tS_n) / M^n(t)$ für alle $n \in \mathbb{N}$ bildet ein Martingal. Wir prüfen

$$\begin{aligned} E[Z_n] &= E[\exp(tX_1) \cdots \exp(tX_n)] / M^n(t) \\ &= E \exp(tX_1) \cdots E \exp(tX_n) / [E \exp(tX_1)]^n = 1, \quad \forall n \in \mathbb{N} \end{aligned}$$

und

$$\begin{aligned} E[Z_{n+1} | Z_1, \dots, Z_n] &= E[\exp(tS_n) \cdot \exp(tX_{n+1}) / M^{n+1}(t) | S_1, \dots, S_n] \\ &= \exp(tS_n) E[\exp(tX_{n+1}) | X_1, \dots, X_n] / M^{n+1}(t) = \exp(tS_n) / M^n(t), \quad \forall n \in \mathbb{N}. \end{aligned}$$

BEISPIEL A.12

Sei $X = \{X_t, t \in \mathcal{T}\}$ ein Martingal (Submartingal) und $f(x)$ eine konvexe (konvexe und nichtfallende) Funktion mit $E|f(X_t)| < \infty$ für alle $t \in \mathcal{T}$. Dann ist

$$\{f(X_t), \mathcal{F}_t, t \in \mathcal{T}\}$$

ein Submartingal.

Das Martingal ist ein mathematischer Ausdruck für ein faires Glücksspiel. Der bedingte Erwartungswert des Gewinnes im nächsten Spiel ist sein derzeitiger Wert. Die Martingaltheorie ist elegant und mächtig; sie ist essentiell bei der Bewertung von Finanztiteln. Die Stärke der Martingaltheorie wird durch folgende Erkenntnis besonders deutlich: viele Gleichungen bzw. Ungleichungen erhalten ihre Gültigkeit, wenn die Zeitvariablen durch die Stoppzeiten ersetzt werden.

Definition A.4.12 (Markov-Eigenschaft). Ein \mathbb{F} -adaptierter Prozess X heißt Markov Prozess bezüglich \mathbb{F} , falls für alle $s, t \geq 0$ und jede beschränkte messbare Funktion F gilt:

$$E[F(X_{t+s}) | \mathcal{F}_t] = E[F(X_{t+s}) | X_t] \quad (\text{A.4.6})$$

Mit Worten: eine Vorhersage der Zukunft auf Grundlage der Gegenwart kann durch die zusätzliche Kenntnis der Vergangenheit nicht verbessert werden.

Definition A.4.13 (lokales Martingal). Ein adaptierter Prozess $M = \{M_t, \mathcal{F}_t, t \in \mathcal{T}\}$ heißt lokales Martingal, falls eine steigende Sequenz von Stoppzeiten (oder „lokalisierende Folge“) $\tau_n \uparrow \infty$ existiert, so dass der gestoppte Prozess $M^{\tau_n} = \{M_{t \wedge \tau_n}, \mathcal{F}_t, t \in \mathcal{T}\}$ für jedes $n \geq 1$ ein Martingal ist.

Jedes Martingal ist ein lokales Martingal; es gibt aber lokale Martingale, die keine Martingale sind.²⁶

Definition A.4.14. Eine σ -Algebra \mathcal{P} , die durch einen linksseitig-stetigen adaptierten Prozess erzeugt ist, heißt **vorhersehbarer σ -Algebra**. Ein \mathcal{P} -messbarer Prozess ist **vorhersehbar**.

Die Vorhersehbarkeit ist eine Eigenschaft, die als Gegensatz zur Martingaleigenschaft interpretiert werden kann. Die Stetigkeit von links impliziert, dass der Zustand *jetzt* schon *unmittelbar zuvor* mit Sicherheit bekannt war. Ein vorhersehbarer Prozess ist aber nicht notwendig linksseitig-stetig; er muss lediglich von einer vorhersehbaren σ -Algebra abhängen²⁷.

Ein Prozess, der die Anzahl der Aktien in einer Handelsstrategie beschreibt, *muss* vorhersehbar sein, da die Entscheidung über die Portfoliozusammensetzung *vor* einer möglichen Preisänderung getroffen wird.

Definition A.4.15. Ein stochastischer Prozess $A = \{A_t, \mathcal{F}_t, t \in \mathcal{T}\}$ heißt **steigender Prozess**, falls $A_0 = 0$ für alle $\omega \in \Omega$ und die Trajektorie $A_t(\omega)$, $t \in \mathcal{T}$ eine nicht-negative steigende Funktion ist.

Ein stochastischer Prozess $A = \{A_t, \mathcal{F}_t, t \in \mathcal{T}\}$ von **beschränkter Variation** ist durch Differenz zweier steigender Prozesse darstellbar:

$$A_t = A_t^+ - A_t^-.$$

Der **Variationsprozess** für einen gegebenen Prozess A ist

$$\text{Var}(A_t) := A_t^+ + A_t^-.$$

²⁶ Zu den Integritätsbedingungen damit ein lokales Martingal ein „echtes“ Martingal ist siehe Protter (1990).

²⁷ Duffie (1988), S. 140.

Satz A.4.1 (Doob–Meyer Zerlegung) *Es sei $X = \{X_t, \mathcal{F}_t, t \in \mathbb{R}_+\}$ ein RCLL Submartingal. Dann ist durch ein lokales Martingal M und einen vorhersehbaren steigenden Prozess $A = \{A_t, \mathcal{F}_t, t \in \mathbb{R}_+\}$ mit $A_0 = 0$ die folgende Zerlegung gegeben:*

$$X = M + A. \quad (\text{A.4.7})$$

Diese Zerlegung ist ein Standardsatz in Mathematical Finance und wird hier nicht bewiesen. Sie ist eindeutig; die Prozesse auf der rechten und linken Seite der Gleichung (A.4.7) sind ununterscheidbar. Der Prozess A_t ist \mathcal{F}_{t-1} messbar. Beweis kann in Karatzas und Shreve (1988) bzw. Protter (1990) gefunden werden.

Es sei $M = \{M_t, t \in \mathbb{R}_+\}$ ein *quadratisch integrierbares* Martingal, d.h. $E \sup_{t \geq 0} M_t^2 < \infty$. Dann ist $M^2 = \{M_t^2, t \in \mathbb{R}_+\}$ ein Submartingal. Folglich existiert nach (A.4.7) ein vorhersehbarer steigender Prozess $\langle M \rangle = \{\langle M_t \rangle, t \in \mathbb{R}_+\}$ derart, dass der *kompensierte* Prozess

$$m = M^2 - \langle M \rangle \quad (\text{A.4.8})$$

ein lokales Martingal ist. Im allgemeinen gibt es für alle quadratisch integrierbare Martingale M, N einen eindeutigen vorhersehbaren Prozess $\langle M, N \rangle$ derart, dass

$$MN - \langle M, N \rangle \quad (\text{A.4.9})$$

ein lokales Martingal ist.

Definition A.4.16. *Ein vorhersehbarer Prozess $\langle M, N \rangle$ in (A.4.9) bzw. $\langle M \rangle$ in (A.4.8) heißt **quadratische Kovariation** bzw. **quadratische Variation** oder **Kompensator**.*

Jedes lokale Martingal M , $M_0 = 0$ lässt eine eindeutige Zerlegung $M = M^c + M^d$ zu, wobei M^c ein stetiges und M^d ein rein unstetiges lokales Martingal ist. Im Vergleich mit (A.4.7) heißt $X^c := M^c$ *stetiger Martingalanteil* von X .

Definition A.4.17. *Ein **Semimartingal** $\{X_t, \mathcal{F}_t, t \in \mathbb{R}_+\}$ ist durch seine Zerlegung*

$$X = M + A \quad (\text{A.4.10})$$

definiert. Dabei ist M ein lokales Martingal, A ein adaptierter Prozess mit beschränkter Variation und $A_0 = 0$ in $t = 0$. Der Prozess A kann auch

BEISPIEL A.13

Die quadratische Variation einer Brownschen Bewegung $W = \{W_t, t \in \mathbb{R}_+\}$ mit $E W_t^2 = t$ ist $\langle W \rangle = t$. Daher ist $(W_t^2 - t)$ ein Martingal.

vorhersehbar gewählt werden. In diesem Fall heißt X **Spezialsemimartingal**, die Zerlegung (A.4.10) heißt **kanonische Zerlegung** und sie ist eindeutig.

Semimartingale bilden die breiteste Klasse stochastischer Prozesse, die als gut-definierte Integrator-Prozesse zur Definition eines stochastischen Integrals vernünftigerweise verwendet werden können.

Die Variation

Sei f eine Funktion auf $A \subseteq \mathbb{R}$. Die *Variation* von f auf A ist definiert als

$$V(f; A) := \sup \sum_{j=1}^n |f(x_j) - f(x_{j-1})|,$$

wobei das Supremum über alle endliche Sequenzen $x_0 \leq x_1 \leq \dots \leq x_n$ in A genommen wird. Ist $V(f; A)$ endlich, dann hat f endliche Grenzwerte links und rechts von jedem internen Punkt in A . Da die Anzahl der Subintervalle n ebenfalls endlich ist, kann f höchstens endlich viele unstetige Punkte haben. Oft genügt Stetigkeit von rechts völlig aus, um zusammen mit der Bedingung $V(f; A) < \infty$ eine *Funktion lokal beschränkter Variation* definieren zu können. Das Vorgehen ist ähnlich, wenn f durch einen stochastischen Prozess ersetzt wird.

Wir teilen ein gegebenes Zeitintervall $[0, T]$ in n Subintervalle

$$0 = t_1 < t_2 < \dots < t_k < \dots < t_n = T$$

auf und definieren

$$\delta = \max_k (t_{k+1} - t_k) \rightarrow 0 \quad \text{mit} \quad n \rightarrow \infty.$$

Mit *Variation* $V(X)$ wird das Supremum

$$V(X) = \sup \sum_{k=1}^{n-1} |X(t_{k+1}) - X(t_k)| \quad (\text{A.4.11})$$

über alle solche Partitionen bezeichnet. Ein Prozess X_t mit $V(X) < \infty$ heißt Prozess mit *beschränkter Variation* oder *VF-Prozess (Variation Finité)*. Die Variation eines Wiener Prozesses W_t ist fast sicher unendlich. Der Grenzwert

$$[W]_t = \lim_{\delta \rightarrow 0} \sum_{k=1}^{n-1} (W_{t_{k+1}} - W_{t_k})^2 \quad \text{in Wahrscheinlichkeit} \quad (\text{A.4.12})$$

für einen Wiener Prozess ist dagegen endlich und heißt *quadratische Variation*. Für einen VF-Prozess Y_t ist die quadratische Variation wie folgt definiert:

$$[Y]_t = \sum_{s \leq t} (Y_s - Y_{s-})^2 = \sum_{s \leq t} \Delta Y_s^2. \quad (\text{A.4.13})$$

Die Summe in (A.4.13) kann lediglich über alle Unstetigkeitspunkte berechnet werden. Dabei bezeichnet $\Delta Y_s = Y_s - Y_{s-} = Y - Y_-$ einen Sprung in Y zum Zeitpunkt s .

Satz A.4.2 (i) Ist M ein lokales Martingal, dann existiert $[M]_t$ und ist rechtsseitig-stetig. (ii) Ist M ein quadratisch integrierbares Martingal, dann existiert der Grenzwert

$$\lim_{\delta \rightarrow 0} \sum (M_{t+1} - M_t)^2$$

in L^1 und es gilt: $\mathbb{E} M_t^2 = \mathbb{E} [M]_t$.

Beweis: Ethier und Kurtz (1986), Proposition 2.3.4. □

Dieser Satz weist also die Existenz der quadratischen Variation eines Martingals nach und besagt, dass ihre Berechnung auf die des zweiten Momentes zurückgeführt werden kann.

Für ein stetiges lokales Martingal X schreibt man gewöhnlich $[X] = \langle X \rangle$. Ein klassisches Ergebnis stochastischer Analyse besagt, dass die quadratische Variation eines Wiener Prozesses eine endliche und, wegen willkürlicher Partitionierung, eine pfadunabhängige Größe

$$\langle W \rangle_t = t \quad (\text{A.4.14})$$

ist. Damit hat der Wiener Prozess eine unbeschränkte Variation, aber eine beschränkte quadratische Variation. Dies lässt sich durch Faktorisierung der quadratischen Variation auch leicht zeigen.²⁸ Aus (A.4.12) folgt

²⁸ Kloeden und Platen (1992), S. 74. Gewöhnlich wird eine dyadische Partitionierung $m = 2^n$ unternommen.

$$t \leq \limsup_{n \rightarrow \infty} \max_{k \geq 0} |W_{t_{k+1}} - W_{t_k}| \cdot \sum_{k=1}^{n-1} |W_{t_{k+1}} - W_{t_k}|, \quad (\text{A.4.15})$$

woraus wegen der Pfadstetigkeit

$$\max_{k \geq 0} |W_{t_{k+1}} - W_{t_k}| \rightarrow 0 \quad \text{fast sicher mit } n \rightarrow \infty$$

resultiert

$$\sum_{k=1}^{n-1} |W_{t_{k+1}} - W_{t_k}| \rightarrow \infty \quad \text{fast sicher mit } n \rightarrow \infty.$$

Pfade eines Wiener Prozesses sind fast sicher unbeschränkter Variation und fast überall nicht differenzierbar. Damit existiert auch keine Ableitung dW/dt .

Für Semimartingale X, Y ist die *quadratische Kovariation* definiert als

$$[X, Y]_t := \langle X^c, Y^c \rangle + \sum_{s \leq t} \Delta X_s \Delta Y_s. \quad (\text{A.4.16})$$

Der zweite Term (die Summe) verschwindet, falls X und Y keine simultanen Sprünge aufweisen. Im einzelnen, ist X oder Y stetig, dann gilt: $[X, Y] = \langle X^c, Y^c \rangle$. Falls X oder Y ein VF-Prozess ist, dann gilt: $[X, Y] = \langle X^c, Y^c \rangle = 0$. Ferner kann gezeigt werden, dass $XY = X_0 Y_0 + X_- Y + Y_- X - [X, Y]$ bzw. für lokal quadratisch integrierbare Martingale $[X, Y] = \langle X, Y \rangle$ gilt.

Stochastisches Integral

Es sei $(\Omega, \mathcal{F}, \mathbb{F}, \mathbb{P})$ eine stochastische Basis, $W = \{W_t, t \geq 0\}$ ein Wiener Prozess und \mathcal{H} eine Prozessklasse von $H = \{H_t, t \in T\}$ auf einem endlichen Intervall $[0, T]$ mit

$$\mathbb{E} \left[\int_0^T H_t^2 dt \right] < \infty. \quad (\text{A.4.17})$$

Auf $(\Omega, \mathcal{F}, \mathbb{F}, \mathbb{P})$ gilt definitionsgemäß:

- i. H ist \mathcal{F}_t -adaptiert, d.h. H_t ist \mathcal{F}_t -messbar für alle $0 \leq t \leq T$, und
- ii. die Abbildung $(t, \omega) \mapsto H(t, \omega)$ ist $\mathcal{B} \times \mathcal{F}$ -messbar, wobei \mathcal{B} eine Borel- σ -Algebra auf $[0, T]$ ist.

In dieser Umgebung kann das Itô-Integral $(I) \int_0^T H dW$ für jedes $H \in \mathcal{H}([0, T])$ definiert werden.

Zunächst wird eine Annahme getroffen, dass H ein vorhersehbarer und linksseitig-stetiger Prozess ist. Dies ermöglicht seine Darstellung durch Elementarfunktionen:

$$H(t) = \sum_{k=0}^n \xi_k \mathbb{1}_{(\tau_k, \tau_{k+1}]}(t), \tag{A.4.18}$$

wobei $\tau_0 < \tau_1 < \dots$ eine Partition der $\{\mathcal{F}_t\}$ -Stoppzeiten und ξ_k eine \mathcal{F}_{τ_k} -messbare Zufallsvariable mit $E \xi_k^2 < \infty$ ist. Für alle solche H 's ist das Itô-Integral definiert als

$$(I) \int_0^T H dW := \sum_{k=0}^n \xi_k [W(\tau_{k+1} \wedge t) - W(\tau_k \wedge t)]. \tag{A.4.19}$$

Um die Einschränkung auf elementare Prozesse zu lockern, verwenden wir ohne Beweis die folgenden beiden Eigenschaften.²⁹ Erstens, für ein beliebiges $H \in \mathcal{H}([0, T])$ existiert eine Folge $\{H_m : m \geq 1\}$ elementarer Prozesse H_m derart, dass

$$\lim_{m \rightarrow \infty} E \left[\int_0^T (H(t) - H_m(t))^2 dt \right] = 0. \tag{A.4.20}$$

Die andere Eigenschaft heißt *Itô-Isometrie* und besagt, dass

$$E \left[\int_0^T H dW \right]^2 = E \left[\int_0^T H_t^2 dt \right] \quad \text{für alle } H \in \mathcal{H}([0, T]) \text{ ist.} \tag{A.4.21}$$

Unter Berücksichtigung von (A.4.20) und (A.4.21) ist das Itô-Integral als Grenzwert

$$(I) \int_0^T H dW := \lim_{m \rightarrow \infty} (I) \int_0^T H_m dW \quad \text{in } L^2 \tag{A.4.22}$$

definiert und es ist *nicht* von einer speziellen Folge $\{H_m\}$ abhängig.

Einige Erweiterungen zu dieser Definition sind sinnvoll und möglich. Die Bedingung (A.4.17) kann zu

$$\int_0^T H^2 dt < \infty \quad \text{fast sicher} \tag{A.4.23}$$

²⁹ Dudley und Norvaiša (1998), S. 8.

abgeschwächt werden. Damit braucht $(I) \int_0^T H dW$ nicht mehr quadratisch integrierbar zu sein, sondern es wird zu einer beliebigen \mathcal{F}_t -messbaren Zufallsvariable. Insbesondere lässt sich ein allgemeines Integral $H_t dX_t$ in Bezug auf ein Semimartingal X_t und einen lokal beschränkten vorhersehbaren Prozess H_t auf das Itô-Integral zurückführen.³⁰

Dies hat eine einfache Begründung: für einen RCLL Prozess $Y(t)$ ist der Prozess $Y(t-)$ vorhersehbar und linksseitig-stetig. Mit der Festlegung

$$Y(t) = \sum_{k=0}^n \xi_k \mathbb{1}_{[\tau_k, \tau_{k+1})}(t) \quad (\text{A.4.24})$$

sieht man aus dem Vergleich von (A.4.18) und (A.4.24) eine Identität $H(t) \equiv Y(t-)$. Folglich ist eine allgemeine Definition

$$(I) \int_0^T H dW := \int_0^T Y(t-) dW \quad (\text{A.4.25})$$

des Itô-Integrals über einen RCLL Prozess $Y(t)$ mit der über Elementarfunktionen (A.4.18) konsistent.

Satz A.4.3 (Itô-Formel) Sei X ein Semimartingal und $F(x)$ eine zweimal stetig differenzierbare Funktion, $F \in \mathbb{C}^2$. Dann ist $F(x)$ auch ein Semimartingal derart, dass:

$$F(X_t) - F(X_0) = \int_0^t F'(X_{s-}) dX_s + \frac{1}{2} \int_0^t F''(X_{s-}) d\langle X^c \rangle_s + \sum_{0 < s < t} [F(X_s) - F(X_{s-}) - F'(X_{s-}) \Delta X_s]. \quad (\text{A.4.26})$$

Im stetigen Fall entfällt der Σ -Term und die Formel vereinfacht sich wesentlich.

Satz A.4.4 Ist X in (A.4.3) ein stetiges Semimartingal, dann gilt:

$$F(X_t) - F(X_0) = \int_0^t F'(X_s) dX_s + \frac{1}{2} \int_0^t F''(X_s) d\langle X \rangle_s. \quad (\text{A.4.27})$$

³⁰ Im Finance wird X_t gewöhnlich als Unsicherheitsquelle (Aktienkurs/Rendite) interpretiert. Mit H_t bezeichnet man eine Anlagestrategie beschrieben durch die Gewichte der einzelnen Wertpapiere im Portfolio.

Einen Beweis der Itô-Formel sowie eine Erweiterung auf die endlich-dimensionalen Semimartingale findet man in Protter (1990).

Sei X ein stetiges Semimartingal, $X_0 = 0$. Dann ist nach Itô-Formel

$$\mathcal{E}(X)_t = \exp\left(X_t - \frac{1}{2}\langle X^c \rangle_t\right) \quad (\text{A.4.28})$$

ein Semimartingal. Es heißt stetiges *stochastisches Exponential*. Ein wichtiger Spezialfall aus (A.4.28) ergibt sich für $X = \theta W$, wobei $W = \{W_t, t \geq 0\}$ ein Wiener Prozess und θ eine Konstante ist. Der Prozess $\mathcal{E}(\theta W)$ heißt *geometrische Brownsche Bewegung* und bildet die Grundlage für die Modellierung des Aktienkursverlaufs. In diesem Zusammenhang ist der folgende Standardsatz im Finance von Bedeutung.

Satz A.4.5 (Girsanov) Für jeden stochastischen Prozess $\theta_t, t \in [0, T]$, $T < \infty$ derart, dass die Novikov-Bedingung

$$\mathbb{E}\left[\exp\left(\frac{1}{2}\int_0^t \theta_s^2 ds\right)\right] < \infty \quad \text{f.s.} \quad (\text{A.4.29})$$

erfüllt ist, kann durch

$$\xi_t = \exp\left(-\int_0^t \theta_s dW_s - \frac{1}{2}\int_0^t \theta_s^2 ds\right) \quad (\text{A.4.30})$$

ein Martingal ξ_t^θ definiert werden. In (A.4.30) ist W_t eine Standard Brownsche Bewegung unter dem Wahrscheinlichkeitsmaß \mathbb{P} und folglich ist W_t unter \mathbb{P} ein Martingal. Da ξ_T streng positiv ist, wird durch

$$\frac{d\mathbb{Q}}{d\mathbb{P}}(\theta) = \xi_T^\theta$$

ein äquivalentes Martingalmaß $\mathbb{Q}(\theta)$ definiert. Unter \mathbb{Q} ist der Prozess

$$W_t^\theta = W_t + \int_0^t \theta_s ds$$

ein Martingal.

A.5 Stabile Verteilungen

Stabile Verteilungen bilden eine breite Klasse der Wahrscheinlichkeitsverteilungen mit vielen interessanten Eigenschaften. Sie wurden am

Anfang des letzten Jahrhunderts von Paul Lévy bei der Erforschung von Summen von unabhängigen, identisch verteilten Zufallsvariablen eingeführt. Die Tatsache, dass es mit wenigen Ausnahmen keine geschlossene Formeln für die Verteilungsfunktionen stabiler Verteilungen (das ist eine Verteilungsklasse, nicht eine einzelne Verteilung!) existieren, ist immer noch ein Stolperstein auf dem Weg zu ihrer breiten Anwendung bei den Praktikern³¹.

Es scheint sich aber zu ändern. Außer „klassischen“ Lehrbüchern (Petrov (1975), Zolotarev (1986), Samorodnitsky und Taqqu (1994)) und grundlegenden Einführungsartikeln (Mandelbrot (1963b), Fama (1963a, 1963b)), sind inzwischen einige Literaturquellen zu finden, in den sich die Autoren auch vertieft mit Anwendungen beschäftigen: Janicki und Weron (1994), Mandelbrot (1997), Uchaikin und Zolotarev (1999), Mittnik und Rachev (2000), Meerschaert und Scheffler (2001) und Referenzen dort.

Jede stabile Verteilung lässt sich durch eine charakteristische Funktion mit vier Parametern (Stabilität, Schiefe, Skalierung, und Lage) darstellen. Dies zeigt eine hohe Flexibilität der Verteilung, die gerade bei der Modellierung solcher empirischen Phänomene wie *fat tails* und *excess kurtosis* sehr hilfreich ist. Es ist bekannt, dass die Normalverteilung in solchen Fällen nicht weiter hilft. Im Gegensatz dazu sind stabile Verteilungen wegen ihrer nicht-Gausschen Natur für solche schwierigen Fälle geradezu geschaffen.

Die Grundlage dafür bildet der sogenannte verallgemeinerte zentrale Grenzwertsatz³². Er besagt, dass die geeignet normierte Summe von unabhängigen, identisch verteilten (i.i.d.) Zufallsvariablen, die nicht unbedingt eine endliche (im Unterschied zur Standardnormalverteilung) Varianz haben, mit steigender Anzahl der Summanden nur gegen eine stabile Verteilung konvergieren kann.

Definition A.5.1 (Lévy (1937)). *Die Klasse stabiler Verteilungen besitzt die folgende charakteristische Eigenschaft: Sind X_1 und X_2 i.i.d. Zufallsvariablen, a_1 und a_2 beliebige positive Zahlen, dann existiert für jedes Paar a_1 und a_2 eine positive Zahl a derart, dass $a_1X_1 + a_2X_2$ dieselbe Verteilung haben wie aX_1 .*

³¹ Es bleibt dabei oft übersehen, dass die Verteilungsfunktion einer gewöhnlichen Standardnormalverteilung ebenfalls keine geschlossene Formel hat, lediglich Tabellen und einige effiziente Algorithmen sind verfügbar.

³² Gnedenko und Kolmogorov (1954), Feller (1968), Gnedenko (1991).

Im Jahre 1935 ergänzte Aleksandr Khinchin diesen Begriff, indem er eine zusätzliche reelle Konstante d zum Ausdruck aX_1 zugefügt hat, so dass für $d = 0$ die Verteilung *streng stabil* heißt.

Definition A.5.2 (Samorodnitsky und Taqqu (1994)). Eine Zufallsvariable X ist stabil oder hat eine stabile Verteilung F genau dann, wenn für eine positive Zahl $n \geq 2$ Konstanten $c_n \in \mathbb{R}_+$ und $d_n \in \mathbb{R}$ existieren, dass

$$X_1 + \dots + X_n \stackrel{d}{=} c_n X_1 + d_n. \tag{A.5.1}$$

Mit anderen Worten, bei entsprechender Wahl der Konstanten c_n und d_n konvergieren die Summen $X_1 + \dots + X_n$ nur gegen ein stabiles Verteilungsgesetz (das Symbol $\stackrel{d}{=}$ bezeichnet eine Konvergenz in Verteilung).

In einer Bemerkung gaben P. Lévy und A. J. Khinchin im Jahre 1936 eine endgültige Darstellungen stabiler Verteilungen durch den Logarithmus der charakteristischen Funktion. Eine Funktion

$$\varphi(\theta) = E e^{i\theta X} = \int e^{i\theta x} dF(x)$$

ist die charakteristische Funktion einer stabilen Verteilung genau dann, wenn sie die folgende Darstellung zulässt:

$$\varphi(a_1\theta) \varphi(a_2\theta) = \varphi(c\theta), \tag{A.5.2}$$

bzw. für n Summanden und $d_n \neq 0$

$$\varphi^n(\theta) = \varphi(c_n\theta) \exp(id_n\theta). \tag{A.5.3}$$

Im Falle einer stabilen Verteilung lässt sich also zu jedem Paar positiver Konstanten a_1 und a_2 ein anderes Paar c und d finden, so dass (A.5.2) bzw. (A.5.3) gilt. In einer modernen Darstellung mit vier Parametern gilt folgende

Definition A.5.3 (Samorodnitsky und Taqqu (1994)). Sei $0 < \alpha \leq 2$. Die charakteristische Funktion einer α -stabilen Zufallsvariable ist gegeben durch ($\theta \in \mathbb{R}$):

$$\begin{aligned} \log \varphi(\theta) = & \begin{cases} -\sigma^\alpha |\theta|^\alpha \left\{ 1 - i\beta \operatorname{sign}(\theta) \tan\left(\frac{\pi\alpha}{2}\right) \right\} + i\mu\theta, & \text{für } \alpha \in (0, 1) \cup (1, 2], \\ -\sigma |\theta| \left\{ 1 + i\beta \left(\frac{2}{\pi}\right) \operatorname{sign}(\theta) \ln|\theta| \right\} + i\mu\theta & \text{für } \alpha = 1, \end{cases} \end{aligned} \tag{A.5.4}$$

mit

$$\text{sign}(\theta) = \begin{cases} 1 & \text{für } \theta > 0, \\ 0 & \text{für } \theta = 0, \\ -1 & \text{für } \theta < 0. \end{cases}$$

Durch den Stabilitätsindex (auch Tail-Index oder charakteristischer Exponent) α wird die Form (shape and tails) der Verteilung bestimmt. Es ist bekannt, dass der Normierungsfaktor c_n in (A.5.1) genau $n^{1/\alpha}$ entspricht. Für $\alpha = 2$ ergibt sich die Normalverteilung als Spezialfall; für $\alpha < 2$ ist die theoretische *Varianz unendlich*, bei $\alpha < 1$ existiert auch das erste Moment oder Erwartungswert nicht mehr. Im Allgemeinen existiert das p -te Moment einer stabilen Verteilung genau dann, wenn p kleiner α ist:

$$0 < p < \alpha < 2 \Rightarrow E|X|^p < \infty, \quad \text{Var}(X) = \infty \quad (\text{A.5.5})$$

$$0 < \alpha \leq p < 2 \Rightarrow E|X|^p = \infty, \quad \text{Var}(X) = \infty \quad (\text{A.5.6})$$

$$1 < \alpha < 2 \Rightarrow E|X| = \mu, \quad \text{Var}(X) = \infty \quad (\text{A.5.7})$$

$$0 < \alpha < 1 \Rightarrow E|X| = \infty, \quad \text{Var}(X) = \infty. \quad (\text{A.5.8})$$

α ist der wichtigste der insgesamt vier Parameter, die eine stabile Verteilung charakterisieren. Als nächstes sei $\beta \in [-1, +1]$ betrachtet. Es ist ein Schiefeparameter, eine Größe, die den Grad der Asymmetrie der Verteilung angibt. Im Symmetriefall ist $\beta = 0$. Das Gegenteil, also maximale Asymmetrie oder Schiefe, tritt bei $\beta = -1$ (links) bzw. $\beta = +1$ (rechts) auf. Ungeachtet der Bezeichnung und einiger ähnlicher Eigenschaften, entspricht β nicht einem klassischen Schiefeparameter, den wir als das dritte Moment kennen. Dies hat eine ganz triviale Erklärung: das dritte Moment ist für eine stabile Verteilung wegen der Divergenz (A.5.5) – (A.5.8) nicht definiert. Eine stabile Verteilung mit $\beta = 1$ hat einen streng positiven Träger und heißt *Subordinator*. Schließlich sind σ und μ entsprechend Skalierungs- und Lageparameter. In der Literatur³³ wird eine stabilverteilte Zufallsvariable X gewöhnlich mit $X \sim S_\alpha(\sigma, \beta, \mu)$ abgekürzt und bezeichnet die Verteilungsfunktion einer stabilen Verteilung mit den Parametern $\alpha \in (0, 2]$, $\beta \in [-1, +1]$, $\sigma \in \mathbb{R}_+$ und $\mu \in \mathbb{R}$. Einen Sonderfall stellt $\beta = \mu = 0$ dar. Dadurch wird die *Symmetrisch- α -Stabile*- oder einfach *S α S-Verteilung* definiert. Ihre charakteristische Funktion hat eine einfache Gestalt: $\varphi(\theta) = \exp(-\sigma^\alpha |\theta|^\alpha)$.

³³ Samorodnitsky und Taquq (1994).

Zu fundamentalen Eigenschaften stabiler Verteilungen gehört eine relativ langsame (im Vergleich mit der Normalverteilung) Pareto-ähnliche Konvergenz an den Tails, die als *heavy tails*-Eigenschaft bekannt ist:

$$\mathbb{P}(X > x) \sim c_\alpha (1 + \beta)x^{-\alpha} \quad (\text{A.5.9})$$

oder in einer anderen Schreibweise³⁴

$$\lim_{x \rightarrow \infty} x^\alpha \mathbb{P}(X > x) = C_\alpha \frac{1 + \beta}{2} \sigma^\alpha, \quad (\text{A.5.10})$$

wobei C_α eine Konstante ist. Hier kommt eine wichtige Rolle des Parameters α zum Tragen: das α ist nämlich für das asymptotische Verhalten der Verteilung an den Tails zuständig. Je kleiner α ist, desto schwerer sind die Tails. Die Summe von zwei unabhängigen, α -stabilen Zufallsvariablen ist ebenfalls α -stabil

$$S_\alpha(\sigma_1, \beta_1, \mu_1) + S_\alpha(\sigma_2, \beta_2, \mu_2) \stackrel{d}{=} S_\alpha(\sigma, \beta, \mu) \quad (\text{A.5.11})$$

mit den Parametern

$$\sigma^\alpha = \sigma_1^\alpha + \sigma_2^\alpha, \quad \beta = \frac{\beta_1 \sigma_1^\alpha + \beta_2 \sigma_2^\alpha}{\sigma_1^\alpha + \sigma_2^\alpha}, \quad \mu = \mu_1 + \mu_2. \quad (\text{A.5.12})$$

Eine Rücktransformierte der charakteristischen Funktion (A.5.4) kann im Allgemeinen nicht in analytischer Form angegeben werden. Explizite Formeln sind lediglich für wenige spezielle Parameterwerte bekannt. Dazu zählen

- die *Normalverteilung* $S_2(\sigma, 0, \mu) = N(\mu, 2\sigma^2)$
- die *Cauchy-Verteilung* $S_1(\sigma, 0, \mu)$

$$f(x) = \frac{\sigma}{\pi \left[(x - \mu)^2 + \sigma^2 \right]}$$

- und die *einseitige Lévy-Verteilung* $S_{1/2}(\sigma, 1, \mu)$

$$f(x) = \left(\frac{\sigma}{2\pi} \right)^{1/2} \frac{1}{(x - \mu)^{3/2}} \exp\left(-\frac{\sigma}{2(x - \mu)} \right) \quad \text{für } x > \mu,$$

³⁴ Marinelli, Rachev, Roll und Göppl (1999), S. 4

wenn die entsprechenden Verteilungsdichtefunktionen existieren. Ansonsten bieten sich effiziente numerische Verfahren³⁵ zur Berechnung und graphischen Darstellung der Dichtefunktion einer stabilen Verteilung an. Nolan (2001) verwendet diese Verfahren zur Berechnung des Maximum-Likelihood Schätzers.³⁶ Algorithmen zur Simulation stabiler Verteilungen findet man z.B. in Mittnik und Rachev (2000) und Weron (2001).

Schließlich lohnt sich in Bezug auf stabile Verteilungen ein Blick auf die Normalverteilung. Obwohl die Normalverteilung auch zur Klasse stabiler Verteilungen gehört, nimmt sie als Spezialfall eine besondere Stellung ein: die Normalverteilung ist die einzige mit endlicher Varianz. Genau das ist der historische Grund dafür, dass stabile Verteilungen lange Zeit bei der Modellierung empirischer Phänomene als unattraktiv empfunden wurden. Dabei stellt das divergente zweite Moment eigentlich kein echtes Problem dar: es ist lediglich eine theoretische Größe, die in der Welt von stabilen Verteilungen als Maß der Variabilität schlicht ungeeignet ist. Man verwendet anstatt $n^{1/2}$ einen modifizierten Normierungsfaktor $n^{1/\alpha}$, führt die notwendigen Anpassungen bei der „Kovariabilität“ durch³⁷ und bekommt eine Möglichkeit, Sprünge beliebiger Höhe einwandfrei modellieren zu können. Problematisch ist es, diese Sprünge in „ruhigen Zeiten“ wieder los zu werden! Manchmal scheint die durch stabile Verteilungen erzeugte Irregularität *zu extrem* zu sein, um die tatsächlich stattgefundenen Ausschläge nachzubilden.

Die Normalverteilung ist also zu regulär, der Rest der stabilen Klasse ist zu extrem. Wäre eine Kompromisslösung möglich? Ja, es sind sogar mehrere Möglichkeiten vorhanden, eine *variierende* Irregularität zu modellieren. Nach dem Zerlegungssatz von Lebesgue (A.1.3) besteht ein beliebiges Wahrscheinlichkeitsmaß aus einem absolutstetigen und einem singulären Teil. Eine Mischung dieser Anteile wird als *Lévy Verteilung* bezeichnet und schließt die stabilen Verteilungen als eine Unterklasse mit ein. Das folgende Kapitel soll eine einführende Übersicht über Lévy Verteilungen bzw. Lévy Prozesse geben.

³⁵ Nolan (1997, 2002).

³⁶ Siehe auch Mittnik, Rachev, Doganoglu und Chenyao (1999), Mittnik, Doganoglu und Chenyao (1999).

³⁷ Das „stabile Äquivalent“ der Kovarianz einer α S-Zufallsvariable heißt „Kovariation“ (*covariation*), siehe Mittnik und Rachev (2000) S. 334.

A.6 Lévy Prozesse

Lévy Prozesse, d.h. Prozesse in stetiger Zeit mit stationären unabhängigen Zuwächsen, sind nach Paul Lévy benannt. In den 30-er Jahren des vorherigen Jahrhunderts³⁸ hat er Zusammenhänge zwischen solchen Prozessen und unbeschränkt (oder unendlich) teilbaren Verteilungen beschrieben. Ebenso intensiv hat sich Paul Lévy mit den Verteilungen selbst und deren Struktur beschäftigt (Lévy-Khinchin-Formel bzw. Lévy-Itô-Zerlegung). Lévy Prozesse sind eine wichtige Klasse der Markov-Prozesse und ein natürliches Beispiel für Semimartingale, für die stochastische Kalküle gelten. Die inzwischen klassischen Referenzen sind Bertoin (1996) und Sato (1999). Ganz aktuell und gut lesbar ist Appelbaum (2004). Viele unbeschränkt teilbare Verteilungen besitzen im Unterschied zur Normalverteilung heavy tails. Dies macht sie für Anwendungen im Finance sehr interessant.³⁹

Definition A.6.1 (Sato (1999), Def. 1.6). Ein \mathbb{R}^n -wertiger stochastischer Prozess $\{X_t, t \geq 0\}$ auf einem Wahrscheinlichkeitsraum $(\Omega, \mathcal{F}, \mathbb{P})$ heißt **Lévy Prozess**, falls gilt:

- (i) X_t hat unabhängige Zuwächse;
- (ii) $X_0 = 0$ fast sicher;
- (iii) X_t hat stationäre Zuwächse;
- (iv) der Prozess ist **stochastisch stetig**: für alle $t \geq 0$ und $\varepsilon > 0$ gilt:

$$\lim_{s \rightarrow t} \mathbb{P}[|X_s - X_t| > \varepsilon] = 0.$$

- (v) es existiert ein $\Omega_0 \in \mathcal{F}$ mit $\mathbb{P}(\Omega_0) = 1$ derart, dass für jedes $\omega \in \Omega_0$ die Pfade des Prozesses $X_t(\omega)$ RCLL als Funktion von t sind; Wird diese Bedingung nicht verlangt, heißt X **Lévy Prozess in Verteilung (in law)**.

Definition A.6.2. Eine Wahrscheinlichkeitsverteilung $F(x)$ mit c.f. (charakteristischer Funktion) $\phi(t)$ heißt **unbeschränkt teilbar, (infinitely**

³⁸ Lévy (1937).

³⁹ Barndorff-Nielsen und Shephard (2000), Barndorff-Nielsen u.a. (2001), Overhaus u.a. (2001), Boyarchenko und Levendorskiĭ (2002) und dortige Referenzen.

divisible, i.d.), falls für jede natürliche Zahl n ihre c.f. gleich der n -ten Potenz einer anderen c.f.

$$\phi(t) \equiv (\phi_n(t))^n$$

ist, die ihrerseits ebenfalls unbeschränkt teilbar heißt.

Bemerkung. Ist X_t ein Lévy Prozess in Verteilung, dann ist die Verteilung von X_t für alle $t \geq 0$ unbeschränkt teilbar. Dies ist eine Konsequenz aus (i), (ii) und (iii) Def. A.6.1.

Die charakteristische Funktion der Zuwächse eines Lévy Prozesses

$$\varphi_t(u) := \mathbb{E} \exp[iu(X_{t+t_0} - X_{t_0})]$$

hat eine Halbgruppeneigenschaft

$$\varphi_{t+t'}(u) = \varphi_t(u) \varphi_{t'}(u)$$

und lässt eine natürliche Darstellung

$$\varphi(u) = \exp[t\Psi(u)] \quad (\text{A.6.1})$$

zu. Der Kumulant $\Psi(u)$ in (A.6.1) kann durch die Lévy-Khinchin-Formel dargestellt werden (der Einfachheit halber betrachten wir hier einen eindimensionalen Fall):

$$\Psi(u) = iub - \frac{1}{2}u^2\sigma^2 + \int_{\mathbb{R}} (e^{iux} - 1 - iuh(x)) \Pi(dx). \quad (\text{A.6.2})$$

Hier bezeichnet Π ein Lévy-Maß auf \mathbb{R} , das folgenden Bedingungen genügt:

$$\Pi(\{0\}) = 0, \quad \int_{|x| \leq 1} x^2 \Pi(dx) < \infty \quad \text{und} \quad \int_{|x| > 1} \Pi(dx) < \infty. \quad (\text{A.6.3})$$

Weitere Parameter sind ein reellwertiger Lageparameter b und eine nichtnegative Zahl σ^2 (Varianz der Gausschen Komponente). Die hier getroffene Wahl für das Konvergenzkriterium ist nicht eindeutig und steht in einem engen Zusammenhang mit der Abschneidefunktion $h(x)$. Diese Funktion wird so gewählt, dass sie für kleinere Argumente linear ist und für größere Argumente gegen Null geht. Die Unterscheidung zwischen groß und klein ist dabei völlig willkürlich. Unter Berücksichtigung der Abschätzung

$$|e^{iux} - 1 - iux\mathbb{1}_{|x|\leq 1}| \sim \begin{cases} O(|x|^2) & \text{für } x \rightarrow 0, \\ O(1) & \text{für } |x| \rightarrow \infty \end{cases}$$

für ein gegebenes u ist zum Beispiel die Funktion $h(x) = x\mathbb{1}_{|x|\leq 1}$ eine gebräuchliche Wahl. Für eine andere Regularisierung des Lévy-Maßes muss auch $h(x)$ entsprechend angepasst werden. Das gleiche gilt für b . Durch das *Tripel* $(b, \sigma^2, \Pi(dx))$ ist eine i.d. Verteilung eindeutig bestimmt.

Sei X ein Lévy Prozess derart, dass die c.f. von X_1 Gleichung (A.6.2) genügt. Dann beschreibt Π Sprünge von X . Betrachten wir ein offenes Intervall A auf der reellen Achse ohne 0. Dann ist $\Pi(A)$ endlich und für jedes infinitesimal kleines Intervall $[x, x + dx)$ hat die Anzahl der Sprünge in diesem Intervall, die zu A gehören

$$\eta(x, A) = \#\{s \in [x, x + dx) \mid \Delta X = X(s) - X(s-) \in A\},$$

den Erwartungswert $\Pi(A)dx$. Für disjunkte Intervalle $[x_1, x_1 + dx)$ und $[x_2, x_2 + dx)$ sind die Zufallsvariablen $\eta(x_1, A)$ und $\eta(x_2, A)$ unabhängig.

Die Gleichung (A.6.2) hat folgende Darstellung in der Sprache von Lévy Prozessen. Jeder Lévy Prozess⁴⁰ kann als Summe aus vier unabhängigen Komponenten $\Psi = \Psi^{(1)} + \Psi^{(2)} + \Psi^{(3)} + \Psi^{(4)}$ dargestellt werden (*Lévy-Itô-Zerlegung*).

$$\Psi^{(1)}(u) = iub \tag{A.6.4}$$

entspricht einem deterministischen linearen Prozess (mit konstantem Drift) und steht für den ersten Summand in (A.6.2).

$\Psi^{(2)}(u)$ ist die sogenannte *Brownsche Komponente* und entspricht einer skalierten Brownschen Bewegung σW_t , wobei W_t eine Standard-Brownsche Bewegung ist:

$$\Psi^{(2)} = -\frac{1}{2}u^2\sigma^2. \tag{A.6.5}$$

$\Psi^{(3)}(u)$ kann als charakteristischer Exponent eines Sprungprozesses interpretiert werden, dessen Sprünge im absoluten Wert größer Eins sind. Dies entspricht einem zusammengesetzten Poisson Prozess mit Lévy-Maß $\mathbb{1}_{|x|>1}\Pi(dx)$:

⁴⁰ Die Lévy-Khinchin-Formel deckt *alle* Lévy Prozesse ab.

$$\Psi^{(3)}(u) = \int_{\mathbb{R}} (e^{iux} - 1) \mathbb{1}_{|x|>1} \Pi(dx). \quad (\text{A.6.6})$$

Schließlich steht $\Psi^{(4)}$ für ein rein unstetiges Martingal, dessen Sprünge nie Eins erreichen:

$$\Psi^{(4)}(u) = \int_{\mathbb{R}} (e^{iux} - 1 - iux \mathbb{1}_{|x|\leq 1}) \Pi(dx). \quad (\text{A.6.7})$$

Falls zusätzlich zu (A.6.3) eine weitere Integritätsbedingung

$$\int_{\mathbb{R}} |x| \mathbb{1}_{|x|<1} \Pi(dx) < \infty \quad (\text{A.6.8})$$

erfüllt ist, kann (A.6.7) vereinfacht werden:

$$\Psi^{(4)}(u) = iub' + \int_{\mathbb{R}} (e^{iux} - 1) \mathbb{1}_{|x|<1} \Pi(dx), \quad (\text{A.6.9})$$

$$b' = \int_{\mathbb{R}} x \mathbb{1}_{|x|<1} \Pi(dx). \quad (\text{A.6.10})$$

Dies ist wieder ein zusammengesetzter Poisson Prozess. Damit können $\Psi^{(3)}$ und $\Psi^{(4)}$ zusammengefasst und (A.6.2) umgeschrieben werden:

$$\Psi(u) = iub_0 - \frac{1}{2}u^2\sigma^2 + \int_{\mathbb{R}} (e^{iux} - 1) \Pi(dx), \quad (\text{A.6.11})$$

wobei $b_0 = b - b'$ unter Berücksichtigung von (A.6.2) bzw. (A.6.10) festgelegt wird und heißt *Drift* von X . Nach der Darstellung (A.6.11) sind unendlich viele Sprünge zulässig. Die einzige Bedingung ist, dass diese Sprünge klein genug sind und bilden eine endliche Summe. Sind die Bedingungen (A.6.3) und (A.6.8) erfüllt, kann ein Lévy Prozess als Prozess mit exponential verteilten Wartezeiten zwischen aufeinanderfolgenden Sprüngen interpretiert werden. Die Sprunghöhen selbst sind zufällig mit $\Pi(dx) / \int_{\mathbb{R}} \Pi(dx)$ verteilt.

Durch unabhängige Zuwächse ist die Markov-Eigenschaft der Lévy Prozesse sichergestellt, d.h., gegeben den Gegenwartswert sind Zukunfts- und Vergangenheitswerte unabhängig. Ein Lévy Prozess ist nur dann ein Martingal, wenn $b = 0$ und wenn $\Pi(dx)$ symmetrisch ist und der Integritätsbedingung (A.6.3) genügt und folglich $E|X_t| < \infty$ ist. Jeder Lévy Prozess ist ein Semimartingal und lässt sich als Summe aus einem Martingal und einem Prozess mit beschränkter Variation darstellen. Dies macht die Anwendung der Theorie stochastischer Integration auf Lévy Prozesse unproblematisch. Haben die Pfade eines Lévy

Prozesses auf jedem kompakten Zeitintervall fast sicher beschränkte Variation, so ist dieser Lévy Prozess ein Prozess beschränkter Variation. Ein Lévy Prozess ist ein Prozess beschränkter Variation genau dann, wenn $\sigma^2 = 0$ und beide Bedingungen (A.6.3) und (A.6.8) erfüllt sind.⁴¹

Subordination

Ein *Subordinator* ist ein Lévy Prozess mit Werten in \mathbb{R}^+ und fast sicher steigenden Pfaden. Seine Eigenschaften ermöglichen die Interpretation eines Subordinators als *Zufallszeit*. *Subordination* oder *Unterordnung* ist eine Transformation eines stochastischen Prozesses mittels zufälliger Zeitänderung bedingt durch einen Subordinator, der unabhängig vom ursprünglichen Prozess ist.⁴² Sei X ein beliebiger Lévy Prozess und T ein Subordinator auf dem selben Wahrscheinlichkeitsraum derart, dass X und T unabhängig sind. Wir definieren einen neuen stochastischen Prozess $Z = \{Z_t, t \geq 0\}$ als $Z_t = X_{T_t}$ so, dass $Z_t(\omega) = X_{T_t(\omega)}(\omega)$ für jedes $\omega \in \Omega$ gilt. Dann ist Z ein Lévy Prozess.⁴³ Da sie nichtfallenden Pfade beschränkter Variation sind, genügt das Lévy-Maß $\Pi(dx)$ der Bedingung (A.6.8). Daher nimmt der charakteristische Exponent eines Subordinators die folgende Form an:

$$\Psi(u) = iub + \int_{\mathbb{R}^+} (e^{iux} - 1) \Pi(dx). \quad (\text{A.6.12})$$

Die Laplace Transformation der Wahrscheinlichkeitsverteilung eines Subordinators Z hat folgende Darstellung:

$$\mathbb{E} \exp(-\lambda Z_t) = \exp(-t\Phi(\lambda)), \quad (\text{A.6.13})$$

wobei $\Phi: \mathbb{R}_+ \rightarrow \mathbb{R}_+$ *Laplace Exponent* von Z heißt. Weiterhin gilt:

$$\Phi(\lambda) = \Psi(i\lambda) = b_0\lambda + \int_{\mathbb{R}_+} (1 - e^{-\lambda x}) \Pi(dx). \quad (\text{A.6.14})$$

A.7 Spezialfunktionen

Für detaillierte Informationen zu Spezialfunktionen verweisen wir auf Gradshteyn und Ryzhik (1980).

⁴¹ Bertoin (1996), S. 15.

⁴² Dieses Konzept geht auf Bochner (1955) zurück.

⁴³ Sato (1999), Theorem 30.1

Die Gamma und die Beta Funktionen

Die *Gamma* Funktion

$$\Gamma(x) \equiv \int_0^{\infty} t^{x-1} e^{-t} dt, \quad x > 0 \quad (\text{A.7.1})$$

und die folgende bekannte Identität $\Gamma(x+1) = x\Gamma(x)$ erweitert $n! = \Gamma(n+1)$ auf nichtganze Zahlen. Aus der Formel

$$\Gamma(x)\Gamma(1-x) = \frac{\pi}{\sin(\pi x)} \quad \text{für } x \text{ nicht ganz}$$

folgt $\Gamma(1/2) = \sqrt{\pi}$ für $x = 1/2$. Mit $x^2 = t$ in (A.7.1) ergibt sich

$$\Gamma(m) = 2 \int_0^{\infty} x^{2m-1} e^{-x^2} dx. \quad (\text{A.7.2})$$

Die *Beta* Funktion ist durch die Formel

$$B(x,y) \equiv \int_0^1 t^{x-1} (1-t)^{y-1} dt. \quad (\text{A.7.3})$$

definiert und es gilt

$$B(x,y) = \frac{\Gamma(x)\Gamma(y)}{\Gamma(x+y)}. \quad (\text{A.7.4})$$

Die Hypergeometrische Funktion

Die *hypergeometrische Funktion* $F(a,b,c;x)$ ist als Lösung der *Gauschen hypergeometrischen Differentialgleichung*

$$x(x-1)w'' + [(a+b+1)x-c]w' + abw = 0 \quad (\text{A.7.5})$$

definiert. Sie hat eine Reihen- und eine Integraldarstellung:

$$F(a,b,c;x) = \sum_{k=0}^{\infty} \frac{(a)_k (b)_k}{(c)_k} \frac{x^k}{k!} = 1 + \frac{ab}{c}x + \frac{a(a+1)b(b+1)}{c(c+1)} \frac{x^2}{2!} + \dots, \quad (\text{A.7.6})$$

mit $(a)_0 = 1$, $(a)_k = a(a+1)\dots(a+k)$, $k \in \mathbb{N}$, $|x| < 1$ und $c \neq 0, -1, -2, -3, \dots$

$$F(a,b,c;x) = \frac{\Gamma(c)}{\Gamma(b)\Gamma(c-b)} \int_0^1 t^{b-1} (1-t)^{c-b-1} (1-tx)^{-a} dt, \quad c > b > 0. \quad (\text{A.7.9})$$

BEISPIEL A.14

In der Wahrscheinlichkeitstheorie kommt die hypergeometrische Funktion bei der *Hypergeometrischen Verteilung* vor.

In einer Urne liegen N Kugeln; $K \leq N$ davon sind rot. Man zieht $n \leq K$ Kugeln ohne Zurücklegen. Die Wahrscheinlichkeit, von diesen n Kugeln genau $x \leq n$ rote zu erhalten, ist

$$p_X(x) = \binom{K}{x} \binom{N-K}{n-x} / \binom{N}{n}. \quad (\text{A.7.7})$$

Die erzeugende Funktion

$$\Psi(z) = \mathbb{E} z^X = \sum_{k=0}^{\infty} p_k z^k$$

für die Hypergeometrische Verteilung ist gegeben durch [Parzen 1967]

$$\Psi(z) = \left[\binom{N-K}{n} / \binom{N}{n} \right] F(a, b, c; z) \quad (\text{A.7.8})$$

mit $a = -n$, $b = -K$, $c = N - K - n + 1$.

Das Laplace Integral

Das Laplace Integral spielt eine wichtige Rolle in der Theorie der großen Abweichungen sowie auch allgemein dort, wo man den Wert einer Funktion *abschätzen will*. Für eine reellwertige und analytische Funktion $\psi(x)$ definiert man das Laplace Integral von $\psi(x)$ auf dem Intervall $[a, b]$ durch

$$\int_a^b e^{-n\psi(x)} dx, \quad n \in \mathbb{N}^+. \quad (\text{A.7.10})$$

Damit der Integrand (A.7.10) seinen größten Wert annimmt, muss ψ sein Minimum erreichen unter der Annahme, dass es ein solches Minimum auf $[a, b]$ gibt. Eine Approximation durch das Laplace Integral erfolgt in zwei Schritten.

Zuerst ersetzt man die Funktion ψ durch die drei ersten Terme aus der Taylor-Reihen-Entwicklung rund um das Minimum x_0 :

$$\psi(x) \approx \psi(x_0) + \psi'(x_0)(x - x_0) + \frac{1}{2}\psi''(x_0)(x - x_0)^2. \quad (\text{A.7.11})$$

Da x_0 ein Minimum von ψ ist, gilt⁴⁴ $\psi'(x_0) = 0$ und $\psi''(x_0) > 0$. Als nächstes verwendet man die Approximation (A.7.11) und ändert die

⁴⁴ In einem allgemeinen Fall dürfte es natürlich vorkommen, dass $\psi''(x_0) = 0$ ist. Dann müssten mehr Terme in (A.7.11) berücksichtigt werden und das Verfahren ändert sich geringfügig.

Grenzwerte des Integrals so, dass nur eine ε -Umgebung um den Punkt x_0 abgedeckt wird:

$$\int_a^b e^{-n\psi(x)} dx \approx e^{-n\psi(x_0)} \int_{x_0-\varepsilon}^{x_0+\varepsilon} e^{-n\frac{\psi''(x_0)}{2}(x-x_0)^2} dx. \quad (\text{A.7.12})$$

Dieser Übergang verursacht lediglich exponentiell kleine Fehler. Im zweiten Schritt erweitert man das ε -Intervall auf der rechten Seite von (A.7.12) bis unendlich und stellt fest, dass der Approximationsfehler im Vergleich mit dem Ursprungsintegral (A.7.10) immer noch exponentiell klein ist!

$$\int_a^b e^{-n\psi(x)} dx \approx e^{-n\psi(x_0)} \int_{-\infty}^{\infty} e^{-n\frac{\psi''(x_0)}{2}(x-x_0)^2} dx. \quad (\text{A.7.13})$$

Die rechte Seite von (A.7.13) ist insoweit wertvoll, da es sich dabei bis auf eine multiplikative Konstante um eine Normalverteilungsdichte mit Erwartungswert x_0 und Varianz $\frac{1}{\psi''(x_0)}$ handelt. Ergebnis der Integration von (A.7.13) ist⁴⁵

$$\int_a^b e^{-n\psi(x)} dx = e^{-n\psi(x_0)} \left\{ \sqrt{\frac{2\pi}{n\psi''(x_0)}} + O\left(\frac{1}{n^{3/2}}\right) \right\}. \quad (\text{A.7.14})$$

Diese Formel stellt eine Näherungslösung mittels *quadratischer Approximation* für (A.7.10) im Falle einer *reellwertigen* Funktion ψ dar. Sie gilt auch bei einer Multiplikation der exponentiellen Funktion unter dem Integral (A.7.10) mit einer sich langsam verändernden Funktion $\varphi(x)$. Dies verändert die rechte Seite von (A.7.14) nur um einen konstanten Multiplikator:

$$\int_a^b \varphi(x) e^{-n\psi(x)} dx = \varphi(x_0) e^{-n\psi(x_0)} \{ \dots \}. \quad (\text{A.7.15})$$

Die Erklärung lässt sich auf die getroffenen Annahmen zurückführen: im Vergleich zu $e^{-n\psi(x)}$ ist $\varphi(x)$ so träge, dass eine Taylor-Reihen-Approximation allein mit dem konstanten Glied $\varphi(x_0)$ ausreicht, um eine gute Näherung zu erhalten.

⁴⁵ Jänich (2001), S. 378.

Die Stirling Formel

Die Originalversion der Formel stammt von Abraham De Moivre in "Miscellanea Analytica", 1730):

$$\lim_{n \rightarrow \infty} \frac{n!}{\sqrt{2\pi n} e^{-n} n^{n+1/2}} = 1. \quad (\text{A.7.16})$$

Sie wurde später verfeinert, aber noch im selben Jahr von J. Stirling in "Methodus Differentialis" publiziert. Approximativ gilt für große n

$$n! \sim \sqrt{2\pi n} e^{-n} n^{n+1/2}, \quad (\text{A.7.17})$$

wobei mit $f(x) \sim g(x)$ gemeint ist, dass $f(x)/g(x) \rightarrow 1$ mit $x \rightarrow \infty$. Die Formel lässt sich recht einfach aus der Gamma-Funktion mittels Laplace-Methode herleiten. Aus der Definition der Gamma-Funktion

$$\Gamma(z) = \int_0^{\infty} t^{z-1} e^{-t} dt$$

und der Eigenschaft $\Gamma(n+1) = n!$ resultiert

$$n! = \int_0^{\infty} t^n e^{-t} dt = \int_0^{\infty} e^{n \log t - t} dt. \quad (\text{A.7.18})$$

Um die Laplace-Methode anzuwenden, ändern wir die Integrationsvariable t zu $v = t/n$ und bekommen

$$n! = n^{n+1} \int_0^{\infty} e^{-nh(v)} dv \quad (\text{A.7.19})$$

mit $h(v) = v - \ln v$. Diese Hilfsfunktion $h(v)$ hat ein einziges Minimum in $v = 1$. Für große n ergibt sich dann durch die Taylor-Reihenentwicklung um $v = 1$ und die Anwendung der Laplace-Regel (A.7.14) auf (A.7.19)

$$e^{-n(v - \log v)} \sim e^{-n \left(1 + \frac{(v-1)^2}{2} \right)} \quad (\text{A.7.20})$$

und folglich

$$n! \sim n^{n+1} e^{-n} \int_{-\infty}^{\infty} e^{-\frac{mu^2}{2}} du = \sqrt{2\pi n} e^{-n} n^n, \quad (\text{A.7.21})$$

was mit der Stirling Formel (A.7.17) übereinstimmt.

Symbole O , o und \sim

Bei Untersuchung des Verhaltens einer Funktion $f(x)$ für $x \rightarrow 0$ oder $x \rightarrow \infty$ ist es oft hilfreich, die *Größenordnung* von $f(x)$ mit der Größenordnung einer bekannten einfachen Funktion $g(x)$ zu vergleichen. In solchen Fällen verwendet man in der Literatur gewöhnlich die folgenden Bezeichnungen⁴⁶:

i. Wir schreiben

$$f(x) \sim g(x) \quad \text{für} \quad \lim_{x \rightarrow \infty} f(x)/g(x) = 1$$

und sagen: „ $f(x)$ ist asymptotisch gleich $g(x)$.“

ii. Wir schreiben

$$f(x) = O(g(x))$$

und sagen: „ $f(x)$ ist höchstens der Ordnung $g(x)$ “, falls das Verhältnis $f(x)/g(x)$ für die relevanten Grenzwerte von x beschränkt bleibt.

iii. Es gibt weiterhin das „ o “-Symbol

$$f(x) = o(g(x))$$

mit Bedeutung: „ $f(x)$ ist kleiner Ordnung als $g(x)$ “, falls das Verhältnis $f(x)/g(x)$ für die relevanten Grenzwerte von x gegen Null geht.

Die Symbole O , o und \sim kennzeichnen verschiedene Stufen unserer Kenntnis des asymptotischen Verhaltens einer Funktion:

\sim	wir wissen am meisten
o	wir wissen weniger
O	wir wissen am wenigsten.

Diese Symbole werden oft ohne Bezug auf eine bestimmte Funktion $f(x)$ verwendet. Dabei kann $O(x)$ für „eine beliebige Funktion der Größenordnung höchstens x “ stehen; $O(1)$ bezeichnet eine beschränkte Funktion und man schreibt schließlich $o(1)$ für eine „gegen Null tendierende Funktion.“

⁴⁶ Cramér (1945), S. 122.

A.8 Konvergenzbegriffe in der Stochastik

Sei $X_n, n \in \mathbb{N}$ eine reellwertige Folge von Zufallsvariablen auf dem Wahrscheinlichkeitsraum $(\Omega, \mathcal{F}, \mathbb{P})$ und F_n die entsprechende Verteilungsfunktion.

- i. X_n konvergiert gegen X *fast sicher*, \mathbb{P} -f.s., falls

$$\mathbb{P} \left(\left\{ \omega \in \Omega : \lim_{n \rightarrow \infty} |X_n(\omega) - X(\omega)| = 0 \right\} \right) = 1.$$

Mit anderen Worten,

$$\lim_{n \rightarrow \infty} X_n(\omega) = X(\omega)$$

für alle $\omega \in \Omega - \mathcal{N}$, wobei $\mathcal{N} \in \mathcal{F}$ eine Nullmenge ist: $\mathbb{P}(\mathcal{N}) = 0$.

- ii. X_n konvergiert gegen X *in L^2* , falls

$$\lim_{n \rightarrow \infty} \mathbb{E}(|X_n - X|^2) = 0.$$

- iii. X_n konvergiert gegen X *in Wahrscheinlichkeit*, falls für alle $\varepsilon > 0$

$$\lim_{n \rightarrow \infty} \mathbb{P}(|X_n - X| > \varepsilon) = 0.$$

- iv. X_n konvergiert gegen X *in Verteilung oder schwach*, falls

$$\lim_{n \rightarrow \infty} F_{X_n}(x) \rightarrow F_X(x)$$

an allen Stetigkeitspunkten von F_X .

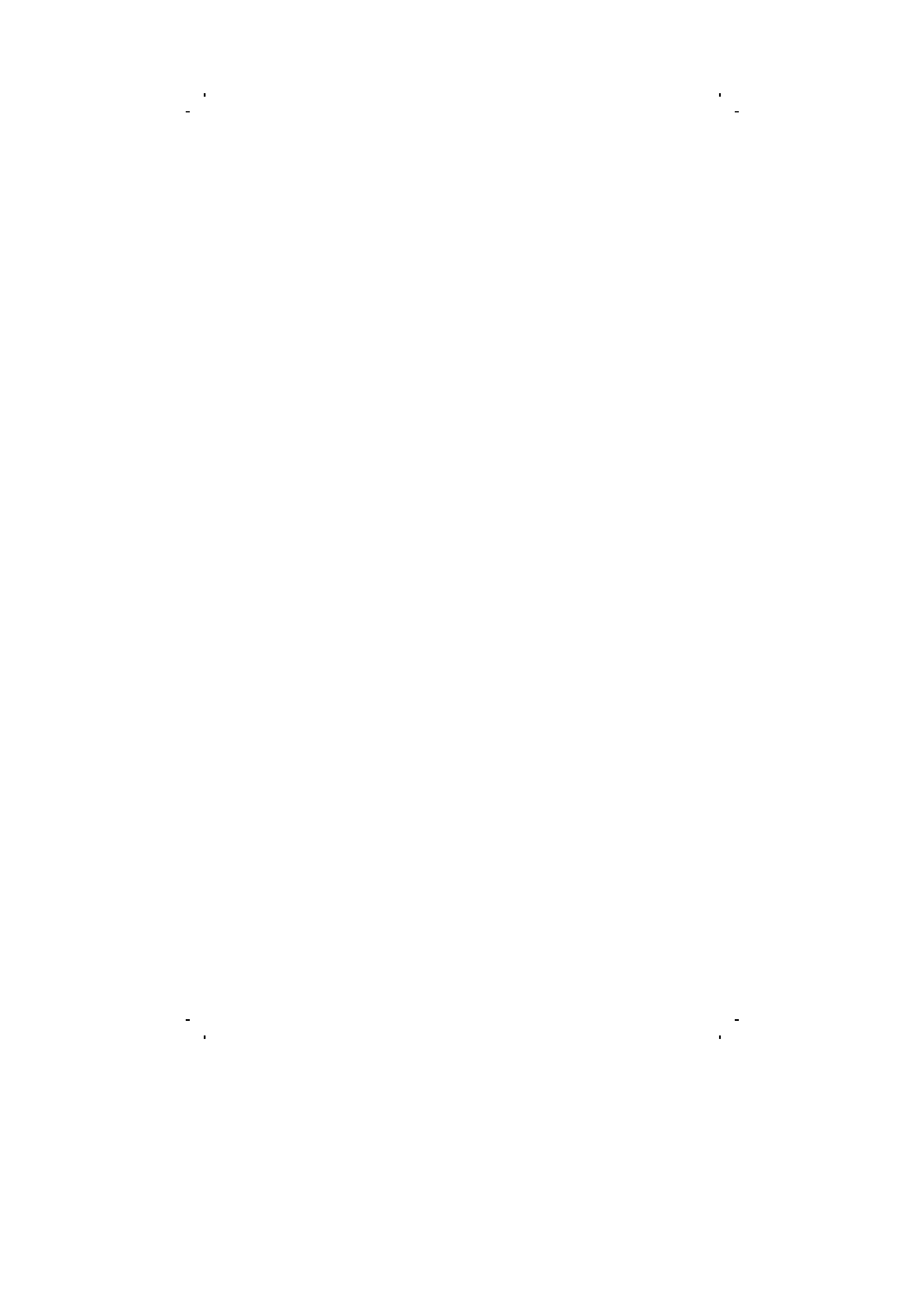
Es gelten folgende Relationen zwischen Konvergenzarten:

$$\begin{aligned} \text{fast sichere Konvergenz} &\implies \text{Konv. in Wahrscheinlichkeit} \implies \text{Konv. in} \\ &\quad \text{Verteilung} \\ L^2\text{-Konvergenz} &\implies \text{Konv. in Wahrscheinlichkeit} \implies \text{Konv. in} \\ &\quad \text{Verteilung.} \end{aligned}$$

Konvergiert X_n gegen X in Wahrscheinlichkeit, dann können wir immer eine Unterfolge finden, die zu \tilde{X} fast sicher konvergiert mit $\mathbb{P}(\tilde{X}) = \mathbb{P}(X)$. Ist der Grenzwert von X_n eine Konstante, dann gilt⁴⁷

$$\text{Konvergenz in Verteilung} \implies \text{Konv. in Wahrscheinlichkeit.}$$

⁴⁷ Durrett (1991).



Literaturverzeichnis

- Aase, K., B. Øksendal, N. Privault und J. Ubøe (2000). White noise generalizations of the Clark-Hausmann-Ocone theorem with application to mathematical finance. *Finance and Stochastics* 4, 465–496.
- Abry, P., P. Flandrin, M. S. Taqqu und D. Veitch (2000). Self-similar network traffic and performance evaluation. In: K. Park und W. Willinger (Hrsg.), *Wavelets for the Analysis, Estimation, and Synthesis of Scaling Data*, New York, S. 39–88. John Wiley & Sons.
- Akigray, V. und G. G. Booth (1988). The stable-law model of stock returns. *Journal of Business & Economic Statistics* 6(1), 51–57.
- Albert, H. (1972). Theorien in den Sozialwissenschaften. In: H. Albert (Hrsg.), *Theorie und Realität. Ausgewählte Aufsätze zur Wissenschaftslehre der Sozialwissenschaften*, S. 3–25. Tübingen: J. C. B. Mohr (Paul Siebeck).
- Alekseev, V. M., V. M. Tikhomirov und S. V. Fomin (1987). *Optimal Control*. New York: Plenum.
- Alemany, P. A. und D. H. Zanette (1994). Fractal random walks from a variational formalism for Tsallis entropies. *Phys. Rev. E* 49, R956.
- Alos, E., O. Mazet und D. Nualart (2000). Stochastic calculus with respect to fractional Brownian motion with Hurst parameter less than $\frac{1}{2}$. *Stochastic Processes and Their Applications* 86(1), 121–139.
- Ané, T. und H. Geman (2000). Order flow, transaction clock, and normality of asset returns. *Journal of Finance* 55, 2259–2284.
- Anh, V. V., Q. M. Tieng und Y. K. Tse (2000). Cointegration of stochastic multifractals with application to foreign exchange rates. *Intl. Trans. in Op. Res.* 7, 349–363.
- Aoki, M. (1996). *New Approaches to Macroeconomic Modeling*. Cambridge UK: Cambridge University Press.
- Appelbaum, D. (2004). *Lévy Processes and Stochastic Calculus*. Cambridge University Press.
- Argoul, F., A. Arneodo, J. Elezgaray und G. Grasseau (1990). Wavelet analysis of the self-similarity of diffusion-limited aggregates and electrodeposition clusters. *Phys. Rev. A* 41(10), 5537–5560.

- Arneodo, A. (1996). *Wavelets: Theory and Application*, Kapitel Wavelet analysis of fractals: From the mathematical concept to experimental reality, S. 349–502. New York: Oxford University Press.
- Arneodo, A., F. Argoul, E. Bacry, J. Elezgaray und J. F. Muzy (1995). *Ondelettes, Multifractales et Turbulences*. Paris: Diterot editeur.
- Bachelier, L. (1900). Théorie de la spéculation. *Annales Scientifiques de l'École Normale Supérieure, 3^e série 17*, 21–86. Reprint in: Cootner (1964), 17–78.
- Ball, R. (1989). What do we know about stock market efficiency? In: R. Guimaraes (Hrsg.), *A Reappraisal of the Efficiency of Financial Markets*, S. 24–55. Heidelberg: Springer.
- Barkoulas, J., C. Baum und N. Travlos (2000). Long memory in the Greek stock market. *Applied Financial Economics 10*, 177–184.
- Barndorff-Nielsen, O., T. Mikosch und S. Resnick (Hrsg.) (2001). *Lévy processes*, Boston. Birkhauser.
- Barndorff-Nielsen, O. und N. Shephard (2000, July). Modelling by Lévy-processes for financial econometrics. Centre for Analytical Finance, University of Aarhus, Denmark.
- Bashkurov, A. G. (2004). Maximum Rényi entropy principle for systems with power-law hamiltonians. *Physical Review Letters 93*, 130601–1–130601–4.
- Bayes, T. (1763). An essay towards solving a problem in the doctrine of chances. *Philos. Trans. Royal Soc. 61.53*, 370–418. London.
- Bender, C. (2003). An Itô formula for generalized functionals of a fractional Brownian motion with arbitrary Hurst parameter. *Stochastic Processes and their Applications 104*, 81–106.
- Bening, V. und V. Korolev (2002). *Generalized Poisson Models and their Applications in Insurance and Finance*. The Netherlands: VSP BV.
- Benth, F. (2003). On arbitrage-free pricing of weather derivatives based on fractional Brownian motion. Working Paper, Dept. of Math., University of Oslo.
- Beran, J. (1994). *Statistics for Long-Memory Processes*. New York: Chapman and Hall.
- Bernoulli, J. (1713). *Ars Conjectandi*. Book 4, Baseae, Impensis Thurnisiorum.
- Bernstein, P. (1992). *Capital Ideas: The Origins of Modern Wall Street*. New York: Free Press.
- Bertoin, J. (1996). *Lévy processes*. Cambridge: Cambridge University Press.
- Bhattacharia, R. N., V. K. Gupta und E. Waymire (1983). The Hurst effect under trends. *Journal of Applied Probability 20*, 649–662.
- Billingsley, P. (1986.). *Probability and Measure* (3. Auflage). New York: Wiley.
- Björk, T. und H. Hult (2005). A note on Wick products and the fractional Black-Scholes model. *Finance & Stochastics 9*(2), 197–209.

- Black, F. (1972). Capital market equilibrium with restricted borrowing. *Journal of Business* 45, 444–455.
- Black, F. (1986). Noise. *Journal of Finance* 41(3), 529–543.
- Bochner, S. (1955). *Harmonic Analysis and the Theory of Probability*. Berkeley: University of Californian Press.
- Bollerslev, T., R. Y. Chou und K. F. Kroner (1992). Arch modeling in finance. a review of the theory and empirical evidence. *Journal of Econometrics* 52, 5–59.
- Borovskikh, Y. V. und V. S. Korolyuk (1997). *Martingale Approximation*. The Netherlands: VSP BV.
- Boyarchenko, S. und S. Levendorskiĭ (2002). *Non-Gaussian Merton-Black-Scholes Theory*. Singapore: World Scientific.
- Bradley, B. O. und M. S. Taqqu (2003). Financial risk and heavy tails. In: S. T. Rachev (Hrsg.), *Handbook of Heavy Tailed Distributions in Finance*, Amsterdam, S. 35–103. Elsevier.
- Branger, N. (2002). *Bewertung nicht redundanter Finanzderivate mittels Entropie und Cross-Entropie*. Deutscher-Universitätsverlag. Dissertation.
- Breiman, L. (1957). The individual ergodic theorem of information theory. *Ann. Math. Stat* 28, 809–811. Korrektur: 31, S. 809–810, 1960.
- Brillouin, L. (1962). *Science and Information Theory*. New York: Academic Press.
- Brock, W. A. (1999). Scaling in economics: a reader's guide. *Industrial and Corporate Change* 8(3), 409–446.
- Brody, D. C., I. R. C. Buckley und I. C. Constantinou (2004). Option price calibration from Rényi entropy. Imperial College Preprint.
- Bryc, W. (2000). Elementary large deviations with applications. Department of Mathematics, University of Cincinnati.
- Bucklew, J. A. (1990). *Large Deviation Techniques in Decision, Simulation and Estimation*. New York: Wiley.
- Bühlmann, H., F. Delbaen, P. Embrechts und A. N. Shiryaev (1996). No-arbitrage, change of measure and conditional Esscher transforms. *CWI Quarterly* 9, 291–317.
- Buiatti, M., P. Grigolini und A. Montagnini (1999). Dynamic approach to thermodynamics of superdiffusion. *Phys. Rev. Lett.* 82, 3383.
- Calvet, L. und A. Fisher (2002). Multifractality in asset returns: Theory and evidence. *The Review of Economics and Statistics* 84(3), 381–406.
- Calvet, L., A. Fisher und B. B. Mandelbrot (1997). Large Deviation theory and the distribution of price changes. Cowles Foundation discussion paper No. 1165, Yale University.

- Campbell, J. Y., A. W. Lo und A. C. MacKinlay (1997). *The Econometrics of Financial Markets*. Princeton, NJ: Princeton University Press.
- Carmona, P., L. Coutin und G. Montseny (1999). Stochastic integration with respect to fractional Brownian motion. Preprint.
- Chaitin, G. J. (1987). *Algorithmic Information Theory*. Cambridge: Cambridge University Press.
- Cheridito, P. (2001). *Regularizing Fractional Brownian Motion with a View towards Stock Price Modelling*. Ph. D. thesis, ETH, Zürich.
- Cheridito, P. (2003). Arbitrage in fractional Brownian motion models. *Finance and Stochastics* 7(4), 533–553.
- Chernoff, H. (1952). A measure of asymptotic efficiency for tests of hypothesis based on the sum of observations. *Annals of Math. Stat.* 23, 493–507.
- Cherny, A. S. und V. P. Maslov (2004). On minimization and maximization of entropy in various disciplines. *Theory Probab. Appl.* 48(3), 447–464.
- Cheung, Y. und K. Lai (1995). A search for long memory in international stock market returns. *Journal of International Money and Finance* 14, 597–615.
- Cheung, Y. W., K. S. Lai und M. Lain (1993). Are there long cycles in foreign stock returns? *Journal of International Financial Markets* 3, 33–47.
- Chung, K. (1974). *A Course in Probability Theory*. New York: Academic Press.
- Clark, P. (1970). *A Subordinated Stochastic Process Model of Cotton Futures Prices*. Ph. D. thesis, Harvard University, Cambridge, MA.
- Clark, P. (1973). A subordinated stochastic process model with finite variance for speculative prices. *Econometrica* 41(1), 135–155.
- Clarkson, R. (2002). A fractal probability distribution for financial risk applications. In: *XXVIIth International Congress of Actuaries*.
- Cohn, D. (1980). *Measure Theory*. Birkhäuser.
- Collier, J. (1986). Entropy in evolution. *Biology and Philosophy* 1, 5–24.
- Cont, R. (2001). Empirical properties of asset returns: stylized facts and statistical issues. *Quantitative Finance* 1, 223–236.
- Cootner, P. H. (Hrsg.) (1964). *The Random Character of Stock Market Prices*, Cambridge, MA. MIT Press.
- Corazza, M., A. Malliaris und C. Nardelli (1997). Searching for fractal structure in agricultural futures markets. *The Journal of Futures Markets* 17, 433–473.
- Cover, T. M. und J. A. Thomas (1991). *Elements of Information Theory*. New York: Wiley.
- Cowles, A. (1933). Can stock market forecast forecast? *Econometrica* 1, 309–324.

- Cowles, A. (1944). Stock market forecasting. *Econometrica* 12(3/4), 206–214.
- Cowles, A. und H. E. Jones (1937). Some a posteriori probabilities in stock market action. *Econometrica* 5, 280–294.
- Cramér, H. (1938). Sur un nouveau théorème-limite de la théorie des probabilités. *Actualités Scientifiques et Industrielles* 736, 5–23.
- Cramér, H. (1945). *Mathematical Methods of Statistics*. Uppsala: Almqvist and Wiksells.
- Crato, N. (1994). Some international evidence regarding the stochastic behavior of stock returns. *Applied Financial Economics* 4, 33–39.
- Cutland, N. J., P. E. Kopp und W. Willinger (1995). Stock price returns and the Joseph effect: a fractional version of the Black-Scholes model. In: *Seminar on Stochastic Analysis, Random Fields and Applications (Ascona, 1993)*, Basel, S. 327–351. Birkhäuser.
- Dacorogna, M. M., R. Gençay, U. A. Müller, R. B. Olsen und O. V. Pictet (2001). *An Introduction to High-Frequency Finance*. New York: Academic Press.
- Dasgupta, A. (1997). *Fractional Brownian motion: Its properties and applications to stochastic integration*. Ph. D. thesis, Dept. of Statistics, University of North Carolina at Chapel Hill.
- Decreusefond, L. und A. S. Üstünel (1999). Stochastic analysis of the fractional Brownian motion. *Potential Analysis* 10, 177–214.
- Delbaen, F. und W. Schachermayer (1994). A general version of the Fundamental Theorem of Asset Pricing. *Math. Annalen* 300, 463–520.
- Delbaen, F. und W. Schachermayer (2000). Applications to Mathematical Finance. In: W. Johnson und J. Lindenstrauss (Hrsg.), *Handbook of the Geometrie of Banach Spaces*, S. 367–391.
- Dembo, A. und O. Zeitouni (1993). *Large Deviations Techniques and Applications*. Boston: Jones and Barlett.
- Deuschel, J.-D. und W. D. Strock (1989). *Large Deviations*. Boston: Academic Press.
- Dimson, E., P. Marsh und M. Staunton (2001). *Millenium Book II: 101 years of investment returns*. ABN AMRO: London Business School.
- Dimson, E. und M. Mussavian (2000). Market efficiency. *The Current State of Business Disciplines* 3, 959–970.
- Ding, M. und W. Yang (1995). Distribution of the first return time in fractional Brownian motion and its application to the study of on-off intermittency. *Phys. Rev. E* 52, 207–213.
- Ding, Z., C. W. J. Granger und R. F. Engle (1993). A long memory property of stock returns and a new model. *Journal of Empirical Finance* 1, 83–106.
- Dohmen, T. J. (2002). Building and using economic models: a case study analysis of the IS-LL model. *Journal of Economic Methodology* 9(2), 191–212.

- Doukan, P., G. Oppenheim und M. Taqqu (Hrsg.) (2003). *Theory and Applications of Long-Range Dependence*, Boston. Birkhäuser.
- Du, K. (2004). Efficient markets hypothesis and behavioral finance: A Kuhnian review. http://papers.ssrn.com/sol3/papers.cfm?abstract_id=617406.
- Dudley, R. M. und R. Norvaiša (1998). An introduction to p -variation and Young integrals. Marphysto, Centre for Mathematical Physics and Stochastics, University of Aarhus, Denmark.
- Duffie, D. (1988). *Security Markets*. San Diego: Academic Press.
- Duncan, T. E., Y. Hu und B. Pasik-Duncan (2000). Stochastic calculus for fractional Brownian motion I: theory. *SIAM Journal on Control and Optimization* 38(2), 582–612.
- Durrett, R. (1991). *Probability: Theory and Examples*. Belmont, CA: Wadsworth.
- Eberling, W., J. Freund und F. Schweitzer (1998). *Komplexe Strukturen: Entropie und Information*. Stuttgart: Teubner.
- Eggleston, H. G. (1949). The fractional dimension of a set defined by decimal properties. *J. Math. Oxford Ser.* 20, 31–36.
- Einstein, A. (1905). Über die von der Molekularkinetischen Theorie der Wärme geforderte Bewegung von in ruhenden Flüssigkeiten suspendierten Teilchen. *Annalen der Physik* 17, 549.
- Ekeland, I. und R. Temam (1976). *Convex Analysis and Variational Problems*. Amsterdam: North Holland.
- Elliott, R. J. und J. van der Hoek (2003). A general white noise theory and applications to finance. *Math. Finance* 13, 301–330.
- Ellis, R. S. (1985). *Entropy, Large Deviations and Statistical Mechanics*. New York: Springer.
- Ellis, R. S. (1995). An overview of the theory of large deviations and applications to statistical mechanics. *Scandinavian Actuarial Journal* 1, 97–142.
- Engle, R. (1982). Autoregressive conditional heteroskedasticity, with estimates of the variance of United Kingdom inflation. *Econometrica* 50, 987–1007.
- Ethier, S. und T. Kurtz (1986). *Markov Processes: Characterization and Convergence*. New York: Wiley.
- Everets, C., R. Hendrych, P. Singer und H.-O. Peitgen (1999). Fraktale Geometrie von Börsenzeitreihen: Neue Perspektiven ökonomischer Zeitreihenanalysen. In: K. Mainzer (Hrsg.), *Komplexe Systeme und nichtlineare Dynamik in Natur und Gesellschaft*, S. 400–419. Berlin: Springer.
- Everets, C. J. und B. B. Mandelbrot (1992). Multifraktale Maße. In: H.-O. Peitgen, H. Jürgen und D. Saupe (Hrsg.), *Bausteine des Chaos: Fraktale*. Springer.
- Falconer, K. J. (1993). *Fraktale Geometrie*. Heidelberg: Spektrum Akademischer Verlag. Original engl. Ausgabe 1990 bei John Wiley & Sons.

- Fama, E. (1963a). *The distribution of daily differences of stock prices: a test of Mandelbrot's stable Paretian hypothesis*. Ph. D. thesis, Graduate School of Business, University of Chicago.
- Fama, E. (1963b). Mandelbrot and the stable paretian hypothesis. *Journal of Business* 36, 420–429.
- Fama, E. (1965). The behavior of stock market prices. *Journal of Business* 38, 34–105.
- Fama, E. (1970). Efficient capital markets: A review of theory and empirical work. *Journal of Finance* 25, 383–417.
- Fama, E. (1976). *Foundations of Finance*. New York: Basic Books.
- Fama, E. (1991). Efficient capital markets, II. *Journal of Finance* 46, 1575–1617.
- Fama, E. (1998). Market efficiency, long-term returns, and behavioral finance. *Journal of Financial Economics* 49, 283–306.
- Fama, E., L. Fisher, M. C. Jensen und R. Roll (1969). The adjustment of stock prices to new information. *International Economic Review* 10, 1–21.
- Fama, E. und K. French (1988). Permanent and temporary components of stock prices. *Journal of Political Economy* 96, 246–273.
- Farmer, J. D. und A. W. Lo (1999, August). Frontiers in finance: Evolution and efficient markets. *Proc. Natl. Acad. Sci. USA* 96, 9991–9992.
- Feller, W. (1951). The asymptotic distribution of the range of sums of independent variables. *Annals of Mathematics and Statistics* 22, 427–432.
- Feller, W. (1957). *An Introduction to Probability Theory and its Applications*, Band I. New York: Wiley.
- Feller, W. (1968). *An Introduction to Probability Theory and its Applications*, Band II. New York: Wiley.
- Feynman, R. P. (2002). *Sechs physikalische Fingerübungen* (2. Auflage). München: Piper. Die Originalausgabe erschien 1995 unter dem Titel "Six Easy Pieces" bei Helix Books, Reading, Massachusetts.
- Fillol, J. (2003). Multifractality: Theory and evidence an application to the french stock market. *Economics Bulletin* 3(31), 1–12.
- Fisher, A., L. Calvet und B. B. Mandelbrot (1997). Multifractality of Deutsche Mark/US Dollar exchange rates. Cowles Foundation discussion paper No. 1166, Yale University.
- Flandrin, P. (1992, March). Wavelet analysis and synthesis of fractional Brownian motion. *IEEE Transactions on Informations Theory* 38(2), 910–916.
- Flandrin, P., P. Bognat und P.-O. Amblard (2002, Feb.). From stationarity to self-similarity, and back: Variations on the Lamperti transformation. Technical report, <http://www.ens-lyon.fr/flandrin/publirecentes.html>.

- Föllmer, H. und A. Schied (2002). *Stochastic Finance. An Introduction in Discrete Time*. Berlin: de Gruyter.
- Fournié, E., J.-M. Lasry, J. Lebuchoux und P.-L. Lions (2001). Applications of Malliavin calculus to Monte-Carlo methods in finance. II. *Finance and Stochastics* 5, 201–236.
- Frankfurter, G. und E. McGoun (2000). Market efficiency or behavioral finance: The nature of the debate. *The Journal of Psychology and Financial Markets* 1(3&4), 200–210.
- Frittelli, M. (2000). The minimal entropy martingale measure and the valuation problem in incomplete markets. *Math. Finance* 10, 39–52.
- Fujiwara, T. und Y. Miyahara (2003). The minimal entropy martingale measures for geometric Lévy processes. *Finance & Stochastics* 7, 509–531.
- Gärtner, J. (2001). Maß- und Integrationstheorie. <ftp://ftp.math.tu-berlin.de/pub/Lehre/Masstheorie/SS01/>.
- Geman, H., D. Madan und M. Yor (2000). Asset prices are Brownian motion: Only in business time. In: M. Avellaneda (Hrsg.), *Quantitative Analysis in Financial Markets*, S. 103–146. World Scientific.
- Gerber, H. U. und E. S. W. Shiu (1994). Option pricing by Esscher transforms. *Transactions of the Society of Actuaries* 4, 99–191.
- Geweke, J. und S. Porter-Hudak (1983). The estimation and application of long memory time series models. *Journal of Time Series Analysis* 4, 221–238.
- Ghashghaie, S., W. Breymann, J. Peinke, P. Talkner und Y. Dodge (1996). Turbulent cascades in foreign exchange markets. *Nature* 381, 767–770.
- Gibbard, A. und H. R. Varian (1978). Economic models. *The Journal of Philosophy* 75, 664–677.
- Gilbert, A., W. Willinger und A. Feldmann (1999, April). Scaling analysis of conservative cascades, with applications to network traffic. *IEEE Trans. on IT* 45(3), 971–991.
- Gnedenko, B. V. und A. N. Kolmogorov (1954). *Limit distributions for Sums of Independent Random Variables*. Reading, MA: Addison–Wesley. English translation by K. L. Chung.
- Gnedenko, B. W. (1991). *Einführung in die Wahrscheinlichkeitstheorie*. Berlin: Akademie Verlag.
- Gnedenko, B. W. und V. Y. Korolev (1996). *Random Summation: limit theorems and applications*. Boca Raton, Florida: CRC Press.
- Goll, T. und L. Rüschendorf (2001). Minimax and minimal distance martingale measures and their relationship to portfolio optimization. *Finance Stoch.* 5, 557–581.
- Gordon, S. (1991). The methodology of modelling. In: S. Gordon (Hrsg.), *The History and Philosophy of Social Science*, London, S. 100–110. Routledge.
- Grandshteyn, I. S. und I. M. Ryzhik (1980). *Table of Integrals, Series and Products*. New York: Academic Press.

- Granger, C. W. J. und R. Joyeux (1980). An introduction to long-memory time series models and fractional differencing. *Journal of Time Series Analysis* 1, 15–29.
- Grassberger, P. (1983). Generalized dimensions of strange attractors. *Phys. Lett.* 97, 227–230.
- Gray, R. M. (1988). *Probability, Random Processes, and Ergodic Properties*. Springer Verlag.
- Greene, M. T. und B. D. Fielitz (1977). Long-term dependence in common stock returns. *Journal of Financial Economics* 5, 339–349.
- Gripenberg, G. und I. Norros (1996). On the prediction of fractional Brownian motion. *Appl. Prob.* 33, 400–410.
- Grossman, S. J. (1976). On the efficiency of competitive stock markets where traders have diverse information. *Journal of Finance* 31, 573–585.
- Grossman, S. J. und J. E. Stiglitz (1980). On the impossibility of informationally efficient markets. *American Economic Review* 70, 393–408.
- Grünwald, P. und P. Vitanyi (2003). Kolmogorov complexity and information theory with an interpretation in terms of questions and answers. *J. Logic, Language, and Information* 12(4), 497–529.
- Guasoni, P. (2004). No arbitrage with transaction costs, fractional Brownian motion and Markov processes. Preprint Department of Mathematics, Boston University. <http://math.bu.edu/people/guasoni/>.
- Gulinsky, O. V. und A. Y. Veretennikov (1993). *Large Deviations for Discrete-Time Processes with Averaging*. Utrecht: VSP.
- Hakansson, N. H. (1992). Eintrag in Portfolio Analysis. New Palgrave Dictionary of Money and Finance, Peter Newman (ed.) London: MacMillan Press, 3, 150–153.
- Hartley, R. V. L. (1928, July). Transmission of information. *Bell Syst. Tech. J.* 7, 535–563.
- Hasley, T., M. Jensen, L. Kadanoff, I. Procaccia und B. Shraiman (1986). Fractal measures and their singularities: The characterization of strange sets. *Phys. Rev. A* 33(2), 1141–1151.
- Hausdorff, F. (1918). Dimension und äußeres Maß. *Mathematische Annalen* 79, 157–179.
- Havrda, J. und F. Charvat (1967). Quantification method of classification processes. *Kybernetika* 3, 30–35.
- Hawkins, D. (1988). The spirit of play. In: Necia Grant Cooper (Hrsg.), *From Cardinals to Chaos*, Cambridge. Cambridge University Press.
- Henry, M. und P. Zaffaroni (2003). The long-range dependence paradigm for macroeconomics and finance. In: P. Dounkan, G. Oppenheim und M. Taqqu (Hrsg.), *Theory and Applications of Long-Range Dependence*, Boston, S. 418–438. Birkhäuser.
- Hentchel, H. G. und I. Procaccia (1983). The infinite number of generalized dimensions of fractals and strange attractors. *Physica D* 8, 435–444.
- Hesse, C. (2003). *Angewandte Wahrscheinlichkeitstheorie*. Braunschweig: Vieweg.

- Heyde, C. C. (1999). A risky asset model with strong dependence through fractal activity time. *J. Appl. Prob.* 36, 1234–1239.
- Holden, H., B. Øksendal, J. Ubøe und T. Zhang (1996). *Stochastic Partial Differential Equations: a modelling, white noise functional analysis approach*. Boston: Birkhäuser.
- Holley, R. und E. Waymire (1992). Multifractal dimensions and scaling exponents for strongly bounded random cascades. *The Annals of Applied Probability* 2(4), 819–845.
- Holton, G. A. (1992). Time: The second dimension of risk. *Financial Analyst Journal* 48(6), 38–45.
- Hosking, J. R. M. (1981). Fractional differencing. *Biometrika* 68, 165–176.
- Howe, J. S., D. W. Martin und B. G. Wood, Jr. (1999). Much ado about nothing. Long-term memory in Pacific Rim equity markets. *International Review of Financial Analysis*. 8(2), 139–151.
- Hu, Y. (2001). Heat equations with fractional white noise potentials. *Applied Mathematics & Optimization* 43, 221–243.
- Hu, Y. und B. Øksendal (2003). Fractional white noise calculus and applications to finance. *Infin. Dimens. Anal. Quantum Probab. Relat. Top.* 6, 1–32.
- Hu, Y., B. Øksendal und A. Sulem (2003). Optimal consumption and portfolio in a Black-Scholes market driven by fractional Brownian motion. *Infin. Dimens. Anal. Quantum Probab. Relat. Top.* 6(4), 519–536.
- Hurst, H. E. (1951). Long-term storage capacity of reservoirs. *Transactions of the American Society of Civil Engineerings* 116, 770–779.
- Hurst, S. R., E. Platen und S. T. Rachev (1997). Subordinated market index models: a comparison. *Finan. Engin. Japan. Markets* 4, 97–124.
- Hurst, S. R., E. Platen und S. T. Rachev (1999). Option pricing for a logstable asset price model. In: S. Mittnik und S. T. Rachev (Hrsg.), *Stable Models in Finance*. Pergamon Press.
- Jacod, J. und A. Shiryaev (1987). *Limit Theorems for Stochastic Processes*. Berlin: Springer.
- Jaffard, S. (1999). Mathematical tools for multifractal signal processing. In: J. S. Byrnes (Hrsg.), *Signal Processing for Multimedia*, S. 111–127. IOS Press.
- Jänich, K. (2001). *Analysis für Physiker und Ingenieure*. Springer.
- Janicki, A. und A. Weron (1994). *Simulation and Chaotic Behavior of α -stable Stochastic Processes*. New York: Marcel Dekker.
- Jaynes, E. T. (1957). Information theory and statistical mechanics. *Phys. Reviews* 106, 620–630.
- Jaynes, E. T. (1982). On the rationale of Maximum-Entropy methods. In: *Proc. IEEE* 70, S. 939–952.
- Jeffreys, H. (1961). *Theory of Probability* (3. Auflage). Oxford: Clarendon Press.

- Jensen, J. L. (1995). *Saddlepoint Approximations*. Oxford: Clarendon Press.
- Jensen, M. C. (1978). Some anomalous evidence regarding market efficiency. *Journal of Financial Economics* 6, 95–101.
- Jonson, R. W. und J. E. Shore (1983). Comments on and correction to ‘axiomatic derivation of the principle of Maximum Entropy and the principle of Minimum Cross-Entropy’. *IEEE Trans. Inf. Theory* 29, 942–943.
- Kahane, J. P. und J. Peyrière (1976). Sur certaines martingales de Benoit Mandelbrot. *Adv. in Math.* 22, 131–145.
- Karatzas, I. und S. E. Shreve (1988). *Brownian Motion and Stochastic Calculus*. Springer.
- Kendall, M. G. (1953). The analysis of economic time series. Part 1. Prices. *Journal of the Royal Statistical Society* 96, 11–25.
- Kloeden, P. E. und E. Platen (1992). *Numerical Solution of Stochastic Differential Equations*. Berlin: Springer.
- Kolmogorov, A. N. (1933). *Grundbegriffe der Wahrscheinlichkeitsrechnung*. Berlin: Springer. Neuauflage 1977.
- Kolmogorov, A. N. (1940). The wiener helix and other interesting curves in hilbert space. *Doklady Acad. of Sciences of USSR* 26, 115–118. (Russian).
- Kolmogorov, A. N. (1941). The local structure of turbulence in an incompressible fluid at very high Reynolds numbers. In: S. K. Freidlander und L. Topper (Hrsg.), *Turbulence*, New York, S. 151–155. Interscience Publishers. Originalpublikation in: *Dokl. Acad. Nauk USSR*, 31 (1941), S. 99–101.
- Kolmogorov, A. N. (1962). A refinement of previous hypotheses concerning the local structure of turbulence in viscous incompressible fluid at high Reynolds number. *J. Fluid Mech.* 13, 82.
- Krämer, W., P. Sibberstein und C. Kleiber (2002). Long memory vs. structural change in financial time series. *Allgemeines Statistisches Archiv* 86, 83–96.
- Kuhn, T. S. (1970a). Reflections on my critics. In: I. Lakatos und A. Misgrave (Hrsg.), *Criticism and the Growth of Knowledge*, Cambridge, UK. Cambridge University Press.
- Kuhn, T. S. (1970b). *The Structure of Scientific Revolutions*. Chicago: The University of Chicago Press.
- Kullback, S. (1959). *Information Theorie and Statistics*. New York: Wiley.
- Kullback, S. und R. A. Leibler (1951). On information and sufficiency. *Ann. Math. Stat.* 22, 79–86.
- Lakatos, I. (1970). Falsification and the methodology of scientific research programmes. In: I. Lakatos und A. Musgrave (Hrsg.), *Criticism and the growth of knowledge*. Cambridge, UK: Cambridge University Press.

- Lakatos, I. (1978). The methodology of scientific research programs. In: J. Worall und G. Currie (Hrsg.), *Philosophical Papers*, Band I. Cambridge: Cambridge University Press.
- Lamperti, J. (1962). Semi-stable stochastic processes. *Transactions of the American Mathematical Society* 104, 62–78.
- Landsberg, P. T. (Hrsg.) (1971). *Problems in thermodynamics and statistical physics*, London. Pion.
- Laplace, P. S. (1774, 1814). *Essai Philosophique sur les Probabilités*. Paris. Die 6-te Ausgabe erschien in englischer Sprache 1951 bei Dover Publications, New York.
- Lawrence, A. J. und N. T. Kategoda (1977). Stochastic modelling of riverflow time series (with discussion). *J. Roy. Statist. Soc. (Ser. A)* 140, 1–47.
- LeBaron, B. (2001). Stochastic volatility as a simple generator of apparent financial power laws and long memory. *Quantitative Finance* 1, 621–631.
- Leeper, E., C. Sims und T. Zha (1996). What does monetary policy do? *Brookings Papers on Economic Activity* 2, 1–63.
- LeRoy, S. (1973). Risk aversion and the martingale property of stock prices. *International Economic Review* 14, 436–446.
- Lévy, P. (1937). *Théorie de l'Addition des Variables Aléatoires*. Paris: Gauthier-Villars. Zweite Ausgabe: 1954.
- Li, M. und P. Vitanyi (1993). *An Introduction to Kolmogorov Complexity and Its Applications*. Berlin: Springer-Verlag.
- Lin, S. J. (1995). Stochastic analysis of fractional Brownian motions. *Stochastics and Stochastic Reports* 55, 121–140.
- Lindstrøm, T., B. Øksendal und J. Ubøe (1992). Wick multiplication and Itô-Skorohod stochastic differential equations. In: S. Albeverio, J. E. Fenstad, H. Holden und T. Lindstrøm (Hrsg.), *Ideas and Methods in Mathematical Analysis, Stochastics and Applications*, S. 183–206. Cambridge University Press.
- Lintner, J. (1965). The valuation of risk assets and the selection of risky investments in stock portfolios and capital budgets. *Review of Economics and Statistics* 47, 13–37.
- Lipster, R. S. und A. N. Shiryaev (1989). *Theory of Martingales*. Dordrecht: Kluwer Acad. Publ.
- Lo, A. und A. C. MacKinlay (1988). Stock market prices do not follow random walks: Evidence from a simple specification test. *Review of Financial Studies* 1, 44–66.
- Lo, A. und A. C. MacKinlay (2002). *A Non-Random Walk Down Wall Street*. Princeton University Press.
- Lo, A. W. (1997). Fat tails, long memory, and the stock market since the 1960s. *Economic Notes* 2, 213–246.
- Lo, W. A. (1991). Long-term memory in stock market prices. *Econometrica* 59, 1279–1313.

- Lucas, R. (1978). Asset prices in an exchange economy. *Econometrica* 46, 1429–1445.
- Lux, T. (1996). The stable paretian hypothesis and the frequency of large returns: an examination of major German stocks. *Applied Financial Economics* 6, 463–475.
- Lux, T. (2001). Turbulence in financial markets: the surprising explanatory power of simple cascade models. *Quantitative Finance* 1, 632–640.
- Malkiel, B. G. (2000). *A Random Walk Down Wall Street* (7. Auflage). W. W. Norton & Company.
- Mallat, S. (1999). *A Wavelet Tour of Signal Processing*. New York: Academic Press.
- Mallat, S. und L. Hwang (1992). Singularity detection and processing with wavelets. *IEEE Transactions on Information Theory* 38(2), 617–642.
- Malliavin, P. (1997). *Stochastic Analysis*. Springer.
- Mandelbrot, B. (2005). Die tatsächlichen Risiken sind wesentlich größer, als wir annehmen. Das Interview in FAZ von 4. Juni 2005.
- Mandelbrot, B. und M. Taqqu (1979). Robust R/S analysis of long run serial correlation. In: *42nd Session of the International Statistical Institute, Manila, Book 2*, S. 69–99.
- Mandelbrot, B. und J. Wallis (1969). Robustness of the rescaled range R/S in the measurement of noncyclic long run statistical dependence. *Water Resources Research* 5, 967–988.
- Mandelbrot, B. B. (1963a). New methods in statistical economics. *Journal of Political Economy* 71, 421–440.
- Mandelbrot, B. B. (1963b). The variation of certain speculative prices. *Journal of Business* 36, 394–419.
- Mandelbrot, B. B. (1967). The variation of some other speculative prices. *Journal of Business* 40, 393–413.
- Mandelbrot, B. B. (1971). When can price be arbitrated efficiently? A limit to the validity of the random walk and martingale models. *Review of Economics and Statistics* 53, 225–236.
- Mandelbrot, B. B. (1972). Statistical methodology for non-periodic cycles: from the covariance to R/S analysis. *Annals of Economic and Social Measurement* 1, 259–290.
- Mandelbrot, B. B. (1974). Intermittent turbulence in self-similar cascades: divergence of high moments and dimension of the carrier. *Journal of Fluid Mechanics* 62(2), 331–358.
- Mandelbrot, B. B. (1975). Limit theorems of the self-normalized range for weakly and strongly dependent processes. *Zeitschrift für Wahrscheinlichkeit und verwandte Gebiete* 31, 271–285.
- Mandelbrot, B. B. (1982). *The Fractal Geometry of Nature*. San Francisco: W. H. Freeman and Co.
- Mandelbrot, B. B. (1987). *Die Fraktale Geometrie der Natur*. Basel: Birkhäuser.

- Mandelbrot, B. B. (1990). A class of multinomial multifractal measures with negative (latent) values for the “dimension” $f(\alpha)$. In: L. Pietronero (Hrsg.), *Fractals' Physical Origin and Properties*, S. 3–29. Springer.
- Mandelbrot, B. B. (1997). *Fractals and Scaling in Finance: Discontinuity, Concentration, Risk*. Berlin: Springer.
- Mandelbrot, B. B. (1999a, Februar). A multifractal walk down wall street. *Scientific American* 2, 50–53.
- Mandelbrot, B. B. (1999b). *Multifractals and 1/f Noise: Wild Self-Affinity in Physics*. Berlin: Springer.
- Mandelbrot, B. B. (1999c). Renormalization and fixed points in finance, since 1962. *Physica A* 263, 477–487.
- Mandelbrot, B. B. (2001a). Scaling in financial prices: I. Tails and dependence. *Quantitative Finance* 1, 113–123.
- Mandelbrot, B. B. (2001b). Scaling in financial prices: II. Multifractals and the star equation. *Quantitative Finance* 1, 124–130.
- Mandelbrot, B. B. (2001c). Scaling in financial prices: III. Cartoon Brownian motions in multifractal time. *Quantitative Finance* 1, 427–440.
- Mandelbrot, B. B. (2001d). Scaling in financial prices: IV. Multifractal concentration. *Quantitative Finance* 1, 641–649.
- Mandelbrot, B. B., A. Fisher und L. Calvet (1997). The multifractal model of asset returns. Cowles Foundation discussion paper No. 1164, Yale University.
- Mandelbrot, B. B. und R. L. Hudson (2004). *The (mis)Behavior of Markets: A Fractal View of Risk, Ruin and Reward*. New York: Basic Books.
- Mandelbrot, B. B. und H. Taylor (1967). On the distribution of stock price differences. *Operat. Res.* 15, 1057–1062.
- Mandelbrot, B. B. und van Ness (1958). Fractional Brownian motions, fractional noises and applications. *SIAM Rev.* 10, 422–437.
- Mansfield, P., S. Rachev und G. Samorodnitsky (2001). Long strange segments of a stochastic process. *The Annals of Applied Probability* 11(3), 878–921.
- Marinelli, C., S. T. Rachev, R. Roll und H. Göppel (1999). Subordinated stock price models: Heavy tails and long-range dependence in the high-frequency deutsche bank price record. In: G. Bol (Hrsg.), *Data Mining and Computational Finance*, S. 69–94. Physica.
- Marshall, A. (1890). *Principles of Economics*. Macmillan and Co., Ltd. Verfügbar via Internet: <http://www.econlib.org/library/Marshall/marPContents.html> oder <http://www.marxists.org/reference/subject/economics/marshall/index.htm>.
- Matteo, T. D., T. Aste und M. M. Dacorogna (2004). Using the scaling analysis to characterize financial markets. Finance 0402014, Economics Working Paper Archive at WUSTL. <http://ideas.repec.org/p/wpa/wuwpfi/0402014.html>.

- Matteo, T. D., T. Aste und M. M. Dacorogna (2005). Long-term memories of developed and emerging markets: Using the scaling analysis to characterize their stage of development. *Journal of Banking and Finance* 29, 827–851.
- Mauldin, J. (2004). *Bull's Eye Investing: Targeting Real Returns in a Smoke and Mirrors Market*. John Wiley & Sons.
- McCloskey, D. N. (1990). Storytelling in economics. In: D. Lavoie (Hrsg.), *Economics and Hermeneutics*, London. Routledge.
- McMillan, B. (1953). The basic theorems of information theory. *Ann. Math. Stat.* 24, 196–216.
- Meerschaert, M. und H. P. Scheffler (2001). *Limit Theorems for Sums of Independent Random Vectors: Heavy Tails in Theory and Practice*. New York: Wiley.
- Meneveau, C. und K. R. Sreenivasan (1987). Simple multifractal cascade model for fully developed turbulence. *Phys. Rev. Lett.* 59, 1424–1427.
- Meneveau, C. und K. R. Sreenivasan (1991). The multifractal nature of turbulent energy dissipation. *J. Fluid Mech.* 224, 429–484.
- Metzler, R. und J. Klafter (2000). The random walk's guide to anomalous diffusion: a fractional dynamics approach. *Phys. Reports* 339(1), 1–77.
- Mittnik, S., T. Doganoglu und D. Chenyao (1999). Computing the probability density function of the stable paretian distribution. *Math. Comput. Modelling* 29, 235–240.
- Mittnik, S., S. Rachev, T. Doganoglu und D. Chenyao (1999). Maximum likelihood estimation of stable paretian models. *Math. Comput. Modelling* 29, 275–293.
- Mittnik, S. und S. T. Rachev (1993). Modelling asset returns with alternative stable distributions. *Econometric Reviews* 12, 261–330.
- Mittnik, S. und S. T. Rachev (2000). *Stable Paretian Models in Finance*, Band 12. New York: Wiley.
- Molchan, G. M. (1999). Maximum of fractional Brownian motion: probabilities of small values. *Commun. Math. Phys.* 205(1), 97–111.
- Monroe, I. (1978). Processes that can be embedded in Brownian motion. *The Annals of Probability* 6(1), 42–56.
- Morgan, M. S. (2001). Models, stories and the economic world. *Journal of Economic Methodology* 8(3), 361–384.
- Müller, U. A., M. M. Dacorogna, R. D. Davé, O. V. Pictet, R. B. Olsen und J. R. Ward (1993). Fractals and intrinsic time — a challenge to econometricians. Research Report UAM. 1993-08-16, Olsen & Associates, Zürich.
- Muth, J. A. (1961). Rational expectations and the theory of price movements. *Econometrica* 29(6), 315–335.

- Necula, C. (2002). Option pricing in a fractional Brownian motion environment. Working Paper, Doctoral School of Finance and Banking, Academy of Economic Studies, Bucharest.
- Nolan, J. P. (1997). Numerical calculation of stable densities and distribution functions. *Comm. Statist. Stochastic Models* 13, 759–774.
- Nolan, J. P. (2001). Maximum likelihood estimation of stable parameters. In: O. Barndorff-Nielsen, T. Mikosch und S. Resnick (Hrsg.), *Lévy processes*. Boston: Birkhauser.
- Nolan, J. P. (2002). *Stable Distributions: Models for Heavy Tailed Data*. Boston: Birkhauser.
- Nordmeier, V. (1999). *Zugänge zur nichtlinearen Physik am Beispiel fraktaler Wachstumsphänomene: Ein generisches Fraktal-Konzept*. Münster: Lit. Universität Essen, Dissertation.
- Norros, I., E. Valkeila und J. Virtamo (1999). An elementary approach to a Girsanov formula and other analytical results on fractional Brownian motions. *Bernoulli* 5(4), 571–587.
- Norvaiša, R. (2000). Modelling of stock price changes: A real analysis approach. *Finance and Stochastics* 4, 343–369.
- Nualart, D. (1995). *The Malliavin Calculus and Related Topics*. Springer.
- Nuzman, C. und H. V. Poor (2000). Linear estimation of self-similar processes via Lamperti's transformation. *Journal of Applied Probability* 37(2), 429–452.
- Obukhov, A. M. (1941). Spectral energy distribution in a turbulent flow. *Dokl. Akad. Nauk SSSR* 32, 22–24.
- Obukhov, A. M. (1949). Structure of the temperature field in a turbulent flow. *Izv. Acad. Nauk SSSR Geogr. I Geofiz.* 13, 55.
- Obukhov, A. M. (1962). Some specific features of atmospheric turbulence. *J. Fluid Mech.* 13, 77.
- Øksendal, B. (1997). An introduction to Malliavin calculus with applications to economics. Working Paper, University of Oslo.
- Olsen, R. B. (1994, Dezember). Foreword. In: *Adaptive Computational Methods in Finance and Trading. Report of the Interek Group*, Zürich.
- Olsen, R. B., M. M. Dacorogna, U. A. Müller und O. V. Pictet (1992). Going back to the basics – rethinking market efficiency. Internal document RBO. 1992-09-07, Olsen & Associates.
- Oreskes, N., K. Shrader-Frechette und K. Belitz (1994). Verification, validation, and confirmation of numerical models in the earth sciences. *Science* 263(5147), 641–646.
- Ormerod, P. (1998). *Butterfly Economics: A New General Theory of Social and Economic Behavior*. Faber and Faber.
- Ortobelli, S., S. Rachev, S. Stoyanov und A. Biglova (2004). The proper use of risk measures in portfolio theory. Technical report, Universität Karlsruhe.

- Osborne, M. F. M. (1959). Brownian motion in the stock market. *Operations Research* 7, 145–173. Reprint in: Cootner (1964), 100–128.
- Overhaus, M., A. Ferraris, T. Knudsen, R. Milward, L. Nguyen-Ngoc und G. Schindlmayr (2001). *Equity Derivatives: Theory and Applications*. New York: John Wiley & Sons.
- Pagan, A. (1996). The econometrics of financial markets. *Journal of Empirical Finance* 3, 15–102.
- Pareto, V. (1897). *Cours d'economie politique*. London: .
- Parisi, G. und U. Frisch (1985). On the singularity structure of fully developed turbulence. In: M. Ghil, R. Benzi und G. Parisi (Hrsg.), *Turbulence and Predictability in Geophysical Fluid Dynamics, Proceedings International School of Physics 'E. Fermi', 1983, Varenna, Italy*, Amsterdam, S. 84–87. North-Holland.
- Park, K. und W. Willinger (2000). *Self-similar network traffic and performance evaluation*. New York: Wiley.
- Parzen, E. (1967). *Stochastic Processes*. San Francisco: Holden-Day.
- Peitgen, H.-O., H. Jürgens und D. Saupe (Hrsg.) (1992). *Bausteine des Chaos: Fraktale*. Springer.
- Peters, E. E. (1991). *Chaos and Order in the Capital Markets: A New View of Cycles, Prices and Market Volatility*. New York: Wiley.
- Peters, E. E. (1994). *Fractal Market Analysis: Applying Chaos Theory to Investment and Economics*. New York: Wiley.
- Petersen, K. E. (1989). *Ergodic Theory*. Cambridge University Press.
- Petrov, V. V. (1975). *Sums of independent Random Variables*. Berlin: Springer.
- Pipiras, V. und M. S. Taqqu (2000). Integration questions related to fractional Brownian motion. *Probability Theory and Related Fields* 118, 251–291.
- Pipiras, V. und M. S. Taqqu (2001). Are classes of deterministic integrands for fractional Brownian motion on an interval complete? *Bernoulli* 7(6), 873–897.
- Pollard, D. (2002). *A User's Guide to Measure Theoretic Probability*. Cambridge, UK: Cambridge University Press.
- Popper, K. (1949). Naturgesetze und theoretische Systeme. In: S. Moser (Hrsg.), *Gesetz und Wirklichkeit, Internationale Hochschulwochen des österreichischen College, Alpbach-Tirol, 21.8.–9.9.1948*. Innsbruck-Wien: Tyrolia-Verlag.
- Popper, K. (1972). Die Zielsetzung der Erfahrungswissenschaft. In: H. Albert (Hrsg.), *Theorie und Realität. Ausgewählte Aufsätze zur Wissenschaftslehre der Sozialwissenschaften*, S. 29–41. Tübingen: J. C. B. Mohr (Paul Siebeck).
- Poterba, J. und L. Summers (1988). Mean reversion in stock prices: Evidence and implications. *Journal of Financial Economics* 22, 27–59.

- Prato, D. und C. Tsallis (1999, August). Nonextensive foundation of Lévy distributions. *Phys. Rev. E* 60(2), 2398–2410.
- Protter, P. (1990). *Stochastic Integration and Differential Equations*. Berlin: Springer.
- Rachev, S. und G. Samorodnitsky (1993). Option pricing formulae for speculative prices modelled by subordinated stochastic processes. *Serdica–Bulgaricae Mathematicae Publicationes* 19, 176–190.
- Rachev, S. T. und L. Rüschendorf (1994). Models for option prices. *Theory Probab. Appl.* 39, 150–190.
- Rangarajan, G. und M. Ding (2000). First passage time distribution for anomalous diffusion. *Phys. Let. A* 273, 322–330.
- Rényi, A. (1961). On measures of entropy and information. In: *Proc. Fourth Berkeley Symp. Math. Stat. and Probability*, Band 1, Berkeley, CA, S. 547–561. University of California Press.
- Riedi, R. (2003). Multifractal processes. In: P. Dounkan, G. Oppenheim und M. Taqqu (Hrsg.), *Theory and Applications of Long-Range Dependence*, Boston, S. 625–716. Birkhäuser.
- Roberts, A. W. und D. E. Varberg (1973). *Convex Functions*. Academic Press.
- Roberts, H. V. (1959). Stock–market ‘patterns’ and financial analysis: Methodological suggestions. *Journal of Finance* 14, 1–10.
- Rockafellar, R. T. (1970). *Convex Analysis*. Princeton University Press.
- Rogers, L. C. G. (1997). Arbitrage with fractional Brownian motion. *Mathematical Finance* 7, 95–105.
- Roll, R. (1994). What every CEO should know about scientific progress in economics: what is known and what remains to be resolved. *Financial Management* 23(2), 69–75.
- Ross, S. (1976). The arbitrage theory of capital asset pricing. *Journal of Economic Theory* 13, 341–360.
- Rubinstein, M. (2001). Rational markets: Yes or No? The affirmative case. *Financial Analyst Journal* 49, 15–29.
- Salopek, D. M. (1998). Tolerance to arbitrage. *Stochastic Processes and their Applications* 76, 217–230.
- Samko, S. G., A. A. Kilbas und O. I. Marichev (1993). *Fractional Integrals and Derivatives*. London: Gordon and Breach Science.
- Samorodnitsky, G. und M. S. Taqqu (1994). *Stable Non-Gaussian Random Prozesses*. New York: Chapman and Hall.
- Samuelson, P. A. (1964). Rational theory of warrant pricing. In: Cootner (1964), 506–532.

- Samuelson, P. A. (1965). Proof that properly anticipated prices fluctuate randomly. *Industrial Management Review* 6, 41–49.
- Sato, K. (1999). *Lévy Processes and Infinitely Divisible Distributions*. Cambridge: Cambridge University Press.
- Sauer, A. (1994). *Faktormodelle und Bewertung am deutschen Aktienmarkt*. Frankfurt am Main: Knapp. Universität Karlsruhe, Dissertation.
- Schachermayer, W. (2001). Optimal investment in incomplete markets when wealth may become negative. *Ann. Appl. Probab.* 11, 694–734.
- Schertzer, D. und S. Lovejoy (1983). The dimension and intermittency of atmospheric dynamics. In: B. Launder (Hrsg.), *Turbulent Shear Flows 4*, S. 7–33. Springer-Verlag.
- Schleifer, A. (2000). *Inefficient Markets*. Oxford: Oxford University Press.
- Schroeder, M. (1991). *Fractals, Chaos, Power Laws*. New York: Freeman and Company.
- Schwert, G. W. (2002). Anomalies and market efficiency. Simon School of Business, University of Rochester.
- Seuront, L., F. Schmitt, Y. Lagadeuc, D. Schertzer und S. Lovejoy (1999). Universal multifractal analysis as a tool to characterize multiscale intermittent patterns: example of phytoplankton distribution in turbulent coastal waters. *Journal of Plankton Research* 21(5), 877–922.
- Shannon, C. E. (1948). A mathematical theory of communication (parts I and II). *Bell System Technical Journal* XXVII, 379–423, 623–656.
- Shannon, C. E. und W. Weaver (1949). *The Mathematical Theory of Communication*. Urbana: University of Illinois Press. Deutsche Ausgabe: *Mathematische Grundlagen der Informationstheorie*, Oldenburg, München, 1976.
- Sharp, W. (1964). Capital asset prices: A theory of market equilibrium under conditions of risk. *Journal of Finance* 19, 425–442.
- Sharpe, W., G. Alexander und J. Bailey (1999). *Investments* (6. Auflage). Englewood Cliffs: Prentice Hall.
- Shefrin, H. (2001). *Behavioral Finance*. UK: Edward Elgar Publishing, Inc.
- Shields, P. (1998, Oktober). The interactions between ergodic theory and information theory. *IEEE Trans. Info. Theory* 44, 2079–2093.
- Shiller, R. (1989). *Market Volatility*. Cambridge, Massachusetts: The MIT Press.
- Shiryayev, A. N. (1998). On arbitrage and replication for fractal models. Research Report Nr. 20, MaPhySto, Department of Mathematical Sciences, University of Aarhus, Denmark.
- Shiryayev, A. N. (1999a). *Essentials of Stochastic Finance*. Singapore: World Scientific.
- Shiryayev, A. N. (1999b). Kolmogorov and the Turbulence. MaPhySto Publication: MPS-misc 1999-12. MaPhySto, University of Aarhus, <http://www.maphysto.dk/>.

- Shlesinger, M., J. Klafter und G. Zumofen (1999). Above, below and beyond Brownian motion. *Am. J. Phys.* 67(12), 1253–1259.
- Shore, J. E. und R. W. Jonson (1980). Axiomatic derivation of the principle of Maximum Entropy and the principle of Minimum Cross-Entropy. *IEEE Trans. Inf. Theory* 26, 26–37.
- Shwartz, A. und A. Weiss (1995). *Large Deviations for Performance Analysis*. London: Chapman & Hall.
- Sinai, Y. G. (1997). Distribution of maximum of fractional Brownian motion. *Russian Math. Surveys* 52(2), 359–378.
- So, M. (2000). Long-term memory in stock market volatility. *Applied Financial Economics* 10, 519–524.
- Sottinen, T. (2001). Fractional Brownian motion, random walks and binary market models. *Finance & Stochastics* 5(3), 343–355.
- Sottinen, T. und E. Valkeila (2001, Oktober). Fractional Brownian motion as a model in finance. University of Helsinki. Department of Mathematics; PP 302.
- Souza, A. M. C. und C. Tsallis (1997,). Student's t - and r -distributions: Unified derivation from an entropic variational principle. *Physica A* 236(1–2), 52–57.
- Stanley, H. E. und V. Plerou (2001). Scaling and universality in economics: empirical results and theoretical interpretation. *Quantitative Finance* 1, 563–567.
- Steele, J. M. (2001). *Stochastic Calculus and Financial Applications*. New York: Springer-Verlag.
- Stevens, S. S. (1951). Mathematics, measurement, and psychophysics. In: S. S. Stevens (Hrsg.), *Handbook of Experimental Psychology*, New York. Wiley.
- Summers, L. (1986). Does the stock market rationally reflect fundamental values? *Journal of Finance* 41, 591–601.
- Sun, X., H. Chen, Z. Wu und Y. Yuan (2001). Multifractal analysis of Hang Seng index in Hong Kong stock market. *Physica A* 291, 553–562.
- Suppe, F. (1998). Suppe vs. Oreskes et al. Verification, Validation, and Confirmation of Simulation Models. May 7, 1998 PHIL 250.
- Takayasu, H. (1990). *Fractals in the physical sciences*. Manchester: Manchester University Press.
- Taleb, N. (2004). *Foiled by Randomness: The Hidden Role of Chance in Life and in the Markets* (2. Auflage). Thomson Texere.
- Taqqu, M. (2001). Bachelier and his times: A conversation with Bernard Bru. *Finance and Stochastics* 5, 3–32.
- Taqqu, M. (2003). Fractional brownian motion and long-range dependence. In: P. Dounkan, G. Oppenheim und M. Taqqu (Hrsg.), *Theory and Applications of Long-Range Dependence*, Boston, S. 5–38. Birkhäuser.

- Teverovsky, V., M. Taqqu und W. Willinger (1998). A critical look at Lo's modified R/S statistic. *Journal of Statistical Planning and Inference* 80, 211–227.
- Touchette, H. (2002). When is a quantity additive, and when is it extensive? *Physica A* 305, 84–88.
- Tribus, M. (1969). *The Principle of Maximum Entropy*. New York: Pergamon Press.
- Tsallis, C. (1988). Possible generalization of Boltzmann-Gibbs statistics. *Journal of Statistical Physics* 52(1–2), 479–487.
- Tsallis, C. (2001). Nonextensive statistical mechanics and thermodynamics: Historical background and present status. In: S. Abe und Y. Okamoto (Hrsg.), *Nonextensive Statistical Mechanics and Its Applications*, Berlin, S. 3–80. Springer.
- Tsallis, C., S. V. F. Levy, A. M. C. Souza und R. Maynard (1995). Statistical-mechanical foundation of the ubiquity of Lévy distributions in nature. *Phys. Rev. Lett.* 75, 3589.
- Uchaikin, V. und V. Zolotarev (1999). *Chance and Stability: Stable Distributions and Their Applications*. Utrecht: VSP.
- Vervaat, W. (1987). Properties of general self-similar processes. *Bull. Int. Stat. Inst.* 52, 199–216.
- Walter, C. (2001). Searching for scaling laws in distributional properties of price variations: a review over 40 years. In: *Transactions of the 11th AFIR International Colloquium*, Band 2, Toronto, S. 653–677.
- Walter, N. (2004). Wochin geht der dollar? *Börse am Sonntag* 29/04, 3.
- Weron, R. (2001). Lévy-stable distributions revisited: tail index > 2 does not exclude the Lévy-stable regime. *International Journal of Modern Physics C* 12(2), 209–223.
- Weron, R. (2002). Estimating long-range dependence: finite sample properties and confidence intervals. *Physica A* 312, 285–299.
- Wiener, N. (1923). Differential space. *Journal of Mathematical Physics. Math. Inst. Tech.* 2, 131–174.
- Williams, D. (1991). *Probability with Martingales*. Cambridge University Press.
- Willinger, W., M. Taqqu und V. Teverovsky (1999). Stock market prices and long-range dependence. *Finance and Stochastics* 3, 1–13.
- Working, H. (1934). A random-difference series for use in the analysis of time series. *Journal of American Statistical Association* 29, 11–24.
- Yajima, Y. (1985). On the estimation of long memory time series models. *Australian Journal of Statistics* 27, 303–320.
- Zanette, D. H. und P. A. Alemany (1995). Thermodynamics of anomalous diffusion. *Phys. Rev. Lett.* 75, 366–369.
- Zhang, P. G. (1998). *Exotic Options* (2. Auflage). Singapore: World Scientific.

Zipf, G. K. (1949). *Human behavior and the principle of least effort*. Cambridge, MA: AddisonWesley Press.

Zolotarev, V. M. (1986). *One-dimensional Stable Distributions*, Band 12. Providence, RI: Amer. Math. Soc.