Einheitswurzeltests bei bedingt heteroskedastischen Innovationen mit einer Anwendung auf deutsche Strompreisindizes

Dissertation

zur Erlangung des Doktorgrades

der Naturwissenschaften der Universität Dortmund

Dem Fachbereich Statistik der Universität Dortmund

vorgelegt von

Evgenij Hasanov

Juni 2004

Erstgutachter: Prof. Dr. Walter Krämer

Zweitgutachter: Prof. Dr. Götz Trenkler

Tag der mündlichen Prüfung: 20. Oktober 2004

Inhaltsverzeichnis

1 Einleitung **3**

2 Lineare stochastische Prozesse und Einheitswurzeltests **5**

2.1 Mixing- und NED-Prozesse 5

2.2 Stationarität und lineare Prozesse 9

2.2.1 ARMA- und ARIMA-Prozesse 10

2.2.2 Langes Gedächtnis und fraktionale Integration 13

2.3 Einheitwurzeltests 15

2.3.1 Tests von Dickey und Fuller 16

2.3.2 Alternative Einheitwurzeltests 24

3 Modellierung von Volatilitäten **27**

3.1 ARCH- und GARCH-Prozesse 28

3.2 Integrierte GARCH-Prozesse 32

3.3 Asymmetrische GARCH-Prozesse 37

3.4 Eigenschaften der GARCH-Prozesse 40

3.5 Der GRS-ARCH-Prozess 43

4 Tests auf stochastische Trends bei bedingter Heteroskedastie **49**

4.1 Einheitwurzeltests und GARCH-Innovationen 49

4.2 Einheitswurzeltests und GRS-ARCH-Prozess 51

5 Strompreisprozesse und -volatilitäten **66**

5.1 Organisation und Struktur des Stromhandels 67
5.1.1 Marktsegmente 67
5.1.2 Börsenhandel 68
5.1.3 Gehandelte Produkte und Marktliquidität 72
5.2 Modellierung des Preisprozesses 73
5.2.1 Modell mit deterministischem Trend 75
5.2.2 Autoregressives Modell mit stochastischem Trend 80

6 Zusammenfassung und Ausblick 83

Abkürzungen und Symbole 86

Literaturverzeichnis 88

Kapitel 1

Einleitung

Die vorliegende Arbeit führt zwei in den letzten Jahrzehnten intensiv beforschte Fragenkomplexe der empirischen Wirtschaftsforschung zusammen: die Erkennung von stochastischen Trends und die Modellierung von Volatilitäten.
Betrachtet man die These des effizienten Marktes, so gehören die so genannten Einheitwurzeltests zum Standardwerkzeug bei der Datenanalyse. Mit diesen Testverfahren, die von Fuller (1976), Dickey und Fuller (1979), Philipps und Perron (1988) vorgeschlagen und von zahlreichen Autoren weiter entwickelt wurden, überprüft man die Random-Walk-Hypothese, aber auch breiter gefasste Behauptungen über das instationäre Verhalten interessierender Größen.
Geht man zur Beschreibung der Marktdynamik über, so führt in der Regel kein Weg an der Modellierung des Schwankungsverhaltens interessierender Größen vorbei. Üblicherweise werden hierfür Modelle der GARCH-Familie eingesetzt, deren zahlreiche Varianten als Verallgemeinerungen der von Engle (1982) und Bollerslev (1986) eingeführten Ansätze angesehen werden können. Trotz einer kaum überschaubaren Literatur zu diesen Themen werden sie bis heute meist separat betrachtet, obwohl sowohl stochastische Trends als auch bedingte Heteroskedastie typisch für viele Finanz- und Preiszeitreihen sind. Nur wenige Beiträge, darunter Kim und Schmidt (1993), Ling und Li (1998), Seo (1999) beschäftigen sich mit den beiden Phänomenen gleichzeitig. Die folgende Untersuchung zielt darauf ab, diese Lücke weiter zu schließen, indem die asymptotische Verteilung der Dickey-Fuller-Teststatistiken bei Vorliegen gewisser Volatilitätsmuster analytisch hergeleitet wird. Ferner werden zur praktischen Durchführung der Tests die durch Simulationen ermittelten kritischen Werte angeführt. Im Mittelpunkt der theoretischen Ausarbeitung steht dabei das von Giraitis, Robinson und Surgailis (2000) eingeführte lineare ARCH-Modell, das als einziges unter den

bisher bekannten Modellen der GARCH-Klasse langfristige Strukturen in den Volatilitäten und die schwache Stationarität der zugrunde liegenden Reihe in sich vereinen kann. Ob dieses Modell auch Vorteile bei der Datenanpassung mit sich bringt - eine Frage, die noch keine Beachtung in der Literatur fand - soll anhand einer empirischen Untersuchung geklärt werden.

Als Anwendungsgebiet für die diskutierten Test- und Modellverfahren dienen europäische Strompreisindizes. Während ihre Rolle als Indikatoren für die Preisentwicklung im Zuge der Liberalisierung am europäischen Energiemarkt immer wichtiger wird, sind komplexe zeitreihenanalytische Auswertungen unter Einbeziehung moderner methodischer Ansätze bislang ausgeblieben. Da die Strompreise aufgrund zahlreicher spezifischer Faktoren (marktdominierende Stellung weniger großer Marktteilnehmer, begrenzte Durchleitungskapazitäten, sich verändernde Handelsrahmenbedingungen u. a.) im Allgemeinen größeren Schwankungen unterliegen als etwa die Aktienkurse, rückt insbesondere die Frage nach geeigneten Modellen zur Volatilitätenbeschreibung in den Vordergrund. Im Rahmen der Datenauswertung sind deswegen zwei Punkte wesentlich. Erstens sollen Modelle gefunden werden, die die Strompreisvolatilitäten gut abbilden, und zweitens sind die entsprechenden Volatilitätsstrukturen bei den Aussagen über die Stationarität von Strompreisraten besonders zu beachten.

Kapitel 2

Lineare stochastische Prozesse und Einheitswurzeltests

Zum Einstieg in den theoretischen Teil der Arbeit werden zum einen einige allgemeine Eigenschaften von diskreten stochastischen Prozessen betrachtet, wobei der Schwerpunkt auf ihrem asymptotischen Abhängigkeitsverhalten liegt. Zum anderen werden grundlegende Modellansätze zur Beschreibung serieller linearer Abhängigkeiten eingeführt. Da, wo es möglich ist, findet eine Einordnung der Modelle in das vorgestellte asymptotische Konzept statt.

2.1 Mixing- und NED-Prozesse

Unter einem *diskreten stochastischen Prozess* $\{X_t\}$ versteht man eine Folge der Zufallsvariablen $X_t(\omega)$ mit $t \in \mathbb{Z}$, $\omega \in \Omega$. Als eine *Zeitreihe* $\{x_t\}$ wird eine Realisation von $\{X_t\}$ bzw. bei empirischen Untersuchungen ein Ausschnitt aus dieser Realisation mit der Länge n: $\{x_1, x_2, \ldots, x_n\}$ bezeichnet. Da in der Praxis die meisten zeitlich geordneten Daten serielle Abhängigkeiten aufweisen, ist es von besonderem Interesse, diejenigen stochastischen Prozesse anzuschauen, die ebenfalls verschiedene Abhängigkeitsmuster enthalten. Hier begegnet man der Frage, inwieweit man gewöhnliche statistische Methoden bei den Aussagen über stochastische Prozesse einsetzten darf. Dies ist von Fall zu Fall zu überprüfen. Sowohl das Gesetz der großen Zahlen als auch der zentrale Grenzwertsatz, auf denen die meisten statistischen Schätz- und Testverfahren aufbauen, gehen in ihren Aussagen im Standardfall von unabhängigen Zufallsvariablen aus. Deswegen ist es von enormer Bedeutung festzustellen, ob und wenn ja, welcher Art Abhängigkeitsstrukturen erlaubt sind, damit die genannten Sätze ihre Geltung beibehalten und folglich die Standardmethodik sinnvoll eingesetzt werden kann.

Im Zusammenhang mit dieser Problematik unterscheidet man zwischen dem kurzfristigen und dem langfristigen Charakter der Zeitabhängigkeiten. Falls die Abhängigkeit zweier Zufallsvariablen X_t und X_{t+h}, die zeitlich genügend weit voneinander entfernt liegen, verschwindend gering ist, spricht man von der asymptotischen Unabhängigkeit des zugehörigen stochastischen Prozesses. Dabei liefert eine aktuelle Beobachtung x_t keinen feststellbaren Beitrag zur Erklärung der in großen zeitlichen Abständen h liegenden Realisationen der entsprechenden Prozessvariablen X_{t-h} bzw. X_{t+h}. Eine Möglichkeit, diese Eigenschaft formal auszudrücken, bieten sogenannte *Mixing*-Prozesse. Laut Gallant und White (1988) unterscheidet man zwischen zwei alternativen Formen dieser Prozesse:

Definition 1 (Mixing-Prozesse)

Sei für $s \leq t$ $\mathcal{F}_s^t = \sigma(X_s, \ldots, X_t)$. Dann nennt man $\{X_t\}$ strong mixing oder α-mixing, falls

$$\alpha_m := \sup_t \sup_{A,B} |P(A \cap B) - P(A)P(B)| \stackrel{m\to\infty}{\longrightarrow} 0\,, \tag{2.1}$$

und uniform oder ϕ-mixing, falls

$$\phi_m := \sup_t \sup_{A,B} |P(B|A) - P(B)| \stackrel{m\to\infty}{\longrightarrow} 0 \tag{2.2}$$

mit $A \in \mathcal{F}_{-\infty}^t,\ B \in \mathcal{F}_{t+m}^\infty,\ P(A) > 0.$

Verglichen miteinander, handelt es sich bei (2.2) um eine stärkere Forderung: aus $P(B|A) = P(A \cap B)/P(A)$ und mit $0 < P(A) \leq 1$ für alle A folgt $\phi_m \geq \alpha_m$. Somit stellen die uniform mixing-Prozesse nur eine Teilmenge der strong mixing-Prozesse dar, auch wenn die Bezeichnung es anders suggeriert.

Die angeführte Definition erlaubt, das asymptotische Abhängigkeitsverhalten der Reihenelemente zu quantifizieren. Als ein geeigneter Grad dient die Rate, mit der α_m bzw. ϕ_m gegen Null konvergieren. So heißt $\{X_t\}$ *strong mixing der Ordnung* $-a$, falls $\alpha_m = O(m^{-a-\varepsilon})$ für ein beliebiges $\varepsilon > 0$. Analoge Bezeichnung gilt auch für die uniform mixing-Prozesse.

Die mixing-Eigenschaft erlaubt unter gewissen Einschränkungen an die Ordnung a weiterhin die Anwendung des Zentralen Grenzwertsatzes bzw. des Funktionalen Zentralen

Grenzwertsatzes[1], sowie der Gesetze der großen Zahlen für die Prozessvariablen X_t. Für Einzelheiten siehe z. B. Davidson (2002a), Theorem 1.1.
Zu den Vorteilen der mixing-Prozesse zählt ihre Modellfreiheit und die Eigenschaft, dass jede messbare Funktion von mixing-Prozessen wieder mixing der gleichen Ordnung ist. Dies erlaubt einen flexiblen Einsatz dieses Konzepts. Allerdings sind die Beziehungen (2.1) bzw. (2.2) direkt nur schwer nachweisbar. Dazu müssen Suprema über alle möglichen Zufallsereignisse entlang der gesamten Zeitachse untersucht werden, was die Spezifikation der betreffenden Verteilungen erfordert. Probleme ergeben sich auch dann, wenn man Prozesse betrachtet, die die gesamte Vergangenheit bzw. Zukunft anderer Prozesse (etwa die der Zufallseffekte) beinhalten. Gerade dieser Fall trifft bei ökonometrischen Modellen häufig auf, wenn man eine Variable in Bezug zu ihrer eigenen Vergangenheit setzt. In solchen Fällen greift die Definition 1 manchmal selbst bei offensichtlich kurzfristigen Abhängigkeitsmustern nicht immer. Das viel zitierte Beispiel dazu stammt von Andrews (1984). Er zeigt, dass ein autoregressiver Prozess erster Ordnung $\{X_t\}$ mit $X_t = \rho X_{t-1} + \varepsilon_t$, $0 < \rho < 1/2$ und unabhängigen Bernoulli-verteilten Innovationen ε_t keine mixing-Eigenschaft besitzt. Als hinreichend für die strong mixing-Eigenschaft wird daher eine gewisse glatte Gestalt der Dichtefunktion der Innovationen ε_t vorausgesetzt, womit etwa die diskreten Verteilungen ausgeschlossen werden. Genaue Bedingungen formuliert Davidson (1994) für strong mixing im Theorem 14.9 und für uniform mixing im Theorem 14.14. Durch das letztere wird u. a. die Normalverteilung ausgeschlossen. Folglich bleiben viele Prozesse, deren serielle Abhängigkeiten kurzfristiger Natur sind und für große Zeitabstände nahezu verschwinden, durch die Definition 1 nicht erfasst.

All dies führt zu einer eingeschränkten praktischen Relevanz der mixing-Prozesse und ruft nach einer alternativen Formulierung der asymptotischen Unabhängigkeit in der Zeitreihentheorie. Ein Konzept, das auf der mixing-Eigenschaft aufbaut und sich in der letzten Zeit mehr und mehr durchsetzt, wird im Folgenden vorgestellt. Die Rede ist von dem Begriff der so genannten *Near Epoch Dependence* (kurz: *NED*), das u.a. auf Billingsley (1968) zurückgeht und von McLeish (1975) weiterentwickelt wurde. Mit den wichtigen Anwendungen der NED-Prozesse beschäftigen sich Gallant und White (1988) und Davidson (2000) und (2002a).
Das Vorgehen besteht darin, parallel zum eigentlich interessierenden Prozess $\{X_t\}$ einen Prozess $\{U_t\}$ zu betrachten, der in einer funktionalen Beziehung zu $\{X_t\}$ steht. Genauer gesagt, schaut man sich die Abhängigkeitsstruktur von $\{X_t\}$ bezüglich $\{U_t\}$ an.

[1]Für die genaue Formulierung des funktionalen Zentralen Grenzwertsatzes siehe Seite 17.

Wie die Bezeichnung *Near Epoch Dependence* verrät, muss bei einem NED-Prozess die aktuelle Zufallsgröße X_t für jedes t maßgebend durch die zeitlich naheliegenden Variablen $U_{t-m}, \ldots, U_{t+m}$ bestimmt werden. Der Einfluss weiter liegender Vergangenheit oder Zukunft von $\{U_t\}$ auf X_t wird somit als vernachlässigbar angesehen. Formal kann diese Idee mit Hilfe der L_p-Norm $\|\cdot\|_p := \sqrt[p]{E(\cdot)^p}$, $p \geq 1$ wie folgt aufgefasst werden:

Definition 2 (NED-Prozesse)
Sei $\{X_t\}$ ein stochastischer Prozess mit $E(X_t^p) < \infty$ für alle t, so dass $X_t(\ldots, U_{t-1}, U_t, U_{t+1}, \ldots)$ für jedes t als messbare Funktion von $\{U_t\}$ darstellbar ist. Sei weiterhin $\mathcal{F}_s^t = \sigma(U_s, \ldots, U_t)$ für $s \leq t$ und $a > 0$. Dann heißt $\{X_t\}$ ein L_p-NED-Prozess der Ordnung $-a$ bzgl. $\{U_t\}$, falls für alle t

$$\|X_t - E(X_t|\mathcal{F}_{t-m}^{t+m})\|_p \leq \nu_m \;\; \textit{und} \;\; \nu_m = O(m^{-a-\varepsilon}) \;\; \textit{für beliebiges}\, \varepsilon > 0,\; m \to \infty. \quad (2.3)$$

Durch ν_m wird zwar nicht direkt das Abhängigkeitsmuster in $\{X_t\}$ beschrieben, sondern der dynamische Zusammenhang zwischen $\{X_t\}$ und $\{U_t\}$. Man kann aber auf das Abhängigkeitsverhalten der Variablen X_t schließen, indem man eine entsprechende Annahme über $\{U_t\}$ macht. Besteht etwa $\{U_t\}$ aus stochastisch unabhängigen Variablen, so gibt v_m unmittelbar Auskunft über das Abklingen serieller Abhängigkeiten in dem zugehörigen NED-Prozess $\{X_t\}$ mit zunehmendem Zeitabstand. Allgemeiner lässt sich sagen, dass die NED-Eigenschaft bezüglich eines mixing-Prozesses gleichbedeutend mit auf lange Sicht verschwindenden seriellen Abhängigkeiten in $\{X_t\}$ ist. Ob diese Eingrenzung der Abhängigkeitsstruktur die Gültigkeit üblicher Grenzwertsätze gewährleistet, hängt von der Ordnung $-a$ ab. Analog zu den mixing-Prozessen gibt a Auskunft über die Langfristigkeit zeitlicher Zusammenhänge: je größer a, desto schneller nimmt der Einfluss vergangener bzw. künftiger Werte von $\{U_t\}$ auf das aktuelle X_t ab. Wird X_t durch eine endliche Anzahl von U_t komplett abgebildet, so ist X_t $\mathcal{F}_{t-l}^{t+l}$-messbar für ein $l < \infty$, und $\nu_m = 0$ für alle $m > l$. Die NED-Ordnung ist somit unendlich. Bei geometrisch abnehmenden Abhängigkeiten ($\nu_m = O(k^m)$, $0 < k < 1$) setzt man ebenfalls $a = \infty$ und spricht sinngemäß von der *exponentiellen* oder *geometrischen* NED-Eigenschaft. Ausführliche Untersuchungen asymptotischer Eigenschaften der NED-Prozesse nehmen De Jong und Davidson (2000) sowie Davidson (2002a) vor. Bedingt durch die spätere Anwendung kann in dieser Arbeit eine Einschränkung auf unabhängig identisch verteilte U_t und $p = 2$ vorgenommen werden. In diesem Spezialfall erweist sich $a \geq 1/2$ als hinreichend für die Gültigkeit der zentralen Grenzwertsätze und folglich auch des schwachen Gesetzes der großen Zahlen.

Es bleibt noch anzumerken, dass die NED- im Gegensatz zu der mixing-Eigenschaft nicht invariant gegenüber beliebigen Transformationen f mit $f(X_t) < \infty$ ist. Insbesondere kann sich dann die Ordnung $-a$ des resultierenden Prozesses ändern. Einzelheiten dazu findet man bei Gallant und White (1988) im Corollar 4.3.

2.2 Stationarität und lineare Prozesse

Um stochastische Prozesse durch Zeitreihenmodelle abbilden zu können, formuliert man bestimmte Restriktionen an die Variablen X_t, die das Gesamtbild einer Zeitreihe betreffen. Eine zentrale Rolle kommt in diesem Zusammenhang dem Begriff der *Stationarität* zu. Man unterscheidet zwischen verschiedenen Arten von Stationarität, je nach dem, ob gesamte Verteilungen oder nur bestimmte Verteilungsmomente betrachtet werden.

Definition **3 (stationäre Prozesse)**

Ein stochastischer Prozess $\{X_t\}$ *heißt*

- *streng stationär, falls für alle* $k > 0$ *die gemeinsame Verteilung von* $(X_t, X_{t+1}, \ldots, X_{t+k})$ *unabhängig von* t *ist;*
- *mittelwertstationär, falls* $E(X_t) =: \mu$ *konstant für alle* t *ist;*
- *kovarianzstationär, falls* $Cov(X_t, X_{t+h}) =: \gamma(h)$ *nur vom Abstand* h*, nicht aber von* t *abhängt;*
- *schwach stationär, falls er mittelwert- und kovarianzstationär ist.*

Wie die Benennung suggeriert, stellen streng stationäre Prozesse eine Teilmenge der schwach stationären dar. Dies ist aber nur dann wahr, wenn die Varianz von X_t existiert, und trifft gerade für die im nächsten Kapitel behandelten bedingt heteroskedastischen Prozesse im Allgemeinen nicht zu. Bollerslev et al. (1994) unterstreichen, dass für solche Prozesse Stationarität in Momenten stärkere Annahmen erfordert als Stationarität in der gesamten Verteilung. Dies hängt damit zusammen, dass man je nach Modellart und Parameterwerten Verteilungen mit unendlichen unbedingten Varianzen begegnet. Wie später gezeigt wird, erweist sich die Existenz der Varianzen unterstellende schwache Stationarität in bedingt heteroskedastischen Prozessen manchmal als hinreichend für die strenge Stationarität.

Um gleich zum Einstieg Komplikationen zu vermeiden, werden in diesem Kapitel nur Verteilungen mit endlichen zweiten Momenten zugelassen, so dass die streng stationären Prozesse automatisch schwach stationär sind. Ferner werden grundsätzlich stationäre Prozesse mit $\mu = 0$ betrachtet, die bei der Modellierung zentrierter Datenreihen eingesetzt werden.

2.2.1 ARMA- und ARIMA-Prozesse

Einen Grundbaustein für jeden stationären Prozess stellt das sogenannte *weiße Rauschen* dar. Als weißes Rauschen wird ein spezieller Prozess $\{\eta_t\}$ bezeichnet, deren Elemente, im Weiteren auch *Zufallsschocks* oder *Schocks* genannt, unabhängig identisch verteilt (kurz: u.i.v.) mit Erwartungswert 0 und Varianz σ^2 sind.
Aufbauend auf dem weißen Rauschen definiert man *lineare Prozesse*. Existiert ein weißes Rauschen $\{\eta_t\}$, so dass gilt:

$$X_t = \sum_{i=0}^{\infty} c_i \eta_{t-i}, \quad c_i \in \mathrm{IR}\ \forall\ i,$$

dann handelt es sich bei $\{X_t\}$ um einen linearen Prozess mit Erwartungswert 0. Die serielle Abhängigkeitsstruktur von $\{X_t\}$ wird somit vollständig durch die Koeffizientenfolge $\{c_i\}$ festgelegt. Besitzt ein linearer Prozess endliche Varianz, so ist er genau dann stationär, wenn die Koeffizienten c_i quadratisch summierbar sind.
Fordert man statt u.i.v. lediglich unkorrelierte Schocks, so wird der zugehörige Prozess $\{\check{\eta}_t\}$ *schwaches weißes Rauschen* genannt. Nach dem Theorem von Wold kann jeder schwach stationäre Prozess ohne deterministische Komponente $\{X_t\}$ als eine Lösung von

$$X_t = \sum_{i=0}^{\infty} c_i \check{\eta}_{t-i}, \ \sum_{i=1}^{\infty} c_i^2 < \infty, \ c_0 = 1$$

dargestellt werden, vgl. Davidson (2000), S. 90.
Weißes Rauschen bildet die Grundlage zur Definition von *autoregressiven moving average Prozessen* (kurz: *ARMA*), mit denen man standardmäßig versucht, vermutete lineare Abhängigkeiten innerhalb einer Zeitreihe zu erfassen. Konkret bezeichnet man $\{X_t\}$ als einen ARMA(p, q)-Prozess, falls es ein weißes Rauschen $\{\eta_t\}$ gibt, so dass gilt:

$$X_t - \rho_1 X_{t-1} - \ldots - \rho_p X_{t-p} = \eta_t - \theta_1 \eta_{t-1} - \ldots - \theta_q \eta_{t-q}. \tag{2.4}$$

Betrachtet man die Schocks η_t als Prozesskomponenten von $\{X_t\}$, so werden sie häufig als *Prozessinnovationen* oder einfach *Innovationen* bezeichnet. Formal lässt sich eine Innovation in einem linearen Prozess $\{X_t\}$ als $\eta_t = X_t - E(X_t|\mathcal{F}_{t-1})$ auffassen.
Um lineare Kombinationen verzögerter Prozessvariablen kürzer darzustellen, wird im Weiteren häufig der Backshift-Operator L mit $L^k(\{X_t\}) := \{X_{t-k}\}$, $k \in \mathbb{Z}$ benutzt. Damit können die linearen Filter mit den Koeffizienten ρ_i bzw. θ_i aus (2.4) als $\rho(L)$ bzw. $\theta(L)$ bezeichnet werden, und ein ARMA-Prozess erhält die Form

$$(1 - \rho(L))\{X_t\} = (1 - \theta(L))\{\eta_t\}. \tag{2.5}$$

Für eine Konstante $X_t = X \in \mathbb{R}$ für alle t gilt entsprechend

$$\rho(L)\{X\} = \rho(1)\{X\} = X \sum_{i=1}^{p} \rho_i.$$

Durch die Annahme, dass die zu den Filtern $\rho(L)$ bzw. $\theta(L)$ gehörenden z-Transformierten

$$\rho(z) = \rho_1 z - \ldots - \rho_p z^p \;\text{ und }\; \theta(z) = \theta_1 z - \ldots - \theta_q z^q, \;\; z \in \mathbb{C}\,,$$

keine gemeinsamen Nullstellen haben und alle Wurzeln von $1 - \rho(z)$ betragsmäßig größer 1 sind, wird ferner stets die Eindeutigkeit einer ARMA-Darstellung garantiert. Die zugehörige MA(∞)-Darstellung hat dann die Form $X_t = (1-\rho(L))^{-1}(1-\theta(L))\{\eta_t\}$, wobei die zum Filter $(1-\rho(L))^{-1}(1-\theta(L))$ gehörenden Koeffizienten absolut summierbar sind. Übersteigen alle Wurzeln des Polynoms $1 - \theta(z)$ im Betrag den Wert 1, so kann der zugehörige ARMA-Prozess als AR(∞) umgeschrieben werden. Er heißt dann *invertierbar*.
Kennzeichnend für stationäre ARMA-Prozesse ist ein schnelles Abklingen der seriellen Abhängigkeiten mit zunehmendem Abstand. Als Maß hierfür dienende Autokovarianzen fallen betragsmäßig exponentiell ab und sind absolut summierbar:

$$\sum_{h=1}^{\infty} |Cov(X_t, X_{t+h})| < \infty. \tag{2.6}$$

Diese Eigenschaft wird als *kurzes Gedächtnis* des zugehörigen Prozesses aufgefasst, denn der Einfluss von X_t auf die künftige Reihenentwicklung ist zeitlich eng begrenzt und somit von kurzfristiger Natur. An diesem Verhalten stellt man eine offensichtliche Ähnlichkeit zu mixing- bzw. NED-Prozessen fest. Während nicht alle stationären

ARMA-Prozesse die mixing-Eigenschaft besitzen, können sie nach der Behauptung 2.1 von Davidson (2002a) generell als exponentielle L_2-NED-Prozesse bzgl. der eigenen Innovationen $\{\eta_t\}$ angesehen werden.

In der empirischen Wirtschaftsforschung begegnet man meistens instationären Reihen. Selbst nach einer Bereinigung von eventuellen deterministischen Komponenten lässt sich eine empirische Reihe oft erst nach der Differenzenbildung, bezeichnet mit Δ oder $(1-L)$, als ein stationärer Prozess mit kurzem Gedächtnis modellieren. Anders ausgedrückt, kann man eine empirische Reihe häufig als einen Prozess $\{Y_t\}$ darstellen, der durch das Kumulieren bzw. Integrieren von stationären Variablen X_t mit $VarX_t \neq -2\sum_{h=1}^{\infty} Cov(X_t, X_{t+h})$ zustandekommt:

$$X_t = Y_t - Y_{t-1} = \Delta Y_t = (1-L)\{Y_t\} \text{ für alle } t \in \mathbb{Z} \implies Y_t = \sum_{i=0}^{t} X_i.$$

Daher nennt man den Prozess $\{Y_t\}$ *integriert*. Handelt es sich bei $\{X_t\}$ um einen ARMA(p,q)-Prozess wie in (2.5), so nennt man den zugehörigen Partialsummenprozess $\{Y_t\}$ demnach *ARIMA*(p,q). Er hat die Form

$$(1-\rho(L))(1-L)\{Y_t\} = (1-\theta(L))\{\eta_t\}. \tag{2.7}$$

Aus dieser Darstellung wird ersichtlich, dass ein ARIMA-Prozess als ein ARMA-Prozess angesehen werden kann, dessen z-Transformierte eine *Einheitswurzel*, d. h. eine Nullstelle bei 1 hat. Denkbar ist auch ein Prozess, bei dem man erst nach der zweifachen Differenzenbildung zu Stationärität und absolut summierbaren Autokovarianzen gelangt, was einer zweifachen Einheitswurzel entspricht. Einen solchen Prozess nennt man integriert der Ordnung 2. Allgemein wird für integrierte Prozesse die zugehörige Ordnung d, auch Differenzenparameter genannt, angegeben. Während ein stationärer ARMA-Prozess trivialerweise als integriert der Ordnung 0 (kurz: I(0)) aufgefasst werden kann, wird $\{Y_t\}$ aus (2.7) als I(1) bezeichnet. Den einfachsten Spezialfall eines I(1)-Prozesses - den ARIMA(0,1,0)-Prozess $Y_t = Y_{t-1} + \eta_t$ nennt man *Random Walk*.
Fast das gesamte Spektrum praktisch relevanter Reihen bewegt sich zwischen I(0)- und I(1)-Prozessen. Deswegen ist es von besonderem Interesse, nach der am Beispiel ARMA erfolgten Kurzbetrachtung der I(0)-Prozesse nun festzustellen, welche charakteristischen Eigenschaften I(1) auszeichnen.
Als Erstes fällt auf, dass durch die Konstruktion eines integrierten Prozesses als einer Folge von partiellen Summen die Stationarität ausgeschlossen wird. Eine Prozessvari-

anz existiert nicht, denn $Var(Y_t)$ bleibt nicht konstant, sondern wächst offensichtlich mit zunehmendem t über alle Grenzen:

$$Var\left(\sum_{i=1}^{t} X_i\right) = t\,\sigma^2 + 2\sum_{i=1}^{t}\sum_{j=1}^{i-1} Cov(X_i, X_j) \;\text{ mit }\; \sigma^2 = Var(X_i). \tag{2.8}$$

Die Variabilität von $\{Y_t\}$ wird somit durch die Varianz des zugehörigen I(0)-Prozesses direkt beeinflusst, bei unkorrelierten X_i sogar eindeutig bestimmt. In diesem Zusammenhang nennt man σ^2 häufig auch *Volatilität*[2] von $\{Y_t\}$. Eine anschauliche Folge der dargestellten Varianzzunahme besteht darin, dass die einzelnen Realisationen eines I(1)-Prozesses sich im Mittel von ihrem Erwartungswert 0 auf unüberschaubare Zeit entfernen. Eine weitere Folge betrifft die Vorhersagbarkeit: es sind keine nichttrivialen Prognosen über eine künftige Beobachtung y_{t+h} möglich. Dies liegt daran, dass jeder Schock η_t über das Y_t komplett in die Prozessvariable Y_{t+h} eingeht. Seine Wirkung setzt sich somit nach beliebig vielen Perioden unvermindert weiter fort, was sich in den Autokorrelationen widerspiegelt:

$$Corr(Y_t, Y_{t+h}) = \frac{Cov(Y_t, Y_{t+h})}{\sqrt{\mathrm{Var}Y_t}\sqrt{\mathrm{Var}Y_{t+h}}} \longrightarrow 1 \;\text{ für alle } h \in \mathbb{Z}, \quad \text{falls } t \to \infty.$$

Diese Beständigkeit bildet den Gegenpol zu kurzfristigen Einflüssen in den I(0)-Prozessen. Aufgrund des dargestellten Verhaltens der I(1)-Prozesse sagt man, dass sie einem *stochastischen Trend* folgen.

2.2.2 Langes Gedächtnis und fraktionale Integration

Wie seit langem bekannt, können nicht alle vorkommenden Abhängigkeitsmuster entweder auf kurzfristige oder auf permanente Wirkungen innerhalb eines Prozesses zurückgeführt werden. Denkbar ist der Fall, bei dem der Einfluss eines jeden Schocks zwar nicht bis ins Unendliche reicht, aber recht lange vor seinem Verschwinden bestehen bleibt. Dieses Verhalten, zunächst in den hydrologischen Datenreihen erkannt, erscheint auch für einige Wirtschaftsdaten und ihre Kombinationen plausibel, siehe u. a. Hassler und Wolters (1995), Krämer (2002). Es kann weder durch I(0) noch durch I(1)-Prozesse abgebildet werden. Visuell machen sich solche langandauernden

[2]Dieser Begriff wird häufig zur allgemeinen Charakterisierung des Schwankungsverhaltens einer Reihe benutzt. In dieser Arbeit wird bei stationären Prozessen Volatilität und Varianz als Synonyme verwendet, in den nächsten Kapiteln werden bedingte Varianzen $Var(X_t|\mathcal{F}_{t-1})$ als Volatilitäten bezeichnet.

Abhängigkeiten im Verlauf einer Zeitreihe an umkehrenden lokalen Trends und Zyklen in Abwesenheit globaler Trends sowie am Verharren der Beobachtungen auf hohem bzw. niedrigem Niveau über längere Zeiträume bemerkbar. Die Autokovarianzen bzw. Autokorrelationen, die ein normiertes Maß für serielle lineare Abhängigkeiten darstellen, zeichnen sich für zunehmende Zeitabstände durch eine langsame, z. B. hyperbolische Annäherung an die Null aus. Man spricht in diesem Zusammenhang von *langem Gedächtnis* des zugehörigen Prozesses. Formal lässt sich das wie folgt ausdrücken:[3]

Definition 4 **(langes Gedächtnis)**

Seien $c > 0$ und $0 < a < 0.5$ konstante reelle Zahlen. Dann heißt ein stationärer stochastischer Prozess $\{X_t\}$ ein Prozess mit langem Gedächtnis, falls gilt:

$$\lim_{h\to\infty} \frac{Corr(X_t, X_{t+h})}{c\, h^{2a-1}} = 1. \tag{2.9}$$

Aus dieser Definition lässt sich leicht folgern, dass die Reihe der absoluten Autokovarianzen divergiert:

$$\sum_{h=0}^{\infty} |Cov(X_t, X_{t+h})| = \infty\,, \tag{2.10}$$

und somit langes und kurzes Gedächtnis einander ausschließen. Gelegentlich wird langes Gedächtnis über diese Divergenz definiert, siehe etwa Baillie (1996). Allerdings sind (2.9) und (2.10) nicht äquivalent.

Die Frage nach der Modellierung langfristiger Abhängigkeiten beantworten Granger und Joyeux (1980) und Hosking (1981). Sie schlagen unabhängig voneinander vor, in den integrierten Prozessen die Ganzzählichkeitsbedingung für den Differenzenparameter d fallen zu lassen. Mit solchen *fraktional integrierten* Prozessen lässt sich dann u. a. auch langes Gedächtnis beschreiben. Als einfachstes Beispiel dazu sei ein Prozess $\{\epsilon_t\}$ betrachtet, den man analog zum Random Walk als

$$(1-L)^d \epsilon_t = \eta_t$$

definiert, aber die Bedingung $d = 1$ durch $0 < d < 1$ ersetzt. Der Prozess $\{\epsilon_t\}$ wird *fraktionales Rauschen* genannt. Der invertierbare fraktionale Filter $(1-L)^d$ ist über die

[3] Die Abhängigkeitsstrukturen lassen sich auch im Frequenzbereich charakterisieren, so dass eine äquivalente Definition von langem Gedächtnis über die Spektraldichte erfolgen kann vgl. z. B. Robinson (1995).

Binomialentwicklung definiert: $(1-L)^d = \sum_{k=0}^{\infty} \psi_k L^k$. Die Koeffizienten ψ_k berechnen sich mit Hilfe der Gammafunktion $\Gamma(k)$ als

$$\psi_k = \frac{\Gamma(k-d)}{\Gamma(k+1)\Gamma(-d)} = \prod_{i=1}^{k} \frac{i-1-d}{i} \quad \text{für } k \geq 1 \quad \text{und } \psi_0 = 1\,. \tag{2.11}$$

Offensichtlich kann man fraktionales Rauschen als einen AR(∞)-Prozess auffassen, dessen Gewichte ψ_k für große k proportional zu $k^{-(1+d)}$ sind und nur langsam gegen Null konvergieren. Zur guten Approximation der Variablen ϵ_t sind deswegen sehr viele vergangene Prozesselemente notwendig, was die Erfassung des langfristigen Verhaltens ermöglicht.

Fraktionales Rauschen ist stationär für $d < \frac{1}{2}$. Einen Beweis dazu findet man z. B. bei Hassler (1993), S. 28. Ferner erhält man mit Hilfe der Stirling-Approximation für genügend große h:

$$Corr(\epsilon_t, \epsilon_{t+h}) = \frac{\Gamma(1-d)\Gamma(h+d)}{\Gamma(d)\Gamma(h+1-d)} \approx \frac{\Gamma(1-d)}{\Gamma(d)}\, h^{2d-1}. \tag{2.12}$$

Somit liegt nach der Definition 4 im Fall $0 < d < \frac{1}{2}$ langes Gedächtnis vor.

Durch den Einsatz des fraktionalen Filters $(1-L)^d$ in der Gleichung (2.7) lassen sich ARIMA- zu *fraktionalen ARIMA-*, oder kurz zu *ARFIMA-Prozessen* erweitern. Sie können sowohl die langzeitige Zusammenhänge innerhalb einer Reihe mittels des Gedächtnisparameters d abbilden als auch kurzzeitige Korrelationen durch die ARMA-Filter $\rho(L)$ bzw. $\theta(L)$ wiedergeben. Ein ARFIMA-Prozess $\{X_t\}$ kann somit als absolut summierbare Filterung eines fraktionalen Rauschens aufgefasst werden:

$$\{X_t\} = (1-\rho(L))^{-1}(1+\theta(L))\{\epsilon_t\} = (1-\rho(L))^{-1}(1+\theta(L))(1-L)^{-d}\{\eta_t\}.$$

Da die ARMA-Parameter nur kurzfristige Abhängigkeiten steuern, bleiben die Stationaritäts- und Gedächtniseigenschaften von fraktionalem Rauschen, als einem ARFIMA(0,d,0)-Prozess in Abhängigkeit von d auch für einen beliebigen ARFIMA(p, d, q)-Prozess erhalten.

2.3 Einheitwurzeltests

Aus den dargestellten Gegensätzen in den Eigenschaften und dem Verhalten von I(0)- und I(1)-Prozessen lässt sich das empirische Interesse an statistischen Verfahren ableiten, die zwischen diesen Prozesstypen unterscheiden. Nimmt man etwa, wie in vielen

Modellen der Fall, eine ökonomische Größe als konstant an, so sollte dies zumindest durch einen stationären Verlauf der entsprechenden Zeitreihe gestützt werden. Geht man von den auf lange Sicht vernachlässigbaren Abhängigkeiten in den Daten bzw. von vorübergehendem Charakter der Außeneinwirkungen aus, so sollte man sich zunächst vergewissern, dass keine zeitlich unbegrenzten Zusammenhänge wie bei Random Walk dieser Annahme widersprechen. Abgesehen davon wird das Bedürfnis nach geeigneten Stationaritätstests durch das Phänomen der sogenannten *Scheinkorrelationen* (engl.: *spurious regression*, vgl. Granger und Newbold (1974)) bei Zusammenhangsanalysen massiv verstärkt. Dabei handelt es sich um die Tatsache, dass die gewöhnliche KQ-Schätzung im linearen Regressionsmodell viel zu häufig signifikante Zusammenhänge zwischen in Wirklichkeit unabhängigen Größen erkennt, sofern diese einem stochastischen Trend folgen. Das sind nur einige Gründe für die fast unüberschaubare Menge an Literatur, die sich mit dem Thema der Tests auf Stationarität beschäftigt. Übersichtsartikel von Hassler (1994) sowie Phillips und Xiao (1998) gewähren einen Einblick in die Vielfalt der methodischen Ansätze und Anwendungen dieser Verfahren, die mit dem Sammelbegriff *Einheitswurzeltests* bezeichnet werden. Der Name geht darauf zurück, dass die meisten von ihnen auf oder gegen eine Einheitswurzel in den z-Transformierten testen, die zu den linearen Filtern aus (2.5) gehören.

Im Rahmen dieser Arbeit stehen die klassischen Tests von Dickey-Fuller im Mittelpunkt, wobei die zentrale Frage lautet:

Wie verändern sich die zugehörigen Teststatistiken unter der Annahme bedingter Heteroskedastie?

Bevor diese Frage im nächsten Kapitel untersucht wird, erläutert der folgende Abschnitt die Funktionsweise dieser Tests im Standardfall.

2.3.1 Tests von Dickey und Fuller

Den von Fuller (1976) und Dickey und Fuller (1979) vorgeschlagenen Testverfahren, die im Weiteren kurz als DF-Tests bezeichnet werden, liegt der Regressionsansatz zugrunde. Zunächst sei angenommen, dass $\{Y_t\}$ einem AR(1)-Prozess folgt: $Y_t = \rho Y_{t-1} + \eta_t$. Betrachte folgendes Testproblem:

$$H_0 : Y_t \text{ ist I(1)} \Leftrightarrow \rho = 1 \ \ \textit{gegen} \ \ H_1 : Y_t \text{ ist I(0)} \Leftrightarrow |\rho| < 1. \tag{2.13}$$

Über die Verteilung des Kleinste-Quadrate-Schätzers für ρ im stationären AR(1)-Modell für eine Reihe der Länge n ist bekannt:

$$\sqrt{n}(\hat{\rho}_{KQ} - \rho) \xrightarrow{d} N(0, 1 - \rho^2).$$

Offensichtlich kann dies nicht unter der Nullhypothese gelten. In diesem Fall besitzt der interessierende Ausdruck folgende Form:

$$\hat{\rho}_{KQ} - 1 = \frac{\sum\limits_{t=1}^{n} Y_{t-1}Y_t - \sum\limits_{t=1}^{n} Y_{t-1}^2}{\sum\limits_{t=1}^{n} Y_{t-1}^2} = \frac{\sum\limits_{t=1}^{n} Y_{t-1}\eta_t}{\sum\limits_{t=1}^{n} Y_{t-1}^2} \tag{2.14}$$

Dabei enthalten sowohl der Zähler als auch der Nenner partielle Summen unabhängiger Zufallsvariablen η_i in Gestalt von $Y_t = \sum_{i=1}^{t} \eta_i$. Während es noch möglich ist, für den normierten Zähler die asymptotische Verteilung unter H_0 direkt herzuleiten:

$$\begin{aligned} Y_t^2 = (Y_{t-1} + \eta_t)^2 \quad &\Rightarrow \quad Y_{t-1}\eta_t = \frac{1}{2}(Y_t^2 - Y_{t-1}^2 - \eta_t^2) \\ &\Rightarrow \quad \frac{1}{n}\sum_{t=1}^{n} Y_{t-1}\eta_t = \frac{1}{2n}(Y_n^2 - Y_0^2 - \sum_{t=1}^{n} \eta_t^2), \end{aligned}$$

und daher mit einem festen $Y_0 = 0$

$$\frac{1}{n}\sum_{t=1}^{n} Y_{t-1}\eta_t \xrightarrow{d} \frac{Var(\eta_t)}{2}\chi_1^2 - \frac{Var(\eta_t)}{2}, \tag{2.15}$$

gelangt man zu einer Verteilungsaussage für den Nenner erst durch eine Erweiterung des Asymptotikbegriffs, die im Weiteren skizziert wird.

Bedingt durch das Varianzverhalten wie in (2.8) kann man keine einheitliche Grenzverteilung für die Folge der partiellen Summen $Y_t/\sqrt{n} = \sum\limits_{i=1}^{t} \eta_i/\sqrt{n},\ t = 1, \ldots, n,\ n \to \infty$ angeben. Um dennoch eine Aussage über das asymptotische Verhalten von Y_t zu formulieren, muss man sie simultan an allen Zeitpunkten $t = 1, \ldots, n$ in der Suche nach einem Grenzprozess betrachten, der die Stochastik von $\{Y_t\}$ am besten wiedergibt. Ganz wichtig ist hierbei, dass bei der asymptotischen Behandlung der Übergang von einem unendlich langen Beobachtungsintervall $[0, n]$ hin zu einer unendlich hohen Beobachtungsfrequenz $1/n$ vollzogen wird. Man geht somit von bisher angenommenen $t \in \mathbb{Z}$ zu $t \in [0; 1]$, d. h. von der diskreten zur stetigen Zeitmessung über. Eine unendliche Stichprobe erhält man dabei, indem $\{Y_t\}$ an möglichst vielen Punkten innerhalb einer festen Zeitspanne $[a, b] \subset \mathbb{R}_+$ angeschaut wird. Auf diese Weise wird

man mit einem stochastischen Prozess konfrontiert, der überabzählbar viele Zufallsvariablen Y_t enthält. Um ihr gemeinsames Konvergenzverhalten zu beschreiben, bedarf es einer Definition der Verteilungsasymptotik, die über den Raum IR hinaus gilt. Sie soll ermöglichen, das Grenzverhalten nicht nur für Folgen von Zufallsvariablen, sondern auch für zufällige Funktionenfolgen zu erfassen.
Den theoretische Rahmen dafür findet man bei Billingsley (1968). Er entwickelt ein geschlossenes Konzept der Verteilungskonvergenz für die Funktionenräume $C[0,1]$ - die Menge aller auf dem Intervall $[0,1]$ definierten reellwertigen stetigen Funktionen, und $D[0,1]$ - alle rechtsseitig stetigen Funktionen mit linksseitigen Grenzwerten (so genannte *cadlag*-Funktionen) auf $[0,1]$. Nähere Erläuterungen zu dieser Konvergenzart, die in folgenden Verteilungsaussagen auch mit $\xrightarrow{d}$ bezeichnet wird, findet man bei Hall und Heyde (1980).
Besonders wichtig für asymptotische Fragestellungen ist die Möglichkeit, den Zentralen Grenzwertsatz für Funktionenräume zu verallgemeinern. Diese schon früher erwähnte Verallgemeinerung heißt *Funktionaler Zentraler Grenzwertsatzes* (kurz: *FZGWS*). Die Grenzverteilung wird dabei durch eine zufällige Funktion $W \in C[0,1]$ vorgegeben, die punktweise betrachtet wie folgt definiert werden kann:

Definition 5 (Wiener-Prozess / Brownsche Bewegung)
Sei Ω die Menge aller Zufallsereignisse. Ein stochastischer Prozess $W := \{W(r,\omega) : [0,1] \times \Omega \to \mathrm{IR}\}$ mit

- $P(W(0) = 0) = 1$,
- $W(s)$ *und* $W(r) - W(s)$ *sind stochastisch unabhängig für alle* $0 \leq s \leq r \leq 1$,
- $W(r) - W(s) \sim N(0,(r-s))$

heißt Wiener-Prozess. *Alternativ wird W auch* Brownsche Bewegung *genannt.*

Der Wiener Prozess stellt einen Spezialfall der so genannten *Martingale* dar. Ein Martingal $\{M_t\}$ ist ein stochastischer Prozess, der als eine stetige Erweiterung des Random Walk angesehen werden kann. Setzt man die Existenz des Erwartungswertes $E|M_t|$ für alle $t \in \mathrm{IR}_+$ voraus, so wird ein Martingal durch die Beziehung

$$E(M_t|\mathcal{F}_s) = M_s \text{ fast sicher für alle } s < t \tag{2.16}$$

definiert.
Eine äquivalente Martingaldefinition kann über seine Zuwächse $MD_{t,s} := M_t - M_s$,

auch *Martingaldifferenzen* genannt, erfolgen. Die Bedingung

$$E(MD_{t,s}|\mathcal{F}_s) = 0 \text{ fast sicher für alle } s < t \tag{2.17}$$

ist äquivalent zu (2.16). Weißes Rauschen besteht ebenfalls aus speziellen Martingaldifferenzen, wie man leicht nachprüft.

Als stetige stochastische Prozesse spielen Martingale und Martingaldifferenzen eine zentrale Rolle bei der Formulierung der Grenzwertaussagen auf Funktionenräumen. Konkret stellt etwa der Wiener-Prozess W den Grenzprozess für die Partialsummen beliebiger standardisierter Zufallsschocks dar, deren zeitlicher Definitionsbereich von $[0, n]$ auf $[0, 1]$ umskaliert wird. Dies besagt der FZGWS in seiner Standardform:

Satz 1 (FZGWS)
Seien η_i, $i = 1, \ldots, n$ u.i.v. mit $E(\eta_i) = 0$ und $E(\eta_i^2) = \sigma^2 < \infty$. Für eine zufällige Funktion Y_n auf $[0, 1]$ mit $Y_n(r) = \sum_{i=1}^{[nr]} \eta_i/\sqrt{n}\sigma$, $r = t/n$, $t = 1, \ldots, n$, gilt: $Y_n \xrightarrow{d} W$.

Der FZGWS ist auch als der *Satz von Donsker* bekannt, der schon 1951 zum gleichen asymptotischen Ergebnis kam. Seine Beweisführung spielt sich dabei im $C[0, 1]$ ab, und zu $\sum_{i=1}^{[nr]} \eta_i$ werden asymptotisch vernachlässigbare Terme $(nr - [nr])\eta_{[nr]+1}$ hinzuaddiert, um $Y_n(r)$ stetig werden zu lassen. Billingsley (1968) zeigt im Kapitel 3 die Konvergenz $Y_n \xrightarrow{d} W$ direkt für $Y_n(r) \in D[0, 1]$.

Wichtig für den Nachweis der Verteilungskonvergenz im nächsten Kapitel ist die Tatsache, dass die ursprüngliche u.i.v.-Forderung sich als nicht notwendig erweist. Nach dem aktuellen Kenntnisstand gilt der FZGWS in Bezug auf das Abhängigkeitsmuster schon für einen L_2-NED-Prozess der Ordnung $-1/2$. Ferner dürfen die Prozessvariablen η_t auch heteroskedastisch sein: $E(u_t^2) = \sigma_t^2$, unter der Bedingung, dass $E\left(\sum_{t=1}^{n} u_t\right)^2/n \to \sigma^2 < \infty$ (vgl. Davidson (2002a), Theorem 1.2).

Hilfreich bei der Formulierung der Konvergenzaussagen ist ferner, dass der Satz über stetige Abbildungen (engl.: *Continious Mapping Theorem*, kurz: *CMT*) auch auf Funktionenräumen gilt. So stellen Hall und Heyde (1980) im Theorem A.3 fest, dass die Verteilungskonvergenz bei stetigen Abbildungen erhalten bleibt: für $Y_n, Y \in C[0, 1]$ folgt aus $Y_n \xrightarrow{d} Y$ für jede messbare Funktion $g : C[0, 1] \to \mathsf{IR}$, dass $g(Y_n) \xrightarrow{d} g(Y)$. Das CMT lässt sich auch direkt für die partiellen Summen formulieren. Diese Erweiterung betrifft nicht nur die Gestalt von zufälligen Funktionen Y_n, die nun aus dem Raum $D[0, 1]$ sind, sondern auch die der Abbildung $g : D[0, 1] \to \mathsf{IR}$. Das Theorem 14.2.3 von Davidson (2000) lässt für g einen sprunghaften Verlauf zu. Solange den

Unstetigkeitsstellen die Wahrscheinlichkeit 0 zugeordnet wird, berührt dies nicht die Verteilungskonvergenz. Genauer:

$$Y_n \xrightarrow{d} Y \;\Rightarrow\; g(Y_n) \xrightarrow{d} g(Y) \tag{2.18}$$

für jede auf $D[0,1]\backslash\bar{D}$ stetige Abbildung g mit $\bar{D} \subset D[0,1]$ und $P(\bar{D}) = 0$.

Mit diesem Instrumentarium, der Äquivalenz $\int_{(t-1)/n}^{t/n} dr = 1/n$ und der Bezeichnung $Var(\eta_t) = \sigma^2$ findet man nun eine asymptotische Verteilung für den geeignet normierten Nenner aus (2.14):

$$\frac{1}{n^2}\sum_{t=1}^{n} Y_{t-1}^2 = \sum_{t=1}^{n} \int\limits_{(t-1)/n}^{t/n} \left(\frac{\sum_{i=1}^{[nr]}\eta_i}{\sqrt{n}}\right)^2 \xrightarrow{d} \sigma^2 \int\limits_0^1 W^2(r)dr. \tag{2.19}$$

Für den Zähler erhält man mit dem FZGWS und dem CMT die Beziehung $\frac{1}{n}\sum_{t=1}^{n} Y_{t-1}\eta_t \xrightarrow{d} \int_0^1 W(r)dW(r)$, deren Äquivalenz zu (2.15) mit dem Ito-Lemma gezeigt werden kann.

Da nach der Definition des Wiener-Prozesses $W(1)$ standardnormal- und folglich $W^2(1)$ χ_1^2-verteilt ist, ergibt sich zusammen mit (2.15) unter der Nullhypothese folgende Verteilungskonvergenz:

$$n(\hat{\rho}_{KQ} - 1) \xrightarrow{d} \frac{W^2(1) - 1}{2\int_0^1 W^2(r)dr}\,. \tag{2.20}$$

Die Quantile der Grenzverteilung konnten lange Zeit nicht analytisch berechnet werden und wurden daher durch Simulationen ermittelt. Sie sind bei Fuller (1976) in der Tabelle 8.5.1 angeführt und lassen auf eine Verteilung schließen, die gegenüber der Normalverteilung etwas rechtschief ist und einen negativen Median hat. Dies wird analytisch von Abadir (1993) bestätigt, der im Theorem 3.1 eine exakte Dichtefunktion angibt. Da sie eine mathematisch äußerst komplizierte Gestalt besitzt, ist dieses Ergebnis nur aus theoretischer Sicht interessant. In der Praxis benutzt man nach wie vor die seit Fuller (1976) bekannten kritischen Werte. Dabei schließt man den Fall von $\rho > 1$ a-priori aus und behandelt (2.13) als ein einseitiges Testproblem, d. h. bei zu kleinen Werten der Teststatistik $n(\hat{\rho}_{KQ} - 1)$ wird H_0 abgelehnt und man entscheidet sich für die Stationarität von $\{Y_t\}$.

Sehr verbreitet ist eine alternative Testvariante, bei der der Schätzer $(\hat{\rho}_{KQ} - 1)$ analog zu einem t-Test mit seinem Standardfehler normiert wird. Zur Ermittlung der

Grenzverteilung geht man ähnlich wie oben vor und erhält unter H_0:

$$\tau = \frac{\sqrt{\sum_{t=1}^{n} Y_{t-1}^2}}{\sqrt{n^{-1}\sum_{t=1}^{n}\hat{\eta}_t^2}}(\hat{\rho}_{KQ}-1) \quad \xrightarrow{d} \quad \frac{\sqrt{\sigma^2\int_0^1 W^2(r)dr}}{\sigma}\frac{W^2(1)-1}{2\int_0^1 W^2(r)dr}$$
$$= \quad \frac{W^2(1)-1}{2\sqrt{\int_0^1 W^2(r)dr}}. \tag{2.21}$$

Die mittels Simulationen ermittelten Quantile für diese Grenzverteilung sind in der Tabelle 8.5.2 bei Fuller (1976) angegeben. Die Dichtefunktion zu dieser Grenzverteilung stammt ebenfalls von Abadir (1995).
Für die letzte Testversion spricht die bequemere Anwendung: der Wert der Prüfgröße als t-Statistik wird bei jeder Standardregressionsrechnung vom Datenverarbeitungsprogramm ausgegeben, somit erfordert das Testen auf Einheitswurzel keinerlei gesonderten Berechnungen. Andererseits ist der KQ-Schätzer unter der Nullhypothese skaleninvariant, wie die Beziehung (2.20) zeigt, und die Normierung mit dem Standardfehler daher überflüssig. Sie führt zu einer langsameren Konvergenzrate $\sqrt{n}$ gegenüber n in (2.20), vgl. Davidson (2000), S. 348ff. Aus diesem Grund schneidet die erste Testvariante bei endlichen Reihenlängen in der Regel etwas besser ab.

Für die praktischen Zwecke ist die Nullhypothese im Rahmen eines AR(1)-Prozesses, d. h. die Hypothese eines reinen Random Walks, zu restriktiv. Ersetzt man allerdings das weiße Rauschen η_t durch autokorrelierte Störgrößen $\tilde{\eta}_t$:

$$Y_t = \rho Y_{t-1} + \tilde{\eta}_t\,, \tag{2.22}$$

so verletzt man die Asymptotik und kann die angeführten Tests nicht ohne weiteres sinnvoll anwenden. Diese Problematik kann speziell an einem integrierten MA(1)-Prozess veranschaulicht werden:

$$Y_t = Y_{t-1} + \tilde{\eta}_t = Y_{t-1} + \eta_t - \theta\eta_{t-1}.$$

Wie an der Filterschreibweise $(1-L)\{Y_t\} = (1-\theta L)\{\eta_t\}$ sichtbar wird, heben sich der Differenzen- und der MA-Operator für θ nahe Eins gegenseitig auf, so dass $\{Y_t\}$ vom weißen Rauschen kaum zu unterscheiden ist. Daher lehnen die DF-Tests die I(1)-Hypothese zu Unrecht ab. Bei relativ kurzen Reihen (n=100) wird das Testniveau selbst für $\theta = 0{,}5$ extrem stark verzerrt, wie die Simulationen von Schwert (1989) demonstrieren.

Allgemein zeigen Phillips und Perron (1988) im Theorem 1, wie für autokorrelierte $\tilde{\eta}_t$ die Grenzverteilungen der DF-Teststatistiken unter H_0 von der Varianz und den Autokovarianzen von $\{\tilde{\eta}_t\}$ abhängen. Genauer wird die Abhängigkeit vom Spektrum von $\{\tilde{\eta}_t\}$ an der Stelle 0 gezeigt, das als $Sp(0) = Var(\tilde{\eta}_t) + 2\sum_{h=1}^{\infty} Cov(\tilde{\eta}_t, \tilde{\eta}_{t+h})$ definiert wird. Daraus folgert Hassler (1994), dass eine negative Summe der Autokorrelationen wie im betrachteten Beispiel zur tatsächlichen Überschreitung des Testniveaus führt. Bei überwiegend positiven Autokorrelationen schneiden die DF-Tests dagegen zu konservativ ab.

Diese Verzerrung des nominalen Testniveaus kann durch die Berücksichtigung der Autokorrelationsstruktur von $\{\tilde{\eta}_t\}$ vermieden werden. So schlagen Phillips und Perron (1988) eine Korrektur der DF-Teststatistiken um den konsistent geschätzten Wert des Spektrums von $\{\tilde{\eta}_t\}$ an der Stelle 0 vor. Dann behalten die Grenzverteilungen aus (2.20) bzw. (2.21) ihre Gültigkeit. Leider funktioniert diese Lösung nur asymptotisch: in den Simulationsstudien von Schwert (1989) und DeJong et al. (1992) erweist sich die Korrektur von Phillips und Perron sowohl für MA- als auch für AR-Struktur in $\{\tilde{\eta}_t\}$ bei den Stichproben der Länge 100 bis 1000 als unzureichend. Daher wird in der Praxis häufiger folgender Ansatz gewählt, um im Modell (2.22) auf eine Einheitswurzel zu testen.

Schon Fuller (1976) schlägt vor, bei AR(p)-Variablen $\tilde{\eta}_t$, die einem AR(p+1)-Prozess $\{Y_t\}$ entsprechen, zum Testen die KQ-Schätzung $\hat{\Psi}_{KQ}$ aus dem erweiterten Regressionsmodell

$$\Delta Y_t = \Psi Y_{t-1} + \sum_{j=1}^{p} \psi_j \Delta Y_{t-j} + \eta_t \tag{2.23}$$

zu verwenden. Auf diese Form, die als *Dickey-Fuller-Transformation* bekannt ist, lässt sich jeder AR(p+1)-Prozess bringen, wobei $\Psi = \sum_{i=1}^{p} \rho_i - 1 = \rho(1) - 1$ und $\psi_j = \sum_{i=j+1}^{p} \rho_i$ gilt. Das Testproblem „$H_0 : \{Y_t\}$ ist I(1) gegen $H_1 : \{Y_t\}$ ist I(0)“ ist somit äquivalent zu „$H_0 : \Psi = 0$ gegen $H_1 : \Psi < 0$“. Im Übrigen geht man wie oben beschrieben vor, denn die Teststatistiken $\hat{\Psi}_{KQ}$ bzw. $\hat{\Psi}_{KQ}/\sqrt{\widehat{Var}(\hat{\Psi}_{KQ})}$ haben die gleiche Grenzverteilung wie in (2.20) bzw. (2.21).

Diese Idee wird von Said und Dickey (1984) auf einen beliebigen invertierbaren ARMA-Prozess $\{\tilde{\eta}_t\}$ übertragen. Durch Invertierung lässt er sich auf die AR(∞)-Form bringen und mit einem reinen autoregressiven Prozess endlicher Ordnung p approximieren. Allerdings muss p genügend groß sein, damit die Grenzverteilungen der Teststatistiken approximativ weiter gelten, und das nominale Testniveau eingehalten werden kann.

Als unangenehmer Nebeneffekt einer hohen autoregressiven Ordnung p tritt der Verlust an Güte auf: die Anzahl der in die Berechnungen eingehenden Beobachtungen reduziert sich um p, und es müssen p zusätzliche Variablen in die Regressionsgleichung aufgenommen werden. Dies führt zu beträchtlichen Wahrscheinlichkeiten für den Fehler 2. Art, vor allem im Bereich $|\rho(1)| > 0.9$. Diese Testversionen, die in der Lage sind, generell „H_0: ARIMA$(p,1,q)$ gegen H_1: ARMA(p,q)“ mit dem DF-Ansatz zu überprüfen, heißen *erweiterte* (engl.: *augmented*) *DF-Tests*.

Erweitert wurden beide Versionen des DF-Tests auch, um nichtzentrierte I(0)-Prozesse der Form

$$Y_t = m + \rho(L)\{Y_t\} + \eta_t \quad \text{mit} \quad m \in \mathrm{IR}, \ |\rho(1)| < 1 \ \text{ und } \ E(Y_t) = \frac{m}{1-\rho(1)} \ \forall\, t \tag{2.24}$$

von den so genannten Reihen mit Drift zu unterscheiden, bei denen im gleichen Modell $\rho(1)=1$ und folglich $E(Y_t)=mt$. Bei der Regression mit dem Absolutglied ergibt sich für die KQ-Schätzung

$$\hat\rho_{KQ} - 1 = \frac{\sum\limits_{t=1}^{n}(Y_{t-1}-\overline{Y})\eta_t}{\sum\limits_{t=1}^{n} Y_{t-1}^2 - n\overline{Y}^2} \qquad \text{(beachte: } \overline{Y} = 1/n\sum_{t=1}^{n} Y_{t-1}\text{)}.$$

Mit dem FZGWS und dem CMT gilt unter H_0:

$$\frac{1}{n}\sum_{t=1}^{n}\overline{Y}\eta_t = \frac{1}{n^{3/2}}\sum_{t=1}^{n}Y_{t-1}\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\eta_t \ \xrightarrow{d}\ \sigma^2 W(1)\int\limits_0^1 W(r)dr, \tag{2.25}$$

denn

$$\frac{1}{n^{3/2}}\sum_{t=1}^{n}Y_{t-1} = \sum_{t=1}^{n}\int\limits_{(t-1)/n}^{t/n}\frac{\sum_{i=1}^{[nr]}\eta_i}{\sqrt{n}} \ \xrightarrow{d}\ \sigma\int\limits_0^1 W(r)dr, \tag{2.26}$$

und analog

$$\frac{1}{n}\overline{Y}^2 = \left(\frac{1}{n}\sum_{t=1}^{n}\frac{Y_t}{\sqrt{n}}\right)^2 \ \xrightarrow{d}\ \sigma^2\left(\int\limits_0^1 W(r)dr\right)^2.$$

Führt man diese Ergebnisse mit (2.15) und (2.19) zusammen, so erhält man unter der Nullhypothese die Verteilungskonvergenz

$$n(\hat\rho_{KQ}-1) \ \xrightarrow{d}\ \frac{0.5(W^2(1)-1) - W(1)\int_0^1 W(r)dr}{\int_0^1 W^2(r)dr - \left(\int_0^1 W(r)dr\right)^2}. \tag{2.27}$$

Analog zeigt man für die Nullverteilung der t-Statistik:

$$\tau = \frac{\sqrt{\sum_{t=1}^{n} Y_{t-1}^2 - n\overline{Y}^2}}{\sqrt{n^{-1}\sum_{t=1}^{n}\hat{\eta}_t^2}}(\hat{\rho}_{KQ} - 1) \xrightarrow{d} \frac{0.5(W^2(1)-1) - W(1)\int_0^1 W(r)dr}{\sqrt{\int_0^1 W^2(r)dr - \left(\int_0^1 W(1)dr\right)^2}}. \tag{2.28}$$

2.3.2 Alternative Einheitwurzeltests

Die Palette existierender Einheitwurzeltests erschöpft sich bei weitem nicht mit dem hier präsentierten klassischen DF-Ansatz und seinen Erweiterungen. Im Folgenden wird ein kurzer Überblick über die wichtigsten alternativen Lösungskonzepte zur Aufdeckung von stochastischen Trends gegeben, wobei auf die Darstellung technischer Details weitgehend verzichtet wird.

Zunächst sind Verfahren hervorzuheben, die H_0 und H_1 im Testproblem (2.13) vertauschen und somit gegen eine Einheitswurzel testen. Dazu gehört der *KPSS-Test* von Kwiatkowski et al. (1992) ebenso wie die Prozeduren von Tanaka (1990) und Saikkonen und Luukkonen (1993), die auf eine Einheitswurzel im MA-Operator der Differenzen ΔY_t testen. Dabei wird im Modell $\Delta Y_t = (1-\theta L)\{\tilde{\eta}_t\}$ das Testproblem

$$H_0 : \theta = 1 \Leftrightarrow Y_t \text{ ist I(0)} \quad \text{gegen} \quad H_1 : |\theta| < 1 \Leftrightarrow Y_t \text{ ist I(1)}$$

betrachtet. Eine ausführliche Darstellung dieser Verfahren mit den Grenzverteilungen der Teststatistiken ist bei Hatanaka (1996) im Abschnitt 3.4 bzw. 6.3 zu finden.

Eine weitere Klasse der Einheitswurzeltests entstand quasi als Nebenprodukt bei der Entwicklung der Theorie fraktional integrierter Prozesse. Für verschiedene semiparametrische Schätzverfahren, die den (nicht notwendigerweise ganzzahligen) Differenzparameter d separat von den ARMA-Parametern im Frequenzbereich schätzen, sind die asymptotischen Verteilungen von $\hat{d}$ bekannt. Die entsprechenden Tests haben eine Reihe von Vorteilen. Dazu zählt in erster Linie ihre Flexibilität: bei der Formulierung des Testproblems hat man freie Hand und kann sowohl auf als auch gegen eine Einheitwurzel testen, die dem Wert $d = 1$ entspricht. Ferner ist es wegen des semiparametrischen Charakters der Schätzprozedur nicht nötig, sich auf eine Modellordnung festzulegen, d. h. das kurzfristige Reihenverhalten zu spezifizieren. Da es sich bei den Grenzverteilungen um Normalverteilungen handelt, sind die Tests außerdem einfach zu handhaben. Als nachteilig erweist sich hingegen, dass bei der praktischen Durchführung der Schätzverfahren für d in endlichen Stichproben einige offene Fragen bestehen, die sich in der Testgüte bzw. der Einhaltung des Testniveaus niederschlagen.

Auf eine Modellspezifikation für kurzfristige Reihendynamik verzichten ebenfalls die auf den Beobachtungsrängen basierenden nichtparametrischen Einheitwurzeltests von Breitung und Gourieroux (1997), Hasan und Koenker (1997) und Breitung (2002).

Im Zusammenhang mit der intensiv diskutierten Problematik des langen Gedächtnisses gehen mehrere Autoren der Frage nach den Eigenschaften der Einheitswurzeltests beim Vorliegen fraktional integrierter Prozesse nach. Für die erweiterten DF-Tests zeigt Krämer (1998) die Konsistenz gegen fraktionale Alternativen. Gleiches weisen für den KPSS-Test Lee und Schmidt (1996) und Lee und Amsler (1997) nach. Allerdings führt eine Zunahme der autoregressiven Ordnung p in endlichen Stichproben laut experimentellen Ergebnissen von Hassler und Wolters (1994) zu einer stark sinkenden Güte der erweiterten DF-Tests. Für die praktischen Zwecke verspricht daher die Anwendung der angesprochenen auf den semiparametrischen Schätzungen von d beruhenden Testprozeduren bessere Resultate bei der Trennung von I(1) und I(d) bzw. I(0) und I(d). Eine vergleichende Simulationsstudie dazu stammt von Schweiger (1999).

Im Hinblick auf die empirischen Fragestellungen sind Methoden zur Diskriminierung zwischen stochastischen und deterministischen Trends sowie zwischen stochastischen Trends und Strukturbrüchen besonders interessant. In beiden Fällen handelt es sich um die Unterscheidung zwischen dem beschriebenen permanenten Einfluss eines Schocks und nicht nachhaltigen stochastischen Einwirkungen in Form von stationären Schwankungen um eine deterministische Funktion, wozu die dargestellten Standardprozeduren nicht in der Lage sind. Nach Perron (1988) bzw. Perron (1990) werden etwa die einfachen DF-Tests die falsche I(1)-Hypothese in großen Stichproben nie verwerfen können. Einen Verlust an Güte für die erweiterten DF-Tests stellen Montanes und Reyes (2000) fest. Die gleiche Gestalt der deterministischen Funktion für alle Zeitpunkte macht es allerdings im ersten Fall möglich, die klassischen Einheitwurzeltests relativ einfach zu modifizieren, indem die vermutete deterministische Trendvariable in die Regressionsgleichung (2.23) einbezogen wird. Bei einem linearen Zeittrend etwa besitzt $\hat{\Psi}_{KQ}$ aus

$$\Delta Y_t = \alpha t + \Psi Y_{t-1} + \sum_{j=1}^{p} \psi_j \Delta Y_{t-j} + \eta_t$$

die gleichen Eigenschaften wie oben. Hingegen gestaltet sich die Untersuchung der Prozesseigenschaften unter Zulassung von Strukturbrüchen viel komplexer, so dass auf die entsprechende Darstellung verzichtet und auf die einschlägige Literatur verwiesen wird. Für einen bekannten Bruchpunkt siehe die Testvorschläge von Perron (1990) mit den Korrekturen von Perron und Vogelsang (1992) sowie von Park und Sung (1994) und

Saikkonen und Lütkepohl (2001). Ist der Zeitpunkt des eventuellen Strukturbruchs unbekannt, so wird er im Rahmen einer zweistufigen Prozedur zunächst geschätzt, und die resultierenden Residuen dann auf Instationarität getestet. Mit diesem Problem beschäftigt sich eine ganze Ausgabe des Journal of Business and Economics Statistics (Band 10, Nr. 3, 1992). Unter neueren Beiträgen sind die Tests von Amsler und Lee (1995) sowie von Saikkonen und Lütkepohl (2002) hervorzuheben.

Zusammenfassend lässt sich zur Auswahl der Einheitwurzeltests sagen, dass kein Testtyp in allen praktisch relevanten Lagen systematisch besser als die anderen abschneidet. Zwar erscheinen die neueren nicht- und semiparametrischen Prozeduren aufgrund der weitgehenden Modellfreiheit attraktiver, weil sie die Gefahr der Niveauverzerrung bzw. des Güteverlustes wegen fehlspezifizierter Modelle reduzieren, doch sind sie bei richtiger Modellspezifikation etwa den erweiterten DF-Tests gegenüber unterlegen, vgl. Breitung (2002). Je nach der kurzfristigen Autokorrelationstruktur in der Reihe empfiehlt sich eher, auf (DF-Tests, unter H_0 eine Einheitswurzel im AR-Polynom) oder gegen (KPSS-Test bzw. Test auf eine Einheitswurzel im MA-Polynom) Instationarität zu testen.

Häufig unterstellt man anstatt einer additiven eine multiplikative autoregressive Struktur und wendet die dargestellten linearen Modelle auf logarithmierte Daten an. Dieser Ansatz sollte stets gründlich überlegt und begründet werden. Während lineare Transformationen einer Zeitreihe den Wert des Integrationsparameters d unverändert lassen, ist seit Granger und Hallman (1988) bekannt, dass durch nichtlineare Transformationen d verändert wird. Mit den Auswirkungen des Logarithmierens auf die Einheitswurzeltests befassen sich Davies und Krämer (2002). Sie zeigen analytisch und experimentell, dass, angewendet auf die Logarithmen eines I(1)-Prozesses, sich die DF-Tests als zu konservativ erweisen, und beim Testen von Ausgangsdaten anstelle der I(1)-Logarithmen das Niveau nicht kontrollierbar ist.

Ferner ist es wichtig, bei der Testdurchführung auf die korrekte Trendmodellierung zu achten. Schließlich sollten bei vermuteten Strukturbrüchen nur speziell konstruierte Tests verwenden werden, die in der Lage sind, dieses Phänomen und I(1)-Charakter eines Prozesses auseinander zu halten.

Im Kapitel 4 werden speziell die DF-Tests bei der Untersuchung von Einflüssen bedingter Heteroskedastie auf Einheitswurzeltests betrachtet. Diese Fokussierung hängt damit zusammen, dass es sich bei diesem Testtyp um die mit Abstand bekanntesten und in der Praxis am häufigsten verwendeten analytischen Methoden zur Aufdeckung stochastischer Trends handelt.

Kapitel 3

Modellierung von Volatilitäten

Ein charakteristisches Merkmal von Kapitalmarktdaten, wie etwa Aktienrenditen, Änderungsraten der Wechselkurse etc. stellen bekanntlich die sich im Laufe der Zeit abwechselnden Phasen größerer und geringerer Schwankungen dar. Als typisches Beispiel hierfür lässt sich die Reihe täglicher DAX-Renditen in Abbildung 1 anführen.

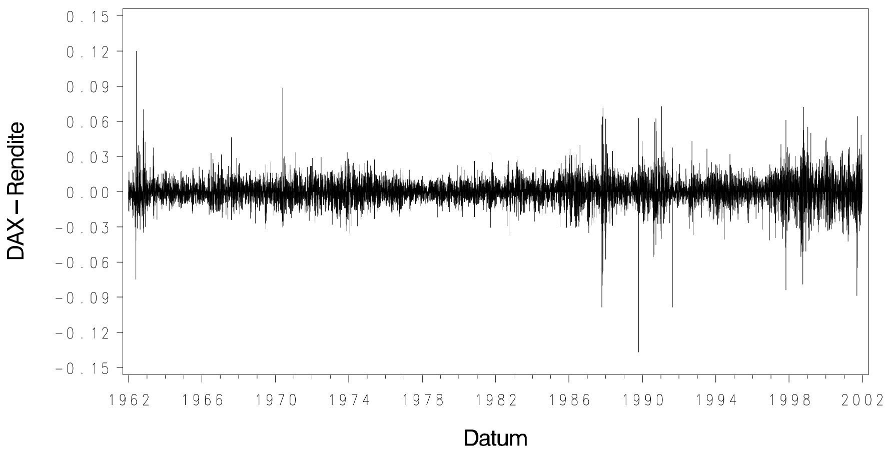

Abbildung 1: DAX-Tagesrenditen

Im Zusammenhang mit den u. a. 1962-63, 1987-88, 1990-91, 1997-98 beobachteten Häufungen stärkerer Ausschläge (Volatilitätscluster) von DAX-Tagesrenditen steht ein weiteres auffälliges Merkmal der Änderungsraten verschiedener Marktdaten. Verglichen mit normalverteilten Daten zeichnen sich ihre Häufigkeitsverteilungen oft durch massivere Ränder und eine gleichzeitige stärkere Besetzung der Verteilungsmitte aus, vgl. Abbildung 2, und somit durch eine höhere Kurtosis. Solche Verteilungen werden daher als hochgipflig, leptokurtisch oder als Verteilungen mit schweren Rändern bezeichnet.

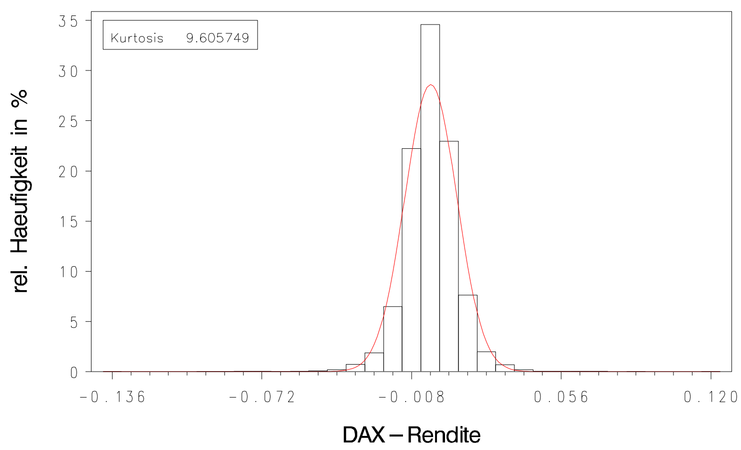

Abbildung 2: DAX-Tagesrenditen: Häufigkeitsverteilung

Die beschriebenen empirischen Eigenschaften der Renditen und anderer Änderungsraten am Kapitalmarkt werden auch am Rohstoffmarkt beobachtet, wie die Studie von Brunetti und Gilbert (2000) zeigt. Man kann sie durch bestehende Autokorrelationen in den quadrierten Reihenwerten erklären. Die bedingten Varianzen (Volatilitäten) $Var(R_t|\mathcal{F}_{t-1})$ des zugrunde liegenden stochastischen Prozesses $\{R_t\}$ sind dabei als zeitabhängig anzusehen und können durch autoregressive Modelle erfasst werden.

3.1 ARCH- und GARCH-Prozesse

Die autoregressiven bedingt heteroskedastischen Prozesse (engl: *Autoregressive Conditional Heteroscedastic*, kurz: *ARCH*) zur Modellierung von im Zeitablauf variierenden Schwankungsintensitäten gehen auf Engle (1982) zurück. Er stellt Elemente eines Innovationenprozesses $R_t = X_t - E(X_t|\mathcal{F}_{t-1})$, die unter der Effizienzmartkthypothese als Renditen interpretierbar sind, als Produkte

$$R_t = \varepsilon_t \, \sigma_t \tag{3.1}$$

dar. Dieser Ansatz wird bei allen folgenden Modellvarianten verfolgt. Die Zufallsvariablen σ_t werden in der Regel als fast sicher positiv und $\mathcal{F}_{t-1}$-messbar vorausgesetzt, die Zufallsschocks ε_t seien weißes Rauschen mit $E\varepsilon_t^2 = 1$ für alle t. Somit sind die Innovationseigenschaften $ER_t = 0$ für alle t und $E(R_t \cdot R_s) = 0$ für alle $t \neq s$ erfüllt.
Eine wichtige Eigenschaft von (3.1) besteht darin, dass die Verteilung von R_t eine höhere Wölbung besitzt als die der ε_t. Seien etwa ε_t normalverteilt, d. h. $E\varepsilon_t^4 = 3$ für

alle t, und die Kurtosis von R_t definiert: $ER_t^4 < \infty$, dann gilt mit der Ungleichung von Jensen:

$$\frac{E(R_t - ER_t)^4}{\left(Var(R_t)\right)^2} = \frac{ER_t^4}{(ER_t^2)^2} = \frac{3\, E\sigma_t^4}{(E\sigma_t^2)^2} \geq 3.$$

Somit wird es mit dem Ansatz (3.1) möglich, der Hochgipfligkeit in empirischen Verteilungen Rechnung zu tragen.

Natürlich kann man versuchen, auf eine hohe empirische Kurtosis von $\{R_t\}$ schon bei der Verteilungsannahme für ε_t einzugehen. So kann man statt einer Standardnormalverteilung eine stärker gewölbte, ggf. auch asymmetrische Verteilung voraussetzen, deren Kurtosis durch (3.1) weiter erhöht wird. Dem Thema der alternativen Verteilungen widmen sich zahlreiche Autoren, wie etwa Bollerslev (1987), Baillie und Bollerslev (1989), Nelson (1991), Palm und Vlaar (1997). So erweisen sich bei verschiedenen Anwendungen die t-Verteilung und die GED-Verteilung (engl.: *generalized error distribution*) von Nelson (1991) als vorteilhaft gegenüber der Normalverteilung. Neuere Vorschläge auf diesem Gebiet betreffen die Abbildung asymmetrischer Strukturen in den Volatilitäten, siehe Fernandez und Steel (1998) sowie Lambert und Laurent (2001).

Setzt man σ_t in Bezug zur vergangenen Entwicklung von $\{R_t\}$, so können die zeitabhängigen bedingten Varianzen $Var(R_t|\mathcal{F}_{t-1}) = \sigma_t^2(R_{t-1}, R_{t-2}, \ldots)$ durch eine geeignete Wahl der funktionalen Form von σ_t^2 modelliert werden. Dieses Vorgehen ist bei allen Modellen des GARCH-Typs gemeinsam.

Den historischen Ausgangspunkt stellt der schon erwähnte ARCH(q)-Prozess dar, in dem die bedingte Varianz als Linearkombination der quadrierten Prozessvariablen ausgedrückt wird:

$$\sigma_t^2 = \omega + \sum_{i=1}^{q} \alpha_i R_{t-i}^2. \tag{3.2}$$

Ein den Gleichungen (3.1) und (3.2) genügender Prozess $\{R_t\}$ erlaubt es, die Volatilitätscluster zu beschreiben: ein betragsmäßig hoher Wert von R_{t-1} zieht tendenziell ebenfalls höhere Werte für σ_t^2 und folglich auch für $|R_t|$ nach sich.

Damit σ_t^2 in jedem Zeitpunkt positiv bleibt, setzt man $\omega > 0$ und $\alpha_i \geq 0$, $i = 1, \ldots, q$ voraus. Diese übliche Einschränkung des Parameterraums ist allerdings nicht notwendig, siehe Nelson und Cao (1992).

Interessiert man sich für die Stationaritätseigenschaft eines ARCH(q)-Prozesses, so betrachtet man die unbedingte Varianz von R_t. Laut Bollerslev et al. (1994) existiert sie genau dann, wenn $\omega > 0$ und $\sum_{i=1}^{q} \alpha_i < 1$:

$$Var(R_t) = E(R_t^2) = \frac{\omega}{1 - \sum_{i=1}^{q} \alpha_i}\,.$$

In diesem Fall ist ein ARCH(q)-Prozess schwach stationär, genauer ein schwaches weißes Rauschen.
Die Konstruktion der ARCH(q)-Prozesse ermöglicht somit prinzipiell die Erfassung oben dargestellter Auffälligkeiten in den Daten, ohne die Stationaritätseigenschaft aufzugeben. Allerdings führte ihre Anwendung zur Erkenntnis, dass häufig erst durch die Wahl einer hohen Ordnung q und die damit verbundene Schätzung zahlreicher Parameter eine befriedigende Beschreibung der Volatilitäten erreicht werden kann.
Zur Lösung dieses Problems nimmt Bollerslev (1986) vergangene bedingte Varianzen in die Gleichung (3.2) als erklärende Variablen auf:

$$\sigma_t^2 = \omega + \sum_{i=1}^{q} \alpha_i R_{t-i}^2 + \sum_{j=1}^{p} \beta_j \sigma_{t-j}^2\,. \tag{3.3}$$

Erfüllt ein Prozess $\{R_t\}$ die Gleichungen (3.1) und (3.3), so heißt er *generalisierter ARCH-Prozess der Ordnung (p,q)*, kurz: *GARCH(p,q)*. Diese Verallgemeinerung erfolgt in Analogie zur Erweiterung der AR-Prozesse zu ARMA: jeder GARCH(p,q)-Prozess kann als ein ARMA($\max(p,q),p$)-Prozess in den zweiten Momenten aufgefasst werden. Dazu definiert man die Innovationen in Quadraten durch

$$\nu_t = R_t^2 - \sigma_t^2.$$

Wie im vorigen Kapitel wird zwecks einfacherer Darstellung vom Backshift-Operator L Gebrauch gemacht. Damit lässt sich (3.3) als

$$\sigma_t^2 = \omega + \alpha(L)\{R_t^2\} + \beta(L)\{\sigma_t^2\}$$

umschreiben und weiter auf die ARMA-Form

$$[1 - \alpha(L) - \beta(L)]\{R_t^2\} = \omega + [1 - \beta(L)]\{\nu_t\} \tag{3.4}$$

bringen. Ähnlich wie beim Übergang von der ARMA(p,q) zur AR(∞)-Darstellung lässt sich ein GARCH(p,q)-Prozess zu einem ARCH-Prozess unendlicher Ordnung umschreiben. Dazu muss der Filter $(1 - \beta(L))$ invertierbar sein, d. h. die Nullstellen der zugehörigen z-Transformierten $1-\beta(z) = 1-\beta_1 z - \ldots - \beta_p z^p,\ z \in \mathbb{C}$ müssen außerhalb des Einheitskreises liegen.

Der Einsatz von ARMA-Prozessen zur Modellierung bedingter Erwartungswerte einer empirischer Datenreihe führt in der Regel bei etwa gleich bleibender Anpassungsgüte zu einer geringeren Anzahl von Parametern im Vergleich zu AR-Modellen, siehe z. B. Schlittgen und Streitberg (1997). Dies gilt auch für bedingte Varianzen: man erzielt häufig eine sparsamere Parametervergabe, wenn statt eines ARCH- ein GARCH-Ansatz zur Volatilitätenbeschreibung gewählt wird.

Genauso wie bei ARCH-Prozessen gewährleisten positives ω und nichtnegative Koeffizienten α_i und β_j in (3.3) eine positive bedingte Varianz $\sigma_t^2 > 0$ und damit die Existenz eines GARCH(p, q)-Prozesses $\{R_t\}$. Dies zu fordern wäre allerdings zu restriktiv: ausgehend von der ARCH(∞)-Form

$$\sigma_t^2 = \omega[1-\beta(1)]^{-1} + \alpha(L)[1-\beta(L)]^{-1}\{R_t^2\} \tag{3.5}$$

reicht es, wenn $\omega[1-\beta(1)]^{-1} > 0$ und alle Koeffizienten des unendlichen Filters $\alpha(L)[1-\beta(L)]^{-1} \geq 0$ sind. Genauer formulierte hinreichende Bedingungen für die Existenz eines GARCH(1,q) bzw. GARCH(2,q)-Prozesses führen Nelson und Cao (1992) an.

Genau dann, wenn die Prozessparameter die Bedingung

$$\sum_{i=1}^{q} \alpha_i + \sum_{j=1}^{p} \beta_j < 1 \tag{3.6}$$

erfüllen, ist $\{R_t\}$ schwach stationär und besitzt die Varianz

$$\sigma^2 = E(R_t^2) = \frac{\omega}{1 - \sum\limits_{i=1}^{q} \alpha_i - \sum\limits_{j=1}^{p} \beta_j}\,.$$

In diesem Fall klingen die Autokorrelationen in quadrierten Reihenwerten $R_t^2 = \sigma_t^2 \varepsilon_t^2$, die unter (3.1) den Autokorrelationen der Volatilitäten $\{\sigma_t^2\}$ entsprechen, exponentiell ab. Dies folgt direkt aus (3.4) und den Eigenschaften eines ARMA-Prozesses. Die Volatilität wird somit zu jedem Zeitpunkt durch die unmittelbare Vergangenheit maßgeblich gesteuert. Für die Prognosen bedeutet das, dass aktuelles Marktgeschehen nur auf kurze Sicht relevant ist. So gilt etwa für die Varianzprognose in einem GARCH(1,1)-Modell:

$$E(\sigma_{t+h}^2|\mathcal{F}_t) = (\alpha_1 + \beta_1)^h \sigma_t^2 + \omega \sum_{k=0}^{h-1} (\alpha_1 + \beta_1)^k. \tag{3.7}$$

Darüber hinaus hängt die durch einen GARCH-Prozess modellierte bedingte Varianz zwar von der Stärke vorangehender Kursausschläge, nicht aber von deren Richtung ab. Da σ_t^2 in (3.3) über R_t^2 nur von der Größe des Zufallsschocks ε_t, nicht aber von dessen Vorzeichen beeinflusst wird, bewirken betragsmäßig gleiche Kursausschläge nach oben und nach unten gleiche Veränderungen in den Volatilitäten.

Diese zwei Eigenschaften der GARCH-Prozesse erweisen sich manchmal als zu starr, um einige auf dem Kapitalmarkt auftretende und theoretisch untermauerte Volatilitätsmuster hinreichend gut abzubilden. Dazu gehören z. B. solche intensiv diskutierten Renditeeigenschaften wie lang andauernde Abhängigkeiten (*Gedächtnis*, vgl. Abschnitt 2.2.2) und *Asymmetrie* in den bedingten Varianzen.

Viele Untersuchungen der Kapitalmarktdaten zeigen, dass lange andauernde serielle Abhängigkeiten gerade für das Langzeitverhalten der Volatilitäten typisch sind, siehe u. a. Baillie et al. (1996), Lobato und Savin (1998).

Unter Asymmetrie in den Volatilitäten versteht man eine ungleich starke Auswirkung unterschiedlich gerichteter vergangener Kursausschläge auf die aktuelle bedingte Varianz. So ist z. B. bei den Aktienrenditen anzunehmen, dass heftige Kursbewegungen nach unten die Volatilität viel stärker erhöhen als betragsmäßig gleiche Kurssteigerungen. Für diese negative Korrelation zwischen den Renditen und den bedingten Varianzen existiert auch eine theoretische Deutung, die als Hebelwirkungshypothese (engl.: *leverage hypothesis*) bekannt ist. Man geht davon aus, dass das Eigenkapital eines Unternehmens in der Regel stärker als das Fremdkapital von einem Kursverfall betroffen wird. Dadurch wird das Gesamtrisiko von einem geringeren Eigenkapitalanteil getragen, was wiederum eine volatilere Kursentwicklung nach sich zieht.

Zur besseren Anpassungsgüte bei der Volatilitätsmodellierung ist man folglich auf Modelle angewiesen, die auf evtl. vorliegende Asymmetrie bzw. langfristige Abhängigkeiten in den Daten eingehen. Darauf zielen die meisten Vorschläge zur Erweiterung der GARCH-Modellpalette ab. Eine Auswahl der wichtigsten bzw. neuesten Modellvarianten wird in den nächsten Abschnitten vorgestellt.

3.2 Integrierte GARCH-Prozesse

Besonders bei hochfrequenten Zeitreihen begegnet man häufig dem Problem, dass die Summe geschätzter Parameter bei der GARCH-Anpassung sehr nah am Wert 1 liegt, vgl. z. B. die empirische Studie von French et al. (1987). Erst durch solche Parameterkonstellationen lässt sich eine hohe empirische Wölbung hinreichend gut modellieren.

Dies stellt offensichtlich die Bedingung (3.6) und somit die schwache Stationarität des zugehörigen Prozesses in Frage. Ein großer Anteil solcher Reihen unter den empirischen Daten begründet das Interesse, auf die instationären GARCH-Prozesse näher einzugehen.
Ähnlich wie beim Übergang von ARMA zu ARIMA-Prozessen führt das Zulassen einer Einheitswurzel für die z-Transformierte des Filters $1 - \alpha(L) - \beta(L)$ (vgl. die Darstellung (3.4)) zu den *integrierten GARCH-Prozessen* (kurz: IGARCH). Eingeführt von Engle und Bollerslev (1986), können sie als I(1)-Prozesse für $\{R_t^2\}$ interpretiert werden. Man spricht in diesem Zusammenhang auch von einem stochastischen Trend in den Volatilitäten. In der Literatur wird für IGARCH(p, q)-Prozesse oft folgende ARMA-Darstellung gewählt, die die Existenz der Einheitswurzel durch das explizite Anführen des Differenzenoperators $(1 - L)$ herausstellt:

$$\phi(L)(1 - L)\{R_t^2\} = \omega + [1 - \beta(L)]\{\nu_t\}. \tag{3.8}$$

Der Filter $\phi(L)$ hat hierbei die Ordnung $(\max(p, q) - 1)$.
Ein IGARCH-Prozess besitzt keine endliche Varianz, wie man aus (3.3) leicht folgern kann, ist aber dennoch streng stationär, vgl. Bougerol und Picard (1992).
Als seine wichtigste theoretische Eigenschaft in Bezug auf Prognosen wird in der Literatur häufig die unverminderte Wirkung eines jeden Schocks ε_t auf alle künftigen Volatilitäten herausgestellt, vgl. Gourieroux (1997) bzw. Baillie et al. (1996). Am Beispiel eines IGARCH(1,1)-Prozesses ($\alpha_1 = 1 - \beta_1$) folgert man aus Gleichung (3.7) mit $\alpha_1 + \beta_1 = 1$

$$E(\sigma_{t+h}^2|\mathcal{F}_t) = \sigma_t^2 + \omega h. \tag{3.9}$$

Die optimale Vorhersage für σ_{t+h}^2 steigt also mit zunehmendem Prognosehorizont h über alle Grenzen linear an. In diesem Zusammenhang nennt man IGARCH-Prozesse persistent. Andererseits zeigt Nelson (1990), dass ein IGARCH(1,0)-Prozess $\{R_t : R_t = |R_{t-1}|\,\varepsilon_t\}$ fast sicher gegen Null konvergiert, was mit der Persistenzthese unvereinbar ist. Diesen Widerspruch erklären Bollerslev et al. (1994) damit, dass persistente Wirkungen in GARCH-Prozessen nicht eindeutig sind. Während $E(\sigma_t^k|\mathcal{F}_s)$ für genügend hohe Momente und $t \to \infty$ divergiert, wird für genügend kleine k Konvergenz gegen eine von $\mathcal{F}_s$ unabhängige Konstante festgestellt. Daher muss beim Gebrauch des Persistenzbegriffs stets präzisiert werden, auf Momente welcher Ordnung man sich bezieht.
Der neuste Beitrag zu dieser Diskussion stammt von Davidson (2002b). Sein Theorem

3.3 zeigt, dass ein streng stationärer bedingt heteroskedastischer Prozess auch dann eine Art mixing-Eigenschaft besitzt, wenn seine kumulierten ARCH(∞)-Koeffizienten den Wert 1 übersteigen. Insbesondere sind aufgrund dieser Behauptung die IGARCH-Prozesse frei von jeglichen langandauernden Wirkungen. Dies widerspricht offensichtlich der traditionellen Interpretation der I(1)-Abhängigkeitsstruktur in bedingten Varianzen eines IGARCH-Prozesses. Davidson (2002b) erklärt dieses Paradoxon damit, dass in Prozessen mit unendlicher unbedingter Varianz $\sigma^2 = E(\sigma_t^2)$ die bedingte Varianz σ_t^2 als Zufallsvariable nicht wohldefiniert ist. Daher kann die Gleichung (3.9) nicht sinnvoll interpretiert werden.
Heuristisch kann man das Fehlen persistenter Varianzstrukturen am oben angeführten einfachen Beispiel eines IGARCH(1,0)-Prozesses damit begründen, dass ein betragsmäßig hinreichend kleiner Wert des Zufallsschocks ε_t ausreicht, um die vorausgegangenen Volatilitäten „auszulöschen“. Da die Realisationen von normal- bzw. t-verteilten ε_t um den Wert Null sehr dicht liegen, verhindert die Multiplikation $|R_{t-1}|\,\varepsilon_t$ eine nachhaltige Wirkung der in R_{t-1} enthaltenen Vergangenheit auf die künftigen Volatilitäten. Somit wird das Auftreten langfristiger Abhängigkeiten immer wieder unterbrochen.

In der Praxis beobachtet man allerdings häufig empirische Autokorrelationen quadrierter Reihenwerte, die für große Zeitabstände relativ langsam abnehmen. Die schon betrachteten DAX-Tagesrenditen bieten ein Beispiel für solches Verhalten.

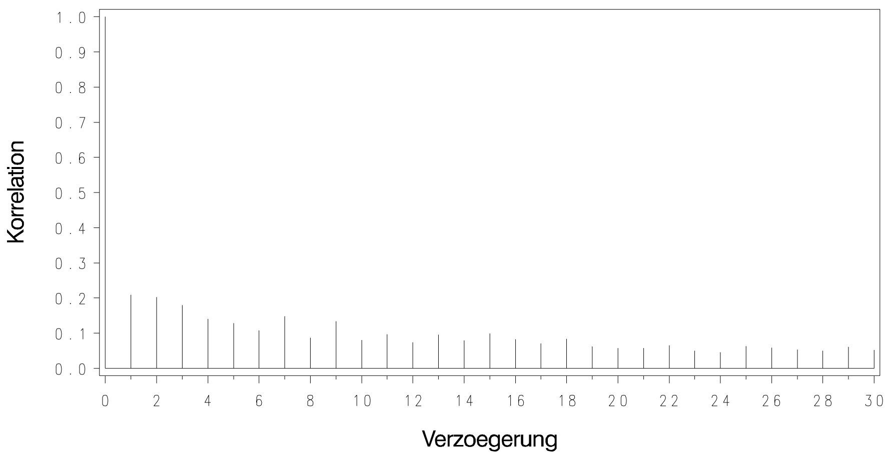

Abbildung 3: empirische Autokorrelationen quadrierter DAX-Tagesrenditen

Laut Abbildung 3 scheinen auch sehr weit voneinander liegende quadrierte Renditen miteinander zu korrelieren. Dies deutet auf länger andauernde Abhängigkeiten in der Varianzstruktur, die im Zusammenhang mit der Modellierung bedingter Erwartungswerte im Kapitel 2 als langes Gedächtnis bezeichnet wurden. Sie sind weder durch reine GARCH- noch durch IGARCH-Prozesse modellierbar.

Baillie et al. (1996) schlagen zur Lösung dieses Problems das Konzept der fraktionalen Integration vor. Dieses Vorgehen lehnt sich an die Modellierung bedingter Erwartungswerte durch ARFIMA-Prozesse an (siehe S. 14). Ein *fraktional integrierter GARCH(p,q)-Prozess der Ordnung* d, kurz: $FIGARCH(p,d,q)$ für R_t wird definiert, indem man an die Darstellung (3.8) für die bedingte Varianz der IGARCH-Prozesse anknüpft. Die einfache Differenzenbildung in (3.8) wird durch den fraktionalen Differenzenfilter $(1-L)^d$ ersetzt:

$$\phi(L)(1-L)^d\{R_t^2\} = \omega + [1-\beta(L)]\{\nu_t\}, \;\; 0 \leq d \leq 1. \tag{3.10}$$

Die Nullstellen der Polynome $\phi(z)$ und $1-\beta(z)$ sind dabei betragsmäßig größer 1. Aufgelöst nach σ_t^2, ergibt sich aus (3.10) die übliche GARCH-Form:

$$\sigma_t^2 = \omega + [1-\beta(L)-\phi(L)(1-L)^d]\{R_t^2\} + \beta(L)\{\sigma_t^2\}.$$

Mit Hilfe der ARCH(∞)-Darstellung

$$\sigma_t^2 = \omega[1-\beta(1)]^{-1} + \lambda(L)\{R_t^2\}, \;\; \lambda(L) := [1-(1-\beta(L))^{-1}\phi(L)(1-L)^d] \tag{3.11}$$

geben Laurent und Peters (2002) hinreichende Bedingungen für die Existenz eines FIGARCH(1,1)-Prozesses wie folgt an:

$$\omega > 0, \quad \beta_1 - d \geq \phi_1 \geq 1 - d/2, \quad d\left(\phi_1 - \frac{1-d}{2}\right) \geq \beta_1(\phi_1 - \beta_1 + d).$$

Bedingt durch die Konstruktion (2.11) summieren sich die Koeffizienten $\tilde{\psi}_k(d)$ des fraktionalen Differenzenfilters $(1-L)^d$ für $d > 0$ zu Null auf. Folglich erhält man mit $\lambda(1) = 0$ wie schon im IGARCH-Fall eine unendliche unbedingte Varianz, wodurch die schwache Stationarität ausgeschlossen wird. Eine direkte Übertragung der Vorgehensweise von Bougerol und Picard (1992) für IGARCH-Prozesse auf die entsprechenden Koeffizienten der ARCH(∞)-Darstellung (3.10) zeigt aber, dass ein FIGARCH-Prozess streng stationär ist. In Bezug auf die Stationarität besitzen die FIGARCH-Prozesse somit die gleichen Eigenschaften wie die IGARCH-Prozesse.
Anders beurteilt man das langfristige Verhalten: ein Zufallsschock ε_t bleibt zwar noch lange in der Zukunft wirksam, sein Einfluss, gemessen durch $\tilde{\psi}_k(d)$, nimmt aber mit

wachsender Entfernung k kontinuierlich ab und wird in der genügend weit liegenden Zukunft verschwindend gering, vgl. (2.11). Aus diesem Grund messen Baillie et al. (1996) den FIGARCH-Prozessvariablen eine im Vergleich zu IGARCH weniger ausgeprägte Langzeitstruktur bei. Demnach nehmen diese eine Position zwischen GARCH und IGARCH ein. In Analogie zu ARFIMA-Prozessen spricht man vom *langen Gedächtnis* in bedingten Varianzen.

Hingegen weist Davidson (2002b) im Rahmen der angesprochenen Diskussion über die IGARCH-Eigenschaften darauf hin, dass direkte Parallelen zu Modellen für bedingte Erwartungswerte mißverständlich sein können. Insbesondere kann es irreführend sein, im Zusammenhang mit FIGARCH-Prozessen von langem Gedächtnis in bedingten Varianzen zu sprechen. Diesen Einwand stützt Davidson auf das Lemma 3.1 von Giraitis et al. (2000), das zeigt, dass die Autokovarianzen quadrierter Prozessvariablen absolut summierbar sind, falls $E\varepsilon^4 < \infty$. Im gleichen Beitrag schlägt Davidson daher vor, bei FIGARCH-Prozesses den Begriff „hyperbolisches Gedächtnis“ zu benutzen. Er führt die Schwierigkeiten in der FIGARCH-Interpretation darauf zurück, dass der Parameter d sowohl das Langzeitverhalten beschreibt als auch die unbedingten Momente und deren Existenz direkt beeinflusst. Daher kann es vorteilhaft sein, bei der Modellierung dieser Prozesseigenschaften zu differenzieren und das Momenten- bzw. Langzeitverhalten durch separate Parameter steuern zu lassen. Zu diesem Zweck führt Davidson eine neue Variante der bedingt heteroskedastischen Prozesse - den *hyperbolischen GARCH*, kurz: *HYGARCH*, indem er den FIGARCH-Ansatz verallgemeinert. In der Gleichung für die bedingte Varianz (3.11) erhält der Filter $\lambda(L)$ die Form

$$\lambda(L) = 1 - (1 - \beta(L))^{-1}\phi(L)(1 + \zeta(1 - L)^d - \zeta), \quad \zeta > 0. \tag{3.12}$$

An einigen Spezielfällen wird nun illustriert, wie sich der neu eingeführte Parameter ζ auf Stationaritäts- und Momenteneigenschaften auswirkt. So entsprechen Parameterwerte $\zeta = 0$ bzw. $\zeta = 1$ dem schwach stationären GARCH- bzw. dem FIGARCH-Prozess. Ebenfalls einen schwach stationären GARCH-Prozess erhält man im Falle $d = 0$. Für $d = 1$ reduziert sich (3.12) zu $\lambda(L) = 1-(1-\beta(L))^{-1}\phi(L)(1-\zeta L)$, das Polynom $\lambda(z)$ bekommt somit eine zusätzliche Wurzel. In Abhängigkeit von ζ kann dann $\{R_t\}$ entweder schwach stationär bleiben($\zeta < 1$), oder er wird zu IGARCH ($\zeta = 1$). Ist $\zeta > 1$, so kann der zugehörige Prozess wegen unendlicher Varianz nicht schwach stationär sein, wohl aber streng stationär. Weitere Details sowie zwei Anwendungsbeispiele zur Modellierung der Wechselkursänderungen findet man bei Davidson (2002b). Eine ausführliche Diskussion des HYGARCH-Modells mit weiteren Anwendungsbeispielen

und vergleichenden Analysen stammt von Schoffer (2003).
Auch wenn der HYGARCH-Ansatz sich vorteilhaft gegenüber den FIGARCH-Modellen erweisen sollte, erscheint die Argumentation von Davidson gegen langes Gedächtnis in FIGARCH-Prozessen nicht ganz berechtigt zu sein. Dies wird auf der Seite 43 erläutert.
Eine Alternative zur Erfassung lange andauernden Abhängigkeiten in den Volatilitäten bieten *LMSV-Prozesse* (engl.: *long memory stochastic volatility*) von Breidt et al. (1998). Das zugehörige Grundmodell unterscheidet sich leicht von (3.1), zur Beschreibung langsam abnehmender Autokorrelationen wird aber ebenfalls der fraktionale Filter $(1-L)^d$ benutzt.

3.3 Asymmetrische GARCH-Prozesse

In den bisher betrachteten Modellen reagiert die bedingte Varianz gleichermaßen auf positive und negative vergangene Abweichungen vom Mittelwert. Dies erscheint wegen der angesprochenen Hebelwirkungshypothese zumindest für Aktienrenditen häufig nicht plausibel. Eine relativ einfache Möglichkeit, eventuell auftretende asymmetrische Strukturen in den Volatilitäten zu erfassen, besteht darin, in das Modell (3.3) verzögerte Prozessvariablen aufzunehmen, etwa durch die Gleichung

$$\sigma_t^2 = \omega + \sum_{i=1}^{q} \alpha_i (R_{t-i} - \kappa)^2 + \sum_{i=j}^{p} \beta_j \sigma_{t-j}^2, \quad \kappa \neq 0,$$

vgl. Bollerslev et al. (1994). Für einen *asymmetrischen ARCH-Prozess* erster Ordnung erhält man dann

$$\sigma_t^2 = (\omega + \alpha\kappa^2) + \alpha R_{t-1}^2 - 2\alpha\kappa R_{t-1} + \beta\sigma_{t-1}^2,$$

Falls $\kappa > 0$, verringert sich die Volatilität nach Kurssteigerungen und nimmt nach Kursrückgängen zu.
Eine wesentlich breitere Anwendung findet jedoch der *exponentielle GARCH-Prozess*, kurz: *EGARCH* von Nelson (1991). Nach seiner Definition folgen die Zufallsvariablen R_t aus (3.1) einem EGARCH(p,q)-Prozess genau dann, wenn für ihre bedingten Varianzen gilt:

$$\log \sigma_t^2 = \omega + \sum_{i=1}^{q} \alpha_i \left(\xi \varepsilon_{t-i} + |\varepsilon_{t-i}| - E|\varepsilon_{t-i}| \right) + \sum_{j=1}^{p} \beta_j \log \sigma_{t-j}^2 \, . \tag{3.13}$$

Hier erlaubt der Parameter ξ, asymmetrische Effekte zu modellieren. Falls $\xi = 0$, spielt das Vorzeichen des Schocks ε_{t-i} bei der Volatilitätsmodellierung keine Rolle; ähnlich wie im GARCH-Modell beeinflusst nur die Stärke der Abweichungen $|\varepsilon_t| - E|\varepsilon_t|$ das weitere Schwankungsverhalten der Reihe. Durch negative Werte von ξ kann der Leverage-Effekt erfasst werden. So erhöhen bei $-1 < \xi < 0$ positive ε_{t-i} die Volatilität in einem geringeren Maß als negative. Falls $\xi < -1$, kommt es sogar zu entgegengesetzten Wirkungen: während Ausschläge nach oben die künftige Volatilität mindern, wird sie durch Ausschläge nach unten verstärkt. Eine auffallende Besonderheit der Modellgleichung (3.13) besteht darin, dass nicht die bedingte Varianz direkt, sondern ihr Logarithmus als Zielvariable auftritt. Dadurch wird gewährleistet, dass σ_t^2 unabhängig von den Werten der Prozessparameter immer positiv ist. Somit bedarf es für die Existenz eines EGARCH-Prozesses keinerlei Einschränkungen des Parameterraums. Dies erleichtert numerische Anpassungsverfahren, lässt flexiblere dynamische Strukturen zu und zählt deshalb zu den Vorteilen der EGARCH-Modellierung, vgl. Hamilton (1994), S. 668. Offenbar kann die Gleichung (3.13) als eine ARMA(p, q)-Darstellung für $\log \sigma_t^2$ angesehen werden. Daraus leitet Nelson (1991) Bedingungen für die Stationarität und die Existenz höherer Momente in einem EGARCH-Prozess her.

Einen anderen Ansatz bei der Darstellung asymmetrischer Volatilitätsmuster verfolgt das häufig angewendete Modell von Glosten, Jagannathan und Runkle (1993) (im Weiteren: *GJR-GARCH*). Eng verwandt hiermit ist auch das *TGARCH-Modell* (T steht für „threshold“), vgl. Zakoian (1994). In beiden Fällen wird der Einfluss vergangener Kursbewegungen auf Volatilität in Abhängigkeit von ihrer Richtung modelliert, d. h. man benutzt unterschiedliche Parameter für positive und negative Werte von R_{t-i}. Diese Modelle lassen sich mit Hilfe der Indikatorfunktion I in der folgenden Gleichung darstellen:

$$\sigma_t^\delta = \omega + \sum_{i=1}^{q} \left[\alpha_i I_{\{R_{t-i}>0\}} |R_{t-i}|^\delta + \tilde{\alpha}_i I_{\{R_{t-i}\le 0\}} |R_{t-i}|^\delta\right] + \sum_{j=1}^{p} \beta_j \sigma_{t-j}^\delta\,, \tag{3.14}$$

wobei $\delta = 1$ im GJR- und $\delta = 2$ im TGARCH-Modell. Das GJR-Modell unterscheidet sich somit vom TGARCH nur durch die Exponente δ. Um bei empirischen Fragestellungen sich nicht von vorne herein auf eine dieser Modelle festlegen zu müssen und dadurch eine möglicherweise passendere Modellvariante auszuschließen, liegt es nahe, δ als einen zusätzlichen Modellparameter zu betrachten und aus den Daten zu schätzen. Diese Idee, kombiniert mit dem asymmetrischen Ansatz, führt zum *APARCH*-Modell (engl.: *asymmetric power ARCH*) von Ding et al. (1993). Die Reihe $\{R_t\}$ folgt einem

APARCH(p, q)-Prozess, falls

$$\sigma_t^\delta = \omega + \sum_{i=1}^{q} \alpha_i(|R_{t-i}| - \xi_i R_{t-i})^\delta + \sum_{j=1}^{p} \beta_j \sigma_{t-j}^\delta \,, \tag{3.15}$$

wobei $\delta > 0$, $\omega > 0$, $\alpha_i \geq 0, |\xi_i| < 1$ für alle i und $\beta_j \geq 0$ für alle j. Die Flexibilität der Datenanpassung mittels APARCH zeigt sich darin, dass gleich mehrere zur GARCH-Familie gehörende Modellvarianten als Spezialfälle abgedeckt werden:

- ARCH, falls $\delta = 2$, $\xi_i = 0$ für alle i und $\beta_j = 0$ für alle j;
- GARCH, falls $\delta = 2$ und $\xi_i = 0$ für alle i;
- GARCH von Taylor (1986), falls $\delta = 1$ und $\xi_i = 0$ für alle i;
- GJR-GARCH, falls $\delta = 2$;
- TGARCH, falls $\delta = 1$;
- NARCH (engl.: *non-linear ARCH*) von Higgins und Bera (1992), falls $\xi_i = 0$ für alle i und $\beta_j = 0$ für alle j;
- Log-ARCH von Geweke (1986), falls $\delta \to 0$.

Die Anpassungsgüte der GARCH-Modellpalette kann weiter gesteigert werden, indem man die gleichzeitige Erfassung langfristiger und asymmetrischer Volatilitätsstrukturen ermöglicht. So erweitern Bollerlev und Mikkelsen (1996) das EGARCH- und Tse (1998) das APARCH-Modell durch die aus dem letzten Abschnitt bekannte Faktorisierung des Filters $1 - \beta(L) = \phi(L)(1 - L)^d$ zu *FIEGARCH* bzw. *FIAPARCH*. Die Volatilitäten werden dann durch die Gleichungen

$$\phi(L)(1 - L)^d\{\log \sigma_t^2\} = \omega + \alpha(L)\{\xi\varepsilon_t + |\varepsilon_t| - E|\varepsilon_t|\} \quad \text{bzw.}$$

$$\phi(L)(1 - L)^d\{\sigma_t^\delta\} = \omega + \alpha(L)\{(|r_t| - \xi r_t)^\delta\}$$

modelliert. Anzumerken bleibt, dass man bei der Darstellung des FIAPARCH-Modells auf die Filterschreibweise mit L angewiesen ist. Sie ist nur möglich, wenn im Unterschied zu (3.15) ein einheitlicher Asymmetrieparameter ξ vorausgesetzt wird.

3.4 Eigenschaften der GARCH-Prozesse

Wie die obige Betrachtung zeigt, wurde die GARCH-Modellpalette in den 90-er Jahren durch zahlreiche Verallgemeinerungen erheblich erweitert. Neuere Arbeiten auf diesem Gebiet untersuchen die stochastischen Eigenschaften verschiedener Prozessvarianten. So formulieren Carrasco und Chen (2002) hinreichende Bedingungen für die strenge Stationarität von acht verschiedenen Prozessen der GARCH(1,1)-Familie. Sie leiten ihre Ergebnisse über eine mixing-Eigenschaft her, die der uniform-mixing aus (2.2) ähnelt und die strenge Stationarität impliziert. Für die konkrete Gestalt der Bedingungen wird auf die übersichtliche Darstellung bei Carrasco und Chen (2002), Tabelle 1 verwiesen. He et al. (2002) konzentrieren sich auf der Untersuchung des EGARCH(1,1)-Prozesses und geben seine Kurtosis und die Gestalt der Autokovarianzfunktion für verschiedene Potenzen von $|R_t|$ an.

Im Unterschied zu dem Vorgehen, einzelne Modellvarianten separat zu betrachten, versuchen andere Autoren, durch eine einheitliche Darstellung eine möglichst breite Fülle autoregressiver bedingt heteroskedastischer Prozesse zu erfassen, um dann allgemeine Eigenschaften in Bezug auf Stationarität und die Existenz höherer Momente zu untersuchen. Leider treten hierbei Unterschiede in der Begriffsverwendung auf. Neben dem Innovationenprozess $\{R_t\}$ aus (3.1) werden mitunter auch die absoluten Werte bzw. ihre Potenzen $\{|R_t|^\delta\}$, $\delta > 0$ als allgemeine ARCH-Prozesse bezeichnet. Dies macht die formulierten Aussagen nicht direkt vergleichbar. Bei der folgenden Betrachtung der Beiträge von He und Teräsvirta (1999), Giraitis, Kokoszka und Leipus (2000) (im weiteren GKL (2000)), Kazakevicius und Leipus (2002), Ling und McAleer (2002a) und (2002b) kommt es daher entscheidend darauf an, von welcher ARCH-Definition die Autoren jeweils ausgehen.
Als Spezialfall sei zunächst die Prozessordnung $p = q = 1$ betrachtet. Dann können ARCH-Prozesse aus (3.1) mit verschiedenen autoregressiven Varianzausdrücken wie etwa (3.2), (3.3), (3.14) u. a. durch die Gleichung

$$\sigma_t^\delta = g(\varepsilon_{t-1}) + f(\varepsilon_{t-1})\sigma_{t-1}^\delta \quad \text{mit} \quad \delta > 0 \text{ und } g, f : \mathrm{IR} \to \mathrm{IR}_+ \tag{3.16}$$

wiedergegeben werden. Ausgehend von dieser Darstellung zeigen He und Teräsvirta (1999) und Ling und McAleer (2002a), dass unter den Annahmen $E|\varepsilon_t|^{k\delta} < \infty$ und $E[g(\varepsilon_t|)]^{k\delta} < \infty$ die Bedingung

$$E[f(\varepsilon_t)]^k < 1 \quad \text{für alle } t, \quad k \in \mathbb{Z}, k > 0 \tag{3.17}$$

notwendig und hinreichend für die Existenz des (unbedingten) $k\delta$-ten Momentes von $R_t = \sigma_t \varepsilon_t$ ist. Ferner beweisen Ling und McAleer (2002a) im Theorem 2.1, dass die strenge Stationarität von $\{R_t\}$ durch

$$E|\varepsilon_t|^{a\delta} < \infty, \quad E[g(\varepsilon_t|)]^a < \infty \text{ und } E[f(\varepsilon_t)]^a < 1$$

für ein beliebiges $a \in (0, 1]$ gewährleistet wird.
Zur Illustration sei der gewöhnliche GARCH(1,1)-Prozess aus (3.3) betrachtet. Dann ist $g(\varepsilon_t) \equiv \omega > 0$ und $f(\varepsilon_t) \equiv \alpha_1 \varepsilon_{t-1}^2 + \beta_1$. Somit ist $E[f(\varepsilon_t)]^a < 1$ äquivalent zu der von Nelson (1990) und Bougerol und Picard (1992) aufgestellten notwendigen und hinreichenden Bedingung für die strenge Stationarität

$$E \log(\alpha_1 \varepsilon_{t-1}^2 + \beta_1) < 0. \tag{3.18}$$

Interessiert man sich für die Existenz der unbedingten Varianz und folglich für die schwache Stationarität, so ergibt sich aus (3.17) die bekannte Bedingung $\alpha_1 + \beta_1 < 1$, vgl. (3.6).
Die angeführten Stationaritätsbedingungen lassen sich allerdings nicht ohne weiteres auf Prozesse beliebiger Ordnung übertragen. Ling und McAleer (2002b) lösen dieses Problem für den APARCH-Prozess, der zahlreiche Modelle als Spezialfälle einschließt (siehe S.39) und deswegen von großem Interesse ist. Schon bei der Entwicklung des APARCH-Ansatzes zeigten Ding et al. (1993), dass die Beziehung

$$\sum_{i=1}^{q} \alpha_i E(|\varepsilon_t| - \xi_i \varepsilon_t)^\delta + \sum_{j=1}^{p} \beta_j < 1 \Longrightarrow E|R_t|^\delta < \infty \tag{3.19}$$

gilt. Für diesen Fall erhält man damit

$$E(\sigma^\delta) = \frac{\omega}{1 - \sum_{i=1}^{q} \alpha_i (|\varepsilon_t| - \xi_i \varepsilon_t)^\delta + \sum_{j=1}^{p} \beta_j}.$$

Im Theorem 3.1 zeigen Ling und McAleer (2002b), dass die unter (3.19) formulierte Bedingung sowohl notwendig als auch hinreichend für die strenge und (im Falle $\delta = 2$) auch für die schwache Stationarität eines speziellen APARCH(p, q)-Prozesses mit $\xi_i = \xi$ für alle i ist. Ferner formulieren sie im Theorem 3.2 eine notwendige und hinreichende Bedingung für die Existenz höherer Momente ($E|R_t^{k\delta}| < \infty$), die sich aus der Lösung eines Eigenwertproblems ergibt.

Alternative Formulierungen der Stationaritätsbedingungen für Prozesse der GARCH-Familie führen GKL (2000) an. Ihre Resultate beziehen sich allerdings auf Beträge von R_t bzw. ihre Potenzen, die im Weiteren mit X_t bezeichnet werden. Konkret bezeichnen sie $\{X_t\}$ als ARCH(∞)-Prozess, wenn es eine Folge nichtnegativer u.i.v. Variablen ϵ_t gibt, so dass

$$X_t = \left(b_0 + \sum_{i=1}^{\infty} b_i X_{t-i} \right) \epsilon_t \;\text{ für alle } t \in \mathbb{Z}. \tag{3.20}$$

Dabei sind $b_i \geq 0$ für alle i. Bezogen auf die bisher betrachteten Zufallsschocks sind sie als $\epsilon_t = f(\varepsilon_t)$ mit $f : \mathbb{R} \to \mathbb{R}_+$ darstellbar. Laut GKL (2000) ist $\{X_t\}$ streng stationär mit einem endlichen Erwartungswert, falls

$$E\epsilon_t \sum_{i=1}^{\infty} b_i < 1. \tag{3.21}$$

Dann lässt sich X_t in eine Volterra-Reihe entwickeln:

$$X_t = b_0\epsilon_t + b_0\epsilon_t \sum_{k=1}^{\infty} \sum_{i_1,\ldots,i_k=1}^{\infty} b_{i_1} b_{i_2} \ldots b_{i_k} \epsilon_{t-i_1} \epsilon_{t-i_1-i_2} \ldots \epsilon_{t-i_1-i_2-\ldots-i_k}. \tag{3.22}$$

Verstärkt man die Bedingung (3.21) zu

$$\sqrt{E\epsilon_t^2} \sum_{i=1}^{\infty} b_i < 1, \tag{3.23}$$

so erhält man für X_t eine endliche Varianz und folglich die schwache Stationarität. Als notwendig für die strenge Stationarität formulieren Kazakevicius und Leipus (2002) ausgehend von (3.20) die Bedingung

$$E \log \epsilon_t + \log \sum_{i=1}^{\infty} b_i < 0.$$

Sie ist allerdings nicht hinreichend, wie man am Beispiel eines GARCH(1,1)-Prozesses durch den Vergleich mit (3.18) feststellt.

Die bislang anhand einzelner Spezialfälle betrachtete langfristige Abhängigkeitsstruktur lässt sich mit Hilfe der Darstellung (3.20) ebenfalls für allgemeine Prozesse der GARCH-Familie untersuchen. GKL (2000) kommen zum generellen Ergebnis, dass der Abfall der Koeffizienten b_j auf das asymptotische Verhalten der Autokovarianzfunktion direkt übertragbar ist:

$$b_j \sim j^{-k} \implies Cov(X_t, X_{t+h}) \sim h^{-k}, \;\; k > 1. \tag{3.24}$$

Das Zeichen $\sim$ ist als „bis auf eine multiplikative Konstante asymptotisch gleich“ zu verstehen: $x_j \sim y_j \Leftrightarrow x_j/y_j \to C < \infty$ für $j \to \infty$.
Für die Behauptung (3.24) wird (3.23) angenommen, um die Existenz der Autokovarianzen festzustellen. Aus (3.24) folgert man direkt

$$\sum_{h=0}^{\infty} Cov(X_t, X_{t+h}) < \infty. \tag{3.25}$$

Mit Hinblick auf die Diskussion über das lange Gedächtnis in bedingten Varianzen eines FIGARCH-Prozesses (siehe S.36) muss nochmals betont werden, dass eine endliche Autokovarianzensumme die Annahme von (3.23) erfordert. Aus $E\epsilon_t^4 > E\epsilon_t^2 = 1$ folgt, dass dazu die Bedingung $\sum_j b_j < 1$ erfüllt sein muss. Dadurch wird der Fall $\sum_j b_j = 1$, wie im FIGARCH-Modell gegeben, ausgeschlossen und die Summierbarkeit der FIGARCH-Autokovarianzen kann nicht mit (3.24) begründet werden. Somit ist der FIGARCH-Prozess entgegen der Argumentation von Davidson (2002b) durchaus mit langem Gedächtnis (aber nicht mit schwacher Stationarität) vereinbar. Dagegen besitzen beispielsweise ARCH-Prozesse, deren Quadrate schwach stationär sind, kein langes Gedächtnis in bedingten Varianzen. Trotzdem können auch relativ lang andauernde Abhängigkeiten in $\{R_t^2\}$ erfasst werden: wenn die Prozessparameter b_j sich so verhalten, dass der Exponent k aus (3.24) nahe an 1 liegt, flachen die Autokovarianzen nur langsam ab.

Neben der Betrachtung der Autokorrelationen kann man das im Abschnitt 2.1 eingeführte NED-Konzept heranziehen, um serielle Abhängigkeitsstrukturen zu charakterisieren. Seit Hansen (1991) ist bekannt, dass GARCH(1,1)-Prozesse unter der hinreichenden Bedingung $\|\alpha + \beta\varepsilon_t\|_p < 1$ die geometrische L_p-NED-Eigenschaft bzgl. des zugehörigen weißen Rauschens besitzen. Davidson (2002a) verallgemeinert diese Aussage im Theorem 2.3 zu der Behauptung, dass ein schwach stationärer GARCH(p, q)-Prozess L_1-NED und bei Existenz vierter Momente auch L_2-NED geometrischer Ordnung ist.

3.5 Der GRS-ARCH-Prozess

In diesem Abschnitt wird das von Giraitis, Robinson und Surgailis (2000) (im Weiteren: GRS) stammende Modell zur Beschreibung bedingter Varianzen eingeführt, das im weiteren Verlauf der Arbeit eine zentrale Rolle spielen wird. Die Autoren nennen ihr Modell *lineares ARCH*. Diese Bezeichnung kann aber auch für andere Prozesse verwendet werden, vgl. etwa (3.20). Um das neue Modell von den anderen Ansätzen,

die eine ähnliche Darstellungsform aufweisen können, abzugrenzen, wird es *GRS-ARCH* genannt.
Die bedingte Volatilität wird für den GRS-ARCH-Prozess $\{R_t\}$, $R_t = \varepsilon_t \sigma_t$ durch die Gleichung

$$\sigma_t = b_0 + \sum_{i=1}^{\infty} b_i R_{t-i}, \quad b_0 \neq 0 \tag{3.26}$$

gesteuert. Besonders hervorzuheben ist, dass keine weiteren Einschränkungen des Parameterraums vorgenommen werden, d. h. die b_i , $i \geq 1$ dürfen auch negativ sein. Dies verspricht eine flexiblere Datenanpassung im Vergleich zu vielen anderen Modellen der GARCH-Klasse. Ein GRS-ARCH-Prozess hat genau dann eine schwach stationäre Lösung, wenn

$$B := \sum_{i=1}^{\infty} b_i^2 < 1. \tag{3.27}$$

In diesem Fall kann man analog zum Modell (3.20) zur äquivalenten Darstellung von $\{R_t\}$ die Volterra-Reihe verwenden:

$$R_t = b_0 \varepsilon_t + b_0 \varepsilon_t \sum_{k=1}^{\infty} \sum_{i_1,\ldots,i_k=1}^{\infty} b_{i_1} b_{i_2} \ldots b_{i_k} \varepsilon_{t-i_1} \varepsilon_{t-i_1-i_2} \ldots \varepsilon_{t-i_1-i_2-\ldots-i_k}. \tag{3.28}$$

Dies zeigen GRS im Theorem 2.1. Die strenge Stationarität folgt dann direkt aus (3.21). Einfaches Nachrechnen liefert folgende Ausdrücke für die Verteilungsparameter des Prozesses:

$$E(\sigma_t) = b_0\,; \quad Var(R_t) = E(\sigma_t^2) = \frac{b_0}{1 - \sum\limits_{i=1}^{\infty} b_i^2}\,; \quad Cov(\sigma_t, \sigma_{t+h}) = \frac{b_0^2 \sum\limits_{i=1}^{\infty} b_i b_{i+h}}{1 - \sum\limits_{i=1}^{\infty} b_i^2}. \tag{3.29}$$

Das GRS-ARCH-Modell zeichnet sich dadurch aus, dass es das ganze Spektrum zeitlicher Abhängigkeiten innerhalb einer Reihe abbilden kann: von kurzfristigen bis hin zu sehr lange andauernden. In dieser Hinsicht bietet es eine Alternative zum FIGARCH-Ansatz. Im Vergleich dazu besteht ein wichtiger Vorteil von GRS-ARCH in der Fähigkeit, langes Gedächtnis in den Volatilitäten zu erfassen, ohne auf die Stationarität der Innovationen R_t zu verzichten. Auf die Untersuchung dieses Aspekts konzentrieren sich GRS (2000) bei der Einführung des Prozesses. So zeigen sie, vereinfachend ausgedrückt, dass bei schwacher Stationarität für $0 < \delta < 1$ gilt:

$$b_i \sim i^{-\frac{\delta+1}{2}} \Longrightarrow Cov(\sigma_t, \sigma_{t+h}) \sim h^{-\delta} \text{ für } i \to \infty. \tag{3.30}$$

Im Theorem 2.2 weisen sie nach, dass dieses Ergebnis sich auf die Autokovarianzen quadrierter Prozessvariablen für normalverteilte Schocks ε_t unter der Zusatzannahme $\sum_i b_i^2 < 1/\sqrt{3}$ bzw. unter einer etwas strengeren Forderung $15\sqrt{E\varepsilon_t^4}\sum_i b_i^2 < 1$ bei einer beliebigen Verteilung von ε_t übertragen lässt. Dann folgert man aus (3.30):

$$Cov(R_t^2, R_{t+h}^2) \sim \frac{b_0^2}{\left(1 - \sum\limits_{i=1}^{\infty} b_i^2\right)^2} \, h^{-\delta}. \tag{3.31}$$

Die Bedingung (3.30) erinnert an das Verhalten der fraktionalen Gewichte $\tilde{\psi}_i$ in 2.11. Tatsächlich stimmen die b_i in diesem Fall bis auf eine multiplikative Konstante mit den Koeffizienten eines ARFIMA($p, \frac{\delta-1}{2}, q$)-Prozesses überein. Für den Spezialfall $p = q = 0$ ist die Darstellung (3.26) identisch mit

$$\sigma_t = b_0 + (1 - L)^{(\delta-1)/2}\{R_t\}. \tag{3.32}$$

Somit wird deutlich, dass ein GRS-ARCH-Prozess auf eine ähnliche Weise das langfristige Verhalten beschreiben kann wie ein fraktional integrierter ARMA-Prozess. Da hierbei die Volatilität über die Größe σ_t aus (3.1) und nicht direkt über σ_t^2 modelliert wird, kann die Prozessvarianz endlich bleiben und somit die schwache Stationarität beibehalten werden. Es ist allerdings nicht korrekt, den GRS-ARCH-Prozess dem ARFIMA-Prozess in irgendeiner Form gleichzusetzen, denn die Restgrößen $(R_t - \sigma_t)$, die sich nach der Anwendung des ARFIMA-Filters ergeben, sind im Allgemeinen autokorreliert und stellen somit keine Innovationen dar.
Als ein weiteres Ergebnis zeigen GRS, dass die normierten partiellen Summen von Quadraten und höheren Potenzen eines GRS-ARCH-Prozesses mit $b_i \sim i^{-\frac{\delta+1}{2}}$ unter bestimmten Voraussetzungen gegen die fraktionale Brownsche Bewegung konvergieren. Die genaue Gestalt der Grenzverteilung gibt ihr Theorem 2.3 wieder.
Bei der Einführung des Modells richten GRS den Blick hauptsächlich auf das Langzeitverhalten des zugehörigen Prozesses, denn vor allen Dingen in diesem Bereich besitzt es offensichtlich vorteilhafte Eigenschaften. Gleichzeitig weisen die Autoren auf die bislang unerforschten Fragen im Zusammenhang mit der GRS-ARCH-Beschreibung der Kurzzeitstrukturen in den Volatilitäten hin. Die folgenden Ausführungen verfolgen unter anderem das Ziel, diese Lücke sowohl durch analytische Aussagen als auch durch empirische Untersuchungen zu schließen.
Als erster Schritt in diese Richtung wird der GRS-ARCH-Prozess auf die NED-Eigenschaft untersucht. Das folgende Lemma stellt fest, unter welchen Annahmen

für die Parameter b_i der Prozess eine für den FZGWS ausreichende NED-Struktur aufweist. Auf diese Erkenntnis ist man angewiesen, wenn im nächsten Kapitel die Grenzverteilungen für die Teststatistiken der Einheitswurzeltests hergeleitet werden.

Lemma 1

Sei $\{R_t\}$ ein schwach stationärer GRS-ARCH-Prozess mit $|b_i| \sim i^{-1-\vartheta}$, $\vartheta > 0$. Dann sind $\{R_t\}$ und der zugehörige Volatilitätenprozess $\{\sigma_t\}$ L^2-NED der Ordnung $-1/2$ bezüglich $\{\varepsilon_t\}$.

Beweis:

Wir betrachten zunächst σ_t und erhalten für den GRS-Prozess nach m rekursiven Schritten aus (3.1) und (3.26):

$$\begin{aligned} \sigma_t &= b_0 + \sum_{i_1=1}^{\infty} b_{i_1}\varepsilon_{t-i_1}\sigma_{t-i_1} = b_0 + b_0\sum_{i_1=1}^{\infty} b_{i_1}\varepsilon_{t-i_1} + b_0\sum_{i_1=1}^{\infty}\sum_{i_2=1}^{\infty} b_{i_1}b_{i_2}\varepsilon_{t-i_1}\varepsilon_{t-i_1-i_2} \\ \ldots &+ b_0\sum_{i_1=1}^{\infty}\ldots\sum_{i_m=1}^{\infty} b_{i_1}\ldots b_{i_m}\varepsilon_{t-i_1}\ldots\varepsilon_{t-i_1-\ldots-i_m} \\ &+ \sum_{i_1=1}^{\infty}\ldots\sum_{i_{m+1}=1}^{\infty} b_{i_1}\ldots b_{i_{m+1}}\varepsilon_{t-i_1}\ldots\varepsilon_{t-i_1-\ldots-i_{m+1}}\sigma_{t-i_1-\ldots-i_{m+1}}. \end{aligned} \tag{3.33}$$

Es ist zu beachten, dass der Ausdruck

$$S(i_1,\ldots i_k) := \sum_{i_1=1}^{\infty}\ldots\sum_{i_k=1}^{\infty} b_{i_1}\ldots b_{i_k}\varepsilon_{t-i_1}\ldots\varepsilon_{t-i_1-\ldots-i_k}$$

für $i_1+\ldots+i_k \leq m$ $\mathcal{F}_{t-m}^{t+m}$-messbar ist. Daher sind bei der Berechnung von $\left\|\sigma_t - E(\sigma_t|\,\mathcal{F}_{t-m}^{t+m})\right\|_2$ nur solche Komponenten $S(i_1,\ldots i_k)$ zu berücksichtigen, die weiter als m Perioden zurück in der Vergangenheit liegende Zufallsschocks, d. h. $\varepsilon_{t-i_1\ldots-i_k}$ mit $i_1+\ldots+i_k>m$ enthalten. Wegen unabhängig verteilter ε_t gilt für solche Terme

$$E\left(S(i_1,\ldots i_k)|\mathcal{F}_{t-m}^{t+m}\right) = 0.$$

Somit erhält man mit Hilfe der Indikatorfunktion I und der Minkowski-Ungleichung folgende Darstellung:

$$\begin{aligned} \left\|\sigma_t - E(\sigma_t|\,\mathcal{F}_{t-m}^{t+m})\right\|_2 &\leq \left\|b_0\sum_{k=1}^{m} S(i_1,\ldots i_k)I_{\{i_1+\ldots+i_k>m\}}(i_1,\ldots,i_k)b_{i_1}\ldots b_{i_k}\right\|_2 \\ &+ \left\|\sum_{i_1=1}^{\infty}\ldots\sum_{i_{m+1}=1}^{\infty} b_{i_1}\ldots b_{i_{m+1}}\varepsilon_{t-i_1}\ldots\varepsilon_{t-i_1-\ldots-i_{m+1}}\sigma_{t-i_1-\ldots-i_{m+1}}\right\|_2. \end{aligned} \tag{3.34}$$

Verbunden mit der stochastischen Unabhängigkeit der ε_t impliziert $E\varepsilon_t^2 = 1$, dass

$$E(S(i_1,\dots,i_k))^2 = E\Big(\sum_{i_1=1}^{\infty}\dots\sum_{i_k=1}^{\infty} b_{i_1}\dots b_{i_k}\varepsilon_{t-i_1}\dots\varepsilon_{t-i_1-\dots-i_k}\Big)^2 = \sum_{i_1=1}^{\infty}\dots\sum_{i_k=1}^{\infty} b_{i_1}^2\dots b_{i_k}^2,$$

und

$$E\Big(\sum_{i_1=1}^{\infty}\dots\sum_{i_{m+1}=1}^{\infty} b_{i_1}\dots b_{i_{m+1}}\varepsilon_{t-i_1}\dots\varepsilon_{t-i_1-\dots-i_{m+1}}\sigma_{t-i_1-\dots-i_{m+1}}\Big)^2 = \sum_{i_1=1}^{\infty}\dots\sum_{i_{m+1}=1}^{\infty} b_{i_1}^2\dots b_{i_{m+1}}^2\, E\sigma_t^2.$$

Das asymptotische Verhalten von b_i^2 folgt direkt aus der Annahme $|b_i| \sim i^{-1-\vartheta}$:

$$b_i^2 \sim i^{-1-\tilde\vartheta} \quad \text{mit} \quad \tilde\vartheta := 1 + 2\vartheta > 1.$$

Mit diesen Ergebnissen und unter Berücksichtigung der Folgerung

$$i_1 + \dots + i_k > m \;\Rightarrow\; [\max\{i_1,\dots i_k\}] > [m/k]$$

kann man die Komponenten von (3.34) wie folgt abschätzen:

$$\sum_{i_1=1}^{\infty}\dots\sum_{i_k=1}^{\infty} b_{i_1}^2\dots b_{i_k}^2 I_{\{i_1+\dots+i_k>m\}}(i_1,\dots,i_k) \le B^{k-1}\sum_{[m/k]+1}^{\infty} b_i^2$$

$$\le B^{k-1} C \sum_{[m/k]+1}^{\infty}\left(\frac{1}{i}\right)^{1+\tilde\vartheta} = O(k^{\tilde\vartheta} m^{-\tilde\vartheta} B^{k-1}),$$

$$\sum_{i_1=1}^{\infty}\dots\sum_{i_{m+1}=1}^{\infty} b_{i_1}^2\dots b_{i_{m+1}}^2\, E\sigma_t^2 = B^{m+1}E\sigma_t^2.$$

Schließlich ist wegen $B < 1$ (siehe (3.27))

$$\left\|\sigma_t - E(\sigma_t|\,\mathcal{F}_{t-m}^{t+m})\right\|_2 = O\Big(m^{-\tilde\vartheta/2}\sum_{k=1}^{m} k^{\tilde\vartheta/2}B^{(k-1)/2}\Big) + o(1) = O(m^{-\tilde\vartheta/2}).$$

Aus der Definition von $\tilde\vartheta > 1$ folgt, dass die NED-Ordnung $-1/2$ beträgt, vgl. die Definition 2.
Die Beziehung $\{R_t\} = \{\sigma_t\varepsilon_t\}$ und $E\varepsilon_t^2 = 1$ liefert direkt, dass $\{R_t\}$ ebenfalls L^2-NED bzgl. $\{\varepsilon_t\}$ mit der gleichen Ordnung ist. □

Als Folge des Lemmas hält man fest, dass das asymptotische Abklingen der Koeffizienten b_i mit der Rate $i^{-1-\vartheta}$, $\vartheta > 0$, hinreichend für den FZGWS ist. Die Beziehung

$b_i \sim 1/i$ stellt die Schwelle beim Übergang zu länger andauernden Abhängigkeiten dar, die der Standardasymptotik nicht mehr genügen. Diese Behauptung steht im Einklang mit den Ergebnissen von GRS, vgl. (3.30). Nehmen b_i für wachsende i dagegen schneller ab, verstärkt dies den Kurzfristcharakter der seriellen Abhängigkeiten und erhöht damit die NED-Ordnung des GRS-ARCH-Prozesses.

Kapitel 4

Tests auf stochastische Trends bei bedingter Heteroskedastie

Dieses Kapitel behandelt die zentrale Frage der vorliegenden Untersuchung: inwieweit halten die Standardtests auf Einheitswurzel unter autoregressiven Volatilitätsstrukturen das Niveau ein? Auch wenn diese für die Praxis extrem relevante Fragestellung nicht neu ist - schon Kim und Schmidt (1993) beschäftigen sich in einer Simulationsstudie mit diesem Thema - sind die meisten Erkenntnisse experimenteller Natur. Konkret ist bisher nur der Fall der GARCH(p, q)-Innovationen durch die analytische Untersuchung von Seo (1999) abgedeckt. Auf deren Ergebnis wird kurz eingegangen, bevor der Einfluss der GRS-ARCH-Innovationen auf die DF-Tests detailliert betrachtet wird.

4.1 Einheitwurzeltests und GARCH-Innovationen

Den Ausgangspunkt stellen der erweiterte DF-Ansatz und die Dickey-Fuller-Transformation (2.23) eines AR(k+1)-Prozesses dar, wobei weißes Rauschen $\{\eta_t\}$ durch einen schwach stationären GARCH(p, q)-Prozess $\{R_t\}$ wie in (3.3) ersetzt wird. Angenommen, die GARCH-Schocks aus (3.2) sind standardnormalverteilt. Dann besitzten die auf ML-Schätzer $\hat{\Psi}_{ML}$ basierenden DF-Teststatistiken unter der Nullhypothese $H_0: \Psi = 0$, der Stationaritätsbedingung (3.6), Momentenexistenz $E|R_t|^8 < \infty$ für alle t sowie zwei weiteren Regularitätsbedingungen zum Verhalten von Prozessparametern nach Seo (1999) folgende Grenzverteilungen:

$$T\hat{\Psi} \xrightarrow{d} \frac{\int\limits_0^1 W_1(r)dW_1(r) + \sqrt{E\Big(\frac{\sigma^2}{\sigma_t^2}\Big) + 2\sum\limits_{i=1}^{\infty}\gamma_i^2 E\Big(\frac{\sigma^2\varepsilon_{t-i}^2}{\sigma_t^4}\Big) - 1}\int\limits_0^1 W_1(r)dW_2(r)}{\Big[E\Big(\frac{\sigma^2}{\sigma_t^2}\Big) + 2\sum\limits_{i=1}^{\infty}\gamma_i^2 E\Big(\frac{\sigma^2 R_{t-i}^2}{\sigma_t^4}\Big)\Big]\Big(1-\sum\limits_{j=1}^{\infty}\psi_j\Big)^{-1}\int\limits_0^1 W_1^2(r)dr} \tag{4.1}$$

$$\tau \xrightarrow{d} \frac{\int\limits_0^1 W_1(r)dW_1(r) + \sqrt{E\Big(\frac{\sigma^2}{\sigma_t^2}\Big) + 2\sum\limits_{i=1}^{\infty}\gamma_i^2 E\Big(\frac{\sigma^2 R_{t-i}^2}{\sigma_t^4}\Big) - 1}\int\limits_0^1 W_1(r)dW_2(r)}{\sqrt{E\Big(\frac{\sigma^2}{\sigma_t^2}\Big) + 2\sum\limits_{i=1}^{\infty}\gamma_i^2 E\Big(\frac{\sigma^2\varepsilon_{t-i}^2}{\sigma_t^4}\Big)\int\limits_0^1 W_1^2(r)dr}}. \tag{4.2}$$

Dabei ist $\sigma^2 = Var(\varepsilon_t) = E(\sigma_t^2)$, W_1 und W_2 sind zwei unabhängige Wiener-Prozesse, γ_i gehören zum ARCH(∞)-Filter $\gamma(L) = \alpha(L)(1-\beta(L))^{-1}$ und ψ_j zum AR(k)-Filter $\psi(L)$ aus (2.23).

Ohne bedingt heteroskedastische Effekte ist für beide Teststatistiken die gleiche Verteilungsasymptotik wie in den gewöhnlichen DF-Tests zu beobachten, denn $\gamma_i = 0$ für alle i und $\sigma^2/\sigma_t^2 = 1$ für alle t. Durch einfache äquivalente Umformungen der Grenzverteilung in (4.2) kann man die GARCH-Einflüsse mit einem einzigen Parameters ξ ausdrücken:

$$\tau \equiv \tau(\xi) \xrightarrow{d} \frac{\xi\int\limits_0^1 W_1(r)dW_1(r) + \sqrt{1-\xi^2}\int\limits_0^1 W_1(r)dW_2(r)}{\sqrt{\int\limits_0^1 W_1^2(r)dr}} \tag{4.3}$$

mit $\xi := \left(2\sum\limits_{i=1}^{\infty}\gamma_i^2 E\Big(\frac{\sigma^2 R_{t-i}^2}{\sigma_t^4}\Big) + E\Big(\frac{\sigma^2}{\sigma_t^2}\Big)\right)^{-1/2}$. Diese Darstellung ist gegenüber (4.2) leichter zu interpretieren. Ist keine GARCH-Struktur zu verzeichnen, so ist $\xi = 1$. Sind GARCH-Parameter dagegen von Null verschieden, so ist $\sum\gamma_i^2 > 0$ und wegen der Ungleichung von Jensen $E(\sigma^2/\sigma_t^2) > 1$. Daher gilt allgemein $0 < \xi \leq 1$, so dass auch die zweite Komponente $\int_0^1 W_1(r)\,dW_2(r)\Big/\sqrt{\int_0^1 W_1^2(r)dr}$ die interessierende Grenzverteilung beeinflusst. Bei näherem Betrachten stellt man fest, dass es sich dabei um eine N(0,1)-verteilte Zufallsvariable handelt. Da W_1 und W_2 unabhängige Wiener-Prozesse sind, ist der Zähler normalverteilt mit dem Erwartungswert 0. Mit den Rechenregeln für stochastische Intergrale gilt: $E\,(\int_0^1 W_1(r)dW_2(r))^2 = E\,\int_0^1 W_1^2(r)dr$. Die Grenzverteilung für τ kann daher als eine Linearkombination zwischen der DF-Verteilung aus (2.21) und der Standardnormalverteilung angesehen werden. Dabei tritt der Parameter ξ als eine Art Gewicht auf: mit stärkeren GARCH-Effekten wird ξ kleiner und der Anteil der zweiten Normalverteilungskomponente gewichtiger, während der Einfluss der DF-Verteilung abnimmt. Im Grenzfall $\xi \to 0$ nähert sich die Verteilung der Teststatistik τ der Standardnormalverteilung an. Dies wird durch die von Seo (1999) in der Tabelle 3 angeführten Simulationsergebnisse bestätigt. So steigen die Quantile der τ-Grenzverteilung mit fallendem ξ monoton und sind im Extremfall $\xi = 0$ identisch mit denen der Standardnormalverteilung. Für die Praxis schlägt Seo (1999) vor, für die Auswahl der richtigen kritischen Werte $\hat{\xi}$ zu verwenden, indem die entsprechenden Komponenten konsistent geschätzt werden, und für $\sum_i \gamma_i^2 E(\sigma^2 R_{t-i}^2/\sigma_t^4)$ eine genügend hohe Summationsgrenze einzusetzen ist.

4.2 Einheitswurzeltests und GRS-ARCH-Prozess

Es wird im Weiteren ein AR($k+1$)-Prozess betrachtet, seine Innovationen R_t folgen nun einem schwach stationären GRS-ARCH-Prozess wie im Lemma 1. Das komplette Modell sieht demnach wie folgt aus:

$$\begin{aligned}
\Delta Y_t &= \Psi Y_{t-1} + \sum_{j=1}^{k} \psi_j \Delta Y_{t-j} + R_t, \quad R_t = \sigma_t\, \varepsilon_t, \\
\sigma_t &= b_0 + \sum_{i=1}^{\infty} b_i R_{t-i}, \ b_0 \neq 0, \ \varepsilon_t \text{ sind u.i.N(0,1)-verteilt.} \qquad (4.4)
\end{aligned}$$

Zu untersuchen ist das Testproblem $H_0 : \Psi = 0$ gegen $H_1 : \Psi < 0$.
Anhand Simulationen soll überprüft werden, inwieweit die Standardtests das Niveau auch in empirischen Untersuchungen, d. h. bei endlich langen Reihen einhalten.
Es werden jeweils 5000 Random Walk-Prozesse der Länge $n = 1000$ mit unterschiedlich parametrisierten GRS-ARCH-Innovationen simuliert. Die Länge für den Parametervektor $(b_0, b_1, \ldots)'$ wird zu Simulationszwecken auf 4 begrenzt. In jedem Simulationslauf erzeugt man zunächst 1050 stardardnormalverteilte Zufallsvariablen ε_t, die, multipliziert mit σ_t wie im Modell (4.4), Innovationen R_t ergeben. Um den Startwert für σ_1 zu erhalten, werden im ersten Schritt alle benötigten R_{1-i} gleich Null gesetzt, so dass $\sigma_1 = b_0$ und $R_1 = b_0\, \varepsilon_1$. Im zweiten Schritt geht R_1 in die Berechnung mit ein, $\sigma_2 = b_0 + b_1 R_1$. Allgemein werden in einem GRS-ARCH-Modell nach $m+1$ Schritten die Parameter $b_0, b_1, \ldots, b_m$ sukzessive in die Bestimmung von σ_t einbezogen. Um dieser Einschwingungsphase in der simulierten Reihe Rechnung zu tragen, werden anschließend die ersten 50 Werte aus dem Innovationenprozess $\{R_t\}$ entfernt. Aus den restlichen Innovationen $R_{51}, \ldots, R_{1050}$ bildet man für $s = 2, \ldots, 1001$ Partialsummen

$Y_s = \sum_{i=51}^{50+(s-1)} R_i$ und erhält auf diese Weise mit $\{Y_s\}$ einen Random Walk-Prozess der Länge 1000 mit GRS-ARCH-Innovationen.
Dieser Prozess wird dem gewöhnlichen DF-Test (Teststatistik wie in (2.21)) unterzogen. Die ensprechenden Ablehnhäufigkeiten der zutreffenden Nullhypothese gibt die Tabelle 1 an. Daraus wird ersichtlich, dass mit stärkeren bedingt heteroskedastischen Effekten das nominale Testniveau von 5% zunehmend überschritten wird.

Parametervektor $(b_0, b_1, b_2, b_3)'$	$\left\lvert\sum_{i=1}^{3} b_i\right\rvert$	relative Ablehnhäufigkeit (t-Statistik< -1.94)
$(1\ 0.05\ 0.03\ 0.02)'$	0.1	0.054
$(1\ 0.25\ 0.15\ 0.05)'$	0.5	0.056
$(1\ 0.50\ 0.30\ 0.10)'$	0.9	0.069

Tabelle 1: empirisches Testniveau des DF-Tests unter GRS-ARCH-Innovationen

Es liegt daher nahe, als einen Grund für das insgesamt zu häufige Ablehnen der I(1)-Hypothese die durch die KQ-Schätzung von Ψ vernachläßigte Dynamik in bedingten Varianzen zu vermuten. Deswegen kann versucht werden, das Abschneiden der Einheitwurzeltests zu verbessern, indem ähnlich wie in Seo (1999) die Teststatistiken auf $\hat{\Psi}_{ML}$ aufgebaut werden. Dazu ist die Gestalt von $\hat{\Psi}_{ML}$ aus dem für AR- und ARCH-Parameter gemeinsamen Schätzvektor $\hat{\theta}_{ML} = (\hat{\Psi}, \hat{\psi}_1, \ldots, \hat{\psi}_k, \hat{b}_0, \hat{b}_1, \ldots \hat{b}_\infty)'$ zu ermitteln. Das Ergebnis ist im Lemma 2 angeführt. Lemma 3 stellt die Grenzverteilungen für die Komponenten von $\hat{\Psi}_{ML}$ fest, die schließlich im Satz 2 zu asymptotischen Verteilungsaussagen über die kompletten Teststatistiken zusammengeführt werden.

Den Ausgangpunkt für das folgende Vorgehen bildet die mittlere Log-Likelihoodfunktion

$$L_n(\theta) = \frac{1}{n}\sum_{t=1}^{n} l_t(\theta) \ \text{ mit } \ l_t(\theta) = -\frac{1}{2}\left(\log 2\pi + \log \sigma_t^2(\theta) + \frac{R_t^2(\theta)}{\sigma_t^2(\theta)}\right)$$

Mit den Modellgleichungen (4.4) ergeben sich folgende Ableitungen:

$$\frac{\partial l_t(\theta)}{\partial \Psi} = -\frac{1}{\sigma_t} b(L)\{Y_{t-1}\}\,(\varepsilon_t^2 - 1) + \frac{R_t Y_{t-1}}{\sigma_t^2} \qquad \text{und}$$

$$\frac{\partial^2 l_t(\theta)}{\partial \Psi^2} = -\frac{1}{\sigma_t^2}\left[(b(L)\{Y_{t-1}\})^2(\varepsilon_t^2 - 1) - 4b(L)\{Y_{t-1}\}\,\varepsilon_t\, Y_{t-1} + Y_{t-1}^2\right]. \quad (4.5)$$

Der ML-Schätzvektor, dem die Reihenlänge n zugrunde liegt, wird im Weiteren mit $\hat{\theta}_n = \arg\max_{\theta} L_n(\theta)$ bezeichnet, die wahren Parameterwerte sind mit einem * versehen. Gesucht werden die asymptotischen Verteilungen für die Teststatistiken $n\hat{\Psi}_n$ bzw. $\tau_n = n\hat{\Psi}_n/\widehat{s.e.}(n\hat{\Psi}_n)$, deren Konstruktion nach dem Wald-Prinzip erfolgt. Der Standardfehler $s.e.(\hat{\theta}_n)$ wird dabei mit Hilfe der Fisher-Information $I(\theta)$ angegeben.
Um die interessierenden Verteilungen angeben zu können, seien folgende Regularitätsbedingungen vorausgesetzt:

R1: $E|\,R_t|^6 < \infty$;

R2: $1 - \psi(L)$ ist invertierbar, so dass für $\varphi(L) := (1 - \psi(L))^{-1}$ gilt: $\sum_{j=1}^{\infty} j|\varphi_j| < \infty$.

Außerdem ist im betrachteten Modell ausgeschlossen, dass σ_t zu Null und folglich $|1/\sigma_t|$ unendlich wird. Genauer impliziert die Modellannahme $b_0 \neq 0$, dass es zu jedem Parametervektor θ ein $u > 0$ exisitiert, so dass $|\sigma_t| = |b_0 + b(L)R_t| > u$. Mit $\tilde{u} := \inf_{\theta} u > 0$ erhält man

$$\sup_{1\le t\le n} \sup_{\theta} |1/\sigma_t| < 1/\tilde{u} < \infty. \tag{4.6}$$

Ein wichtiger Punkt ist ferner eine eventuelle Superkonsistenz von $\hat{\Psi}_n$: unter der Nullhypothese strebt $\hat{\Psi}_n$ mit wachsender Stichprobe, wie im Kapitel 2 erläutert, schneller gegen den wahren Parameterwert als bei Stationarität. Die Varianz des Schätzers verhält sich wie $1/n^2$: $n(\hat{\Psi} - \Psi_0) = O_p(1)$. Dies berücksichtigt man durch die Normierung $diag(n, \sqrt{n}, \ldots)(\hat{\theta}_n - \theta^*) = O_p(1)$ für den gesamten Schätzvektor $\hat{\theta}_n$. Unter Einhaltung von R1 und R2 ist man in der Lage, folgende Aussage zu formulieren:

Lemma 2
Für die erste Komponente eines konsistenten ML-Schätzers $\hat{\theta}_n$ gilt unter H_0:

$$n\hat{\Psi}_n = \frac{1}{n}\sum_{t=1}^{n}\frac{\partial l_t(\theta^*)}{\partial \Psi} \Bigg/ -\frac{1}{n^2}\sum_{t=1}^{n}\frac{\partial^2 l_t(\theta^*)}{\partial \Psi^2} + o_p(1)\,. \tag{4.7}$$

Das Lemma stützt sich auf die Behauptung 5f.2ii von Rao (1973). Sie besagt, dass unter gewissen Regularitätsbedingungen gilt:

$$\sqrt{n}(\theta_n - \theta^*)I(\theta^*) - \frac{1}{\sqrt{n}}\frac{\partial L_n(\theta^*)}{\partial \theta} = o_p(1). \tag{4.8}$$

Die Fisher-Information $I(\theta)$ kann bekanntlich durch die negative Hesse-Matrix $-H_n(\theta)$ konsistent geschätzt werden, wobei $H_n(\theta) = 1/n \sum_{t=1}^{n} \partial^2 l_t(\theta)/\partial\theta\partial\theta'$. Liegt allerdings ein stochastischer Trend vor, so schlägt sich die Superkonsistenz von $\hat{\Psi}$ auch in der Schätzung der Fisher-Information nieder (s. Saikkonnen (1995)). Aus diesem Grund wird analog zum Vorgehen bei Seo (1999) die Hesse-Matrix modifiziert:

$$\widetilde{H}_n(\theta) = \frac{1}{n}\sum_{t=1}^{n} D\frac{\partial^2 l_t(\theta)}{\partial\theta\partial\theta'}D \quad \text{mit} \quad D = diag\left(\frac{1}{\sqrt{n}}, 1, \ldots, 1\right). \tag{4.9}$$

Der formale Beweis des Lemmas besteht aus zwei Teilen, bei denen $\widetilde{H}_n(\theta)$ betrachtet wird. Zum einen ist die Konsistenz für die modifizierte Hesse-Matrix nachzuweisen. Dies läuft auf die routinemäßige Überprüfung der von Rao (1973) formulierten Regularitätsbedingungen für die dritten partiellen Ableitungen von $l_t(\theta)$ hinaus. Dabei muss allerdings bei dem Ableiten nach Ψ der Korrekturfaktor $1/\sqrt{n}$ berücksichtigt werden. Zum anderen muss $\widetilde{H}_n(\theta^*)$ asymptotisch eine blockdiagonale Gestalt aufweisen, so dass der Einfluss weiterer partieller Ableitungen aus (4.9) auf $n\hat{\Psi}_n$ vernachlässigbar wird. Es folgt der komplette Beweis für ein AR(1)-GRS-ARCH(1)-Prozess. Er kann auf Modelle höherer Ordnung übertragen werden. Die zusätzlichen Parameter haben dabei Auswirkungen auf den Umfang des Beweises, nicht aber auf das Vorgehen.

Beweis von Lemma 2:
Es werden nur autoregressive Komponenten erster Ordnung betrachtet, d. h. die Modellgleichungen (4.4) werden zu $\Delta Y_t = \Psi\, Y_{t-1} + R_t$, $R_t = \sigma_t\, \varepsilon_t, ;\ \sigma_t = b_0 + b_1\, R_{t-1}$. Demnach besteht der Parametervektor aus drei Komponenten: $\theta = (\Psi, b_0, b_1)'$. Unter der Nullhypothese ist $Y_t \sim I(1)$, d. h. $\Psi^* = 0$.
Zu zeigen ist zunächst, dass für alle Parameterwerte $\widetilde{H}(\theta)$ konsistent für $\widetilde{H}(\theta^*)$ ist. Dazu muss für alle einzelnen Einträge $\widetilde{H}_{nij}(\theta)$ gelten:

$$|\widetilde{H}_{nij}(\theta) - \widetilde{H}_{nij}(\theta^*)| = o_p(1), \quad i,j = \Psi, b_0, b_1 \tag{4.10}$$

Die Taylor-Reihenentwicklung 1. Ordnung für $\widetilde{H}_{nij}(\theta)$ an der Stelle θ^* ergibt

$$\widetilde{H}_{nij}(\theta) - \widetilde{H}_{nij}(\theta^*) \approx \left(\frac{\partial \widetilde{H}_{nij}(\bar{\theta})}{\partial\theta}\right)'(\theta - \theta^*) = \left(\frac{\partial \widetilde{H}_{nij}(\bar{\theta})}{\partial\theta}\right)' DD^{-1}(\theta - \theta^*), \quad \bar{\theta} \in [\theta, \theta^*].$$

Wegen der Konstruktion von $\widetilde{\Theta}$ ist $D^{-1}(\theta - \theta^*) = diag(\sqrt{n}, 1, 1)(\theta - \theta^*) = o_p(1)$. Damit (4.10) gilt, reicht es folglich zu zeigen, dass

$$\sup_{\theta\in\widetilde{\Theta}} \left|\frac{\partial \widetilde{H}_{nij}(\theta)}{\partial\theta}\right| D = O_p(1).$$

Für die einzelnen partiellen Ableitungen 3. Grades wird daher gefordert:

$$\sup_{1\le t\le n} E\sup_{\theta\in\widetilde{\Theta}} \frac{1}{n^{3/2}} \left|\frac{\partial^3 l_t(\theta)}{\partial\Psi^3}\right| = \sup_{1\le t\le n} E\sup_{\theta\in\widetilde{\Theta}} \frac{1}{n^{3/2}} \left|\frac{1}{\sigma_t^3}\left[\,2(b_1Y_{t-2})^3 - 4(b_1Y_{t-2})^3\,\varepsilon_t^2 \right.\right.$$

$$\left.\left. +14(b_1Y_{t-2})^2\,Y_{t-1}\,\varepsilon_t - 6b_1Y_{t-2}\,Y_{t-1}^2\,\right]\right| < \infty\,, \quad (4.11)$$

$$\sup_{1\le t\le n} E\sup_{\theta\in\widetilde{\Theta}} \frac{1}{n} \left|\frac{\partial^3 l_t(\theta)}{\partial\Psi^2\partial b_0}\right| = \sup_{1\le t\le n} E\sup_{\theta\in\widetilde{\Theta}} \frac{1}{n} \left|\frac{1}{\sigma_t^3}\left[\,4(b_1Y_{t-2})^2\,\varepsilon_t^2 - 12b_1Y_{t-2}\,\varepsilon_t \right.\right.$$

$$\left.\left. +2Y_{t-1}^2 - 2(b_1Y_{t-2})^2\,\right]\right| < \infty,$$

$$\sup_{1\le t\le n} E\sup_{\theta\in\widetilde{\Theta}} \frac{1}{n} \left|\frac{\partial^3 l_t(\theta)}{\partial\Psi^2\partial b_1}\right| = \sup_{1\le t\le n} E\sup_{\theta\in\widetilde{\Theta}} \frac{1}{n} \left|\frac{1}{\sigma_t^3}\left[\,4(b_1Y_{t-2})^2\,R_{t-1}\,\varepsilon_t^2 - 12b_1Y_{t-2}\,R_{t-1}\,Y_{t-1}\,\varepsilon_t \right.\right.$$

$$\left.\left. -2(b_1Y_{t-2})^2\,R_{t-1} + 2Y_{t-1}^2\,R_{t-1}\,\right] - \frac{1}{\sigma_t^2}\left[\,2b_1Y_{t-2}^2\,\varepsilon_t^2 - 2b_1Y_{t-2}^2 - 4Y_{t-1}^2\,\varepsilon_t\,\right]\right| < \infty,$$

$$\sup_{1\le t\le n} E\sup_{\theta\in\widetilde{\Theta}} \frac{1}{\sqrt{n}} \left|\frac{\partial^3 l_t(\theta)}{\partial\Psi\partial b_0^2}\right| = \frac{1}{\sqrt{n}} \left|\frac{1}{\sigma_t^3}\,\left(\,2b_1Y_{t-2} + 6Y_{t-1}\,\varepsilon_t - 12b_1Y_{t-2}\,\varepsilon_t^2\,\right)\right| < \infty,$$

$$\sup_{1\le t\le n} E\sup_{\theta\in\widetilde{\Theta}} \frac{1}{\sqrt{n}} \left|\frac{\partial^3 l_t(\theta)}{\partial\Psi\partial b_1^2}\right| = \sup_{1\le t\le n} E\sup_{\theta\in\widetilde{\Theta}} \frac{1}{\sqrt{n}} \left|\frac{1}{\sigma_t^3}\left(6R_{t-1}\,Y_{t-1}\,\varepsilon_t - 12b_1Y_{t-2}\,R_{t-1}^2\,\varepsilon_t^2 \right.\right.$$

$$\left.\left. +2b_1Y_{t-2}\,R_{t-1}^2\right) - \frac{1}{\sigma_t^2}\,\left(\,2Y_{t-1}R_{t-1} - 6R_{t-1}\,Y_{t-1}\,\varepsilon_t^2\,\right)\right| < \infty,$$

$$\sup_{1\le t\le n} E\sup_{\theta\in\widetilde{\Theta}} \frac{1}{\sqrt{n}} \left|\frac{\partial^3 l_t(\theta)}{\partial\Psi\partial b_0\partial b_1}\right| = \sup_{1\le t\le n} E\sup_{\theta\in\widetilde{\Theta}} \frac{1}{\sqrt{n}} \left|\frac{1}{\sigma_t^3}\left(6R_{t-1}\,Y_{t-1}\,\varepsilon_t - 12b_1Y_{t-2}\,R_{t-1}\,\varepsilon_t^2 \right.\right.$$

$$\left.\left. +2b_1Y_{t-2}\,R_{t-1}\right) + \frac{1}{\sigma_t^2}\,\left(3Y_{t-1}\,\varepsilon_t^2 - Y_{t-1}\,\right)\right| < \infty,$$

$$\sup_{1\le t\le n} E\sup_{\theta\in\widetilde{\Theta}} \left|\frac{\partial^3 l_t(\theta)}{\partial b_0^3}\right| = \sup_{1\le t\le n} E\sup_{\theta\in\widetilde{\Theta}} \left|\frac{1}{\sigma_t^3}\,\left(12\varepsilon_t^2 - 2\right)\right| < \infty,$$

$$\sup_{1\le t\le n} E\sup_{\theta\in\widetilde{\Theta}} \left|\frac{\partial^3 l_t(\theta)}{\partial b_1^3}\right| = \sup_{1\le t\le n} E\sup_{\theta\in\widetilde{\Theta}} \left|\frac{R_{t-1}^3}{\sigma_t^3}\,\left(12\varepsilon_t^2 - 2\right)\right| < \infty,$$

$$\sup_{1\le t\le n} E\sup_{\theta\in\widetilde{\Theta}} \left|\frac{\partial^3 l_t(\theta)}{\partial b_1^2\partial b_0}\right| = \sup_{1\le t\le n} E\sup_{\theta\in\widetilde{\Theta}} \left|\frac{R_{t-1}^2}{\sigma_t^3}\,\left(12\varepsilon_t^2 - 2\right)\right| < \infty,$$

$$\sup_{1\le t\le n} E\sup_{\theta\in\tilde{\Theta}}\left|\frac{\partial^3 l_t(\theta)}{\partial b_0^2\partial b_1}\right| = \sup_{1\le t\le n} E\sup_{\theta\in\tilde{\Theta}}\left|\frac{R_{t-1}}{\sigma_t^3}\left(12\varepsilon_t^2-2\right)\right| < \infty.$$

Um die Gültigkeit der letzten vier Ungleichungen nachzuweisen, greift man die Bedingung R1 auf. Mit der Hölder-Ungleichung und (4.6) erhält man dann beispielsweise für $\partial^3 l_t(\theta)/\partial b_1^3$:

$$\sup_{1\le t\le n} E\sup_{\theta\in\tilde{\Theta}}\left|\frac{R_{t-1}^3}{\sigma_t^3}\left(12\varepsilon_t^2-2\right)\right| < \frac{1}{\upsilon^3}\Big(E|\,R_t|^6\Big)^{1/2}\Big(E(12\varepsilon_t^2-2)^2\Big)^{1/2} < \infty.$$

Bevor die restlichen Ungleichungen einzeln betrachtet werden, wird gezeigt, dass $\sup\limits_{1\le t\le n}\|Y_t/\sqrt{n}\|_6 < \infty$. Als Hilfsmittel dient dabei die Burkholder-Ungleichung. Sie gibt die Grenzen für die absoluten Momente eines Martingals $\{M_t\}$ vor:

$$\left(\frac{1}{18\sqrt{p}\,q}\right)^p E\left|\sum_{i=1}^t (M_i-M_{i-1})^2\right|^{p/2} \le E|M_t|^p \le (18p\sqrt{q})^p E\left|\sum_{i=1}^t (M_i-M_{i-1})^2\right|^{p/2}. \quad (4.12)$$

Die Konstanten p und q sind so zu wählen, dass $1/p+1/q=1,\ p>1$. Zusammen mit der Minkowski-Ungleichung und der Annahme R1 gelangt man zur Abschätzung

$$\begin{aligned}
\sup_{1\le t\le n}\left\|\frac{Y_t}{\sqrt{n}}\right\|_6 &= \frac{1}{\sqrt{n}}\sup_{1\le t\le n}\big(E|\,Y_t|^6\big)^{1/6} \overset{(4.12)}{\le} \frac{108\sqrt{6}}{\sqrt{5n}}\sup_{1\le t\le n}\left(E\left|\sum_{i=1}^t R_i^2\right|^3\right)^{1/6}\\
&= \frac{108\sqrt{6}}{\sqrt{5n}}\sup_{1\le t\le n}\left\|\sum_{i=1}^t R_i^2\right\|_3^{1/2} \overset{Mink.}{\le} \frac{108\sqrt{6}}{\sqrt{5n}}\sup_{1\le t\le n}\sqrt{t}\,\|R_i^2\|_3^{1/2}\\
&= \frac{108\sqrt{6}}{\sqrt{5}}\|R_i\|_6 \overset{\mathsf{R1}}{<} \infty.
\end{aligned} \quad (4.13)$$

Wegen der Monotonie der Funktion $\|\cdot\|_p$ ist somit nachgewiesen, dass $\sup\limits_{1\le t\le n}\|Y_t/\sqrt{n}\|_p<\infty$ für $1\le p\le 6$.

Liegen im Modell (4.4) autoregressive Komponenten höherer Ordnungen vor, so führt der Nachweis über die *Beveridge-Nelson-Zerlegung*. Danach setzt sich ein autoregressiver I(1)-Prozess aus zwei stochastischen Komponenten zusammen - einer permanenten und einer transitorischen. Unter der Nullhypothese erhält man für ΔY_t aus (4.4) mit Hilfe des inversen Filters $\varphi(L)=(1-\psi(L))^{-1}$ eine MA(∞)-Darstellung $\Delta Y_t=\varphi(L)R_t$. Unter der Annahme R2 gilt dann

$$Y_t=\varphi(1)\sum_{i=1}^t R_i-\tilde{\varphi}(L)\{R_t\},\quad \text{wobei}\ \ \tilde{\varphi}_j=\sum_{k=j+1}^{\infty}\varphi_k\,. \quad (4.14)$$

Die Beziehung $\sup\limits_{1\le t\le n} \left\|\varphi(1)/\sqrt{n}\,\sum_{i=1}^{t} R_i\right\|_p < \infty$ lässt sich für $1 \le p \le 6$ durch das gleiche Vorgehen wie in (4.13) zeigen. Der Beitrag der transitorischen Komponente zu $\sup\limits_{1\le t\le n} \|Y_t/\sqrt{n}\|_p$ kann auf lange Sicht vernachlässigt werden, denn

$$\sup_{1\le t\le n}\left\|\frac{\tilde{\varphi}(L)}{\sqrt{n}}\{R_t\}\right\|_6 = \left\|\frac{1}{\sqrt{n}}\sum_{j=1}^{p-2}\tilde{\varphi}_j R_{t-j}\right\|_6 \le \frac{1}{\sqrt{n}}\sum_{j=1}^{\infty}\tilde{\varphi}_j\,\|R_t\|_6 \overset{\text{R1,R2}}{=} o_p(1). \tag{4.15}$$

Auf das Ergebnis (4.13) wird bei der Untersuchung einzelner Summanden in den obigen partiellen Ableitungen Bezug genommen. Exemplarisch beschränkt sich die folgende Betrachtung auf $\partial^3 l_t(\theta)/\partial\Psi^3$, deren Gestalt (4.11) wiedergibt.

$$\sup_{1\le t\le n} E\sup_{\theta\in\tilde{\Theta}}\frac{1}{n^{3/2}}\left|\frac{1}{\sigma_t^3}\,2(b_1Y_{t-2})^3\right| < \frac{2}{\tilde{u}^3}\sup_{1\le t\le n} E\left|\frac{Y_{t-2}^3}{n^{3/2}}\right| = \frac{2}{\tilde{u}^3}\sup_{1\le t\le n}\left\|\frac{Y_{t-2}}{\sqrt{n}}\right\|_1^3 < \infty\,,$$

$$\begin{aligned}
\sup_{1\le t\le n} E\sup_{\theta\in\tilde{\Theta}}\frac{1}{n^{3/2}}\left|\frac{1}{\sigma_t^3}\,4(b_1Y_{t-2})^3\,\varepsilon_t^2\right| &< \frac{4}{\tilde{u}^3}\sup_{1\le t\le n} E\left|\frac{Y_{t-2}^3\,\varepsilon_t^2}{n^{3/2}}\right| \\
&\le \frac{4}{\tilde{u}^3}\sup_{1\le t\le n}\left(E\left|\frac{Y_{t-2}^3}{n^{3/2}}\right|^2\right)^{1/2}\left(E\varepsilon_t^4\right)^{1/2} = \frac{4}{\tilde{u}^3}\sup_{1\le t\le n}\left\|\frac{Y_{t-2}}{\sqrt{n}}\right\|_6^3\left(E\varepsilon_t^4\right)^{1/2} < \infty\,,
\end{aligned}$$

$$\begin{aligned}
\sup_{1\le t\le n} E\sup_{\theta\in\tilde{\Theta}}\frac{1}{n^{3/2}}\left|\frac{1}{\sigma_t^3}\,14(b_1\,Y_{t-2})^2\,Y_{t-1}\,\varepsilon_t\right| &< \frac{14}{\tilde{u}^3}\sup_{1\le t\le n} E\left|\frac{Y_{t-2}^2\,Y_{t-1}\,\varepsilon_t}{n^{3/2}}\right| \\
&\le \frac{14}{\tilde{u}^3}\sup_{1\le t\le n}\left(E\left|\frac{Y_{t-2}^2\,\varepsilon_t}{n}\right|^2\right)^{1/2}\left(E\left|\frac{Y_{t-1}}{\sqrt{n}}\right|^2\right)^{1/2} \\
&\le \frac{14}{\tilde{u}^3}\sup_{1\le t\le n}(E|\,\varepsilon_t|\,^2)^{1/2}\left\|\frac{Y_{t-2}}{\sqrt{n}}\right\|_4^2\left\|\frac{Y_{t-1}}{\sqrt{n}}\right\|_2 < \infty\,,
\end{aligned}$$

analog lässt sich auch $\sup\limits_{1\le t\le n} E\sup\limits_{\theta\in\tilde{\Theta}}\dfrac{1}{n^{3/2}}\left|\dfrac{1}{\sigma_t^3}\,6\,b_1Y_{t-2}\,Y_{t-1}^2\right| < \infty$ zeigen.

Das gleiche Vorgehen kann auf die übrigen partiellen Ableitungen direkt übertragen werden. Somit ist die Konsistenz für die modifizierte Hesse-Matrix $\widetilde{H}_n(\theta)$ gezeigt.

Die Gleichung (4.8) hat für $\theta = (\Psi, b_0, b_1)'$ und für $\hat{I}_n(\theta) = \widetilde{H}_n(\theta)/n$ die Form

$$-\frac{1}{n}\begin{pmatrix}
\frac{1}{n}\sum\limits_{t=1}^{n}\frac{\partial^2 l_t(\theta^*)}{\partial\Psi^2} & \frac{1}{\sqrt{n}}\sum\limits_{t=1}^{n}\frac{\partial^2 l_t(\theta^*)}{\partial\Psi\partial b_0} & \frac{1}{\sqrt{n}}\sum\limits_{t=1}^{n}\frac{\partial^2 l_t(\theta^*)}{\partial\Psi\partial b_1} \\
\frac{1}{\sqrt{n}}\sum\limits_{t=1}^{n}\frac{\partial^2 l_t(\theta^*)}{\partial b_0\partial\Psi} & \sum\limits_{t=1}^{n}\frac{\partial^2 l_t(\theta^*)}{\partial b_0^2} & \sum\limits_{t=1}^{n}\frac{\partial^2 l_t(\theta^*)}{\partial b_0\partial b_1} \\
\frac{1}{\sqrt{n}}\sum\limits_{t=1}^{n}\frac{\partial^2 l_t(\theta^*)}{\partial b_1\partial\Psi} & \sum\limits_{t=1}^{n}\frac{\partial^2 l_t(\theta^*)}{\partial b_1\partial b_0} & \sum\limits_{t=1}^{n}\frac{\partial^2 l_t(\theta^*)}{\partial b_1^2}
\end{pmatrix}\begin{pmatrix} n & 0 & 0 \\ 0 & \sqrt{n} & 0 \\ 0 & 0 & \sqrt{n}\end{pmatrix}\begin{pmatrix}\hat{\Psi} \\ \hat{b}_0 - b_0^* \\ \hat{b}_1 - b_1^*\end{pmatrix}$$

$$-\frac{1}{\sqrt{n}}\begin{pmatrix}\frac{1}{\sqrt{n}}\sum\limits_{t=1}^{n}\frac{\partial l_t(\theta^*)}{\partial\Psi}\\ \sum\limits_{t=1}^{n}\frac{\partial l_t(\theta^*)}{\partial b_0}\\ \sum\limits_{t=1}^{n}\frac{\partial l_t(\theta^*)}{\partial b_1}\end{pmatrix}=o_p(1) \tag{4.16}$$

Daher lässt sich die Gestalt von $n\hat{\Psi}_n$ aus (4.7) nachweisen, indem man zeigt:

$$\frac{1}{n^{3/2}}\sum_{t=1}^{n}\frac{\partial^2 l_t(\theta^*)}{\partial\Psi\partial b_0}=o_p(1)\ \text{ und }\ \frac{1}{n^{3/2}}\sum_{t=1}^{n}\frac{\partial^2 l_t(\theta^*)}{\partial\Psi\partial b_1}=o_p\,(1). \tag{4.17}$$

Als Hilfsmittel dient dabei das Theorem 2.1 von Hansen (1992). Es besagt, dass für einen beliebigen diskreten Prozess Y_t, ein Martingal M_t mit $E(M_{t+1}-M_t)^2<\infty$, $t=1,\dots,n$ und ihre stetigen Grenzprozesse Y, M auf $[0,1]$: $(Y_{[nr]},M_{[nr]})\xrightarrow{d}(X,M)$, $r=t/n$ gilt:

$$\sum_{i=1}^{[nr]}Y_i(M_{i+1}-M_i)=\int_0^r Y_{[nr]}dM_{[nr]}\xrightarrow{d}\int_0^r XdM=O_p(1). \tag{4.18}$$

Betrachte

$$\frac{\partial^2 l_t(\theta)}{\partial\Psi\partial b_0}=\frac{1}{\sigma_t^2}\left(3b_1Y_{t-2}\,\varepsilon_t^2-b_1Y_{t-2}-2Y_{t-1}\,\varepsilon_t\right),$$

$$\frac{\partial^2 l_t(\theta)}{\partial\Psi\partial b_1}=\frac{1}{\sigma_t^2}\left(3b_1Y_{t-2}\,R_{t-1}\,\varepsilon_t^2-b_1Y_{t-2}\,R_{t-1}-2Y_{t-1}\,R_{t-1}\,\varepsilon_t\right)+\frac{1}{\sigma_t}\left(Y_{t-1}-Y_{t-1}\,\varepsilon_t^2\right)$$

Offensichtlich stellen die Prozesse $\{\varepsilon_t/\sigma_t^2\}$ und $\{R_{t-1}\,\varepsilon_t/\sigma_t^2\}$ direkt die an die Folge $\mathcal{F}_t=\sigma(\varepsilon_t,\varepsilon_{t-1,\dots})$ adaptierten Martingaldifferenzen (kurz: MD) dar. Die restlichen Komponenten in den Ableitungen lassen sich mit Hilfe folgender Martingaldifferenzen darstellen:

- $\partial^2 l_t(\theta)/\partial\Psi\partial b_0:\ \frac{1}{\sigma_t^2}\left(3b_1Y_{t-2}\,\varepsilon_t^2-b_1Y_{t-2}\right)=3b_1Y_{t-2}\underbrace{\frac{1}{\sigma_t^2}\left(\varepsilon_t^2-1\right)}_{MD}+\frac{2b_1Y_{t-2}}{\sigma_t^2}$

- $\partial^2 l_t(\theta)/\partial\Psi\partial b_1:$

$$\frac{1}{\sigma_t}\left(Y_{t-1}-Y_{t-1}\,\varepsilon_t^2\right)=-Y_{t-1}\underbrace{\frac{1}{\sigma_t}\left(\varepsilon_t^2-1\right)}_{MD}$$

$$\frac{1}{\sigma_t^2}\left(3b_1Y_{t-2}\,R_{t-1}\,\varepsilon_t^2-b_1Y_{t-2}\,R_{t-1}\right)=3b_1Y_{t-2}\underbrace{\frac{R_{t-1}}{\sigma_t^2}\left(\varepsilon_t^2-1\right)}_{MD}+\frac{2b_1Y_{t-2}\,R_{t-1}}{\sigma_t^2}.$$

Weiter ist $2b_1Y_{t-2}/\sigma^2 = 2b_1Y_{t-2}\underbrace{\left[1/\sigma^2 - E(1/\sigma_t^2)\right]}_{MD} + 2b_1Y_{t-2}E(1/\sigma_t^2)$ und

$$2b_1Y_{t-2}/\sigma^2 = 2b_1Y_{t-2}\underbrace{\left[R_{t-1}/\sigma^2 - E(R_{t-1}/\sigma_t^2)\right]}_{MD} + 2b_1Y_{t-2}E(R_{t-1}/\sigma_t^2).$$

Da wegen R1 $E|R_t|^4 < \infty$ gilt, ergibt schließlich die Anwendung von (4.18) auf die angeführte Zerlegung, dass $\partial^2 l_t(\theta)/\partial\Psi\partial b_0 = O_p(n)$ und $\partial^2 l_t(\theta)/\partial\Psi\partial b_1 = O_p(n)$. Somit ist (4.17) und das Lemma 2 gezeigt. $\square$

Als Nächstes wird untersucht, wie $n\hat{\Psi}_n$ unter der Nullhypothese verteilt ist. Ausgehend von (4.7) werden die Grenzverteilungen für die Komponenten $n^{-1/2}Y_{[nr]}$, $n^{-1}\sum_{t=1}^n \partial l_t(\theta^*)/\partial\Psi$ und $n^{-2}\sum_{t=1}^n \partial^2 l_t(\theta^*)/\partial\Psi^2$ gesucht. Ähnlich wie die Ergebnisse im Standardfall (vgl. die Herleitung von (2.20) bzw. (2.21)) werden die folgenden Verteilungsaussagen mit Hilfe des FZGWS formuliert. Sein Einsatz stützt sich auf das Lemma 1.

Lemma 3

Seien W_1 und W_2 zwei unabhängige Wiener Prozesse, $\{Y_t\}$ folge (4.4) und $\{R_t\}$ erfülle das Lemma 1. Dann gilt unter H_0 sowie den Bedingungen R1, R2 *für $n \to \infty$:*

$$\frac{Y_{[nr]}}{\sqrt{n}} \xrightarrow{d} \varphi(1)\left(\frac{b_0^2}{1-B}\right)^{1/2} W_1, \tag{4.19}$$

$$\frac{1}{n}\sum_{t=1}^{n}\frac{\partial l_t(\theta^*)}{\partial\Psi} \xrightarrow{d} \varphi(1)\left(\frac{b_0^2E(1/\sigma_t^2)}{1-B}\right)^{1/2}\left[\int_0^1 W_1(r)dW_1(r) - \sqrt{2}b\int_0^1 W_1(r)dW_2(r)\right], \tag{4.20}$$

$$-\frac{1}{n^2}\sum_{t=1}^{n}\frac{\partial^2 l_t(\theta^*)}{\partial\Psi^2} \xrightarrow{d} (\varphi(1))^2\,\frac{b_0^2E(1/\sigma_t^2)}{1-B}\int_0^1 W_1(r)^2dr\,, \tag{4.21}$$

wobei $\varphi(1) = \left(1 - \sum_{i=1}^{k}\psi_i\right)^{-1}$ und $b = \left|\sum_{i=1}^{\infty} b_i\right| < \infty$ nach Voraussetzung.

Beweis:

$Y_{[nr]}/\sqrt{n}$:

Aus der Beveridge-Nelson-Zerlegung (4.14) folgt für ein beliebiges $\epsilon > 0$ und $n \to \infty$:

$$P\left(\sup_{r\in[0,1]}\frac{1}{\sqrt{n}}\left|Y_{[nr]} - \varphi(1)\sum_{i=1}^{[nr]}R_i\right| > \epsilon\right) = P\left(\sup_{r\in[0,1]}\frac{1}{\sqrt{n}}\left|\tilde{\psi}(L)\{R_{[nbr]}\}\right| > \epsilon\right) \xrightarrow{(4.15)} 0,$$

d. h. der Term $\tilde{\psi}(L)\{R_{[nr]}\}$ ist für die Verteilung von $Y_{[nr]}/\sqrt{n}$ asymptotisch vernachlässigbar und daher $Y_{[nr]}/\sqrt{n} \xrightarrow{d} \varphi(1)\sum_{i=1}^{t} R_i$.
Die durch das Lemma 1 nachgewiesene L_2-NED-Eigenschaft von $\{R_t\}$ erlaubt, den FZGWS anzuwenden. Danach gilt $\sum_{i=1}^{[ns]} R_i/\sqrt{n} \xrightarrow{d} \sqrt{b_0/1-B}\, W_1$ und mit $\varphi(1) = \left(1-\sum_{i=1}^{k}\psi_i\right)^{-1}$ folgt die Behauptung (4.19).

$\dfrac{1}{n}\displaystyle\sum_{t=1}^{n}\dfrac{\partial l_t(\theta^*)}{\partial \Psi}$:

Hier erhält man

$$\frac{\partial l_t(\theta^*)}{\partial \Psi} = \frac{Y_{t-1}\,\varepsilon_t}{\sigma_t} - \frac{b(L)\{Y_{t-1}\}(\varepsilon_t^2-1)}{\sigma_t}.$$

Für die NED-Prozesse $\{\varepsilon_t/\sigma_t\}$ und $\{(\varepsilon_t^2-1)/\sigma_t\}$ gilt ebenfalls der FZGWS. Mit $E(\varepsilon_t/\sigma_t)^2 = E(1/\sigma_t^2)$ und $E\left((\varepsilon_t^2-1)/\sigma_t\right)^2 = E[(\varepsilon_t^4 - 2\varepsilon_t^2+1)/\sigma_t^2] \overset{E\varepsilon_t^4=3}{=} 2E(1/\sigma_t^2)$ ergibt sich

$$\frac{1}{\sqrt{n}}\sum_{t=1}^{[ns]}\frac{\varepsilon_t}{\sigma_t} \xrightarrow{d} \sqrt{E(1/\sigma_t^2)}\,W_1 \;\text{ und }\; \frac{1}{\sqrt{n}}\sum_{t=1}^{[ns]}\frac{\varepsilon_t^2-1}{\sigma_t} \xrightarrow{d} \sqrt{2\,E(1/\sigma_t^2)}\,W_2.$$

Macht man nun von den Aussagen (4.18) und (4.19) Gebrauch, so erhält man

$$\sum_{t=1}^{n}\frac{Y_{t-1}}{\sqrt{n}}\,\frac{\varepsilon_t}{\sqrt{n}\sigma_t} \xrightarrow{d} \varphi(1)\left(\frac{b_0^2\,E(1/\sigma_t^2)}{1-B}\right)^{1/2}\int_0^1 W_1(r)\,dW_1(r) \;\text{ und}$$

$$\sum_{t=1}^{n}\frac{Y_{t-i-1}}{\sqrt{n}}\,\frac{\varepsilon_t^2-1}{\sqrt{n}\sigma_t} \xrightarrow{d} \varphi(1)\left(\frac{b_0^2\,2E(1/\sigma_t^2)}{1-B}\right)^{1/2}\int_0^1 W_1(r)\,dW_2(r).$$

Mit dem CMT gilt ferner

$$\begin{aligned}\frac{1}{n}\sum_{t=1}^{n} b(L)\{Y_{t-1}\}\frac{\varepsilon_t^2-1}{\sigma_t} &= \frac{1}{n}\sum_{t=1}^{n}\sum_{i=1}^{\infty} b_i Y_{t-i-1}\frac{\varepsilon_t^2-1}{\sigma_t}\\ &\xrightarrow{d} \varphi(1)\sqrt{2}\,b\left(\frac{b_0^2\,E(1/\sigma_t^2)}{1-B}\right)^{1/2}\int_0^1 W_1(r)dW_2(r)\end{aligned}$$

Daraus erhält man mit den Rechenregeln für stochastische Integrale die Beziehung (4.20).

$-\frac{1}{n^2}\sum_{t=1}^{n}\frac{\partial^2 l_t(\theta^*)}{\partial\Psi^2}$:

Zu untersuchen ist

$$-\frac{\partial^2 l_t(\theta^*)}{\partial\Psi^2} = \frac{Y_{t-1}^2}{\sigma_t^2} - \frac{4b(L)\{Y_{t-1}\}\,Y_{t-1}\varepsilon_t}{\sigma_t^2} + \frac{[b(L)\{Y_{t-1}\}]^2(\varepsilon_t^2-1)}{\sigma_t^2}$$

komponentenweise. Mit (4.19) und CMT folgt

$$\frac{1}{n^2}\sum_{t=1}^{n} Y_{t-1}^2 \xrightarrow{d} (\varphi(1))^2\,\frac{b_0^2}{1-B}\int_0^1 W_1(r)^2 dr. \tag{4.22}$$

Nun wird Y_{t-1}^2/σ_t^2 zentriert, so dass man Martingaldifferenzen

$$Y_{t-1}^2/\sigma_t^2 - E\left(Y_{t-1}^2/\sigma_t^2\middle|\,\mathcal{F}_{t-1}\right) = Y_{t-1}^2/\sigma_t^2 - Y_{t-1}^2\,E\big(1/\sigma_t^2\big)$$

bekommt. Durch die Anwendung von (4.18) erhält man ferner

$$1/n^2\sum_{t=1}^{n} Y_{t-1}^2/\sigma_t^2 = 1/n^2\sum_{t=1}^{n} Y_{t-1}^2 E(1/\sigma_t^2) + o_p(1)$$

woraus mit (4.22) unmittelbar folgt:

$$\frac{1}{n^2}\sum_{t=1}^{n}\frac{Y_{t-1}^2}{\sigma_t^2} \xrightarrow{d} (\varphi(1))^2\,\frac{b_0^2 E(1/\sigma_t^2)}{1-B}\int_0^1 W_1(r)^2 dr\,.$$

Die weiteren Komponenten beeinflussen die Grenzverteilung von $1/n^2\sum_{t=1}^{n}\partial^2 l_t(\theta^*)/\partial\Psi^2$ nicht, da sie in Wahrscheinlichkeit gegen Null streben. Genauer gesagt, besitzen sie die Ordnung $O_p(n^{-3/2})$. Diese Aussage resultiert unter Einsatz von CMT direkt aus (4.18) mit Martingaldifferenzen ε_t/σ_t^2 bzw. $(\varepsilon_t^2-1)/\sigma_t^2$. Somit ist (4.21) gezeigt. □

Mit den obigen Ergebnissen können die Grenzverteilungen für $n\hat{\Psi}_n$ und die t-Statistik $\tau_n = n\hat{\Psi}_n/\widehat{s.e.}(n\hat{\Psi}_n)$ angegeben werden. Schätzt man die Kovarianzmatrix $Cov(\hat{\theta})$ durch die modifizierte Hesse-Matrix $\widetilde{H}(\theta^*)$, so erhält man

$$\widehat{s.e.}(n\hat{\Psi}_n) = \left(1/n^2\sum_{t=1}^{n} -\frac{\partial^2 l_t(\theta^*)}{\partial\Psi^2}\right)^{-1/2}.$$

Die Lemmas 2 und 3 führen zur folgenden Endaussage über die asymptotische Verteilung der interessierenden Teststatistiken im Modell (4.4):

Satz **2**

Unter den Bedingungen und mit den Bezeichnungen des Lemmas 3 gilt unter H_0 für $n \to \infty$:

$$n\hat{\Psi}_n \xrightarrow{d} \frac{\int_0^1 W_1(r)dW_1(r) - \sqrt{2}b\int_0^1 W_1(r)dW_2(r)}{\varphi(1)\left(\frac{b_0^2 E(1/\sigma_t^2)}{1-B}\right)^{1/2} \int_0^1 W_1(r)^2 dr} \tag{4.23}$$

$$\tau_n \xrightarrow{d} \frac{\int_0^1 W_1(r)\,dW_1(r) - \sqrt{2}\,b\int_0^1 W_1(r)\,dW_2(r)}{\sqrt{\int_0^1 W_1(r)^2 dr}}. \tag{4.24}$$

Der Beweis dieses Satzes folgt direkt aus den Lemmas 2 und 3.

Erwartungsgemäß führt das Fehlen heteroskedastischer Strukturen ($b_i = 0$ für $i \geq 1$ und $\sigma = \sigma_t$ für alle t) zu Standardasymptotik der DF-Tests. Bei Heteroskedastie sind die Grenzverteilungen in (4.23) und (4.24) eine Kombination aus der DF- und der Normalverteilung. Im Gegensatz zu GARCH verwandelt sich die Grenzverteilung der t-Statistik τ unter GRS-ARCH-Innovationen mit absolut wachsenden Parametern $|b_i|$ allerdings nicht in die Standardnormalverteilung. Stärkere heteroskedastische Effekte bewirken stattdessen eine höhere Varianz und somit kleinere kritische Werte, so dass es seltener zur Ablehnung der I(1)-Hypothese kommt.

Lässt man für Y_t einen von Null abweichenden Erwartungswert wie in (2.24) zu, so basieren die DF-Tests auf den zentrierten Variablen $Y_t^* = Y_t - 1/n\sum_{t=1}^n Y_t$, vgl. auch die Ausführungen im Abschnitt 2.3.1. In einem autoregressiven Prozess mit GRS-ARCH-Innovationen wie (4.4) gilt dann unter der I(1)-Hypothese

$$\frac{Y_{[nr]}^*}{\sqrt{n}} \xrightarrow{d} \varphi(1)\left(\frac{b_0^2}{1-B}\right)^{1/2}\left(W_1(r) - \int_0^1 W_1(r)dr\right). \tag{4.25}$$

Um dies zu zeigen reicht es, den Beweis der Behauptung (4.19) aus dem Lemma 3 um die Verteilungsaussage (2.26) zu erweitern. Manchmal wird der Ausdruck

$$W(r) - \int_0^1 W(r)dr =: W^*(r)$$

in der Literatur als *zentrierte Brownsche Bewegung* bezeichnet.

Baut man den ML-Schätzvektor θ_n^* auf zentrierten Variablen Y_t^* auf, so sind die Grenzverteilungen aus dem Satz 2 folgendermaßen zu korrigieren:

$$n\hat{\Psi}_n \xrightarrow{d} \frac{\int_0^1 W_1^*(r)dW_1(r) - \sqrt{2}b\int_0^1 W_1^*(r)dW_2(r)}{\varphi(1)\left(\frac{b_0^2 E(1/\sigma_t^2)}{1-B}\right)^{1/2}\int_0^1 W_1^*(r)^2 dr} \tag{4.26}$$

$$\tau_n \xrightarrow{d} \frac{\int_0^1 W_1^*(r)\,dW_1(r) - \sqrt{2}\,b\int_0^1 W_1^*(r)\,dW_2(r)}{\sqrt{\int_0^1 W_1^*(r)^2 dr}}. \tag{4.27}$$

Somit wird das asymptotische Verhalten der Teststatistiken durch die entsprechend zentrierte Brownsche Bewegung wiedergegeben.

Die durch Simulationen ermittelten Quantile der Verteilungen von τ_n für zentrierte und nichtzentrierte I(1)-Prozesse mit GRS-ARCH-Innovationen sind in der Tabelle 2 enthalten. Grundlage für die Simulation bildet die von Hatanaka (1996) im Kapitel 7.3 angeführte Darstellung für die Teststatistik τ aus (2.21) im gewöhnlichen DF-Test unter der Random Walk-Nullhypothese. Danach erzeugt man in jedem Simulationslauf unabhängige standardnormalverteilte Zufallsvariablen ε_t, $t = 1, \ldots, n$ und bildet die Teststatistik

$$\tau = \frac{(n-1)^{-1}\sum_{t=2}^{n}\left(\sum_{s=1}^{t-1}\varepsilon_s\right)\varepsilon_t}{(n-1)^{-2}\sum_{t=2}^{n}\left(\sum_{s=1}^{t-1}\varepsilon_s\right)^2}.$$

Durch k-maliges Wiederholen dieser Prozedur erhält man eine empirische Verteilung für τ, deren Quantile als kritische Werte beim Testen eingesetzt werden.

Basierend darauf wird nun in jedem Simulationslauf aus zwei Reihen unabhängiger standardnormalverteilter Zufallsvariablen ε_t und ϵ_t der Länge n der modifizierte Ausdruck

$$\frac{(n-1)^{-1}\sum_{t=2}^{n}\left(\sum_{s=1}^{t-1}\varepsilon_s\right)\varepsilon_t - \sqrt{2}b(n-1)^{-1}\sum_{t=2}^{n}\left(\sum_{s=1}^{t-1}\varepsilon_s\right)\epsilon_t}{(n-1)^{-2}\sum_{t=2}^{n}\left(\sum_{s=1}^{t-1}\varepsilon_s\right)^2}$$

berechnet, um die Grenzverteilung nach (4.24) zu simulieren. Analog dazu wird asymptotische Verteilung für Random Walk mit Drift simuliert.

Im Folgenden werden nur GRS-ARCH-Parametrisierungen berücksichtigt, die unter der I(0)-Alternative für die schwache Stationarität hinreichend sind, vgl. die Bedingung (3.27). Die Länge simulierter Reihen n beträgt 1000. Zur Quantilsbestimmung werden pro Parameterkonstellation und Modellannahme (mit und ohne Drift) jeweils 5000 Simulationsläufe durchgeführt.

$\left\lvert\sum_{i=1}^{\infty} b_i\right\rvert$	0.1-Quantil		0.05-Quantil		0.01-Quantil	
	mit Drift	ohne Drift	mit Drift	ohne Drift	mit Drift	ohne Drift
0.05	-2.589	-1.618	-2.886	-1.939	-3.419	-2.559
0.10	-2.600	-1.624	-2.907	-1.951	-3.488	-2.569
0.15	-2.632	-1.650	-2.931	-1.979	-3.471	-2.634
0.20	-2.667	-1.672	-2.984	-2.006	-3.606	-2.650
0.25	-2.697	-1.704	-3.026	-2.039	-3.638	-2.717
0.30	-2.689	-1.743	-3.009	-2.092	-3.670	-2.762
0.35	-2.746	-1.791	-3.110	-2.161	-3.815	-2.857
0.40	-2.815	-1.833	-3.181	-2.210	-3.885	-2.934
0.45	-2.888	-1.893	-3.248	-2.290	-3.910	-3.008
0.50	-2.917	-1.937	-3.304	-2.346	-4.071	-3.130
0.55	-2.972	-1.998	-3.385	-2.412	-4.148	-3.218
0.60	-3.024	-2.053	-3.441	-2.508	-4.196	-3.319
0.65	-3.097	-2.132	-3.540	-2.590	-4.365	-3.438
0.70	-3.163	-2.184	-3.627	-2.652	-4.562	-3.563
0.75	-3.243	-2.261	-3.763	-2.766	-4.697	-3.688
0.80	-3.307	-2.325	-3.825	-2.856	-4.834	-3.822
0.85	-3.371	-2.379	-3.899	-2.925	-5.001	-3.939
0.90	-3.486	-2.469	-3.994	-3.051	-5.120	-4.110
0.95	-3.577	-2.536	-4.177	-3.149	-5.337	-4.265

Tabelle 2: Quantile der ML-basierten Einheitswurzelteststatistik τ_n in einem autoregressiven Prozess mit GRS-ARCH-Innovationen unter der I(1)-Hypothese

Es wird nun geprüft, ob für die angeführten kritischen Werte das nominale Testniveau in endlichen Stichproben tatsächlich eingehalten wird. Dazu werden wieder 5000 Random Walk-Prozesse der Länge 500, deren Innovationen einem GRS-ARCH-Prozess folgen, simuliert und auf I(1)-Hypothese getestet. Um die Ergebnisse mit denen des DF-Tests aus der Tabelle 1 zu vergleichen, wird die gleiche Parametrisierung gewählt. Die

Ablehnhäufigkeiten fasst die Tabelle 3 zusammen.

Parametervektor (b_0, b_1, b_2, b_3)	$\left\lvert \sum_{i=1}^{3} b_i \right\rvert$	relative Ablehnhäufigkeit (t-Statistik< -1.94)
(1 0.05 0.03 0.02)	0.1	0.052
(1 0.25 0.15 0.05)	0.5	0.044
(1 0.50 0.30 0.10)	0.9	0.065

Tabelle 3: empirisches Testniveau des ML-basierten Einheitswurzeltests unter GRS-ARCH-Innovationen

Auch wenn das nominale Testniveau von 5% nicht exakt eingehalten wird, so ist bei den empirischen Ablehnhäufigkeiten eine leichte Annäherung an den theoretischen Wert zu beobachten. Insbesondere ist festzustellen, dass das empirische Testniveau nun nicht mehr, wie bei gewöhnlichen DF-Tests, durchgehend über 5% liegt, sondern eher um diesen Wert schwankt. Somit wird bei vorliegenden GRS-ARCH-Innovationen durch die Anwendung angeführter kritischer Werte im Zusammenhang mit ML-basierten Einheitswurzeltests das Testniveau besser eingehalten und nicht mehr systematisch verzerrt.

Kapitel 5

Strompreisprozesse und -volatilitäten

Die in obigen Kapiteln dargestellten statistischen Methoden und Zeitreihenmodelle werden nun zur Beschreibung der Preisentwicklungen am deutschen Strommarkt herangezogen. Im Vordergrund stehen dabei die in Folge der veränderten Rahmenbedingungen stark intensivierten Aktivitäten auf der Großhandelsebene.

Eingangs wird die allgemeine Struktur des Strommarktes erläutert, der sich in folgende große Teilbereiche gliedert:

- in Bezug auf Geschäftsabwicklung - bilateraler Handel und Börsenhandel
- in Bezug auf die Art der gehandelten Produkte - Spotmarkt und Terminmarkt.

Anschließend werden aus dem Bereich „Börsenhandel - Terminmarkt" die Tagesschlussnotierungen für die liquidesten Terminprodukte an der Leipziger Strombörse EEX als Indikator für die langfristige Preisdynamik des Gesamtmarktes ausgewählt und einer ausführlichen statistischen Analyse unterzogen. Am Ende der Analyse wird unter Berücksichtigung bedingt heteroskedastischer Strukturen die Frage beantwortet, ob Strompreise, ähnlich wie Aktienkurse, eine Brownsche Bewegung darstellen, d. h. I(1)-Charakter aufweisen, ob sie möglicherweise einem deterministischen Trend folgen, oder ob sie als stationäre Reihen anzusehen sind.
Die rechentechnische Umsetzung sämtlicher Schätz- und Testverfahren erfolgt mit der *SAS*®-Software, Version 8.02.

5.1 Organisation und Struktur des Stromhandels

Die fortschreitende Deregulierung europäischer Energiemärkte, die in Deutschland 1998 in die Wege geleitet wurde, und die damit verbundenen neu geschaffenen Zugangsmöglichkeiten zu Stromleitungsnetzen für alle Marktteilnehmer führte zu grundlegenden Veränderungen im Handel mit Strom. Zwar wurde Strom auch davor gehandelt, dies geschah jedoch in einem äußert beschränkten Maß. Als Handelspartner traten dabei Stromproduzenten auf, die in benachbarten Versorgungsgebieten über eigene Stromnetze verfügten und in einem vereinbarten Rahmen wechselseitig Strom transferierten, um ihre Kraftwerkeinsätze zu optimieren.
Die radikal veränderten Rahmenbedingungen leiteten eine neue Qualität des Stromhandels ein, der neue wesentliche Funktionen für die Energiewirtschaft übernommen hat. Integrierten Unternehmen, die sowohl über Erzeugungs- als auch über Vertriebskapazitäten verfügen, ist es möglich geworden, über die Martktaktivitäten ihren Ertrag und Risiko zu optimieren. Daneben verfügen reine Erzeuger mit dem offenen Markt über einen Absatzkanal und Vertriebsgesellschaften über eine Beschaffungsquelle, was ihre operative Flexibilität steigert. Das angesprochene Wegfallen genereller Einschränkungen bei der Netznutzung hat schließlich zur Folge, dass es auch für Unternehmen ohne eigene Stromerzeugungs- bzw. Übertragungseinrichtungen möglich ist, am Stromhandel teilzunehmen. So sind gegenwärtig zahlreiche reine Handelsunternehmen am deutschen Strommarkt aktiv, was zur Liquidität des Marktes beiträgt. Im gesamtwirtschaflichen Kontext macht sich Stromhandel durch einen zusätzlichen preisbildenden Aspekt bemerkbar. Er besteht in einer generellen Entwicklung weg vom kostendeckungs- und hin zu marktpreisorientierter Preiskalkulation für Stromprodukte.

5.1.1 Marktsegmente

Bei der Organisation der Geschäftsabwicklung wird zwischen Handelsaktivitäten auf bilateraler Ebene, häufig als OTC-Markt (engl: *over the counter*) bezeichnet, und Stromhandel über die Börse unterschieden. Die wichtigsten Merkmale des Börsen- und des OTC-Handels fasst die Tabelle 4 zusammen. Zwar entfiel 2002 in Deutschland nur ein Anteil von ca. 10% des Gesamtverbrauchs (vgl. Curtius (2003)) auf die Börsengeschäfte, was im Vergleich z. B. zu Skandinavien als gering erscheint. Ähnlich wie in weiten Teilen Europas bestimmen jedoch hier zu Lande die im Unterschied zu OTC- öffentlich zugänglichen Börsennotierungen den jeweiligen Marktpreis, an den sich u. a. vertragliche Preisausgestaltung bei direkten Strombezugsverträgen mit Industrie- und Gewerbekunden anlehnt. Gerade für die Bepreisung der Stromlieferungen in mittel- bis langfristigen Direktverträgen mit großen Endabnehmern, die nach wie vor den Großteil des physischen Bezugs ausmachen, dienen die Börsenpreise somit als Messlatte.

Merkmal	Börse	OTC
Lieferprodukte	nur standardisiert	in hohem Maß standardisiert
physische Lieferungen	ja	ja
finanzieller Ausgleich	ja	nein
Preisfeststellung	im Auktionsverfahren	nach direkter Vereinbarung
Preisveröffentlichung	ja	nein
Clearing-Funktion	ja	nein

Tabelle 4: bilateraler und Börsenhandel: Merkmalsvergleich

Ein im Vergleich zum Gesamtverbrauch relativ geringer Börsenumsatz kann dadurch erklärt werden, dass in Deutschland die meisten Erzeuger auch über Vertriebskapazitäten zu Endverbrauchern verfügen, so dass die entsprechenden Transfers nicht über den Handelsmarkt, sondern intern abgewickelt werden. In dieser Situation werden die Börsennotierungen für eine adäquate Bepreisung nach außen umso wichtiger.

5.1.2 Börsenhandel

Wie die Erfahrung zeigt, führt der Weg der Marktliberalisierung von einer energiewirtschaftlichen Monopolstruktur hin zum freien Wettbewerb unter den Stromanbietern über drei Meilensteine, die im Weiteren kurz erläutert werden.

Netzzugang

Zunächst, wie bei jedem fairen Wettbewerb, müssen gleiche Rahmenbedingungen für alle Teilnehmer geschaffen werden. Für den Stromhandel erweist sich in dieser Beziehung insbesondere der Zugang zu Übertragungsnetzen als kritischer Punkt. Dies resultiert aus zwei Besonderheiten der Ware Strom, die den Handel entscheidend prägen.

Zum einen ist Strom nicht speicherbar, d. h. seine Produktion und Verbrauch müssen

nahezu zeitgleich erfolgen. Zum anderen setzen die existierenden Durchleitungskapazitäten im Übertragungsnetz feste Grenzen für die physisch lieferbaren Strommengen und somit in einem hohen Maß für die gehandelten Volumina. Da eine Lieferung nur erfolgen kann, wenn die Netzkapazität dies zulässt, und eine Kapazitätserweiterung hohe Investitionen verlangt, muss die Netznutzung klar geregelt werden.
Unter dem massiven politischen Druck gelang es der Stromindustrie in Verbändevereinbarungen I und II Ende der 90-er Jahre, eine solche Regelung zu erzielen. Für jede Liefervereinbarung wird (außer der Menge und dem Preis) der Zeitraum, der Übergabenetzknoten und die Stromqualität (firm, nonfirm) spezifiziert. Danach deklarieren alle Marktteilnehmer in einem klar definierten Verfahren die geplante Einspeisung und Entnahme von Energie unter Berücksichtigung relevanter Netzrestriktionen in sogenannten Fahrplänen. Der Netzbetreiber führt die Fahrpläne zusammen und beaufsichtigt ihre Einhaltung. Die Abweichungen zwischen der deklarierten und der tatsächlich durchgeleiteten Energie gleicht der Netzbetreiber aus. Den finanziellen Aufwand dafür lässt er sich durch eine Umlage von Stromendverbrauchern erstatten.

Börsenhandel mit kurzfristigen Lieferungen

Ein klares Regelwerk, das den physischen Stromhandel von der technischen Seite her ermöglicht, führt typischerweise zur starken Intensivierung im bilateralen Stromaustausch zwischen großen Energieproduzenten. Neue Durchleitungsregeln erlauben es, Strom nicht nur mit benachbarten Versorgern, sondern auch überregional und international zu handeln. Dies wird zunächst wahrgenommen, um zusätzliche Spielräume für eine effektivere Nutzung vorhandener Kapazitäten im Kurzfristbereich zu schaffen. Besonders in der kurzfristigen Kraftwerkeinsatzplanung sind die Vorteile des erweiterten Handels offensichtlich, denn hier sind sonstige Möglichkeiten, auf unerwartete Ereignisse wie etwa Kraftwerksausfälle, plötzliche Nachfragespitzen u. ä. flexibel zu reagieren, deutlich begrenzt. Darüber hinaus können durch die Wahl, zur Bedienung bestehender Verträge Energie aus eigener Erzeugung zu nutzen oder am Markt zu kaufen, die Produktionsgrenzkosten gesenkt werden.
Eine wachsende Anzahl neuer Handelsteilnehmer führt zur Belebung des OTC-Markts. Doch gerade im Handel mit Lieferungen für den Folgetag (s. g. *Spothandel*), wo Entscheidungen einerseits besonders schnell getroffen werden müssen, und andererseits immer mehr potenzielle Handelspartner zu kontaktieren und die ensprechenden Handlungsalternativen zu bewerten sind, stösst der bilaterale Handelsansatz bald an seine

Grenzen. Daraus resultiert die Notwendigkeit, einen alternativen Mechanismus für eine möglichst unkomplizierte, transparente und verlässliche Abwicklung der Handelsvorgänge für Spotlieferungen zu schaffen. Diesen Anforderungen kann durch spezielle Handelsplätze am besten entsprochen werden, an denen komplett standardisierte physische Energielieferungen (*Produkte*, gemessen in Megawattstunden, MWh) für den Folgetag laufend gekauft und verkauft werden. Die Standardisierung bezieht sich auf die Zeiträume (z. B. 0.00 Uhr bis 1.00 Uhr), in denen konstante elektrische Leistung (gemessen in Megawatt, MW) bereitgestellt wird.

In Analogie zu etablierten Waren- bzw. Finanzmärkten werden solche Einrichtungen als *Strombörsen* bezeichnet. Die Abwicklung des Handels funktioniert nach börsenüblichen Regeln. So erfogt die Preisbildung nach dem Meistausführungsprinzip, bei dem das größte mögliche Handelsvolumen erzielt wird. Vereinfachend dargestellt, geben die Teilnehmer zu einem bestimmten Zeitpunkt ihre mengen- und preisfixierten Kauf- bzw. Verkaufsgebote für fest definierte Produkte ab. Aus allen eingegangenen Aufträgen ermittelt man die Angebots- und die Nachfragekurven, deren Schnittpunkt den Gleichgewichtspreis für den spezifizierten Produkt angibt. Alle ausführungsfähige Aufträge werden zu diesem Referenzwert abgerechnet.

Börsenhandel mit Terminkontrakten

Aufbauend auf einem funktionierenden Handelsmechanismus im Kurzfristbereich können die Vorteile des liberalisierten Marktes erst durch eine Erweiterung des Handels auf Termingeschäfte voll zum Tragen kommen. Im Unterschied zu Spotgeschäften zählen zu Terminkontrakten Vereinbarungen über Monats-, Quartals- und Jahresstromlieferungen. Zwischen der Vereinbarung und der Aufnahme der Lieferung liegen je nach Produkt wenige Tage bis mehrere Jahre. Üblicherweise werden solche frei handelbare Kontrakte über langfristige Liefer- bzw. Abnahmeverpflichtungen als *Forwards* bezeichnet. Sie ermöglichen den Marktteilnehmern kontinuierliches Portfoliomanagement aktiv zu betreiben, offene Positionen abzusichern und somit das Verhältnis zwischen Ertrag und Risiko zu optimieren. Wegen der Nichtspeicherbarkeit von Strom existieren aber keine unmittelbaren Arbitragemöglichkeiten zwischen Spot- und Forwardgeschäften.

Auch in diesem Marktsegment bieten Strombörsen durch ihre standardisierten Produkte und einheitlichen Handelsbedingungen trotz bestimmter börsenspezifischer Gebühren wesentliche Vorteile gegenüber den OTC-Geschäften. War in der Anfangsphase die Anonymität des Börsenhandels für viele Akteure wichtig, die ihre Lieferverträge ohne Rücksichtnahme auf etablierte Vertragsbeziehungen abschließen konnten, rücken mit fortschreitender Marktentwicklung andere Aspekte in den Vordergrund. So wird der finanzielle Nachteil bei einem eventuellen bonitätsbedingten Ausfall des Handelspartners durch die Clearing-Funktion der Börse nahezu ausgeschlossen. Der Stellenwert dieses kreditrisikominimierenden Merkmals wird durch eine sinkende Bonität bis hin zur Insolvenz einiger Marktteilnehmer in der jünster Vergangenheit unterstrichen. Zudem entstehen durch die Abwicklung über die Börse für Unternehmen, die neu in den Markt eintreten, in der Regel niedrigere Informationskosten als bei OTC-Geschäften.

Neben den einfachen Terminlieferungen zu fest vereinbarten Preisen gewinnt der Handel mit komplexeren Stromprodukten - Derivaten an Bedeutung. Einen Überblick über die unterschiedlichen Formen von Derivaten, die in erster Linie zur Absicherung offener Liefer- oder Bezugsverpflichtungen gegen das Preisrisiko dienen, geben Clewlow und Strickland (2000) im Kap. 5. Zu den gängigsten von ihnen gehören zunächst übliche Kauf- und Verkaufsoptionen auf Standardprodukte, aber auch so genannte Optionsbänder (engl.: option strips), die Optionen auf mehrere aufeinanderfolgende Lieferperioden (z. B. Tage oder Monate eines Jahres) mit jeweiligen mehrfachen Ausübungszeitpunkten beinhalten. Für die Entwicklung des Derivatenhandels ist eine ausreichende Börsenliquidität besonders wichtig, denn nur sie kann eine transparente Bepreisung von Derivaten gewährleisten.

Seit 1993 mit dem norwegischen Nord Pool die erste Strombörse in Europa, an der sich inzwischen auch dänische, schwedische und finnische Konzerne beteiligen, ihren Betrieb aufnahm, hat die Bedeutung des Börsenhandels stark zugenommen. Gegründet, um die durch starke Dominanz ehemaliger Monopolisten geprägten Märkte für Wettbewerber zu öffnen, entwickelt sich der Börsenhandel mit zunehmender Liquidität zum alternativen Vermarktungs- bzw. Beschaffungskanal mit Preisindikatorfunktion für den gesamten Markt.
Momentan wird neben Nord Pool an vier weiteren kontinentaleuropäischen Börsen Strom gehandelt. Im Mai 1999 ging die APX in Amsterdam (*Amsterdam Power Exchange*, z. Z. 39 Teilnehmer) an den Start. In Deutschland entstanden kurz darauf mit der Frankfurter EEX und der Leipziger Power Exchange gleich zwei Börsenplätze. Durch ihre Fusion im Juli 2002 zu EEX (*European Energy Exchange*, z. Z. 111 Teilnehmer) in Leipzig wurde die gesplitterte Liquidität und unterschiedliche Produktpaletten zusammengeführt. Ende 2001 wurde die Powernext in Paris gegründet, EXAA in Graz (*Energy Exchange Austria*, z. Z. 16 Teilnehmer) folgte im März 2002. Im Folgenden

werden die historischen Preise für Forwards mit der höchsten Liquidität an der Leipziger EEX stellvertretend für die Preisentwicklung am deutschen Strommarkt analysiert.

5.1.3 Gehandelte Produkte und Marktliquidität

Bei den an der EEX gehandelten Forwardprodukten wird eine Unterscheidung in Grundlastprodukte (engl.: baseload), die tägliche Lieferung einer konstanten elektrischen Leistung von 0 bis 24 Uhr im jeweiligen Zeitraum darstellen, und Spitzenlastprodukte (engl.: peakload), bei denen eine konstante elektrische Leistung nur werktags zwischen 8 und 20 Uhr zur Verfügung gestellt wird. Gegenwärtig sind an der EEX monats-, quartals- und jahresweise Grund- und Spitzenlastlieferungen für die nächsten 6 Monate, 8 Quartale und 6 Jahre zum Handel zugelassen. Auch wenn die Marktliquidität relativ starken Schwankungen unterliegt, kann man davon ausgehen, dass die umgesetzten Volumina für das jeweilige Produkt mit der zeitlichen Annäherung an den physischen Lieferanfang steigen. Die Energieanteile von Grundlast- und Spitzenlastprodukten am Gesamtumsatz stehen annähernd im Verhältnis von 10 zu 1. Grundsätzlich können Geschäfte für die ersten 3-4 Folgemonate, 3-4 Folgequartale und 3 Folgejahre als liquide betrachtet werden, da die entsprechenden Preisfeststellungen zumindest alle 2-3 Tage erfolgen, während Kontrakte für die weiter in der Zukunft liegenden Zeiträume nur sporadisch gehandelt werden. Dabei werden die Lieferungen für das Folgejahr (Jahr 1) mit Abstand am häufigsten gehandelt. Aufgrund der im Vergleich zu Monats- und Quartalsprodukten langen Lieferperiode wird mit diesen Kontrakten auch ein Großteil der gehandelten Energie umgesetzt.

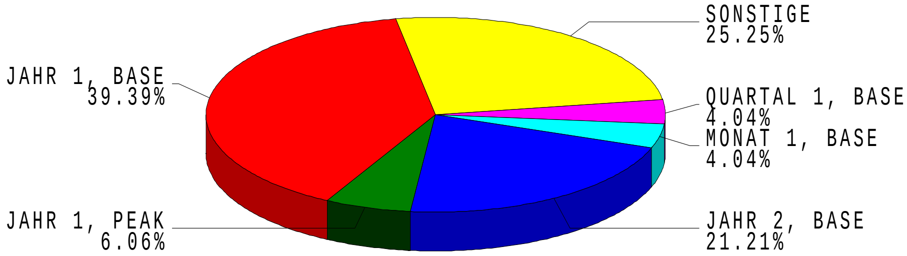

Abbildung 4: Umsatzanteile nach Kontraktarten in %

Eine Auskunft über den Börsenumsatz (inklusive Clearing) verschiedener Produkte gibt die Abbildung 4. Sie enthält Anteile an Energie, die in den ersten 10 Monaten 2003 in verschiedenen Kontrakten gehandelt wurden. Als Basis für die Berechnung dienten die von der EEX veröffenlichten Handelsvolumina in [MW]. Basierend auf dieser Information werden im folgenden Abschnitt die börslichen Terminpreise für Grundlastlieferungen elektrischer Energie im Folgejahr (Jahr 1 Base) als stellvertretend für die Srompreise am deutschen Markt untersucht.

5.2 Modellierung des Preisprozesses

Der Untersuchungszeitraum erstreckt sich über 34 Monate vom 2. Januar 2001 bis 31. Oktober 2003. Bei täglicher Beobachtungsfrequenz verfügt man mit mehr als 700 Beobachtungen über eine ausreichend lange Reihe, so dass der Einsatz asymptotischer Aussagen angemessen erscheint.

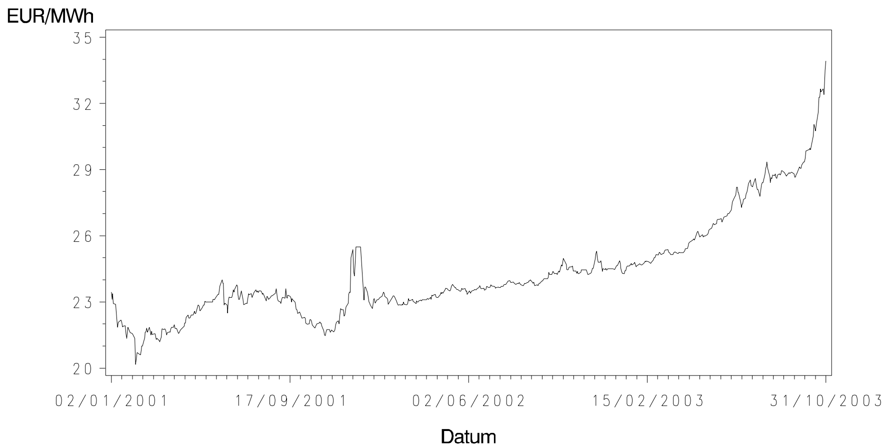

Abbildung 5: Strompreise: Börsentagesschlusskurse für Baseload-Lieferungen Folgejahr

Den Reihenverlauf der Rohpreise (in EUR/MWh) gibt die Abbildung 5 wider. Daran ist zu erkennen, dass eine Phase relativ stark schwankender Preisbewegungen im Bereich zwischen 20 und 24 im Jahre 2001 mit einem steilen Preisanstieg im Dezember endet. Innerhalb weniger Tage klettern die Preise im Dezember 2001 auf ein Zwischenhoch von 25,5 und fallen anschließend ebenso rapide zurück auf 23. Danach setzt ein kontinuierlicher Preisanstieg, der bis Mitte 2003 relativ gleichmäßig verläuft und im zweiten Halbjahr 2003 sich deutlich verstärkt. Am Ende der Beobachtungsperiode im Oktober 2003 brechen die Preise über die Marke von 30 aus und erreichen letzlich 34.

Die zugehörigen deskriptiven Statistiken für den gesamten Zeitraum sowie getrennt nach Kalenderquartalen sind in der Tabelle 5 zusammengefasst.

Periode	arith. Mittel	Standardabw.	Minimum	Maximum
Q1 2001	21.67	0.62	20.18	23.45
Q2 2001	22.75	0.62	21.58	24.00
Q3 2001	22.22	0.24	22.48	23.78
Q4 2001	22.53	1.09	21.48	25.50
Q1 2002	23.08	0.19	22.70	23.68
Q2 2002	23.47	0.17	23.10	23.80
Q3 2002	23.85	0.15	23.63	24.35
Q4 2002	24.52	0.23	24.25	25.30
Q1 2003	24.89	0.29	24.28	25.38
Q2 2003	26.38	0.81	25.20	28.20
Q3 2003	28.58	0.47	27.28	29.35
Okt. 2003	31.29	1.35	29.50	33.93

Tabelle 5: Strompreise: Lage- und Streuungsmaße

Die folgende Analyse bezieht sich auf die natürlichen Logarithmen der Preise. Der in Frage kommende Einsatz des Differenzenfilters Δ auf die Logarithmen ist gut interpretierbar, denn die Differezen der Logarithmen können annäherend als Renditen $R_t = (Y_t/Y_{t-1} - 1)$ angesehen werden:

$$\log Y_t - \log Y_{t-1} = \log(Y_t/Y_{t-1}) = \log(1 + R_t) \approx R_t \quad \text{für relativ kleine } R_t.$$

Überträgt man die Definition 5 auf die Renditen, so spricht man von der *geometrischen* Brownschen Bewegung für den zugrunde liegenden Preisprozess. Es ist zu beachten, dass die im Weiteren gemachten Feststellungen in Bezug auf die Stationarität untersuchter Reihen nicht direkt auf die Preise übertragen werden können, vgl. die Ausführung im Abschnitt 2.3.2. Den zeitlichen Verlauf logarithmierter Preise zeigt die Abbildung 6.

Es werden nun zwei alternative Vorgehensweisen bei der Modellierung dieser Reihe dargestellt. Während im ersten Fall versucht wird, durch klassische Zerlegung vorliegender Reihe in eine deterministische und eine stochastische Komponente die Preise zu beschreiben, verfolgt man im zweiten Fall den Ansatz eines rein stochastischen Preisprozesses.

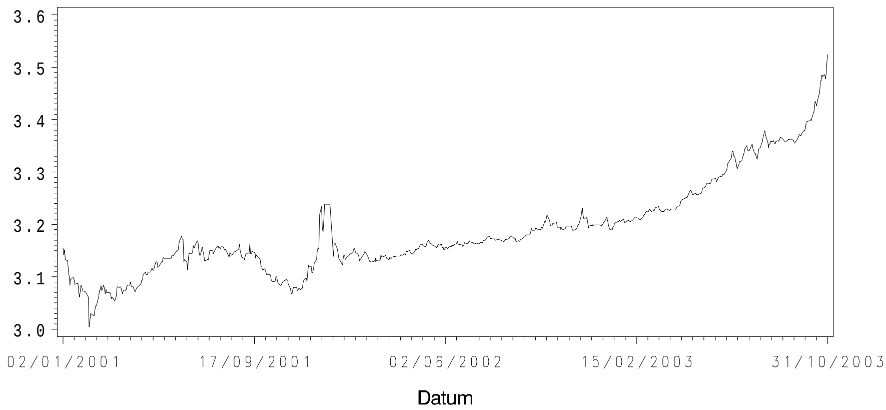

Abbildung 6: Strompreise: natürliche Logarithmen

5.2.1 Modell mit deterministischem Trend

Der Verlauf der Preislogarithmen, wie auch schon der Preispfad in der Abbildung 5 und die Quartalsmittelpreise in der Tabelle 5, suggeriert einen zeitlichen Aufwärtstrend in der gesamten Beobachtungsperiode. Es liegt daher nahe, bei der Modellierung im ersten Schritt eine globale polynomiale Trendkomponente für den logarithmierten Preisprozess zu unterstellen. Bei der folgenden Modellselektion mit Hilfe von Standardkennzahlen (adjustiertes Bestimmtheitsmaß, p-Werte der t-Statistiken) wird durch ein Polynom dritten Grades die beste Anpassung erreicht. Weder ein linearer noch ein quadratischer Ansatz für den Zeittrend können die vorliegenden Daten vergleichbar gut erklären. Polynome höheren Grades verbessern trotz zum Teil betragsmäßig hoher t-Statistiken für Regressionsparameter in höheren Potenzen die durch Informationskriterien gemessene Anpassungsgüte nicht. Für die ausgewählte Regressionsgleichung werden folgende Parameter mit der KQ-Methode geschätzt:

$$\log P_t = \underset{(850.3)}{3.07} + \underset{(15.9)}{7.00 \cdot 10^{-4}}\, t - \underset{(16.0)}{2.33 \cdot 10^{-6}}\, t^2 + \underset{(22.0)}{2.97 \cdot 10^{-9}}\, t^3 + r_t, \quad t = 1, \ldots, 704. \tag{5.1}$$

Unter jedem geschätzten Koeffizienten ist der Betrag der jeweiligen t-Statistik eingetragen. Auch wenn die Standardannahme unabhängig identisch normalverteilter Störgrößen im Modell (5.1) nicht zutrifft, lassen die t-Statistiken auf die Relevanz identifizierter Zeitabhängigkeit schließen. Somit wird empirisch die These gestützt, dass Terminpreise für Jahresstromlieferungen im Beobachtungszeitraum exponentiell ansteigen.

Von Interesse ist nun, inwiefern der geschätzte Anstieg im Zeitverlauf stabil bleibt.

Daher beschäftigt sich die weitere Analyse mit der (unbeobachtbaren) stochastischen Komponente logarithmierter Preise. Stellvertretend für sie wird die Zeitreihe der Regressionsresiduen $\{\hat{r}_t\}$ untersucht.

Die Abbildung 7 enthält die grafische Darstellung von $\hat{r}_t$ sowie die zugehörige Häufigkeitsverteilung und empirische Autokorrelationen in Abhängigkeit vom Verzögerungsparameter (Korrelogramm).

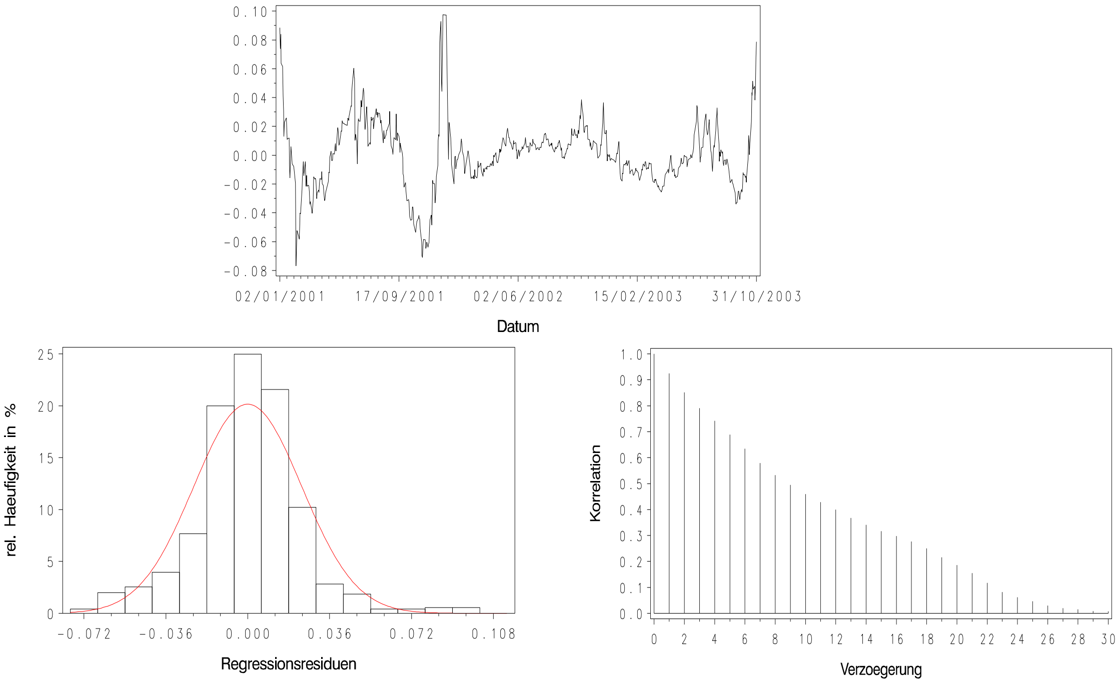

Abbildung 7: Regressionsresiduen: Zeitverlauf, Histogramm, Korrelogramm

Die bereits angesprochenen Phasen starker Preisausschläge 2001 und am Ende der Beobachtungsperiode und damit zusammenhängende schwere Verteilungsränder sind im Reihenverlauf bzw. im Historgramm deutlich zu sehen. Das Korrelogramm vermittelt den Eindruck relativ gleichmäßig abfallender Autokorrelationen in der Residualreihe. Hohe lineare Abhängigkeiten in der Reihenstruktur für kleine Zeitabstände verschwinden nach ca. einem Monat. Das deutliche autoregressive Muster in der Dynamik der Zeitreihe lässt insgesamt auf ein Abhängigkeitsverhalten schließen, das typisch für stark autoregressiv geprägte Prozesse ist.

Der visuelle Eindruck wird durch die Modellauswahl und anschließende Parameterschätzung in einem ARMA-Modell bestätigt. Zur Auswahl der Modellordnung wird das Bayes'sche Informationskriterium (kurz: BIC) eingesetzt:

$$BIC(\hat{\theta}) = -2L(\hat{\theta}) + k \cdot \log n$$

mit $L(\hat{\theta})$: Log-Likelihoodfunktion an der Stelle, k : Anzahl der Modellparameter und n : Reihenlänge. Bei der Modellselektion wird durch die Wahl des Modells mit dem geringsten BIC-Wert aus allen geschätzten ARMA-Modellen gewährleistet, dass unter der Prämisse sparsamer Parametrisierung die bestmögliche Anpassung an die Daten erreicht wird.

Für die untersuchte Residualreihe wird auf diese Weise ein zentrierter AR(1)-Prozess mit dem AR-Koeffizienten 0.94 (t-Wert=68.39) als das beste ARMA-Modell identifiziert. Somit entsteht in Bezug auf die zu prüfende Stationarität von $\{\hat{r}_t\}$ ein uneinheitliches Bild. Einerseits liegt der Schätzwert für den autoregressiven Parameter relativ nah an Eins und deutet somit auf eine I(1)-Eigenschaft der Residualreihe hin. Andererseits wird eine unendlich lang andauernde Wirkung einer jeden Innovation auf die künftige Reihenentwicklung, wie sie für I(1)-Prozesse charakteristisch ist, durch das Korrelogramm nicht bestätigt. Die innerhalb eines Monats verschwindenden Autokorrelationen sind eher als Anzeichen für die Stationarität von $\{\hat{r}_t\}$ zu werten.

Hochinteressant ist nun, wie sich in dieser Situation die Einheitswurzeltests entscheiden. Ihre Ergebnisse sollen den Aufschluss über die Stationarität der Reihe liefern. Da es sich bei der Reihe um Regressionsresiduen mit Erwartungswert 0 handelt, kommt der Dickey-Fuller-Test für den autoregressiven Prozess ohne Drift zum Einsatz. Zusätzlich werden seine Erweiterungen für höhere autoregressive Ordnungen betrachtet.

Wie aus der Tabelle 6 ersichtlich, lehnen beide Testversionen, die auf den Teststatistiken (2.20) bzw. (2.21) basieren, die Nullhypothese der Einheitswurzel für alle untersuchten Modellordnungen zum Niveau 5% ab.

AR-Ordnung	$n(\hat{\rho}-1)$	p-Wert		τ	p-Wert
1	-42.48	0.0018		-4.76	0.0001
2	-45.75	0.0018		-4.73	0.0001
3	-43.57	0.0018		-4.67	0.0002
4	-37.54	0.0018		-4.14	0.001
5	-42.77	0.0018		-4.33	0.0005

Tabelle 6: Ergebnisse des DF-Tests im linearen Modell

Somit wäre die Random-Walk-Hypothese für die Reihe $\{\hat{r}_t\}$ widerlegt. Folglich könnte man einen stabilen exponentiellen Trend im Strompreisverlauf seit 2001 annehmen. Zu bedenken ist allerdings, dass bei der bisherigen Analyse die Anzeichen für bedingte Heteroskedastie in den Residuen unberücksichtigt blieben. Wie in Kapitel 4 gezeigt, hat eine autoregressive Struktur in bedigten zweiten Momenten Auswirkungen auf die

Verteilung der Teststatistik im DF-Test. Sind $\hat{r}_t$ tatsächlich bedingt heteroskedastisch, so kann dies gegebenenfalls dazu führen, dass die eindeutigen Testergebnisse revidiert werden müssen. Um dieser Frage gezielt nachzugehen, wird zunächst die Reihe quadrierter Regressionsresiduen $\{\hat{r}_t^2\}$ untersucht.

Das entsprechende Korrelogramm in der Abbildung 8 zeigt, dass auch in den bedingten Varianzen starke serielle Abhängigkeiten bestehen. Zwar nehmen sie für kleine Zeitabstände etwas schneller ab als in den bedingten Erwartungswerten. Insgesamt bleibt der Einfluss einer aktuellen Beobachtung auf das Schwankungsverhalten der Reihe an nachfolgenden Tagen relativ lang bemerkbar.

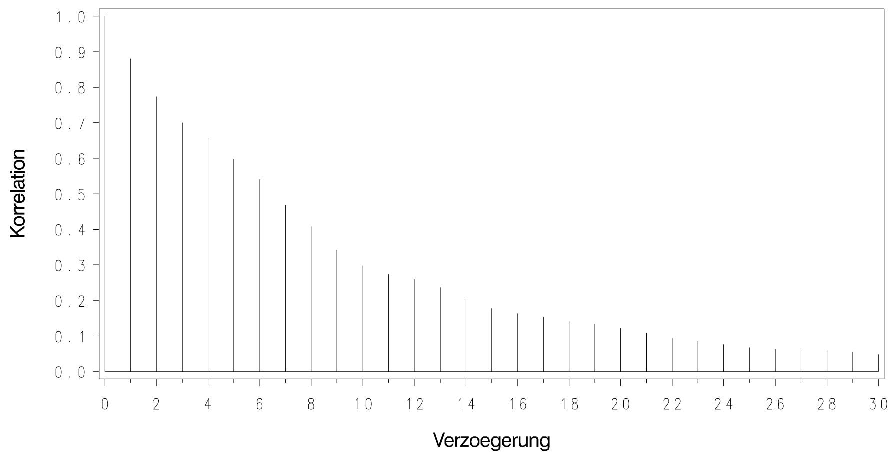

Abbildung 8: Korrelogramm quadrierter Regressionsresiduen

Zur formalen Aufdeckung bedingter Heteroskedastie kann man aus empirischen Autokorrelationen $c(k) = Corr(\hat{r}_t^2, \hat{r}_{t+k}^2)$ die Ljung-Box-Teststatistiken bilden:

$$Q(m) = n(n+2) \sum_{k=1}^{m} \frac{c^2(k)}{n-k} \,.$$

Unter der Nullhypothese unkorrelierter Quadrate $\hat{r}_t^2$ ist die Teststatistik $Q(m)$ nach McLeod und Li (1983) asymptotisch χ_m^2-verteilt. Diese Modifikation des weit bekannten Portmanteau Tests von Box/Pierce ist eins der Standardverfahren bei der Überprüfung, inwiefern eine empirische Reihe bedingt heteroskedastisch ist.

Wählt man $m \leq 24$, so lehnt der Test die Nullhypothese zum 5%-Niveau durchgehend ab. Die p-Werte sind extrem gering ($< 10^3$). Somit bestätigt das Testresultat bedingte Heteroskedastie für die Reihe $\{\hat{r}_t\}$, die durch hohe empirische Autokorrelationen in der Abbildung 8 bereits erkennbar wurde.

Es wird nun versucht, die aufgedeckte autoregressive Struktur in bedingten Varianzen durch das GRS-ARCH-Modell abzubilden. Dabei stellt sich heraus, dass auch hier die Modellordnung von 1 die beste Modellanpassung in Bezug auf das BIC-Informationskriterium liefert. Die simultane ML-Schätzprozedur für bedingte Erwartungswerte und Varianzen von $\hat{r}_t$ ergibt dann folgende Modellgleichungen:

$$\begin{aligned} \hat{r}_t &= 0.94\,\hat{r}_{t-1} + \widetilde{r}_t, \quad \widetilde{r}_t = \sigma_t \varepsilon_t, \\ \sigma_t &= 0.01 + 0.10\,\widetilde{r}_{t-1}, \quad \varepsilon \sim WR. \end{aligned} \tag{5.2}$$

Inwieweit das Modell die bestehenden Autokorrelationen in $\{\hat{r}_t\}$ und $\{\hat{r}_t^2\}$ erklärt, veranschaulicht die Abbildung 9. Darin sind Korrelogramme für die Restgrössen ε_t und die zugehörigen Quadrate zu sehen.

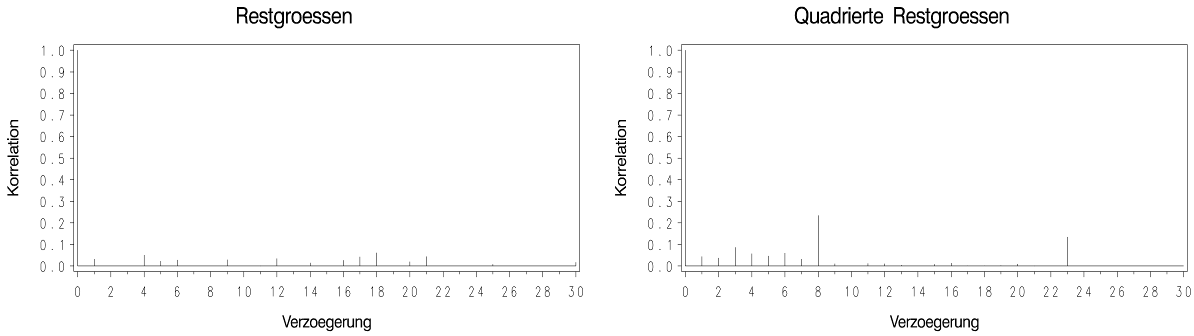

Abbildung 9: Korrelogramme für die Reihen der Restgrößen $\{\varepsilon_t\}$ und $\{\varepsilon_t^2\}$

Um die Unabhängigkeitsannahme für ε_t zu überprüfen, werden die Reihen $\{\varepsilon_t\}$ und $\{\varepsilon_t^2\}$ wiederum dem Portmanteau-Test von Box/Pierce unterzogen.

Für ε_t stellt man fest, dass die Ljung-Box-Statistiken für $m \leq 24$ unter dem theoretischen Median der jeweiligen χ^2_{m-2}-Verteilungen liegen. Nicht so eindeutig fällt das Urteil für die Quadrate ε_t^2. Während $Q(7)$ noch deutlich kleiner als der kritische Wert ist, übersteigen die Ljung-Box-Statistiken für $8 \leq m \leq 52$ die 95%-Quantile entsprechender χ^2_{m-2}-Verteilungen. Dies liegt an der hohen empirischen Autokorrelation der Ordnung 8. Erst langsam wird der Einfluss dieser Autokorrelation durch höhere Freiheitsgrade und somit auch steigende kritische Werte kompensiert, so dass für $m \geq 53$ die Nullhypothese unkorrelierter ε_t^2 zum Niveau 5% nicht mehr verworfen wird.

Die Aufnahme der von $\widetilde{r}_{t-8}$ in die GRS-ARCH-Modellgleichung bewirkt weder in Bezug auf Informationskriterien noch in Bezug auf die Korrelationsstruktur eine Anpassungsverbesserung. Aus fundamentaler Sicht gibt es keine Gründe, bei der Preisbildung für den 8-Werktage-Abstand irgendwelche Besonderheiten oder Abhängigkeitsmuster zu unterstellen. Daher ist es naheliegend, im hohen empirischen Wert für $Corr(\hat{r}_t, \hat{r}_{t+8})$ einen Ausreißer zu sehen, der rein zufälliger Natur ist. Somit erscheint die Unabhängigkeitsannahme vetretbar. Alles in allem liefert das Modell (5.2) eine zufriedenstellende Anpassung an die Reihe der Regressionsresiduen $\{\hat{r}_t\}$.

Nun ist man in der Lage, eine endgültige Entscheidung über die I(1)-Eigenschaft von $\{\hat{r}_t\}$ zu treffen. Dazu ist die erste Gleichung im Modell (5.2) auf die Form wie in (4.4) zu bringen. Die zu $\hat{\Psi} = -0.06$ gehörende t-Statistik τ_n ist dann mit dem kritischen Wert aus der Tabelle 2 zu vergleichen. Aus $|\sum_{b=1}^{\infty} b_i| \approx 0.1$ folgt, dass der kritische Wert zum 5%-Testniveau -2.907 beträgt. Er ist größer als die in der Tabelle 6 angeführte t-Statistik. Somit wird die I(1)-Hypothese verworfen.
Als Fazit ist festzuhalten, dass die stochastischen Abweichungen der Strompreise vom deterministischen exponentiellen Zeittrend (5.1) im betrachteten Zeitraum einem I(d)-Prozess mit $d < 1$ folgen.

5.2.2 Autoregressives Modell mit stochastischem Trend

Verneint man a priori die Existenz eines deterministischen Zeittrends in Strompreisen, so kann man versuchen, die logarithmierten Preise als einen rein stochastischen autoregressiven Prozess mit Drift wie in (2.24) zu modellieren. Zu bedenken ist allerdings die auf Seite 26 angesprochene Testgüteproblematik. Spezifiziert man im Modell keine deterministische Zeitkomponente, obwohl eine vorliegt, so wird der DF-Test sie auch nicht erkennen und die I(1)-Hypothese nicht ablehnen. Die bis hierhin bereits durchgeführte Analyse lässt genau dieses Szenario erwarten. Ob berechtigt oder nicht, wird im Folgenden der Standpunkt vertreten, in den beobachteten Preisen liege keine deteministische Trendkomponente vor.
Das Korrelogramm der Preislogarithmen sieht wie folgt aus:

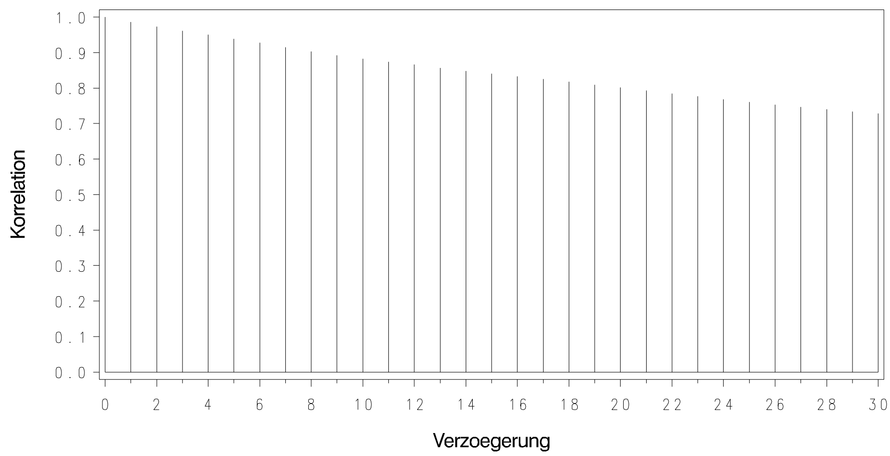

Abbildung 10: Korrelogramm natürlicher Preislogarithmen

Die empirischen Autokorrelationen nehmen äußerst langsam ab und sind selbst für die Ordnung von 30 mit ca. 0.8 noch sehr hoch. Das kann als eine mit zunehmender zeitlicher Entfernung kaum verminderte Wirkung einer jeden Innovation gewertet werden und folglich auf eine Einheitswurzel hinweisen. Das deutliche Anzeichen für die Instationarität wird vom DF-Test wie in (2.28) bestätigt. Die I(1)-Hypothese wird zum 5%-Niveau nicht abgelehnt, der p-Wert liegt nah an 1 bei einer positiven(!) t-Statistik. An diesem eindeutigen Bild ändert auch die bedingte Heteroskedastie nichts. Somit würde man den Schluss ziehen, dass die Preislogarithmen einem I(1)-Prozess folgen und zur weiteren Modellierung eine Differenzenbildung erfordern.
Wendet man den Differenzenfilter Δ auf die Preislogarithmen an, so erhält man, wie bereits erläutert, die Tagesrenditen der Strompreise. Ihren zeitlichen Verlauf sowie die geschätzte Autokorrelationsstruktur in bedingten ersten und zweiten Momenten enthält die Abbildung 11.

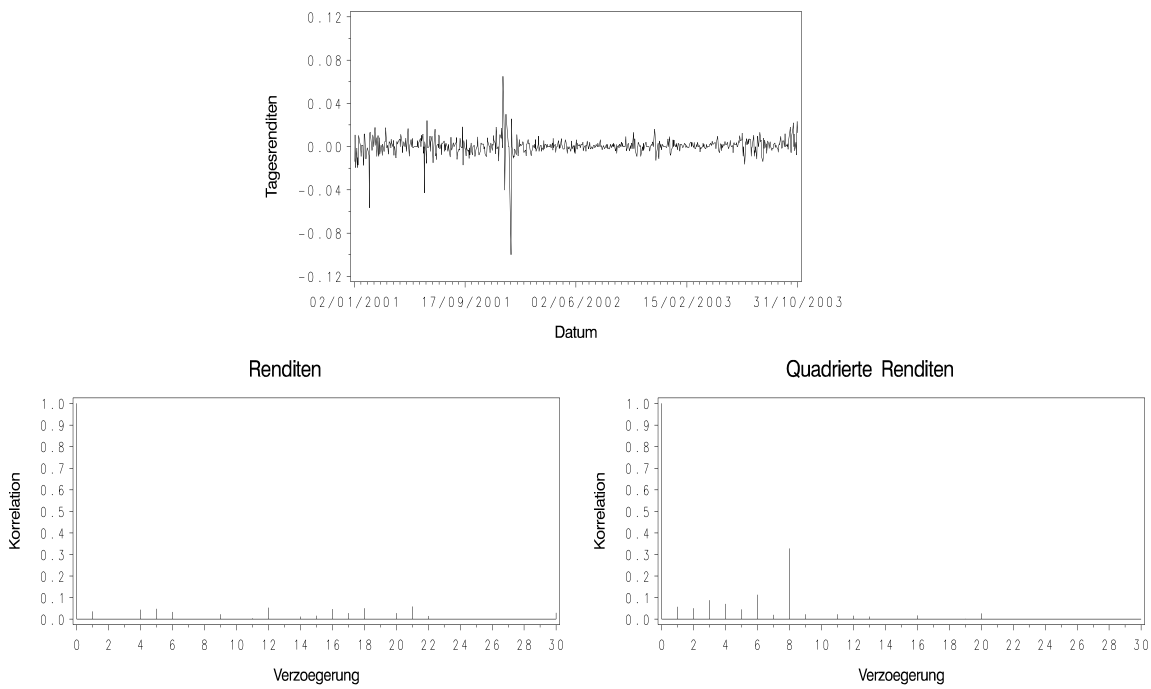

Abbildung 11: Differenzen von Preislogarithmen (Renditen) und zugehörige Korrelogramme

Während in bedingten Erwatrungswerten offenbar keine seriellen Abhängigkeiten bestehen, fällt in bedingten Varianzen, wie schon bei der Annahme einer deterministischen Trendkomponente, die Autokorrelation 8. Ordnung auf. Weder mit GARCH noch mit GRS-ARCH-Ansatz gelingt es, ein statistisch befriedigendes Modell für das geschätzte autoregressive Muster aufzustellen. Dies stützt indirekt die Annahme, dass es sich hier um keinen systematischen Saisoneffekt, sondern eher um den Einfluss einiger weniger Beobachtungen handelt. Mit diesem Abstrich können die Strompreisrenditen als weißes Rauschen angesehen werden. Zum Abschluss zeigt die Abbildung 12 ihre Häufigkeitsverteilung, die im Vergleich zu Normalverteilung eine wesentlich höhere Wölbung aufweist.

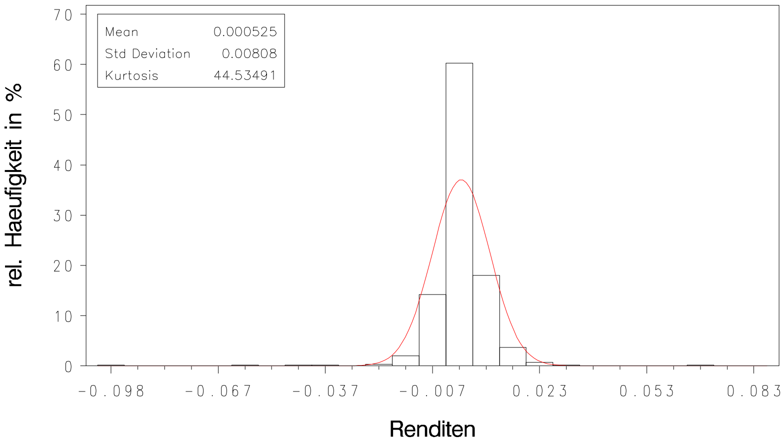

Abbildung 12: Strompreisrenditen: Korrelogramm

Fasst man die empirische Auswertung zusammen, sind folgende Punkte festzuhalten:

- Bei Annahme einer deterministischen Zeitkomponente wird der globale Trend in Preislogarithmen am besten durch ein Polynom dritten Grades erfasst. Die geschätzten stochastischen Abweichungen von diesem Trend besitzen sowohl in bedingten Erwartungswerten als auch in bedingten Varianzen serielle Abhängigkeiten. Unter Berücksichtigung bedingter Heteroskedastie wird für diese Abweichungen die I(1)-Hypothese abgelehnt. Dies lässt zu, sie als stationär anzusehen.

- Lehnt man die Annahme eines deterministischen Trends ab, so stellt man die I(1)-Eigenschaft in den Preislogarithmen fest. Die Preisrenditen können annähernd als unkorreliert in den ersten beiden bedingten Momenten angesehen werden.

- Unabhängig von der Trendannahme wird in bedingten Varianzen mit der hohen empirischen Autokorrelation 8. Ordnung eine nicht modellierbare Auffälligkeit festgestellt. Sie ist vermutlich auf ganz wenige Beobachtungen zurückzuführen, die einen Ausreißercharakter haben.

Kapitel 6

Zusammenfassung und Ausblick

Die vorliegende Arbeit beschäftigt sich mit der Überprüfung der Random-Walk-Hypothese für stochastische Prozesse mit bedingt heteroskedastischen Innovationen. Sowohl bei der Erkennung stochastischer Trends als auch bei der Modellierung autoregressiver Muster in bedingten Varianzen handelt es sich um Themen, die in der Zeitreihentheorie seit Jahrzehnten sehr intensiv untersucht und weiterentwickelt werden. Das Interesse speist sich aus Empirie: schaut man sich die Entwicklung unzähliger Marktdaten im Zeitablauf, so wird man fast unvermeidlich mit beiden Phänomenen konfrontiert. Allerdings widmen sich nur wenige Autoren dem Einfluss bedingt heteroskedastischer Innovationen auf die Nullverteilung in klassischen Einheitswurzeltests. Die vorliegende Untersuchung trägt dazu bei, diese Frage an der Schnittstelle zweier großer Themenbereiche in der empirischen Wirtschaftsforschung weiter zu beleuchten. Nach getrennter Betrachtung beider Themen in Kapiteln 2 und 3 werden sie im Kapitel 4 auf theoretischer Ebene zusammengeführt. Resultierende Verfahren werden in Kapitel 5 zur Beschreibung von Preisen am deutschen Strommarkt eingesetzt.

Kapitel 2 definiert einige wichtige Eigenschaften diskreter stochastischer Prozesse und stellt grundlegende Ansätze zur Modellierung linearer serieller Abhängigkeiten in den Reihenwerten dar. Dabei wird insbesondere auf unterschiedliches langfristiges Abhängigkeitsverhalten eingegangen, wie es für ARMA-Modelle, Reihen mit langem Gedächtnis (ARFIMA-Modelle) und Reihen mit stochastischem Trend (ARIMA-Modelle) typisch ist. Mit Einheitswurzeltests werden dann Verfahren vorgestellt, die basierend auf der ARMA-Parametrisierung die Hypothese eines stochastischen Trends für eine gegebene Reihe überprüfen. Für klassische Dickey-Fuller-Tests werden asymptotische Verteilungen der Teststatistiken unter I(1)-Nullhypothese nachvollzogen. Ein

Überblick über Erweiterungen und alternative Vorgehen beim Testen auf die Einheitswurzel schließt sich an.

Im Kapitel 3 werden Ansätze zur Modellierung autoregressiver Strukturen in bedingten Varianzen betrachtet. Es handelt sich dabei um Modelle der GARCH-Familie, die u. a. Persistenz, Asymmetrie und langes Gedächtnis in bedingten Varianzen abbilden können. Ein besonderer Augenmerk gilt hier den Stationaritätseigenschaften entsprechender Prozesse. Als ein Modell, das langes Gedächtnis und Stationarität in sich vereinen kann, wird der lineare ARCH-Prozess von Giraitis, Robinson und Surgailis (2000), kurz: GRS-ARCH, genauer betrachtet. Für stationäre GRS-ARCH-Prozesse wird im Lemma 1 die NED-Eigenschaft nachgewiesen.

Kapitel 4 behandelt Veränderungen in der asymptotischen Verteilung von Dickey-Fuller-Teststatistiken unter der Nullhypothese, die aufgrund bedingt heteroskedastischer Innovationen entstehen. Als zentrales Ergebnis wird im Satz 2 die entsprechende Grenzverteilung für stationäre GRS-ARCH-Prozesse angegeben. Basierend darauf werden kritische Werte der Teststatistiken simuliert. Sie sind in Tabelle 2 enthalten. Zum Abschluss der theoretischen Betrachtung wird ein Vergleich mit unkorrigierten Dickey-Fuller-Tests angestellt. Weitere Simulationen zeigen, dass beim Testen auf eine Einheitswurzel die neuen kritischen Werte das nominale Niveau insgesamt besser einhalten.

Im Kapitel 5 werden Terminpreise für Grundlast-Stromlieferungen für das jeweils folgende Jahr anhand historischer Börsennotierungen an der Strombörse EEX betrachtet. Sie geben die allgemeine Entwicklung der Großhandelspreise am liberalisierten deutschen Strommarkt wieder. Zur Datenanalyse und Modellierung werden die im theoretischen Teil diskutierten Schätz- und Testverfahren eingesetzt. Im untersuchten Zeitraum zwischen dem 2. Januar 2001 und dem 31. Oktober 2003 kann an die Preislogarithmen ein globaler Zeittrend angepasst werden, der durch einen Polynom dritten Grades dargestellt wird - vgl. Modellgleichung (5.1). Die stochastischen Abweichungen von diesem Trend können, auch unter Berücksichtigung ihrer bedingt heteroskedastischer Struktur, als stationär angesehen werden.

Die erfolgte theoretische Betrachtung von Einheitswurzeltests unter bedingt heteroskedastischen Innovationen lässt sich mindestents in zwei Richtungen erweitern. Einerseits wäre es interessant zu sehen, wie weitere, z. B. asymmetrische Volatilitätsmuster auf die Grenzverteilungen von Teststatistiken in klassichen Einheitswurzeltests wirken. Nicht weniger spannend wäre zu sehen, wie bedingt heteroskedastische Strukturen

Verteilungsaussagen in speziellen Einheitswurzeltests, die etwa zwischen einem stochastischen Trend und einem Strukturbruch unterscheiden, verändern.
In Bezug auf die Entwicklungen am Strommarkt ist besonders die zweite Fragestellung interessant. Laut allgemeiner Einschätzung steht die beobachtete Strompreisentwicklung im Zusammenhang mit der ab 2005 erwarteten Einführung der Emissionszertifikate. Danach werden die bei der Stromerzeugung aus Gas, Stein- und Braunkohle entstehenden CO_2-Emissionen durch Herausgabe entsprechender handelbarer Zertifikate eingeschränkt. Da der gesamte CO_2-Ausstoß stufenweise abgesenkt werden soll, ist davon auszugehen, dass die kostenlose Ausstattung mit Emissionszertifikaten den Bedarf deutscher Stromerzeuger nicht vollständig decken wird. Zusätzliche Zertifikate wären dann am Markt zu beschaffen und würden somit die Erzeugungskosten anheben. Die Unsicherheiten in Bezug auf das Volumen und Aufteilung der Zertifikate auf verschiedene Energieträger waren zuletzt ein wesentlicher Preistreiber am Strommarkt. Mit Bekanntgabe des Nationalen Allokationsplans 2004 würde diese preistreibende Wirkung wegfallen und es könnte zu einem Strukturbruch in der Preisentwicklung kommen.

Abkürzungen und Symbole

$\|\cdot\|_p$	-	L_p-Norm $\sqrt[p]{E(\cdot)^p}$, $p \geq 1$
$\xrightarrow{d}$	-	Verteilungskonvergenz in Funktionenräumen
α_m, ϕ_m	-	Suprema bei der Definition der Mixing-Prozesse
α_i	-	ARCH-Parameter, Koeffizienten des ARCH-Filters
β_i	-	GARCH-Parameter, Koeffizienten des GARCH-Filters
$\gamma(h)$	-	Autokovarianz der Ordnung h: $Cov(X_t, X_{t+h})$
Δ	-	Differenzen-Operator
δ	-	Asymmetrie-Parameter im asymmetrischen Power-GARCH-Prozess
ε_t	-	Innovationen eines stochastischen Prozesses
$\{\epsilon_t\}$	-	fraktionales Rauschen
$\{\eta_t\}$	-	starkes Weißes Rauschen bestehend aus unabhängig identisch verteilten Zufallsvariablen mit Erwartungswert 0 und Varianz σ^2
$\{\check{\eta}_t\}$	-	schwaches Weißes Rauschen bestehend aus unkorrelierten identisch verteilten Zufallsvariablen mit Erwartungswert 0 und Varianz σ^2
$\{\tilde{\eta}_t\}$	-	Prozess autokorrelierter identisch verteilter Zufallsvariablen mit Erwartungswert 0 und Varianz σ^2
τ	-	t-Teststatistik in Dickey-Fuller-Tests
ν_t	-	Innovationen in Quadraten eines bedingt heteroskedastischen Prozesses
ξ	-	Asymmetrie-Parameter im exponentiellen GARCH-Prozess
θ_i	-	moving average-Parameter, Koeffizienten des moving average-Filters $\theta(L)$
ρ_i	-	autoregressive Parameter, Koeffizienten des autoregressiven Filters $\rho(L)$
$\hat{\rho}_{KQ}$	-	Kleinste-Quadrate-Schätzung für ρ
$\hat{\rho}_{ML}$	-	Maximum-Likelihood-Schätzung für ρ
Ψ, ψ_j	-	Parameter eines autoregressiven Prozesses nach der Dickey-Fuller-Transformation
$\omega \in \Omega$	-	Zufallsereignis aus der Menge aller möglicher Ausgänge eines Zufallsexperiments

(F)(I)(G)ARCH	-	(fraktional) (integrierter) (verallgemeinerter) autoregressiver bedingt heteroskedastischer Prozess
AR(F)(I)MA	-	autoregressiver (fraktional) (integrierter) Moving-Average-Prozess
b_i	-	Parameter im GRS-ARCH-Prozess
b	-	Summe der Parameter im GRS-ARCH-Modell: $b = \sum_{i=1}^{\infty} b_i$
B	-	Summe quadrierter Parameter im GRS-ARCH-Modell: $B = \sum_{i=1}^{\infty} b_i^2$
BIC	-	Bayes'sches Informationskriterium
$C[0,1]$	-	Menge aller stetiger reelwertiger Funktionen auf [0,1]
CMT	-	Continious Mapping Theorem
$D[0,1]$	-	Menge aller rechtseitig stetiger Funktionen mit linkseitigem Grenzwert auf [0,1]
d	-	Differenparameter, Ordnung eines integrierten Prozesses
FZGWS	-	Funktionaler Zentraler Grenzwertsatz
$\mathcal{F}_s^t$	-	Sigma-Algebra $\sigma(X_s, \ldots, X_t)$, $s \leq t$
GED	-	General Error Distribution
GRS-ARCH	-	linearer ARCH-Prozess von Giraitis/Robinson/Surgailis
$H_n(\cdot)$	-	Hesse-Matrix
$I(d)$	-	integrierter Prozess der Ordnung d
$l_t(\cdot)$	-	Log-Likelihood-Funktion
$L(\cdot)$	-	Backshift-Operator
$\{M_t\}$	-	Martingal
MD	-	Martingaldifferenz
OTC-Handel	-	Sammelbegriff für außerbörslichen (Strom)Handel
n	-	Reihenlänge, $t = 1, \ldots, n$
R1,R2	-	Regularitätsbedingungen, S. 55
$\{R_t\}$	-	bedingt heteroskedastischer stochastischer Prozess
$\hat{r}_t$	-	Regressionsresiduum
u.i.v.	-	unabhängig identisch verteilt
$\{U_t\}$	-	Referenzprozess bei der Definition von NED
W	-	Wiener-Prozess (Brownsche Bewegung)
$\{X_t\}$, x_t	-	diskreter stochastischer Prozess und Elemente zugehöriger Zeitreihe
$\{Y_t\}$	-	integrierter stochastischer Prozess

Literaturverzeichnis

Abadir, K.M. (1993): "On the asymptotic power of unit root tests", *Econometric Theory* 9, 189-221.

Abadir, K.M. (1995): "The limiting distribution of the t ratio under a unit root", *Econometric Theory* 11, 775-793.

Amsler, C., Lee, J. (1995): "An LM test for a unit root in the presence of a structural change", *Econometric Theory* 11, 359-368.

Andrews, D.W.K. (1984): "Non-strong mixing autoregressive processes", *Journal of Applied Probability* 21, 930-934.

Baillie, R.T. (1996): "Long memory processes and fractional integration in econometrics", *Journal of Econometrics* 73, 5-59.

Baillie, R.T., Bollerslev, T. (1989): "The message in daily exchange rates: A conditional variance tale", *Journal of Business and Economic Statistics* 7, 297-305.

Baillie, R.T., Bollerslev, T., Mikkelsen, H.O. (1996): "Fractionally integrated generalized autoregressive conditional heteroscedasticity", *Journal of Econometrics* 74, 3-30.

Billingsley, P. (1968): *Convergence of Probability Measures*, John Wiley, New York.

Bollerslev, T. (1986): "General autoregressive conditional heteroscedasticity", *Journal of Econometrics* 31, 307-327.

Bollerslev, T. (1987): "A conditional heteroscedastic time series model for speculative prices and rates of return", *Review of Economics and Statistics* 69, 542-547.

Bollerslev, T., Engle, R.F., Nelson, D.B. (1994): "ARCH models", Kapitel 49 in: Engle, R.F., McFadden D.L. (Hrsg.), *Handbook of Econometrics, Volume 4*, Elsevier Science, Amsterdam.

Bollerslev, T., Mikkelsen, H.O. (1996): "Modelling and pricing long-memory in stock market volatility", *Journal of Econometrics* 73, 151-184.

Bougerol, P., Picard, N. (1992): "Stationarity of GARCH processes and of some nonnegative time series", *Journal of Econometrics* 52, 115-127.

Breidt, F.J., Crato, N., de Lima, P. (1998): "On the detection and estimation of long memory in stochastic volatility", *Journal of Econometrics* 83, 325-348.

Breitung, J. (2002): "Nonparametric tests for unit roots and cointegration", *Journal of Econometrics* 108, 343-363.

Breitung, J., Gourieroux, C. (1987): "Rank tests for unit roots", *Journal of Econometrics* 81, 7-28.

Brunetti, C., Gilbert, C. (2000): "Bivariate FIGARCH and Fractional Cointegration", *Journal of Empirical Finance* 7, 509-530.

Carrasco, M., Chen, X. (2002): "Mixing and moment properties of various GARCH and stochastic volatility models", *Econometric Theory* 18, 17-39.

Clewlow, L., Strickland, C. (2000): *Energy Derivatives: Pricing and Risk Management*, Lacima Publications, London.

Curtius, F. (2003): "Die Rolle des Stromhandels für die Energiewirtschaft hat sich verändert", *Energiewirtschaftliche Tagesfragen* 53, 580-584.

Davidson, J. (1994): *Stochastic Limit Theory*, Oxford University Press, Oxford.

Davidson, J. (2000): *Econometric Theory*, Blackwell Publishers, London.

Davidson, J. (2002a): "Establishing conditions for the Functional Cental Limit Theorem in nonlinear and semiparametric time series processes", *Journal of Econometrics* 106, 243-269.

Davidson, J. (2002b): "Moment and memory properties of linear conditional heteroscedasticity models", *Discussion Paper, University of Cardiff.*

Davies, L., Krämer, W. (2002): "Testing for unit roots in the context of misspecified logarithmic random walks", Economic Letters 74, 313-319.

DeJong, D.N., Nankervis, J.C., Savin, N.E., Whiteman, C.H.(1992): "The power of unit root tests in time series with autoregressive errors", *Journal of Econometrics* 53, 323-343.

DeJong, D.N., Davidson, J. (2000): "The functional central limit theorem and weak convergence to stochastic integrals I: Weakly dependent processes", *Econometric Theory* 16, 621-642.

Dickey, D.A., Fuller, W.A. (1979): "Distribution of the estimators for autoregressive time series with a unit root", *Journal of the American Statistical Association* 74, 427-431.

Ding, Z., Granger, C.W.J., Engle, R.F. (1993): "A long memory property of stock market returns and a new model", *Journal of Empirical Finance* 1, 103-106.

Engle, R.F. (1982): "Autoregressive conditional heteroscedasticity with estimation of the variance of UK inflation", *Econometrica* 50, 987-1007.

Engle, R.F., und Bollerslev, T. (1986): "Modelling the persistence of conditional variances", *Econometric Reviews* 5, 1-50.

Engle, R.F., Granger, C.W.J. (1987): "Cointegration and error correction: representation, estimation and testing", *Econometrica* 55, 176-251.

Fernandez, C., Steel, M. (1998): "On Bayesian modelling of fat tails and skewness", *Journal of American Statistical Association* 93, 359-371.

French, K.R., Schwert, G.W., Stambaugh, R.F. (1987): "Expected stock returns and volatility", *Journal of Financial Economics* 19, 3-30.

Fuller, W.A. (1976): *Intoduction to Statistical Time Series*, John Wiley, New York.

Gallant, A.R., White, H. (1988): *A Unified Theory of Inference for Nonlinear Dynamic Models*, Basil Blackwell, Oxford.

Geweke, J. (1986): "Modelling persistence of conditional variances: Comment", *Econometric Reviews* 5, 57-61.

Giraitis, L., Kokoszka, P., Leipus, R. (2000): "Stationary ARCH models: dependence structure and central limit theorem", *Econometric Theory* 16, 3-22.

Giraitis, L., Robinson, P.M., Surgailis, D. (2000): "A model for long memory conditional heteroscedasticity", *Annals of Applied Probability* 10, 1002-1024.

Glosten, L., Jagannathan, R., Runkle, D. (1993): "On the relation between expected return on stocks", *Journal of Finance* 48, 1779-1801.

Gourieroux, C. (1997): *ARCH Models and Financial Applications*, Springer-Verlag, New York.

Granger, C.W.J., Newbold, P. (1974): "Spurious regressions in econometrics", *Journal of Econometrics* 2, 111-120.

Granger, C.W.J., Joyeux, R. (1980): "An introduction to long-memory time series models and fractional differencing", *Journal of Time Series Analysis* 1, 15-29.

Granger, C.W.J., Hallmann, J. (1988): "The algebra of I(1) time series", *Finance of Economics Discussion Series 45, Federal Reserve Board, Washington D.C.*.

Hall, P., Heyde, C.C. (1980): *Martingale Limit Theory and Application*, Academic Press, New York.

Hamilton, J.D. (1994): *Time Series Analysis*, Princeton University Press, Princeton.

Hansen, B.E. (1991): "GARCH(1,1) processes are near-epoch dependent", *Economics Letters* 36, 181-186.

Hansen, B.E. (1992): "Convergence to stochastic integrals for dependent heterogeneous processes", *Econometric Theory* 8, 489-500.

Hasan, M.N., Koenker, R.W. (1997): "Robust rank tests of the unit root hypothesis", *Econometrica* 65, 133-161.

Hassler, U. (1993): *Fraktional integrierte Prozesse in der Ökonometrie*, Haag+Herchen, Frankfurt/Main.

Hassler, U. (1994): "Einheitswurzeltests - Ein Überblick", *Allgemeines Statistisches Archiv* 78, 207-228.

Hassler, U., Wolters, J. (1994): "On the power of unit root tests against fractional alternatives", *Economics Letters* 45, 1-5.

Hassler, U., Wolters, J. (1995): "Long memory in inflation rates: International Evidence", *Journal of Business and Economic Statistics* 13, 37-45.

Hatanaka, M. (1996): *Time-Series-Based Econometrics*, Oxford University Press.

He, C., Teräsvirta, T. (1999): "Properties of moments of a family of GARCH processes", *Journal of Econometrics* 92, 173-192.

He, C., Teräsvirta, T., Malmsten, H. (2002): "Moment structure of a family of first-order Exponential GARCH models", *Econometric Theory* 18, 868-885.

Higgins, M.L, Bera, A.K. (1992): "A class of nonlinear ARCH models", *International Economic Review* 33, 137-158.

Hosking, J.R.M. (1981): "Fractional differencing", *Biometrika* 68, 165-176.

Kazakevicius, V., Leipus, R. (2002): "On stationarity in the ARCH(∞) model", *Econometric Theory* 18, 1-16.

Kim, K., Schmidt, P. (1993): "Unit root tests with conditional heteroscedasticity", *Journal of Econometrics* 59, 287-300.

Krämer, W. (1998): "Fractional integration and the augmented Dickey-Fuller-Test", *Economics Letters* 61, 269-272.

Krämer, W. (2002): "Statistische Besonderheiten von Finanzzeitreihen", *Jahrbücher für Nationalökonomie und Statistik* 222, 2, 210-229.

Kwiatkowski, D., Phillips, P.C.B., Schmidt, P. und Shin, Y. (1992): "Testing the null hypothesis of stationarity against the alternative of a unit root: How sure are we that economic time series have a unit root?", *Journal of Econometrics* 54, 159-178.

Lambert, P., Laurent, S. (2001): "Using GARCH-type models with a skewed student distribution for the innovations", *Discussion Paper 0125, Université Catholique de Louvain, Institut de Statistique.*

Laurent, S., Peters, J.P. (2002): "G@RCH 2.2: An Ox package for estimating and forecasting various ARCH models", *Journal of Economic Surveys* 16, 447-485.

Lee, H.S., Amsler, C. (1997): "Consistency of the KPSS unit root test against fractionally integrated alternative", *Economics Letters* 55, 151-160.

Lee, D., Schmidt, P. (1996): "On the power of the KPSS test of stationarity against fractionally-integrated alternatives", *Journal of Econometrics* 73, 285-302.

Ling, S., McAleer, M. (2002a): "Stationarity and the existence of moments of a family of GARCH processes", *Journal of Econometrics* 106, 109-117.

Ling, S., McAleer, M. (2002b): "Necessary and sufficient moment conditions for the GARCH(r,s) and Asymmetric Power GARCH(r,s) models)", *Econometric Theory* 18, 722-729.

Ling, S., Li, W.K. (1998): "Limiting distributions of maximum likelihood estimators for unstable ARMA models with GARCH errors", *Annals of Statistics* 26, 84-125.

Lobato, I.N., Savin, N.E. (1998): "Real and spurious long-memory properties of stock market data", *Journal of Business and Economic Statistics* 16, 261-268.

McLeod, A., Li, W.K. (1983): "Diagnostic checking ARMA time series models using squared-residual autocorrelations", *Journal of Time Series Analysis* 4, 269 - 273.

McLeish, D.L. (1975): "A maximal inequality and dependent strong laws", *Annals of Probability* 3, 329-339.

Montanes, A., Reyes, M. (2000): "Structural breaks, unit roots and methods for removing the autocorrelation pattern", *Statistics and Probability Letters* 48, 401-409.

Nelson, D.B. (1990): "Stationarity and persistence in the GARCH(1,1) model", *Econometric Theory* 6, 318-334.

Nelson, D.B. (1991): "Conditional heteroskedasticity in asset returns: a new approach", *Econometrica* 59, 349-370.

Nelson, D.B., Cao, C.Q. (1992): "Inequality constraints in the univariate GARCH model", *Journal of Business and Economic Statistics* 10, 229-235.

Palm, F., Vlaar, P. (1997): "Simple diagnostics procedures for modelling financial time series", *Allgemeines Statistisches Archiv* 81, 85-101.

Pantula, S.G. (1986): "Modelling persistence of conditional variances: Comment", *Econometric Reviews* 5, 71-74.

Park, J.Y., Sung, J. (1994): "Testing for unit roots in models with structural change", *Econometric Theory* 10, 917-936.

Perron, P. (1988): "Trends and random walks in macroeconomic time series: Further evidence from a new approach", *Journal of Economic Dynamics and Control* 12, 297-332.

Perron, P. (1990): "Testing for a unit root in a time series with a changing mean", *Journal of Business and Economic Statistics* 8, 153-162.

Perron, P., Vogelsang, T.J. (1992): "Testing for a unit root in a time series with a changing mean: Correction and extensions", *Journal of Business and Economic Statistics* 10, 467-470.

Phillips, P.C.B. (1987): "Time series regression with a unit root", *Econometrica* 55, 277-301.

Phillips, P.C.B., Perron, P. (1988): "Testing for a unit root in time series regression", *Biometrika* 75, 335-346.

Phillips, P.C.B., Xiao, Z. (1998): "A primer on unit root testing", *Journal of Economic Surveys* 12, 423-467.

Rao, C.R., (1973): *Linear Statistical Inference and Its Applications*, 2. Auflage, John Wiley, New York.

Robinson, P.M. (1995): "Gaussian semiparametric estimation of long range dependence", *Annals of Statistics* 23, 1630 - 1661.

Said, S.E., Dickey, D.A. (1984): "Testing for unit roots in ARMA(p,q)-models with unknown p and q", *Biometrika* 71, 599-607.

Saikkonen, P. (1995): "Problems with the asymptotic theory of maximum likelihood estimation in integrated and cointegrated systems", *Econometric Theory* 11, 888-911.

Saikkonen, P., Luukkonen, R. (1993): "Point optimal tests for testing the order of differencing in ARIMA models", *Econometric Theory* 9, 343-362.

Saikkonen, P., Lütkepohl, H. (2001): "Testing for unit roots in time series with level shifts", *Allgemeines Statistisches Archiv* 85, 1-25.

Saikkonen, P., Lütkepohl, H. (2002): "Testing for a unit root in atime series with a level shift at unknown time", *Econometric Theory* 18, 313-348.

Schweiger, J. (1999): *Tests auf langes Gedächtnis in Kointegrationsresiduen*, Diplomarbeit, Fachbereich Statistik, Universität Dortmund.

Schlittgen, R., Streitberg, B. (1997): *Zeitreihenanalyse*, 7. Auflage, Oldenbourg Verlag, München.

Schoffer, O. (2003): *Modellierung von Kapitalmarktrenditen mittels asymmetrischer GARCH-Modelle*, Dissertation, Fachbereich Statistik, Universität Dortmund.

Schwert, G.W. (1989): "Tests for unit roots: A Monte Carlo investigation", *Journal of Business and Economic Statistics* 7, 147-158.

Seo, B. (1999): "Distribution theory for unit root tests with conditional heteroscedasticity", *Journal of Econometrics* 91, 113-134.

Tanaka, K. (1990): "Testing for a moving average unit root", *Econometric Theory* 6, 433-444.

Taylor, S.J., (1986): *Modelling Financial Time Series*, Wiley, New York.

Tse, Y. (1998): "The conditional heteroscedasticity of the Yen-Dollar exchange rate", *Journal of Applied Econometrics* 193, 49-55.

Zakoian, J.-M. (1994): "Threshold heteroscedastic models", *Journal of Economic Dynamics and Control* 18, 931-955.