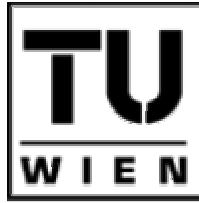


Die approbierte Originalversion dieser Diplom-/Masterarbeit ist an der Hauptbibliothek der Technischen Universität Wien aufgestellt (<http://www.ub.tuwien.ac.at>).

The approved original version of this diploma or master thesis is available at the main library of the Vienna University of Technology (<http://www.ub.tuwien.ac.at/englweb/>).



TECHNISCHE
UNIVERSITÄT
WIEN

VIENNA
UNIVERSITY OF
TECHNOLOGY

DIPLOMARBEIT

Identifizierbarkeit von univariaten GARCH-Prozessen

Ausgeführt am Institut für

Wirtschaftsmathematik

der Technischen Universität Wien

unter der Anleitung von **Univ.Prof. Dr. Manfred DEISTLER**

durch

Bettina Sereinig

Gumpendorferstrasse 83/2/13
1060 Wien

Wien, am _____
Datum

Unterschrift (Student)

Identifizierbarkeit von univariaten GARCH-Prozessen

28. Oktober 2008

Inhaltsverzeichnis

1	Einleitung	3
2	Grundlagen der Wahrscheinlichkeitstheorie und der Zeitreihenanalyse	5
2.1	Wahrscheinlichkeitstheorie	5
2.2	Stochastische Prozesse und Zeitreihen	11
2.2.1	Grundlegende Definitionen	11
2.2.2	Einige Beispiele für stationäre Prozesse	12
2.3	Stationäre Prozesse im Frequenzbereich	15
2.3.1	Lineare Transformation von stationären Prozessen	17
3	Finanzzeitreihen	20
3.1	„Stylized Facts“ von Finanzzeitreihen	21
3.1.1	Returns (X_t) sind unkorreliert, (X_t) und (X_t^2) sind stark auto- korreliert	21
3.1.2	Die bedingte Volatilität der Returns ist nicht konstant	21
3.1.3	Die Verteilung von Returns ist leptokurtisch	21
3.1.4	Leverage Effekte	22
4	Bedingt heteroskedastische Zeitreihen	23
4.1	GARCH(p, q)–Prozesse	25
4.1.1	Existenz von stationären Lösungen	26
4.1.2	Existenz von Momenten höherer Ordnung	31
5	Skalare ARMA–Prozesse und ihre Identifizierbarkeit	35
5.1	Das Spektrum eines (kausalen) ARMA–Prozesses	38
5.2	Identifizierbarkeit von skalaren ARMA–Prozessen	39
6	Der quadrierte GARCH(p, q)–Prozess	43
6.1	Schwache Stationarität von (X_t^2)	44
6.1.1	Bestimmung von $E(X_t^4)$	44
6.1.2	Bedingungen für die schwache Stationarität von (X_t^2)	48
7	Identifizierbarkeit von univariaten GARCH(p, q)–Prozessen	51

1 Einleitung

„*Optimal behavior takes risks that are worthwhile.*“

Frei übersetzt heißt dieser Satz: „Ein optimales Verhalten geht Risiken ein, die es wert sind eingegangen zu werden.“ Er steht zu Beginn eines Artikels, den Robert F. Engle anlässlich der Verleihung des Wirtschaftsnobelpreises an ihn im Dezember 2003 veröffentlicht hat. Robert F. Engle bezeichnet diesen Satz als das zentrale Paradigma im Finanzwesen. Es gibt Risiken, bei denen die Vorteile, die Gewinne, eventuelle Nachteile, meist in Form von Kosten, bei weitem überwiegen, wenn man sie eingeht. Um überhaupt einen Gewinn zu erzielen, muss ein Risiko eingegangen werden, aber nicht jedes Risiko bringt einen angemessenen Gewinn. Welche Risiken sind es also wert, dass man sie eingeht? Die Risiken, also die Verluste, und die Gewinne fallen erst zu einem unbestimmten Zeitpunkt in der Zukunft an, deshalb muss der Erwartungswert der Verluste durch den Erwartungswert der Gewinne ausgeglichen werden. Daher werden bei optimalem Verhalten, speziell in einem optimalen Portfolio die Erwartungswerte der Gewinne maximiert und die der Risiken minimiert.

Schon seit den 50er Jahren des vorigen Jahrhunderts wird das Risiko eines Portfolios durch die Varianz des Gesamtwerts dieses Portfolios ausgedrückt. Unter der Annahme, dass das Risiko minimiert beziehungsweise vermieden wird, wurden optimale Portfolios und Strategien abgeleitet. Zum Beispiel das Capital Asset Pricing Model (CAPM), dem die Annahme zu Grunde liegt, dass alle Investoren identische Ziele und dieselben Informationen haben. Als dann Praktiker diese Strategien anwenden und implementieren wollten, wurden Schätzwerte für die Varianzen gebraucht. Meist wurde die sogenannte *Volatilität*, das ist die positive Quadratwurzel der Varianz (Standardabweichung), angegeben. Es wurde dann sehr rasch erkannt, dass bei Finanzzeitreihen sich die Volatilität im Laufe der Zeit verändert.

Für unterschiedliche Zeiträume wurden verschiedene Volatilitäten gefunden. Oft wurde ein sehr einfacher Ansatz gewählt, nämlich die *Historische Volatilität*. Dabei wird die Stichproben-Standardabweichung über eine bestimmte vergangene Periode als Schätzwert für die Volatilität genommen. Die Frage ist dabei, wie lange diese Periode sein soll. Wird sie zu lange gewählt, werden manche Stichprobenwerte, die schon weit zurückliegen, nicht mehr relevant für die zu schätzende Volatilität sein. Bei einer zu kurzen Periode werden die Schätzwerte für aufeinanderfolgende Zeitpunkte vielleicht zu stark fluktuieren. Außerdem soll eigentlich das Risiko über einen zukünftigen Zeitraum geschätzt werden. Daher werden der Schätzwert für den aktuellen Zeitpunkt wie auch Prognosewerte für zukünftige Zeitpunkte benötigt. Deshalb versuchten die Wissenschaftler fortan, eine Theorie für *dynamische* Volatilitäten zu entwickeln.

1982 stellte Robert F. Engle in dem Artikel „*Autoregressive conditional heteroscedastic*

models with estimates of the variance of United Kingdom inflation“ (Econometrica, 50: S. 987 – 1007) zum ersten Mal ARCH(q)(*autoregressive conditionally heteroscedastic*)–Prozesse vor. In diesem mathematischen Modell stellt die Zufallsvariable X_t zum Beispiel den Wert des Ertrags (Returns) einer risikobehafteten Finanzanlage zum Zeitpunkt t dar. Die bedingte Varianz σ_t^2 von X_t wird unter der Bedingung, dass die vergangenen Werte von X_t und damit auch von σ_t^2 bekannt sind, berechnet. Es wird angenommen, dass sich diese bedingte Varianz als Funktion der letzten q Werte von X_t darstellen lässt. Die bedingte Varianz ändert sich im Laufe der Zeit, während die unbedingte Varianz, also der unbedingte Erwartungswert der bedingten Varianz, gleich bleibt. Durch diese Voraussetzungen können einige der „Stylized Facts“ von Finanzdaten, die in Kapitel 3 beschrieben werden, recht gut in einem mathematischen Modell abgebildet werden. Dieses Modell wurde 1986 von Tim Bollerslev, einem Mitarbeiter von Engle, zu GARCH(p, q)(*generalized autoregressive conditionally heteroscedastic*)–Prozessen erweitert (siehe [1]). Bollerslev setzte voraus, dass die bedingte Varianz zusätzlich noch von den letzten p Werten der bedingten Varianz selbst abhängt.

Die Forschung und Untersuchungen, die seither auf diesem Gebiet gemacht wurden, sind sehr vielfältig, wie die umfangreiche Literatur, die es zu diesem Thema gibt, zeigt. Diese Arbeit beschäftigt sich mit einem kleinen Aspekt davon, nämlich der Identifizierbarkeit von GARCH(p, q)–Prozessen. Es wird das Problem behandelt, unter welchen Bedingungen es möglich ist, aus bestimmten Informationen, die aus den zugrunde liegenden Daten gewonnen werden, eindeutig die Parameter des Modells zu bestimmen.

Im nächsten Kapitel werden Grundlagen der Wahrscheinlichkeitstheorie und der Zeitreihenanalyse wiederholt, die in weiterer Folge verwendet werden. Es wird auch ein kurzer Überblick über die Darstellung von stationären Prozessen im Frequenzbereich und über die lineare Transformation von stationären Prozessen gebracht. Das Kapitel 3 widmet sich Finanzdaten. Es wird gezeigt mit welchen Finanzzeitreihen gearbeitet wird, und welche Eigenschaften diese haben, die, wie schon bemerkt wurde, sich durch GARCH–Prozesse recht gut modellieren lassen. Im vierten Kapitel werden zuerst die bedingt heteroskedastischen Zeitreihen, zu denen die (G)ARCH–Prozesse gehören, allgemein vorgestellt, ehe dann die GARCH–Prozesse eingeführt werden. In Unterkapiteln wird gezeigt, wann stationäre GARCH–Prozesse existieren, beziehungsweise wird die Existenz von höheren Momenten von GARCH–Prozessen untersucht.

Da in Kapitel 6 gezeigt wird, dass sich quadrierte GARCH(p, q)–Prozesse als ARMA($\max(p, q), p$)–Prozesse darstellen lassen, und die Identifizierbarkeit der Parameter des GARCH–Prozesses über die Identifizierbarkeit des ARMA–Prozesses erfolgt, werden im vorherigen Kapitel 5 ARMA–Prozesse allgemein vorgestellt, sowie deren Stationaritätseigenschaften aufgezeigt, das Spektrum eines ARMA–Prozesses wird hergeleitet und die Identifizierbarkeit von skalaren ARMA–Prozessen wird behandelt. Im sechsten Kapitel wird außerdem gezeigt, unter welchen Bedingungen der quadrierte GARCH(p, q)–Prozess schwach stationär ist, und im Zuge dessen wird eine Darstellung für das vierte Moment des GARCH(p, q)–Prozesses hergeleitet. Das letzte Kapitel beinhaltet schließlich die Identifizierbarkeit von GARCH–Prozessen mit den mathematischen Werkzeugen, die in den vorherigen Kapiteln vorgestellt wurden.

2 Grundlagen der Wahrscheinlichkeitstheorie und der Zeitreihenanalyse

Dieses Kapitel enthält einen Überblick über allgemeine mathematische und wahrscheinlichkeitstheoretische Definitionen, Sätze und Bemerkungen, die im weiteren Verlauf verwendet werden. Es ist in drei Teile gegliedert. Im ersten Teil sind Grundlagen aus der Wahrscheinlichkeitstheorie dargestellt, der zweite Teil behandelt Stochastische Prozesse und Grundlagen der Zeitreihenanalyse. Im dritten Teil wird die Spektraldarstellung von stationären Prozessen durchgemacht. Dieses Kapitel soll nur dem leichteren Verständnis der folgenden Kapitel dienen und hat keinen Anspruch auf Vollständigkeit.

2.1 Wahrscheinlichkeitstheorie

In der Wahrscheinlichkeitstheorie geht man von einem *Stichprobenraum* Ω aus, der alle möglichen *Elementarereignisse* eines Zufallsexperiments enthält. Verwendet wird nicht der Stichprobenraum Ω alleine, sondern eine gewisse Klasse von Teilmengen von Ω , eine sogenannte σ -Algebra über Ω .

Definition 2.1.1 (σ -Algebra über Ω) Ein System \mathcal{A} von Teilmengen einer Menge Ω heißt σ -Algebra (über Ω), wenn es folgende Eigenschaften besitzt:

1. $\Omega \in \mathcal{A}$
2. $A \in \mathcal{A} \Rightarrow \bar{A} \in \mathcal{A}$
3. \forall Folgen $(A_n)_{n \in \mathbb{N}}$ von Mengen aus \mathcal{A} liegt $\bigcup_{n=1}^{\infty} A_n$ in \mathcal{A}

Auf der σ -Algebra \mathcal{A} wird ein *Wahrscheinlichkeitsmaß* P definiert.

Definition 2.1.2 (Wahrscheinlichkeitsmaß) Ein Wahrscheinlichkeitsmaß P auf einer σ -Algebra über Ω ist eine Mengenfunktion $P : \mathcal{A} \rightarrow [0, 1]$, die den folgenden drei Axiomen genügt:

1. $P(A) \geq 0 \quad \forall A \in \mathcal{A} \quad$ (Nichtnegativität)
2. Seien $A_i \in \mathcal{A} \quad \forall i = 1, 2, \dots$ mit $A_i \cap A_j = \emptyset$ für $i \neq j$
 $\Rightarrow P(\bigcup_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} P(A_i) \quad$ (σ -Additivität)

3. $P(\Omega) = 1$ (Normierung)

Jede Menge $A \in \mathcal{A}$ wird Ereignis genannt.

- Durch das Wahrscheinlichkeitsmaß P wird jedem Ereignis $A \in \mathcal{A}$ eine *Wahrscheinlichkeit* $P(A)$ zugeordnet, die der Funktionswert von P in $[0, 1]$ ist.
- Die Mengen in \mathcal{A} heißen *messbar*, und sie allein besitzen Wahrscheinlichkeiten.

Definition 2.1.3 (Wahrscheinlichkeitsraum) Das Tripel (Ω, \mathcal{A}, P) wird Wahrscheinlichkeitsraum genannt.

Definition 2.1.4 (Zufallsvariable) (Ω, \mathcal{A}, P) sei ein Wahrscheinlichkeitsraum. Eine eindeutige, endliche Funktion $\xi : \Omega \rightarrow \mathbb{R}$, die $(\mathcal{A}-)$ messbar ist, nennt man eine eindimensionale (reelle) Zufallsvariable. Eine Funktion $\xi : \Omega \rightarrow \mathbb{R}$ heißt $(\mathcal{A}-)$ messbar, wenn $A_x = \{\omega \mid \xi(\omega) \leq x\} \in \mathcal{A}, \forall x \in \mathbb{R}$. Eine n -dimensionale Zufallsvariable ist ein n -dimensionaler Vektor ξ , dessen Komponenten eindimensionale Zufallsvariable sind.

Durch jede Zufallsvariable ξ wird ein neuer Wahrscheinlichkeitsraum $(\mathbb{R}, \mathcal{B}, P^*)$ erzeugt. \mathcal{B} ist dabei die σ -Algebra der Borelmengen von \mathbb{R} .¹ P^* ergibt sich aus P durch

$$P^*(B) = P(\xi^{-1}(B)) \quad \forall B \in \mathcal{B}$$

wobei $\xi^{-1}(B) = \{\omega \in \Omega \mid \xi(\omega) \in B\}$.

Es muss $\xi^{-1}(B) \in \mathcal{A}$ gelten, denn nur für Teilmengen von \mathcal{A} sind Wahrscheinlichkeiten definiert.

Definition 2.1.5 (Wahrscheinlichkeitsverteilung) P^* wird die Wahrscheinlichkeitsverteilung von ξ genannt.

Die Wahrscheinlichkeitsverteilung von ξ wird eindeutig durch die Verteilungsfunktion F_ξ beschrieben.

Definition 2.1.6 (Verteilungsfunktion) Sei ξ eine n -dimensionale Zufallsvariable. Die Funktion $F_\xi : \mathbb{R}^n \rightarrow [0, 1]$ mit $F_\xi(\mathbf{x}) = P(\xi_1 \leq x_1, \xi_2 \leq x_2, \dots, \xi_n \leq x_n)$ heißt die Verteilungsfunktion der Zufallsvariablen ξ .

¹Die σ -Algebra der Borelmengen von \mathbb{R} erhält man, indem man von den geschlossenen Intervallen $[a, b] \subset \mathbb{R}$ ausgeht, und dann alle Mengen dazu gibt, die man für eine σ -Algebra braucht. Das heißt, wegen der Eigenschaften einer σ -Algebra, dass alle Vereinigungen von geschlossenen Intervallen enthalten sind ebenso wie alle offenen Intervalle, denn jedes offene Intervall ist die Vereinigung einer Folge von geschlossenen Intervallen. Außerdem ist dann jede offene Menge als die Vereinigung einer Folge von offenen Intervallen enthalten, genauso wie jede geschlossene Menge als Komplement einer offenen Menge eine Borelmenge ist. Mit anderen Worten ist \mathcal{B} die kleinste σ -Algebra auf \mathbb{R} , die alle geschlossenen Intervalle $[a, b]$ enthält.

Satz 2.1.1 Sei $F_\xi : \mathbb{R}^n \rightarrow [0, 1]$ die Verteilungsfunktion einer Zufallsvariablen ξ . Dann hat F_ξ folgende Eigenschaften:

1. $F(-\infty, x_2, \dots, x_n) = F(x_1, -\infty, x_3, \dots, x_n) = \dots = F(x_1, \dots, x_{n-1}, -\infty) = 0$

$$F(+\infty, +\infty, \dots, +\infty) = 1$$

2. $\forall \mathbf{x} \in \mathbb{R}^n$ und $\forall \mathbf{y} < \mathbf{x}^2$ gilt:

$$\begin{aligned} P(y_1 < \xi_1 \leq x_1, \dots, y_n < \xi_n \leq x_n) &= \\ &= F_\xi(x_1, \dots, x_n) - [F_\xi(x_1, \dots, x_{n-1}, y_n) + \dots + F_\xi(y_1, x_2, \dots, x_n)] + \\ &+ [F_\xi(x_1, \dots, x_{n-2}, y_{n-1}, y_n) + \dots + F_\xi(y_1, y_2, x_3, \dots, x_n)] + \dots + \\ &+ (-1)^{n-1} [F_\xi(x_1, y_2, \dots, y_n) + \dots + F_\xi(y_1, y_2, \dots, y_{n-1}, x_n)] + \\ &+ (-1)^n F_\xi(y_1, \dots, y_n) \geq 0 \end{aligned}$$

3. $\lim_{h \downarrow 0} F_\xi(x_1, \dots, x_i + h, \dots, x_n) = F_\xi(\mathbf{x}) \quad \forall i = 1, \dots, n$ und $\forall \mathbf{x} \in \mathbb{R}^n$

Umgekehrt ist jede Funktion $F(\mathbf{x})$ mit diesen Eigenschaften die Verteilungsfunktion einer Zufallsvariablen ξ .

- Für $n = 1$ wird die Eigenschaft 2 zu

$$F_\xi(x) \geq F_\xi(y) \quad \text{bei } x > y$$

(Monotonie).

- Die dritte Eigenschaft zeigt die komponentenweise rechtsseitige Stetigkeit einer Verteilungsfunktion.

Man unterscheidet zwischen *diskreten* und *stetigen* Zufallsvariablen, die in ihrer reinen Form oder gemischt miteinander auftreten. Im Folgenden werden nur stetige Zufallsvariable betrachtet.

Definition 2.1.7 (stetige Zufallsvariable) Eine Zufallsvariable ξ heißt stetig, wenn eine integrierbare Funktion $f_\xi(\mathbf{u}) = f_\xi(u_1, \dots, u_n)$ existiert, für die

$$F_\xi(\mathbf{x}) = \int_{-\infty}^{x_1} \dots \int_{-\infty}^{x_n} f_\xi(u_1, \dots, u_n) du_n \dots du_1 \quad \forall \mathbf{x} \in \mathbb{R}^n$$

gilt. f_ξ heißt Dichte(funktion) von ξ .

² $\mathbf{y} < \mathbf{x} \Leftrightarrow y_i < x_i \quad \forall i = 1, \dots, n$

Satz 2.1.2 Eine Dichte $f_\xi : \mathbb{R}^n \rightarrow \mathbb{R}^n$ hat folgende Eigenschaften:

1. $f_\xi(\mathbf{u}) \geq 0$ ω -fast überall ³
2. $\int_{-\infty}^{+\infty} \cdots \int_{-\infty}^{+\infty} f_\xi(u_1, \dots, u_n) du_n \cdots du_1 = 1$

Umgekehrt stellt jede nichtnegative, integrierbare Funktion f_ξ , deren Integral über \mathbb{R}^n gleich 1 ist, die Dichte einer Zufallsvariablen ξ dar.

Wenn eine Dichte $f_\xi(\mathbf{u})$ existiert, dann ist die Verteilungsfunktion $F_\xi(\mathbf{x})$ ω -fast überall in \mathbb{R}^n differenzierbar, und man kann

$$f_\xi(\mathbf{u}) = \frac{\partial^n F_\xi(\mathbf{x})}{\partial x_1, \dots, \partial x_n} \Big|_{\mathbf{x}=\mathbf{u}}$$

setzen.

Wenn ξ ein n -dimensionaler Zufallsvektor ist, dann ist jede Komponente $\xi_j, j = 1, \dots, n$, oder jeder s -dimensionale Teilvektor $\xi' = (\xi_1, \dots, \xi_s), 1 \leq s < n$, von ξ wieder eine Zufallsvariable. Die Verteilungen von ξ_j oder ξ' heißen *Randverteilungen* von ξ , und man kann sie aus der Verteilung von ξ folgendermaßen herleiten:

$$F_{\xi'}(\mathbf{x}) = F_\xi(x_1, \dots, x_s, +\infty, \dots, +\infty)$$

$F_{\xi'}$ wird *Randverteilungsfunktion* von ξ' genannt.

Bei stetigen Zufallsvariablen erhält man die *Randdichtefunktion* von ξ' durch

$$f_{\xi'}(u_1, \dots, u_s) = \int_{-\infty}^{+\infty} \cdots \int_{-\infty}^{+\infty} f_\xi(\mathbf{u}) du_{s+1} \cdots du_n$$

Definition 2.1.8 (Erwartungswert einer eindimensionalen Zufallsvariablen ξ)

Sei ξ eine eindimensionale Zufallsvariable auf (Ω, \mathcal{A}, P) . Der Erwartungswert der Zufallsvariablen ξ ist definiert durch

$$E \xi = \int_{\Omega} \xi(\omega) dP(\omega) \text{ } ^4$$

³ „ ω -fast überall“ bedeutet, es gilt auf ganz \mathbb{R}^n , bis auf eine Menge vom Lebesgue-Maß 0.

Das Lebesgue-Maß ω , das auf \mathbb{R} definiert ist, ordnet jeder Borelmenge $B \in \mathcal{B}$ einen Wert aus dem Intervall $[0, \infty)$ oder ∞ zu, sodass

a) $\omega([a, b]) = b - a \quad a \leq b$

b) für eine Folge von disjunkten Mengen $B_1, B_2, B_3, \dots \in \mathcal{B}$ gilt

$$\omega\left(\bigcup_{n=1}^{\infty} B_n\right) = \sum_{n=1}^{\infty} \omega(B_n)$$

⁴ $\int_{\Omega} \xi(\omega) dP(\omega)$ ist das Lebesgue-Integral.

$E\xi$ existiert genau dann, wenn die Zufallsvariable ξ integrierbar ist, das heißt, wenn

$$E|\xi| = \int_{\Omega} |\xi(\omega)| dP(\omega) < \infty$$

gilt.

Definition 2.1.9 (σ -Algebra erzeugt durch ξ) Sei ξ eine Zufallsvariable, die auf dem Ereignisraum $\Omega \neq \emptyset$ definiert ist. Die Menge aller Teilmengen $\{\xi \in B\} = \{\omega | \xi(\omega) \in B\}$ von Ω , $B \in \mathcal{B}$, ist die durch ξ erzeugte σ -Algebra $\sigma(\xi)$.

Definition 2.1.10 (\mathcal{F} -messbar) Sei ξ eine Zufallsvariable, die auf $\Omega \neq \emptyset$ definiert ist. \mathcal{F} sei eine σ -Algebra von Teilmengen von Ω . Wenn jede Menge aus $\sigma(\xi)$ auch in \mathcal{F} enthalten ist, dann heißt die Zufallsvariable ξ \mathcal{F} -messbar.

Eine Zufallsvariable ξ ist genau dann \mathcal{F} -messbar, wenn die Information in \mathcal{F} hinreichend ist, um den Wert von ξ zu bestimmen.

Definition 2.1.11 (unabhängige σ -Algebren, unabhängige Zufallsvariable) Sei (Ω, \mathcal{A}, P) ein Wahrscheinlichkeitsraum, und \mathcal{F} und \mathcal{G} seien σ -Algebren, die in \mathcal{A} enthalten sind. \mathcal{F} und \mathcal{G} sind genau dann unabhängig, wenn

$$P(A \cap B) = P(A) \cdot P(B) \quad \forall A \in \mathcal{F}, \forall B \in \mathcal{B}$$

ξ und η seien Zufallsvariable auf (Ω, \mathcal{A}, P) . ξ und η sind genau dann unabhängig, wenn die beiden von ihnen erzeugten σ -Algebren $\sigma(\xi)$ und $\sigma(\eta)$ unabhängig sind.

Definition 2.1.12 (Varianz, Kovarianz, Korrelationskoeffizient) Sei ξ eine Zufallsvariable, deren Erwartungswert definiert ist.

Die Varianz von ξ ist durch

$$\text{Var}(\xi) = E[(\xi - E\xi)^2]$$

definiert.

Die Standardabweichung von ξ ist $\sqrt{\text{Var}(\xi)}$.

Sei η eine Zufallsvariable und seien $E\xi$, $\text{Var}(\xi)$, $E\eta$ und $\text{Var}(\eta)$ endlich. Die Kovarianz von ξ und η ist durch

$$\text{Cov}(\xi, \eta) = E[(\xi - E\xi)(\eta - E\eta)]$$

definiert.

Der Korrelationskoeffizient von ξ und η ist durch

$$\rho(\xi, \eta) = \frac{\text{Cov}(\xi, \eta)}{\sqrt{\text{Var}(\xi)\text{Var}(\eta)}}$$

gegeben. Wenn $\rho(\xi, \eta) = 0$ ist, dann sind ξ und η unkorreliert.

Bemerkung: Zwei unabhängige Zufallsvariablen sind auch unkorreliert, die Umkehrung gilt nicht.

Abschließend wird noch der bedingte Erwartungswert einer Zufallsvariablen definiert.

Definition 2.1.13 (bedingter Erwartungswert von ξ gegeben \mathcal{F}) Sei (Ω, \mathcal{A}, P) ein Wahrscheinlichkeitsraum, sei $\mathcal{F} \subset \mathcal{A}$ eine σ -Algebra, und die Zufallsvariable ξ sei nichtnegativ oder integrierbar.

Der bedingte Erwartungswert von ξ gegeben \mathcal{F} , bezeichnet mit $E(\xi|\mathcal{F})$, ist eine Zufallsvariable, die folgende Bedingungen erfüllt:

1. $E(\xi|\mathcal{F})$ ist \mathcal{F} -messbar
2. $\int_A E(\xi|\mathcal{F})(\omega) dP(\omega) = \int_A \xi(\omega) dP(\omega) \quad \forall A \in \mathcal{F}$

\mathcal{F} kann auch die von einer anderen Zufallsvariablen erzeugte σ -Algebra sein.

Der folgende Satz stellt einige Eigenschaften von $E(\xi|\mathcal{F})$ dar, die im weiteren Verlauf immer wieder verwendet werden.

Satz 2.1.3 (Ω, \mathcal{A}, P) sei ein Wahrscheinlichkeitsraum und $\mathcal{F} \subset \mathcal{A}$ eine σ -Algebra.

1. ξ und η seien integrierbare Zufallsvariable, und ξ sei \mathcal{F} -messbar. Dann gilt

$$E(\xi\eta|\mathcal{F}) = \xi E(\eta|\mathcal{F})$$

(Man kann also alles, was bekannt ist, aus dem Erwartungswert herausziehen.)

2. Sei $\mathcal{G} \subset \mathcal{F}$ eine σ -Algebra (\mathcal{G} enthält weniger Informationen als \mathcal{F}), und ξ sei integrierbar, dann gilt

$$E[E(\xi|\mathcal{F})|\mathcal{G}] = E(\xi|\mathcal{G})$$

3. ξ sei integrierbar und $\sigma(\xi)$ sei unabhängig von \mathcal{F} , dann gilt

$$E(\xi|\mathcal{F}) = E\xi$$

2.2 Stochastische Prozesse und Zeitreihen

2.2.1 Grundlegende Definitionen

Definition 2.2.1 (Stochastischer Prozess) Ein stochastischer Prozess $(X_t | t \in \mathbb{T})$ ⁵ ist eine Familie von Zufallsvariablen $X_t : \Omega \rightarrow \mathbb{R}^n, \forall t \in \mathbb{T}$, die auf einem Wahrscheinlichkeitsraum (Ω, \mathcal{A}, P) definiert sind. $\mathbb{T} \subset \mathbb{R}$ ist eine Indexmenge.

Ein stochastischer Prozess $(X_t | t \in \mathbb{T})$ kann auf zwei Arten betrachtet werden. Für jedes fixe $t \in \mathbb{T}$ ist $X_t(\omega)$ eine Funktion, die auf der Menge Ω definiert ist. Für jedes fixe $\omega \in \Omega$ ist $X_t(\omega)$ eine auf \mathbb{T} definierte Funktion, wie die folgende Definition zeigt.

Definition 2.2.2 (Trajektorie, Realisation) Die Funktionen $\{X_t(\omega) | \omega \in \Omega \text{ fix}\}$ auf \mathbb{T} heißen Trajektorien oder Realisationen des stochastischen Prozesses $(X_t | t \in \mathbb{T})$.

Die Annahme ist, dass eine Zeitreihe $X_t, t = 1, \dots, T$, durch einen zu Grunde liegenden stochastischen Prozess $(X_t | t \in \mathbb{T})$ erzeugt wird. Die Werte $X_t \in \mathbb{R}^n$ entsprechen den Realisationen $X_t(\omega)$ ($\omega \in \Omega$ fix) der Zufallsvariablen $X_t, t = 1, \dots, T$. Im Fall $n = 1$ wird die Zeitreihe *univariat* oder *skalar*, bei $n > 1$ *multivariat* genannt.

Bei einer Zeitreihe ist die Anordnung der Beobachtungen wichtig und birgt Informationen in sich. Hier wird nur der diskrete Fall $\mathbb{T} = \mathbb{Z}$ behandelt. Die Indexmenge besteht dann aus *äquidistanten* Zeitpunkten. Anstelle von $(X_t)_{t \in \mathbb{Z}}$ wird im Folgenden für eine Zeitreihe kürzer (X_t) geschrieben.

Definition 2.2.3 (Kovarianzfunktion) Sei (X_t) ein stochastischer Prozess mit $E X_t' X_t < \infty, \forall t \in \mathbb{Z}$. Dann wird die Kovarianzfunktion $\gamma_X : \mathbb{Z} \times \mathbb{Z} \rightarrow \mathbb{R}^{n \times n}$ von (X_t) folgendermaßen definiert:

$$\gamma_X(r, s) = \text{Cov}(X_r, X_s) = E \left[(X_r - E X_r)(X_s - E X_s)' \right] \quad r, s \in \mathbb{Z} \quad (2.1)$$

γ_X ist eine $n \times n$ -Matrix, deren Einträge die Funktionen $\gamma_{ij} : \mathbb{Z} \times \mathbb{Z} \rightarrow \mathbb{R}$ sind. $\gamma_{ij}(s, t) = \text{Cov}(X_s^{(i)}, X_t^{(j)}) = E(X_s^{(i)} - E X_s^{(i)})(X_t^{(j)} - E X_t^{(j)})$ beschreibt die *lineare Abhängigkeit* zwischen $X_s^{(i)}$ und $X_t^{(j)}$.

Die Bedingung $E X_t' X_t < \infty, \forall t \in \mathbb{Z}$ sichert die Existenz von $E X_t$ und von γ_X .

Wichtig in der Zeitreihenanalyse sind die *stationären* stochastischen Prozesse. Vereinfacht gesagt sind stationäre Prozesse Prozesse, deren Charakteristika sich im Laufe der Zeit nicht ändern. Es gibt *stark stationäre* und *schwach stationäre* stochastische Prozesse.

Definition 2.2.4 (Starke Stationarität) Ein stochastischer Prozess (X_t) heißt stark stationär, wenn die Wahrscheinlichkeitsverteilung von X_{t_1}, \dots, X_{t_l} und von $X_{t_1+h}, \dots, X_{t_l+h}$ für alle l, t_1, \dots, t_l, h übereinstimmt.

⁵Äquivalente Schreibweise: $(X_t)_{t \in \mathbb{T}}$

Die Eigenschaften eines stark stationären Prozesses sind also *invariant* gegenüber Verschiebungen auf der Zeitachse.

Definition 2.2.5 (Schwache Stationarität) Ein stochastischer Prozess (X_t) heißt schwach stationär, wenn die folgenden Bedingungen erfüllt sind:

1. $EX'_t X_t < \infty \quad \forall t \in \mathbb{Z}$
2. $EX_t = m = \text{konstant} \quad \forall t \in \mathbb{Z}$
3. $\gamma(s, t) = \gamma(s + r, t + r) \quad \forall r, s, t \in \mathbb{Z}$

Bei einem schwach stationären Prozess sind das erste und das zweite Moment invariant gegenüber Verschiebungen auf der Zeitachse.

Bemerkung:

Ein stark stationärer Prozess mit $EX'_t X_t < \infty, \forall t \in \mathbb{Z}$, ist auch schwach stationär. Die Umkehrung gilt nicht.

Die Kovarianzfunktion γ_X eines stationären Prozesses (X_t) hängt nur von der Zeitdifferenz zwischen den beiden Beobachtungen X_s und X_t ab, denn $\gamma_X(s, t) = \gamma_X(s - t, 0)$. $s - t = h$ wird „lag“ genannt, und die Kovarianzfunktion $\gamma_X : \mathbb{Z} \rightarrow \mathbb{R}^{n \times n}$ eines stationären Prozesses kann als Funktion von nur mehr einer Variablen h betrachtet werden:

$$h \mapsto \gamma_X(h) \equiv \gamma_X(h, 0) = \text{Cov}(X_{t+h}, X_t) \quad \forall h, t \in \mathbb{Z} \quad (2.2)$$

$\gamma_{ii} : \mathbb{Z} \rightarrow \mathbb{R}$ wird *Autokovarianzfunktion* des (eindimensionalen) Prozesses $(X_t^{(i)})$, der aus den i . Komponenten besteht, genannt. $\gamma_{ij} : \mathbb{Z} \rightarrow \mathbb{R}, i \neq j$, ist die *Kreuzkovarianzfunktion* zwischen dem i . Komponentenprozess $(X_t^{(i)})$ und dem j . Komponentenprozess $(X_t^{(j)})$.

Wenn (X_t) ein stationärer Prozess ist, dann ist γ_X eine gerade Funktion, denn es gilt:

$$\gamma(-h) = \text{Cov}(X_{t-h}, X_t) = \text{Cov}(X_t, X_{t+h}) = \gamma(h) \quad (2.3)$$

2.2.2 Einige Beispiele für stationäre Prozesse

Im folgenden kurzen Abschnitt werden jene (schwach) stationären Prozesse dargestellt, die im weiteren Verlauf der Arbeit verwendet werden.

Weißes Rauschen:

Ein stochastischer Prozess (ε_t) heißt *Weißes Rauschen*, wenn

$$E \varepsilon_t = 0 \quad \forall t \in \mathbb{Z}$$

und

$$E \varepsilon_s \varepsilon_t^* = \delta_{st} \Sigma$$

erfüllt ist⁶. Dabei ist $\Sigma \geq 0$, das heißt, Σ muss *positiv semidefinit*⁷ sein.

Weißes Rauschen hat kein lineares Gedächtnis und konstante Varianz. Weißes Rauschen ist der elementarste stationäre Prozess, der als Baustein für andere stochastische Prozesse verwendet wird.

ARMA-Prozesse:

Eine lineare Differenzgleichung der Form

$$X_t + a_1 X_{t-1} + \dots + a_p X_{t-p} = b_0 \varepsilon_t + b_1 \varepsilon_{t-1} + \dots + b_q \varepsilon_{t-q} \quad (2.4)$$

wobei $a_i \in \mathbb{R}^{n \times n}$, $\forall i = 1, \dots, p$, $b_j \in \mathbb{R}^{n \times n}$, $\forall j = 1, \dots, q$, (ε_t) ist Weißes Rauschen, heißt ein *ARMA(p,q)-System*.

Eine (schwach) stationäre Lösung von 2.4 wird *ARMA(p,q)-Prozess* genannt.

ARMA(p,q)-Prozesse werden in dieser Arbeit noch öfter verwendet, und ihre (Stationaritäts-)Eigenschaften werden in Kapitel 5 noch genauer beschrieben.

Harmonische Prozesse:

Harmonische Prozesse sind von der Form

$$X_t = \sum_{j=1}^h e^{i\lambda_j t} Z_j \quad \forall t \in \mathbb{Z} \quad (2.5)$$

wobei $Z_j : \Omega \rightarrow \mathbb{C}^n$, $j = 1, \dots, h-1$, komplexwertige, n -dimensionale Zufallsvariable mit der Eigenschaft $E Z_j^* Z_j < \infty$, $j = 1, \dots, h$, sind.

O.B.d.A. wird $\lambda_j < \lambda_{j+1}$, $j = 1, \dots, h-1$, angenommen.

Die λ_j werden *Winkelfrequenzen* genannt, und jeder Periodenlänge T_j wird eine Winkel-
frequenz λ_j durch $\lambda_j = \frac{2\pi}{T_j}$ zugeordnet.

Die Winkelfrequenzen λ_j können auf das Intervall $(-\pi, \pi]$ beschränkt werden, denn jedes beliebige $\lambda_j \in \mathbb{R}$ kann durch $\lambda_j = 2k\pi + \tilde{\lambda}_j$, $k \in \mathbb{Z}$, $\tilde{\lambda}_j \in (-\pi, \pi]$, dargestellt werden. In 2.5 gilt dann, dass $e^{i\lambda_j t} = e^{i(2k\pi + \tilde{\lambda}_j)t} = e^{2k\pi i} e^{i\tilde{\lambda}_j t}$. Unter Berücksichtigung, dass $e^{2k\pi i} = 1$ ist, folgt dann $e^{i\lambda_j t} = e^{i\tilde{\lambda}_j t}$, $\forall t \in \mathbb{Z}$. Da nur diskrete Zeitreihen betrachtet werden, können größere Winkelfrequenzen nicht erkannt werden. Die höchste Frequenz, die in diesem diskreten Zeitraster auftreten kann, ist π , und π wird *Nyquist-Frequenz* genannt.

⁶Kronecker Symbol $\delta_{st} = \begin{cases} 1 & \text{für } s = t \\ 0 & \text{sonst} \end{cases}$

⁷Eine symmetrische Matrix \mathbf{A} heißt *positiv semidefinit*, wenn $\mathbf{x}' \mathbf{A} \mathbf{x} \geq 0$, $\forall \mathbf{x}$, und $\mathbf{x}' \mathbf{A} \mathbf{x} = 0$ für mindestens ein $\mathbf{x} \neq 0$ ist.

Der Ausdruck $e^{i\lambda_j t}$ kann durch $\cos \lambda_j t + i \sin \lambda_j t$ ersetzt werden.

Die l -te Komponente der komplexen Zufallsvariablen Z_j lässt sich durch $|Z_j^{(l)}| e^{i\phi_j^{(l)}}$ darstellen, dabei sind $|Z_j^{(l)}|$ und $\phi_j^{(l)}$ Zufallsvariable.

Mit Hilfe dieser beiden Darstellungen erkennt man, dass ein harmonischer Prozess (X_t) eine Linearkombination von harmonischen Schwingungen mit zufälligen Amplituden $|Z_j^{(l)}|$ und Phasen $\phi_j^{(l)}$ ist.

Damit ein harmonischer Prozess (schwach) stationär ist, müssen drei Bedingungen erfüllt sein:

1. $E Z_j^* Z_j < \infty \quad j = 1, \dots, h$

2. $E Z_j = \begin{cases} 0 & \lambda_j \neq 0 \\ m & \lambda_j = 0 \end{cases}$

Die Zufallsvariablen Z_j , die die Winkelfrequenzen $\lambda_j \neq 0$ gewichten, müssen $E Z_j = 0$ haben.

3. $E Z_j Z_l^* = 0 \quad j \neq l$

Die Gewichte Z_j und Z_l , die zu unterschiedlichen, von 0 verschiedenen Winkelfrequenzen λ_j und λ_l gehören, dürfen keine linearen Abhängigkeiten aufweisen.

Die Herleitung der Bedingungen findet sich zum Beispiel im Skriptum [3].

2.3 Stationäre Prozesse im Frequenzbereich

Definition 2.3.1 (Zeitbereich H des Prozesses (X_t)) Sei $(X_t|t \in \mathbb{Z})$ ein (schwach) stationärer Prozess, dann heißt der von der Menge $\left\{X_t^{(i)} : \Omega \rightarrow \mathbb{C} \mid i = 1, \dots, n, t \in \mathbb{Z}\right\}$ in L_2 erzeugte Hilbertraum der Zeitbereich H des Prozesses (X_t) .

Die Darstellung von (schwach) stationären Prozessen, wie sie im Abschnitt 2.2 beschrieben wurde, wird die Darstellung von stationären Prozessen *im Zeitbereich* genannt.

(Schwach) stationäre Prozesse haben die Eigenschaft, dass sie sich noch auf eine bestimmte andere Art darstellen lassen, die viele Berechnungen und Anwendungen vereinfacht, unter anderem auch das Problem der Identifizierbarkeit. Deshalb enthält der folgende Abschnitt eine kurze Beschreibung der Darstellung von stationären Prozessen *im Frequenzbereich*, die sich aber hauptsächlich auf die für die Identifizierbarkeit wichtigen Aspekte beschränkt. Für eine detaillierte Ausführung dieses Themas und deren Beweise sei auf Brockwell und Davis [2] verwiesen.

Satz 2.3.1 (Spektraldarstellung stationärer Prozesse) Zu jedem (schwach) stationären Prozess (X_t) existiert genau ein zugeordneter Prozess $(z(\lambda)|\lambda \in [-\pi, \pi])$ mit den Eigenschaften

1. $z(-\pi) = 0 \quad z(\pi) = X_0$
2. $\lim_{\varepsilon \downarrow 0} z(\lambda + \varepsilon) = z(\lambda) \quad \forall \lambda \in [-\pi, \pi]$
3. Für $\lambda_1 < \lambda_2 \leq \lambda_3 < \lambda_4$ gilt $E(z(\lambda_2) - z(\lambda_1))(z(\lambda_4) - z(\lambda_3))^* = 0$

sodass gilt

$$X_t = \int_{-\pi}^{\pi} e^{i\lambda t} dz(\lambda) \quad \text{fast sicher}^8 \quad (2.6)$$

$(z(\lambda)|\lambda \in [-\pi, \pi])$ ist ein stochastischer Prozess mit der Indexmenge $[-\pi, \pi]$. Diese Indexmenge besteht nicht aus (diskreten) Zeitpunkten, sondern aus Winkelfrequenzen. $(z(\lambda)|\lambda \in [-\pi, \pi])$ mit den Eigenschaften 1), 2) und 3) wird *Prozess mit orthogonalen Zuwächsen* genannt, und er ist *eindeutig* durch $(X_t|t \in \mathbb{Z})$ festgelegt. Wegen 2) ist der Prozess rechtsstetig, und Eigenschaft 3) bedeutet, dass die Zuwächse des Prozesses über disjunkte Intervalle unabhängig sind.

Das stochastische Integral ist folgendermaßen definiert:

$$\int_{-\pi}^{\pi} e^{i\lambda t} dz(\lambda) = \lim_{n \rightarrow \infty} \sum_{j=1}^n e^{i\lambda_j^n t} (z(\lambda_j^n) - z(\lambda_{j-1}^n)) \quad (2.7)$$

$-\pi = \lambda_0^n < \lambda_1^n < \dots < \lambda_{n-1}^n < \lambda_n^n = \pi$ ist eine Zerlegung des Intervalls $[-\pi, \pi]$.

⁸Diese Darstellung entspricht der Fouriertransformation von (schwach) stationären Prozessen.

Wenn der Grenzwert im quadratischen Mittel für alle Zerlegungsfolgen von $[-\pi, \pi]$ für $n \rightarrow \infty$ existiert, dann ist das Integral gleich diesem Grenzwert und dadurch definiert. $n \rightarrow \infty$ ist gleichbedeutend damit, dass die maximalen Intervalllängen $\max_j(\lambda_j^n - \lambda_{j-1}^n)$ der Zerlegungsfolgen gegen 0 konvergieren.⁹

Aus der Darstellung 2.7 folgt

$$X_t = l.i.m._{n \rightarrow \infty} \sum_{j=1}^n e^{i\lambda_j^n t} (z(\lambda_j^n) - z(\lambda_{j-1}^n)) \quad \forall t \in \mathbb{Z}$$

woraus ersichtlich ist, dass jeder (schwach) stationäre Prozess der Grenzwert einer Folge von harmonischen Prozessen ist. Jeder (schwach) stationäre Prozess ist also durch eine Summe von harmonischen Schwingungen mit zufälligen Amplituden und Phasen approximierbar.

Definition 2.3.2 (Spektrale Verteilungsfunktion) Sei (X_t) ein (schwach) stationärer Prozess mit Spektraldarstellung $X_t = \int_{-\pi}^{\pi} e^{i\lambda t} dz(\lambda)$, dann heißt die durch $F(\lambda) = E z(\lambda) z(\lambda)^*$ definierte Funktion $F : [-\pi, \pi] \rightarrow \mathbb{C}^{n \times n}$ die spektrale Verteilungsfunktion des Prozesses (X_t) .

Nach einer einfachen Berechnung unter Einbeziehung der Eigenschaften eines Prozesses mit orthogonalen Zuwächsen erkennt man

$$F(\lambda_2) - F(\lambda_1) = E (z(\lambda_2) - z(\lambda_1))(z(\lambda_2) - z(\lambda_1))^*$$

für $\lambda_1 < \lambda_2$. Anders ausgedrückt entspricht der Zuwachs der spektralen Verteilungsfunktion über einem Intervall $[\lambda_1, \lambda_2]$ dem zentralen zweiten Moment des Zuwachses des Prozesses mit orthogonalen Zuwächsen über diesem Intervall.

Definition 2.3.3 (Spektrale Dichte) Sei $f : [-\pi, \pi] \rightarrow \mathbb{C}^{n \times n}$, sodass

$$F(\lambda) = \int_{-\pi}^{\lambda} f(\omega) d\omega \quad \omega \dots \text{Lebesgue-Ma\ss}$$

dann heißt f die spektrale Dichte von (X_t) .

Bemerkung: F ist dann absolut stetig bezüglich des Lebesgue-Maßes ω .

Ohne Beweis wird folgender Satz gebracht, der die Eigenschaften einer spektralen Verteilungsfunktion und einer spektralen Dichte beschreibt.

Satz 2.3.2 1. Eine Funktion $F : [-\pi, \pi] \rightarrow \mathbb{C}^{n \times n}$ ist eine spektrale Verteilungsfunktion genau dann, wenn gilt:

⁹Bei einem „normalen“ Riemann-Integral werden die Werte der zu integrierenden Funktion f (hier: $f(\lambda) = e^{i\lambda t}$) mit den Intervalllängen $\lambda_j - \lambda_{j-1}$ gewichtet. Hier wird die Definition des Riemann-Integrals erweitert, indem die Werte von f mit dem Zuwachs des stochastischen Prozesses über dem Intervall $[\lambda_{j-1}, \lambda_j]$ gewichtet werden.

- a) $F(-\pi) = 0 \quad F(\pi) = \gamma(0)$
- b) F ist rechtsstetig
- c) $F(\lambda_2) - F(\lambda_1) \geq 0 \quad \lambda_2 > \lambda_1$

2. $f : [-\pi, \pi] \rightarrow \mathbb{C}^{n \times n}$ ist eine spektrale Dichte genau dann, wenn gilt:

- a) $f \geq 0$ ω -fast überall
- b) $\int_{-\pi}^{\pi} f(\lambda) d\lambda = \gamma(0)$

Satz 2.3.3 (Spektraldarstellung der Kovarianzfunktion) Jede Kovarianzfunktion $\gamma : \mathbb{Z} \rightarrow \mathbb{R}^{n \times n}$ eines stationären Prozesses (X_t) besitzt die Darstellung

$$\gamma(t) = \int_{-\pi}^{\pi} e^{i\lambda t} dF(\lambda) \tag{2.8}$$

wobei F die spektrale Verteilungsfunktion von (X_t) ist.

F ist durch γ eindeutig bestimmt.

Existiert eine spektrale Dichte f , dann gilt

$$\gamma(t) = \int_{-\pi}^{\pi} e^{i\lambda t} f(\lambda) d\lambda \tag{2.9}$$

Umgekehrt kann f durch

$$f(\lambda) = \frac{1}{2\pi} \sum_{t=-\infty}^{\infty} \gamma(t) e^{-i\lambda t} \tag{2.10}$$

aus der Kovarianzfunktion γ bestimmt werden.

Bemerkung:

Hinreichende Bedingungen für die Existenz einer spektralen Dichte sind $\sum_{s \in \mathbb{Z}} \|\gamma(s)\| < \infty$ und die schwächere Bedingung $\sum_{s \in \mathbb{Z}} \|\gamma(s)\|^2 < \infty$

2.3.1 Lineare Transformation von stationären Prozessen

Die *lineare Transformation* von (schwach) stationären Prozessen wird durch die Darstellung im Frequenzbereich auch vereinfacht. Da die lineare Transformation von (schwach) stationären Prozessen auch für das Problem der Identifizierbarkeit wichtig ist, wird im Folgenden ein kurzer Überblick darüber gebracht.

Sei (X_t) ein (schwach) stationärer Prozess.

Der Ausdruck

$$Y_t = \sum_{j=-\infty}^{\infty} a_j X_{t-j} \quad a_j \in \mathbb{R}^{m \times n} \tag{2.11}$$

¹⁰Diese Darstellung entspricht der Fouriertransformation von $\gamma(t)$.

wird eine *lineare Transformation* (manchmal auch *linearer Filter*) von (X_t) genannt.

Es wird die Bedingung

$$\sum_{j=-\infty}^{\infty} \|a_j\| < \infty \quad (2.12)$$

angenommen, um die Existenz der Summe zu gewährleisten. ($\|\cdot\|$ ist eine beliebige Matrixnorm.)

$(a_j|j \in \mathbb{Z})$ wird *Gewichtsfunktion* genannt.

Durch Nachrechnen der Bedingungen für (schwache) Stationarität ist einfach zu sehen, dass auch der Prozess $\begin{pmatrix} X_t \\ Y_t \end{pmatrix}$ (schwach) stationär ist, wenn (X_t) (schwach) stationär ist, und Bedingung 2.12 gilt.

Wenn man die Spektraldarstellung des (schwach) stationären Prozesses (X_t) in die Gleichung 2.11 einsetzt, erhält man

$$\begin{aligned} Y_t &= \sum_{j=-\infty}^{\infty} a_j \int_{-\pi}^{\pi} e^{i\lambda(t-j)} dz_X(\lambda) = \\ &= \int_{-\pi}^{\pi} \left(\sum_{j=-\infty}^{\infty} a_j e^{-i\lambda j} \right) e^{i\lambda t} dz_X(\lambda) \quad \forall t \in \mathbb{Z} \end{aligned}$$

Summe und Integral dürfen vertauscht werden, da das Integral linear und stetig bezüglich der Konvergenz im quadratischen Mittel ist. $z_X(\lambda)$ ist der zu (X_t) gehörige Prozess mit orthogonalen Zuwächsen.

Man definiert die sogenannte *Transferfunktion* $k : [-\pi, \pi] \rightarrow \mathbb{C}^{m \times n}$ durch

$$k(\lambda) = \sum_{j=-\infty}^{\infty} a_j e^{-i\lambda j} \quad (2.13)$$

Durch die Bedingung 2.12 existiert auch $k(\lambda)$, und $k(\lambda)$ ist im Sinne der gewöhnlichen Konvergenz (punktweise) definiert.

Daraus folgt:

$$Y_t = \int_{-\pi}^{\pi} e^{i\lambda t} k(\lambda) dz_X(\lambda) = \int_{-\pi}^{\pi} e^{i\lambda t} dz_Y(\lambda) \quad \forall t \in \mathbb{Z}$$

Zwischen der Gewichtsfunktion $(a_j|j \in \mathbb{Z})$ und der Transferfunktion $k(\lambda)$ besteht ein bijektiver Zusammenhang. $k(\lambda)$ kann durch 2.13 bestimmt werden, umgekehrt gilt

$$a_j = \frac{1}{2\pi} \int_{-\pi}^{\pi} e^{i\lambda j} k(\lambda) d\lambda \quad (2.14)$$

(ohne Beweis)

Satz 2.3.4 (X_t) sei (schwach) stationär mit spektraler Dichte f_X . $Y_t = \sum_{j=-\infty}^{\infty} a_j X_{t-j}$ sei die lineare Transformation von (X_t) . Dann existieren auch die spektrale Dichte f_Y von (Y_t) und das sogenannte Kreuz-Spektrum f_{YX} zwischen (X_t) und (Y_t) , und sie sind gegeben durch

$$f_Y(\lambda) = k(\lambda)f_X(\lambda)k(\lambda)^* \quad (2.15)$$

$$f_{YX}(\lambda) = k(\lambda)f_X(\lambda) \quad (2.16)$$

Bemerkung:

$\begin{pmatrix} X_t \\ Y_t \end{pmatrix}$ ist, wie vorher bemerkt wurde, (schwach) stationär mit spektraler Dichte

$$f_{\begin{pmatrix} X \\ Y \end{pmatrix}} = \begin{pmatrix} f_X & f_{XY} \\ f_{YX} & f_Y \end{pmatrix}$$

Beweis:

$$E Y_t Y_0^* = \gamma_Y(t) = \int_{-\pi}^{\pi} e^{i\lambda t} f_Y(\lambda) d\lambda$$

$$\begin{aligned} E Y_t Y_0^* &= E \left[\left(\sum_{j=-\infty}^{\infty} a_j X_{t-j} \right) \left(\sum_{k=-\infty}^{\infty} a_k X_{-k} \right)^* \right] = \\ &= E \left[\sum_{j=-\infty}^{\infty} \sum_{k=-\infty}^{\infty} a_j X_{t-j} X_{j+k}^* a_{-(j+k)}^* \right] = \\ &= \sum_{j=-\infty}^{\infty} \sum_{k=-\infty}^{\infty} a_j E(X_{t-j} X_{j+k}^*) a_{-(j+k)}^* = \\ &= \sum_{j=-\infty}^{\infty} \sum_{k=-\infty}^{\infty} a_j \gamma_X(t - 2j - k) a_{-(j+k)}^* = \\ &= \sum_{j=-\infty}^{\infty} \sum_{k=-\infty}^{\infty} a_j \int_{-\pi}^{\pi} e^{i\lambda(t-2j-k)} f_X(\lambda) d\lambda a_{-(j+k)}^* = \\ &= \int_{-\pi}^{\pi} \underbrace{\sum_{j=-\infty}^{\infty} a_j e^{-i\lambda j}}_{=k(\lambda)} e^{i\lambda t} f_X(\lambda) \underbrace{\sum_{k=-\infty}^{\infty} e^{-i\lambda(j+k)} a_{-(j+k)}^*}_{=\sum_{k=-\infty}^{\infty} e^{-i\lambda k} a_{-k}^* = k(\lambda)^*} d\lambda = \\ &= \int_{-\pi}^{\pi} k(\lambda) e^{i\lambda t} f_X(\lambda) k(\lambda)^* d\lambda \\ &\Rightarrow \int_{-\pi}^{\pi} e^{i\lambda t} f_Y(\lambda) d\lambda = \int_{-\pi}^{\pi} k(\lambda) e^{i\lambda t} f_X(\lambda) k(\lambda)^* d\lambda \quad \forall t \in \mathbb{Z} \quad (2.17) \end{aligned}$$

Diese Gleichung gilt $\forall t \in \mathbb{Z}$ und die Menge der Funktionen $(e^{i \cdot t})$ formt eine orthogonale Basis im Raum aller quadratisch integrierbaren Funktionen über $[-\pi, \pi]$. Deshalb kann aus 2.17

$$f_Y(\lambda) = k(\lambda)f_X(\lambda)k(\lambda)^*$$

gefolgert werden.

Der Beweis von 2.16 erfolgt analog. \square

3 Finanzzeitreihen

Bei der Untersuchung von Finanzzeitreihen werden Daten analysiert, die auf Finanzmärkten auftreten, wie z.B. der Preis der Aktie einer bestimmten Firma, der Wert eines Aktienindizes, der Umrechnungskurs einer Fremdwährung, der Goldpreis, der Ölpreis und vieles mehr. Finanzzeitreihen sind also Zeitreihen, die die Preise P_t einer Finanzanlage zu einem bestimmten Zeitpunkt t enthalten. Diese Zeitpunkte sind äquidistant, die Zeiträume zwischen den Beobachtungszeitpunkten sind also gleich lang. Allgemein wird aber nicht mit diesen rohen Finanzdaten P_t gearbeitet, sondern es werden die sogenannten „Returns“ oder „log – Returns“ verwendet.

Die Gewinnrate einer Anlage ist durch

$$X_{t+1} = \frac{P_{t+1} - P_t}{P_t} \quad (3.1)$$

gegeben. P_t und P_{t+1} bezeichnen die Preise der Anlage zu den Zeitpunkten t und $t + 1$. X_{t+1} ist also der Gewinn der Anlage pro Einheit des eingesetzten Kapitals und wird (relativer) *Return* genannt. Bei dieser Formel wird angenommen, dass keine Zahlungen zwischen t und $t + 1$, wie Dividendenzahlungen oder ähnliches, geleistet werden. Returns sind im Gegensatz zu absoluten Preisen unabhängig von der Einheit, in der sie beobachtet werden, und dadurch leichter zu vergleichen. Weitere Überlegungen zeigen, dass bei „hinreichend kleinen“ Returns X_{t+1} gilt:¹

$$\ln P_{t+1} - \ln P_t = \ln \frac{P_{t+1}}{P_t} = \ln (1 + X_{t+1}) \approx \ln u|_{u=1} + \frac{X_{t+1}}{1!} \frac{d \ln u}{du} |_{u=1} = X_{t+1} \quad (3.2)$$

Damit sieht man, dass kleine Returns X_{t+1} ungefähr mit der Differenz der logarithmierten Preise übereinstimmen. Die Differenz der logarithmierten Preise $\ln P_{t+1} - \ln P_t$ wird *log – Return* genannt. Log – Returns werden den relativen Returns vorgezogen, da diese bei der Betrachtung über mehrere Perioden additiv sind. Im Folgenden wird nicht mehr zwischen Returns und log – Returns unterschieden.

Eine grundlegende Annahme ist, dass Finanzzeitreihen, also Zeitreihen die Returns enthalten, (stark) stationär sind, zumindest über passende, nicht zu lange Zeiträume. Dies ist auch der Grund dafür, dass Returns anstelle von „rohen“ Finanzdaten, wie z.B. den Aktienpreisen, bei den Untersuchungen verwendet werden. Denn Zeitreihen, die die reinen Preise von Finanzanlagen beinhalten, sind schon offensichtlich nicht stationär. Die Annahme der Stationarität von Finanzzeitreihen ist diskutierbar, dafür sei aber auf Straumann [7] verwiesen. Im weiteren Verlauf wird die Stationarität von Finanzzeitreihen vorausgesetzt.

¹Taylorreihenentwicklung von $f(u) = \ln u$ an der Stelle $u = 1$ (abgebrochen nach dem Term, der die erste Ableitung enthält)

3.1 „Stylized Facts“ von Finanzzeitreihen

Empirische Forschung hat gezeigt, dass Finanzdaten einige besondere charakteristische Merkmale aufweisen, die als *Stylized Facts* bekannt sind.

3.1.1 Returns (X_t) sind unkorreliert, ($|X_t|$) und (X_t^2) sind stark autokorreliert

Return – Daten zeigen oft das Verhalten, dass die Beobachtungen untereinander unkorreliert sind. Mathematisch ausgedrückt bedeutet dies, dass die Werte der Autokorrelationsfunktion für alle Lags $h \neq 0$ vernachlässigbar klein sind, bis vielleicht auf $h = 1$. Dies weist darauf hin, dass Finanzzeitreihen eventuell gut durch Weises Rauschen modelliert werden könnten.

Im Gegensatz dazu weisen Zeitreihen, die die betragsmäßig genommenen Returns ($|X_t|$) oder die quadrierten Returns (X_t^2) beinhalten, ein völlig anderes Verhalten auf. Deren Autokorrelationsfunktion ist für eine große Anzahl von Lags h von 0 verschieden. Sie konvergiert für große h nur sehr langsam gegen 0. Also sind die Prozesse ($|X_t|$) und (X_t^2) sehr stark autokorreliert. Sie besitzen ein langes Gedächtnis, weit zurückliegende Werte haben noch immer einen gewissen Einfluss auf die aktuellen Werte.

3.1.2 Die bedingte Volatilität der Returns ist nicht konstant

Es wird beobachtet, dass Returns, die einen bestimmten Wert haben, von Returns umgeben sind, die einen vergleichbaren Wert haben. Anders ausgedrückt, wenn ein unüblich kleiner oder großer Return auftritt, treten in der nachfolgenden Periode ähnlich kleine oder große Returns auf. Zeiträume mit großen Returns wechseln sich mit Zeiträumen mit kleinen Returns ab. Diese Eigenschaft von Finanzdaten wird *Volatility Clustering* genannt.

Mathematisch bedeutet das, dass die Varianzen der Returns unter der Bedingung, dass die vergangenen Returns gegeben sind, nicht konstant über den betrachteten Zeitraum sind, und dass die zugrunde liegende Zeitreihe *bedingt heteroskedastisch* ist. Die bedingte Volatilität $\sigma_t = (\text{Var}[X_t | X_{t-1}, X_{t-2}, \dots])^{\frac{1}{2}}$ ändert sich also im Laufe der Zeit.

3.1.3 Die Verteilung von Returns ist leptokurtisch

Bei Finanzzeitreihen treten sehr große und sehr kleine Werte öfter auf als bei Daten, denen eine Normalverteilung zu Grunde liegt. Returns haben also eine Verteilung, die „dickere Schwänze“ als die Normalverteilung hat. Die Wahrscheinlichkeiten $P(X_t \leq -x)$ und $P(X_t \geq x)$ für große Werte von x , also die Schwänze der Verteilung, sind größer als die analogen Wahrscheinlichkeiten einer normalverteilten Zufallsvariablen.

Diese Eigenschaft bedeutet, dass größere Verluste, aber auch größere Gewinne vorkommen können, als wenn den Daten eine Normalverteilung zu Grunde liegen würde.

3.1.4 Leverage Effekte

Kurz erwähnt wird folgende Eigenschaft der Volatilitäten von Returns, die empirisch nicht vorliegt, sondern die nur auf Grund von theoretischen Überlegungen zugrunde gelegt wird. Es wird angenommen, dass die Volatilitäten asymmetrisch auf positive oder negative Returns reagieren. Das heißt, dass ein negativer Return eine größere Volatilität aufweist, als ein betragsmäßig gleich großer positiver Return. Der Grund für dieses Verhalten wird *Leverage Effekt* genannt.

4 Bedingt heteroskedastische Zeitreihen

Ein Ziel in der Analyse von Finanzzeitreihen ist es, die im vorigen Kapitel beschriebenen Eigenschaften von Finanzdaten durch ein mathematisches Modell zu erklären und zu beschreiben. Als Standardmodelle gelten die „*bedingt heteroskedastischen Zeitreihen*“. Diese Arbeit untersucht nur univariate Prozesse.

Allgemeine bedingt heteroskedastische Prozesse haben die Form

$$X_t = \mu_t + \sigma_t \varepsilon_t \quad t \in \mathbb{Z} \quad (4.1)$$

wobei $(\varepsilon_t) \sim IID(0, 1)$ ¹ beziehungsweise oft auch $(\varepsilon_t) \sim NID(0, 1)$ ².

(μ_t) und (σ_t) sind stochastische Prozesse bei denen vorausgesetzt wird, dass sie nur von der Vergangenheit abhängen. Sie sind also Funktionen der Vergangenheit des Prozesses (X_t) . Das bedeutet, dass die Zufallsvariablen μ_t und σ_t messbar bezüglich der σ -Algebra

$$\mathcal{F}_{t-1} = \sigma(\varepsilon_j | j \leq t-1) \quad (4.2)$$

sind.

Die Menge \mathcal{F}_t kann als die Information über den Prozess, die zum Zeitpunkt t verfügbar ist, interpretiert werden.

Daraus folgt, dass X_t \mathcal{F}_t -messbar ist, $\forall t \in \mathbb{Z}$, was als Analogie zur Kausalität von ARMA-Prozessen betrachtet werden kann.

Zu bemerken ist, dass

$$\mu_t = E(X_t | \mathcal{F}_{t-1})$$

μ_t ist also der Erwartungswert von X_t unter der Bedingung, dass die Vergangenheit des Prozesses X_t bis zum Zeitpunkt $t-1$ bekannt ist.

Ebenso ist leicht zu sehen, dass

$$Var(X_t | \mathcal{F}_{t-1}) = \sigma_t^2$$

¹ (ε_t) ist eine Folge von identisch verteilten und (stochastisch) unabhängigen Zufallsvariablen ε_t , mit $E \varepsilon_t = 0$ und $Var \varepsilon_t = E \varepsilon_t^2 = 1 \quad \forall t \in \mathbb{Z}$
(*IID*: independent and identically distributed)

²unabhängig und identisch standardnormalverteilt

σ_t^2 ist also die bedingte Varianz von X_t gegeben die Information \mathcal{F}_{t-1} , und $|\sigma_t|$ kann als die Volatilität zum Zeitpunkt t interpretiert werden.

Die Prozesse (μ_t) und (σ_t) sind nicht direkt beobachtbar.

Bedingte Heteroskedastizität tritt auf, wenn (σ_t) ein nicht-trivialer Prozess ist. Daraus leitet sich auch der Name für diese Prozesse ab. σ_t^2 ist die bedingte Varianz der Prozessvariablen X_t zum Zeitpunkt t unter der Bedingung, dass die Vergangenheit bis zum Zeitpunkt $t - 1$ bekannt ist, und σ_t^2 ist variabel, also heteroskedastisch.

Unter der Berücksichtigung, dass (μ_t) und (σ_t) \mathcal{F}_{t-1} -messbar sind und dass $(\varepsilon_t) \sim IID(0, 1)$, berechnet man

$$\text{Var}(X_t) = E\mu_t^2 + E\sigma_t^2 - (E\mu_t)^2$$

Unter der Annahme $\mu_t = \mu = \textit{konstant}$ wird daraus

$$\text{Var}(X_t) = E\sigma_t^2 = E(\text{Var}(X_t | \mathcal{F}_{t-1}))$$

Daher gilt nur unter der Voraussetzung $\mu_t = \mu = \textit{konstant}$, dass die unbedingte Varianz von X_t gleich dem unbedingten Erwartungswert der durch \mathcal{F}_{t-1} bedingten Varianz von X_t ist.

Im Folgenden werden nur Prozesse betrachtet, für die $\mu_t = 0$ ist. Prozesse, deren durch die Information zum Zeitpunkt t bedingter Erwartungswert zum Zeitpunkt t null ist, werden *Innovationen* genannt.

Es gilt also

$$X_t = \sigma_t \varepsilon_t \quad \forall t \in \mathbb{Z}$$

Zusammenfassend gilt für die betrachteten bedingt heteroskedastischen Prozesse:

- $E(X_t | \mathcal{F}_{t-1}) = 0$
- $E X_t = E(E(X_t | \mathcal{F}_{t-1})) = 0$
- $\text{Var} X_t = E(\text{Var}(X_t | \mathcal{F}_{t-1})) = E\sigma_t^2$
- $\forall k \neq 0$ gilt: $\gamma_X(k) = \text{Cov}(X_{t-k}, X_t) = E(E(\sigma_{t-k} \varepsilon_{t-k} \sigma_t \varepsilon_t | \mathcal{F}_{t-1})) =$
 $= E(\sigma_{t-k} \varepsilon_{t-k} \sigma_t E(\varepsilon_t | \mathcal{F}_{t-1})) = 0$

Das heißt, wenn eine stationäre Lösung existiert, dann ist diese Lösung ein Weißes Rauschen.

4.1 GARCH(p, q)–Prozesse

Die sogenannten GARCH(p, q)–Prozesse gehören zu den bedingt heteroskedastischen Prozessen. Diese haben sich als recht effizient erwiesen und werden in der Praxis häufig verwendet. GARCH(p, q)–Prozesse und ihre Eigenschaften werden im folgenden Abschnitt dargestellt.

Die Lösung (X_t) der Gleichungen

$$X_t = \sigma_t \varepsilon_t \quad \forall t \in \mathbb{Z} \quad (4.3)$$

$$\sigma_t^2 = \alpha_0 + \sum_{i=1}^q \alpha_i X_{t-i}^2 + \sum_{j=1}^p \beta_j \sigma_{t-j}^2 \quad \forall t \in \mathbb{Z}^3 \quad (4.4)$$

wobei $(\varepsilon_t) \sim IID(0, 1)$

$$q \in \mathbb{N} = \{1, 2, \dots\}$$

$$p \in \mathbb{N}_0 = \{0, 1, 2, \dots\}$$

$$\alpha_0 > 0$$

$$\alpha_1, \dots, \alpha_{q-1} \geq 0 \quad \alpha_q > 0$$

$$\beta_1, \dots, \beta_{p-1} \geq 0 \quad \beta_p > 0$$

wird ein GARCH(p, q)(*generalized autoregressive conditionally heteroscedastic*)–Prozess mit dem Volatilitätsprozess (σ_t) genannt.

Der Volatilitätsprozess (σ_t) ist nichtnegativ.

GARCH($0, q$)–Prozesse werden ARCH(q)(*autoregressive conditionally heteroscedastic*)–Prozesse genannt.

Die Klasse der ARCH(q)–Prozesse wurde 1982 von Robert F. Engle erstmals vorgestellt. Dabei wird der Unterschied zwischen der unbedingten und der bedingten Varianz des Prozesses hervorgehoben, wobei die bedingte Varianz σ_t^2 als Funktion von vergangenen Werten sich im Laufe der Zeit ändert, während die unbedingte Varianz gleichbleibt. 1986 wurden die ARCH(q)–Modelle von Tim Bollerslev zu GARCH(p, q)–Modellen erweitert, der die Varianz nicht nur von vergangenen Werten des Prozesses, sondern zusätzlich noch von vergangenen Varianz–Werten abhängig macht.

³Anmerkung: Die Verwendung der Parameter p und q wurde [1] angepasst. In manchen anderen Artikeln sind p und q vertauscht.

GARCH(p, q)-Prozesse gehören, wie schon festgestellt wurde, zur Klasse der bedingt heteroskedastischen Prozesse. Deshalb wird vorausgesetzt, dass σ_t nur von der Vergangenheit abhängt. Das heißt, σ_t soll messbar bezüglich der σ -Algebra $\mathcal{F}_{t-1} = \sigma(\varepsilon_j \mid j \leq t-1)$ sein. Ein GARCH(p, q)-Prozess mit dieser Eigenschaft wird *kausal* genannt. Der Prozess (X_t) ist dann \mathcal{F}_t -messbar.

Aus der Voraussetzung, dass (ε_t) eine Folge von unabhängigen Zufallsvariablen ist und dass σ_t \mathcal{F}_{t-1} -messbar ist, folgt die stochastische Unabhängigkeit von σ_t und ε_t , $t \in \mathbb{Z}$. Es gilt: σ_t ist unabhängig von $(\varepsilon_{t+h})_{h \in \mathbb{N}_0}$, und X_t ist unabhängig von $(\varepsilon_{t+h})_{h \in \mathbb{N}}$.

Die Bedingungen, dass alle Koeffizienten $\alpha_1, \dots, \alpha_q, \beta_1, \dots, \beta_p$ nicht-negativ sind, sichert, dass auch σ_t^2 nichtnegativ ist. Daher kann die Volatilität σ_t als die positive Quadratwurzel von σ_t^2 definiert werden.

4.1.1 Existenz von stationären Lösungen

Bollerslev hat in seinem Artikel von 1982 [1], in dem er den GARCH(p, q)-Prozess vorstellte, auch eine Bedingung für die Existenz von schwach stationären Lösungen angegeben. Bedingungen für die Existenz von stark stationären Lösungen und eine Darstellung dieser Lösung wurden aber erst ein paar Jahre später zuerst von Nelson für den GARCH(1, 1)-Prozess und dann von Bougerol und Picard für den allgemeinen Fall aufgezeigt. Der nächste Abschnitt, in dem die Stationaritätseigenschaften behandelt werden, folgt der Darstellung im Kapitel 3 von Straumann [7] und dem Artikel von Lindner [6].

Wenn man von einem „stationären GARCH(p, q)-Prozess“ spricht, meint man, dass der bivariate Prozess (X_t, σ_t) stationär ist.

Schwache Stationarität

Zu Beginn dieses Kapitels wurde bereits $E X_t = 0$, $\forall t \in \mathbb{Z}$, und $Cov(X_t, X_s) = 0$, $\forall t, s \in \mathbb{Z}$, gezeigt. Damit der GARCH(p, q)-Prozess (X_t) schwach stationär ist, muss noch die Bedingung

$$E X_t^2 = Var X_t = E \sigma_t^2 < \infty \quad \forall t \in \mathbb{Z}$$

erfüllt sein.

Satz 4.1.1 (Bollerslev [1]) *Ein GARCH(p, q)-Prozess ist genau dann schwach stationär mit $E X_t = 0$, $Var X_t = E \sigma_t^2 = \frac{\alpha_0}{1 - \sum_{i=1}^q \alpha_i - \sum_{j=1}^p \beta_j}$ und $Cov(X_t, X_s) = 0$ für $t \neq s$, wenn $\sum_{i=1}^q \alpha_i + \sum_{j=1}^p \beta_j < 1$.*

Beweis: Sei $m := \max(p, q)$.

$$\sigma_t^2 = \alpha_0 + \sum_{i=1}^q \alpha_i \sigma_{t-i}^2 \varepsilon_{t-i}^2 + \sum_{j=1}^p \beta_j \sigma_{t-j}^2 =$$

$$= \alpha_0 + \sum_{i_1=1}^m (\alpha_{i_1} \varepsilon_{t-i_1}^2 + \beta_{i_1}) \sigma_{t-i_1}^2$$

$$\text{wobei } \begin{cases} \alpha_{q+1}, \dots, \alpha_m = 0 & \text{falls } q < p \\ \beta_{p+1}, \dots, \beta_m = 0 & \text{falls } q > p \end{cases}$$

$$\begin{aligned} \sigma_t^2 &= \alpha_0 + \sum_{i_1=1}^m (\alpha_{i_1} \varepsilon_{t-i_1}^2 + \beta_{i_1}) [\alpha_0 + \sum_{i_2=1}^m (\alpha_{i_2} \varepsilon_{t-i_1-i_2}^2 + \beta_{i_2})] \sigma_{t-i_1-i_2}^2 = \dots = \\ &= \alpha_0 \left[1 + \sum_{i_1=1}^m (\alpha_{i_1} \varepsilon_{t-i_1}^2 + \beta_{i_1}) + \dots + \sum_{i_1=1}^m (\alpha_{i_1} \varepsilon_{t-i_1}^2 + \beta_{i_1}) \cdots \sum_{i_{k-1}=1}^m (\alpha_{i_{k-1}} \varepsilon_{t-i_1-\dots-i_{k-1}}^2 + \beta_{i_{k-1}}) \right] + \\ &\quad + \sum_{i_1=1}^m (\alpha_{i_1} \varepsilon_{t-i_1}^2 + \beta_{i_1}) \cdots \sum_{i_k=1}^m (\alpha_{i_k} \varepsilon_{t-i_1-\dots-i_k}^2 + \beta_{i_k}) \sigma_{t-i_1-\dots-i_k}^2 = \dots = \\ &= \alpha_0 \sum_{k=0}^{\infty} M(t, k) \end{aligned}$$

mit $M(t, 0) = 1$

$$M(t, 1) = \sum_{i=1}^m (\alpha_i \varepsilon_{t-i}^2 + \beta_i)$$

\vdots

$$M(t, k+1) = \sum_{i=1}^m (\alpha_i \varepsilon_{t-i}^2 + \beta_i) M(t-i, k)$$

$M(t, k) \geq 0$ fast sicher, $M(t, k)$ ist unabhängig von ε_t und $\forall t, s, k$ gilt $E M(t, k) = E M(s, k)$.

$$\begin{aligned} \Rightarrow E M(t, k+1) &= \sum_{i=1}^m E[(\alpha_i \varepsilon_{t-i}^2 + \beta_i) M(t-i, k)] = \\ &= \sum_{i=1}^m (\alpha_i \underbrace{E \varepsilon_{t-i}^2}_{=1} + \beta_i) E M(t-i, k) = \\ &= \sum_{i=1}^m (\alpha_i + \beta_i) E M(t, k) = \dots = \\ &= (\sum_{i=1}^m (\alpha_i + \beta_i))^{k+1} \underbrace{E M(t, 0)}_{=1} = \\ &= (\sum_{i=1}^m (\alpha_i + \beta_i))^{k+1} \\ \Rightarrow E \sigma_t^2 &= \alpha_0 \sum_{k=0}^{\infty} E M(t, k) = \alpha_0 \sum_{k=0}^{\infty} (\sum_{i=1}^m (\alpha_i + \beta_i))^k \end{aligned}$$

Diese unendliche Reihe konvergiert genau dann, wenn

$$\sum_{i=1}^m (\alpha_i + \beta_i) < 1$$

Das heißt, $E(X_t^2) < \infty$ gilt genau dann, wenn die Bedingung

$$\sum_{i=1}^q \alpha_i + \sum_{j=1}^p \beta_j < 1 \tag{4.5}$$

erfüllt ist.

Daraus kann gefolgert werden, dass unter der Bedingung 4.5 eine Lösung von 4.3 und 4.4 schwach stationär ist.

Starke Stationarität

Um die Existenz einer stark stationären Lösung von 4.3 und 4.4 zu untersuchen, wird dieses Problem zuerst in einer anderen Form dargestellt.

Fall $p = q = 1$:

$$\sigma_t^2 = \alpha_0 + \alpha_1 X_{t-1}^2 + \beta_1 \sigma_{t-1}^2 = \alpha_0 + \alpha_1 \sigma_{t-1}^2 \varepsilon_{t-1}^2 + \beta_1 \sigma_{t-1}^2 = \alpha_0 + (\alpha_1 \varepsilon_{t-1}^2 + \beta_1) \sigma_{t-1}^2$$

Man setzt: $A_t = \alpha_1 \varepsilon_t^2 + \beta_1$

$$B_t = \alpha_0$$

$$Y_t = \sigma_{t+1}^2$$

Daraus folgt, dass $(Y_t) = (\sigma_{t+1}^2)$ die Lösung der rekursiven stochastischen Gleichung

$$Y_t = A_t Y_{t-1} + B \quad t \in \mathbb{Z} \quad (4.6)$$

ist. Dabei ist $(A_t) \sim IID$.

Gleichung 4.6 ist die *Zustandsraumdarstellung* des GARCH(1, 1)–Prozesses.

Damit wird das Anfangsproblem, Bedingungen für die starke Stationarität und eine stark stationäre Lösung von 4.3 und 4.4 zu finden, in das äquivalente Problem transformiert, eine stark stationäre Lösung (Y_t) von 4.6 und Bedingungen für deren Existenz zu erhalten.

Der folgende Satz, der von Nelson stammt, löst dieses Problem:

Satz 4.1.2 (Nelson) *Ein GARCH(1, 1)–Prozess mit $\alpha_0, \alpha_1, \beta_1 \geq 0$ hat genau dann eine stark stationäre Lösung, wenn*

$$E |\log A_0| = E |\log(\beta_1 + \alpha_1 \varepsilon_0^2)| < \infty$$

und

$$E(\log A_0) = E [\log(\beta_1 + \alpha_1 \varepsilon_0^2)] < 0 \quad (4.7)$$

Diese Lösung ist eindeutig, und ihre quadrierte Volatilität ist gegeben durch

$$\sigma_t^2 = \alpha_0 + \alpha_0 \sum_{j=1}^{\infty} \left(\prod_{i=1}^j A_{t-i} \right) = \alpha_0 \left[1 + \sum_{j=1}^{\infty} \prod_{i=1}^j (\beta_1 + \alpha_1 \varepsilon_{t-i}^2) \right] \quad t \in \mathbb{Z} \quad (4.8)$$

Dieser Satz wird ohne Beweis angeführt. Darstellungen des Beweises finden sich in Straumann [7] und Lindner [6].

Die Bedingung $E[\log(\beta_1 + \alpha_1 \varepsilon_0^2)] < 0$ ist notwendig und hinreichend für die Existenz von stark stationären GARCH(1, 1)–Prozessen.

Die Anwendung der Jensen’schen Ungleichung⁴ auf die Bedingung 4.7 liefert

$$E[\log(\beta_1 + \alpha_1 \varepsilon_0^2)] \leq \log E(\beta_1 + \alpha_1 \varepsilon_0^2) = \log(\beta_1 + \alpha_1 E\varepsilon_0^2)$$

Da dieser Ausdruck kleiner 0 sein muss, ist $\beta_1 + \alpha_1 E\varepsilon_0^2 < 1$ eine hinreichende Bedingung für die Existenz einer stark stationären Lösung.

Im Fall $(\varepsilon_t) \sim IID(0, 1)$ wird daraus $\alpha_1 + \beta_1 < 1$. Zu dieser hinreichenden Bedingung für die Existenz einer stark stationären Lösung ist anzumerken, dass es stark stationäre Lösungen mit $\alpha_1 + \beta_1 > 1$ gibt. Jedoch gilt dabei immer, dass $\beta_1 < 1$ notwendig ist, während $\alpha_1 > 1$ sein darf.

Im Fall $p, q \geq 2$ definiert man die $(p + q - 1)$ –dimensionalen Vektoren

$$\mathbf{Y}_t = (\sigma_t^2, \dots, \sigma_{t-p+1}^2, X_{t-1}^2, \dots, X_{t-q+1}^2)'$$

und

$$\mathbf{B} = (\alpha_0, 0, \dots, 0)'$$

sowie die $(p + q - 1) \times (p + q - 1)$ –dimensionale Matrix

$$\mathbf{A}_t = \begin{pmatrix} \alpha_1 \varepsilon_t^2 + \beta_1 & \beta_2 & \cdots & \beta_{p-1} & \beta_p & \alpha_2 & \cdots & \alpha_{q-1} & \alpha_q \\ 1 & 0 & \cdots & 0 & 0 & 0 & \cdots & 0 & 0 \\ 0 & 1 & \cdots & 0 & 0 & 0 & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & 1 & 0 & 0 & \cdots & 0 & 0 \\ \varepsilon_t^2 & 0 & \cdots & 0 & 0 & 0 & \cdots & 0 & 0 \\ 0 & 0 & \cdots & 0 & 0 & 1 & \cdots & 0 & 0 \\ 0 & 0 & \cdots & 0 & 0 & 0 & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & 0 & 0 & 0 & \cdots & 1 & 0 \end{pmatrix} \quad (4.10)$$

\mathbf{Y}_t und \mathbf{A}_t werden für alle $t \in \mathbb{Z}$ definiert.

Diese Matrizen stammen von Bougerol und Picard.

⁴ f sei eine konvexe, reelle Funktion definiert auf \mathbb{R} , $E|X| < \infty$

$$\Rightarrow f(EX) \leq E f(X) \quad (4.9)$$

Die Gleichungen 4.3 und 4.4 implizieren

$$\mathbf{Y}_{t+1} = \mathbf{A}_t \mathbf{Y}_t + \mathbf{B} \quad t \in \mathbb{Z} \quad (4.11)$$

wobei $(\mathbf{A}_t, \mathbf{B}) \sim IID$.

Die obige Gleichung ist eine lineare, rekursive, stochastische Gleichung im Zustandsraum \mathbb{R}^{p+q-1} und stellt die Beziehung zwischen der Gegenwart und der Vergangenheit des stochastischen Prozesses (\mathbf{Y}_t) dar. 4.11 ist die *Zustandsraumdarstellung* des GARCH(p, q)–Prozesses.

Man sieht wie im Fall $p = q = 1$, dass 4.3 und 4.4 genau dann eine eindeutige, nichtnegative und stationäre Lösung haben, wenn 4.11 eine eindeutige, nichtnegative und stationäre Lösung hat. Diese Lösung (\mathbf{Y}_t) zeigt in der ersten Koordinate den „aktuellen“ Wert der quadrierten Volatilität σ_t^2 . X_t erhält man über die Gleichung 4.3. Das ursprüngliche Problem ist also wieder dahingehend transformiert worden, eine Lösung von 4.11 zu finden.

Eine vollständige Antwort auf dieses Problem wurde von Bougerol und Picard (1992) gegeben. Ihre Resultate werden hier ohne Beweise gebracht. Eine Darstellung der Beweise findet man in Straumann [7]. In Lindner [6] werden kurze Beweisskizzen gebracht.

Die Existenz einer stark stationären Lösung einer multivariaten, rekursiven, stochastischen Gleichung wird in Abhängigkeit vom sogenannten *Top–Lyapunov–Exponenten* ausgedrückt.

Satz 4.1.3 (Bougerol und Picard) *Sei $\alpha_0 > 0$ und seien $\alpha_i \geq 0$, $i = 1, \dots, q$ und $\beta_j \geq 0$, $j = 1, \dots, p$. Zu den Gleichungen 4.3 und 4.4 existiert genau dann eine eindeutige, stark stationäre Lösung $((X_t, \sigma_t))$, wenn folgende Bedingung gilt:*

$$\rho = \inf_{t \in \mathbb{N}} \left\{ \frac{1}{t+1} E(\log \|\mathbf{A}_0 \cdots \mathbf{A}_{-t}\|) \right\} < 0 \quad (4.12)$$

ρ ist der Top–Lyapunov–Exponent von (\mathbf{A}_t) .

Die zugehörige Lösung (\mathbf{Y}_t) der Zustandsraumgleichung 4.11 lässt sich folgendermaßen darstellen:

$$\mathbf{Y}_t = \mathbf{B} + \sum_{j=1}^{\infty} \left(\prod_{i=1}^j \mathbf{A}_{t-i} \right) \mathbf{B} \quad \omega - \text{fast sicher} \quad (4.13)$$

Dabei ist $\|\cdot\|$ eine Matrizennorm, die durch

$$\|\mathbf{A}\| := \sup_{\mathbf{x} \neq \mathbf{0}} \frac{\|\mathbf{A}\mathbf{x}\|}{\|\mathbf{x}\|}$$

definiert ist. $\mathbf{A} \in \mathbb{R}^{d \times d}$, $\mathbf{x} \in \mathbb{R}^d$ und $\|\cdot\|$ ist eine beliebige Norm in \mathbb{R}^d , da in einem endlich–dimensionalen Vektorraum alle Normen äquivalent sind. ρ ist unabhängig von

der gewählten Vektornorm und Matrizenorm.

Zu bemerken ist, dass \mathbf{Y}_t \mathcal{F}_{t-1} -messbar, $\forall t \in \mathbb{Z}$, ist. Jeder (stark) stationäre GARCH(p, q)-Prozess ist also kausal, das heißt X_t ist \mathcal{F}_t -messbar, $\forall t \in \mathbb{Z}$. σ_t ist \mathcal{F}_{t-1} -messbar und daher unabhängig von ε_t , $t \in \mathbb{Z}$ fix.

Es ist oft nicht möglich, trotz gegebener Koeffizienten α_i und β_j und gegebener Verteilung von ε_t , den Top-Lyapunov-Exponenten ρ für eine Folge von Matrizen zu berechnen. Deshalb muss auf numerische Simulationen zurückgegriffen werden. Es gibt aber spezielle Fälle, in denen es einfacher zu bestimmen ist, ob der Top-Lyapunov-Exponent $\rho < 0$ oder $\rho > 0$ ist.

Satz 4.1.4 *Folgende Bedingungen gelten für den Top-Lyapunov-Exponenten ρ von (\mathbf{A}_t) :*

1. $E(\log \|\mathbf{A}_0 \cdot \mathbf{A}_{-r+1}\|) < 0$ für ein $r \geq 1 \Rightarrow \rho < 0$
2. $\sum_{i=1}^q \alpha_i E(\varepsilon_0^2) + \sum_{j=1}^p \beta_j < 1 \Rightarrow \rho < 0$
3. Seien $\alpha_i, \beta_j > 0, \forall i, j$, und für die Verteilung von ε_0 soll $P(|\varepsilon_0| = 0) = 0$ und $P(|\varepsilon_0| \leq z) < 1, \forall z \geq 0$, gelten. Aus der Bedingung $\sum_{i=1}^q \alpha_i E(\varepsilon_0^2) + \sum_{j=1}^p \beta_j = 1$ folgt $\rho < 0$.⁵
4. $\sum_{j=1}^p \beta_j \geq 1 \Rightarrow \rho > 0$

Wenn also die Summe der Koeffizienten β_j größer oder gleich 1 ist, dann existiert keine stark stationäre Lösung.

Kurz zusammenfassend kann man sagen, dass unter der Voraussetzung $(\varepsilon_t) \sim IID(0, 1)$ die Bedingung $\sum_{i=1}^q \alpha_i + \sum_{j=1}^p \beta_j < 1$ *hinreichend* und dass die Bedingung $\sum_{j=1}^p \beta_j < 1$ *notwendig* für die Existenz einer stark stationären Lösung ist.

4.1.2 Existenz von Momenten höherer Ordnung

Für die Identifizierbarkeit von GARCH(p, q)-Prozessen ist in weiterer Folge speziell die Existenz des vierten Moments der stationären Lösung eines GARCH(p, q)-Systems wichtig. Eine Möglichkeit zur Berechnung des vierten Moments wird in Kapitel 6 gezeigt. Im folgenden Abschnitt werden in knapper Form Bedingungen für die Existenz von Momenten höherer Ordnung von GARCH(p, q)-Prozessen dargestellt.

Sei $P(\varepsilon_0 = 0) < 1$. X_t und σ_t sind stochastisch unabhängig für stationäre, kausale Lösungen; es gilt also $E(X_t^m) = E(\sigma_t^m) E|\varepsilon_t|^m$. Das heißt, das m . Moment von $X_t = \sigma_t \varepsilon_t$ existiert genau dann, wenn $E(\sigma_t^m) < \infty$ und $E|\varepsilon_t|^m < \infty$. Es genügt, sich auf Bedingungen für die Existenz von Momenten höherer Ordnung von σ_t zu beschränken.

⁵GARCH(p, q)-Prozesse mit der Eigenschaft $\sum_{i=1}^q \alpha_i + \sum_{j=1}^p \beta_j = 1$ heißen *integrierte* GARCH(p, q)-Prozesse oder IGARCH(p, q)-Prozesse. Diese Bedingung zeigt, unter welchen Annahmen IGARCH(p, q)-Prozesse eine stark stationäre Lösung haben.

Meist ist die Verteilung von ε_t symmetrisch, wie zum Beispiel im Fall der Normalverteilung. Im Folgenden wird eine symmetrische Verteilung von ε_t vorausgesetzt. Dann sind die ungeraden Momente von ε_t und daher auch von X_t null.

Satz 4.1.5 (Bollerslev [1]; He & Teräsvirta) Sei (X_t, σ_t) ein stark stationärer GARCH(1, 1)-Prozess mit

$$\begin{aligned} X_t &= \sigma_t \varepsilon_t \\ \sigma_t^2 &= \alpha_0 + \alpha_1 X_{t-1}^2 + \beta_1 \sigma_{t-1}^2 \end{aligned}$$

wobei $\varepsilon_t \sim IID(0, 1)$.

Sei $m \in \mathbb{N}$.

Das m . Moment $E(\sigma_t^{2m})$ von σ_t^2 existiert dann und nur dann, wenn $E(\beta_1 + \alpha_1 \varepsilon_0^2)^m < 1$.

$$\mu_j := E(\sigma_t^{2j})$$

μ_m kann rekursiv durch

$$\mu_m = (1 - E(\beta_1 + \alpha_1 \varepsilon_0^2)^m)^{-1} \sum_{j=0}^{m-1} \binom{m}{j} \alpha_0^{m-j} E(\beta_1 + \alpha_1 \varepsilon_0^2)^j \mu_j \quad (4.14)$$

berechnet werden.

Das $2m$. Moment von X_t ist durch

$$E(X_t^{2m}) = \mu_m E(\varepsilon_0^{2m}) \quad (4.15)$$

gegeben.

Die Bedingung für die Existenz von μ_m wird ausgehend von der Gleichung 4.8 mit Hilfe der Minkowski-Ungleichung gezeigt. Mit dem Binomialsatz kann

$$\sigma_t^{2m} = \sum_{k=0}^{m-1} \binom{m}{k} \alpha_0^{m-k} \sigma_{t-1}^{2k} (\beta_1 + \alpha_1 \varepsilon_{t-1}^2)^k + \sigma_{t-1}^{2m} (\beta_1 + \alpha_1 \varepsilon_{t-1}^2)^m$$

einfach nachgerechnet werden. Daraus und mit den Eigenschaften der Stationarität kann die Rekursionsformel für μ_m gezeigt werden. Ein genauer Beweis dieses Satzes wird hier nicht gebracht.

Beispiele:

1. $(\varepsilon_t) \sim IID(0, 1)$

$$m = 1:$$

$\mu_1 = E(\sigma_t^2)$ existiert, wenn

$$E(\beta_1 + \alpha_1 \varepsilon_0^2) = \alpha_1 + \beta_1 < 1$$

gilt.

$$\mu_1 = E(\sigma_t^2) = \frac{\alpha_0}{1 - \alpha_1 - \beta_1}$$

$m = 2$:

$\mu_2 = E(\sigma_t^4)$ existiert, wenn

$$E(\beta_1 + \alpha_1 \varepsilon_0^2)^2 = \beta_1^2 + 2\alpha_1\beta_1 + \alpha_1^2 E(\varepsilon_0^4) < 1$$

gilt.

$$\mu_2 = E(\sigma_t^4) = \frac{\alpha_0^2}{1 - \alpha_1^2 E(\varepsilon_0^4) - 2\alpha_1\beta_1 - \beta_1^2} \cdot \frac{1 + \alpha_1 + \beta_1}{1 - \alpha_1 - \beta_1}$$

2. $(\varepsilon_t) \sim NID(0, 1)$ ⁶

$E(\sigma_t^4)$ existiert, wenn

$$\beta_1^2 + 2\alpha_1\beta_1 + 3\alpha_1 < 1$$

gilt.

$$E(\sigma_t^4) = \frac{\alpha_0^2}{1 - 3\alpha_1^2 - 2\alpha_1\beta_1 - \beta_1^2} \cdot \frac{1 + \alpha_1 + \beta_1}{1 - \alpha_1 - \beta_1}$$

Ling und McAleer haben in einem Satz das Problem der Existenz von höheren Momenten von GARCH(p, q)-Prozessen, $p, q \geq 2$, gelöst. Die Bedingung, die sie dafür angeben, wird mit Hilfe des *Kronecker-Produkts* von Matrizen formuliert.

Definition 4.1.1 (Kronecker-Produkt) $C = (c_{ij})_{i=1, \dots, m, j=1, \dots, n}$ sei eine $(m \times n)$ -Matrix, D sei eine $(p \times r)$ -Matrix. Dann ist das Kronecker-Produkt $C \otimes D$ von den zwei Matrizen C und D eine $(mp \times nr)$ -Matrix, die durch

$$C \otimes D = \begin{pmatrix} c_{11}D & \cdots & c_{1n}D \\ \vdots & \ddots & \vdots \\ c_{m1}D & \cdots & c_{mn}D \end{pmatrix}$$

definiert ist.

Satz 4.1.6 (Ling & McAleer) (X_t, σ_t) sei ein stark stationärer GARCH(p, q)-Prozess mit $\alpha_1 + \beta_1 > 0$. Sei die $(p+q-1) \times (p+q-1)$ -Matrix \mathbf{A}_t definiert wie in 4.10. $m \in \mathbb{N}$. Dann ist das m . Moment von σ_t^2 genau dann endlich, wenn für den Spektralradius ρ der Matrix $E(\mathbf{A}_t^{\otimes m})$ ⁷ gilt:

$$\rho(E(\mathbf{A}_t^{\otimes m})) < 1 \tag{4.16}$$

⁶ $X \sim N(0, 1) \Rightarrow E(X^{2j}) = \prod_{i=1}^j (2i-1)$

⁷Der Spektralradius einer $(n \times n)$ -Matrix $\mathbf{A} \in \mathbb{C}^{n \times n}$ ist gleich dem Wert des betragsmäßig größten Eigenwerts von \mathbf{A} . Wenn $\lambda_1, \dots, \lambda_n$ die Eigenwerte der Matrix \mathbf{A} sind, dann gilt:

$$\rho(\mathbf{A}) = \max_{1 \leq i \leq n} |\lambda_i|$$

Eine Beweisskizze für diesen Satz findet sich in Lindner [6]. Hier wird der Satz ohne Beweis gebracht.

Oft ist es sehr aufwendig oder überhaupt nur numerisch möglich, die Bedingung $\rho(E(\mathbf{A}_t^{\otimes m})) < 1$ zu überprüfen. Einige Wissenschaftler haben eine einfachere, hinreichende Bedingung für die Existenz des m . Moments von σ_t^2 gefunden, die im folgenden Satz (ohne Beweis) beschrieben wird.

Satz 4.1.7 (Giraitis und andere) Sei (X_t, σ_t) ein stark stationärer GARCH(p, q)-Prozess. Sei $m \in [1, \infty)$ und $0 < E|\varepsilon_0^{2m}| < \infty$.

$$\left(\frac{\sum_{i=1}^p \alpha_i}{1 - \sum_{j=1}^q \beta_j} \right)^m E|\varepsilon_0^{2m}| < 1 \quad (4.17)$$

ist eine hinreichende Bedingung für $E(\sigma_t^{2m}) < \infty$.

Zum Abschluss dieses Kapitels soll noch bemerkt werden, dass die bedingte Heteroskedastizität, also die Variabilität von $\sigma_t^2 = \text{Var}(X_t | \mathcal{F}_{t-1})$, die höheren Momente der unbedingten Verteilung von X_t beeinflusst. Im Speziellen hat die unbedingte Verteilung von X_t „fettere Schwänze“ als eine Normalverteilung, was bei der Modellierung von Finanzdaten ja gewünscht wird (vergleiche Kapitel 3).

Ein Maß dafür, wie groß die Abweichung von der Normalverteilung ist, ist die *Kurtosis* κ , die für eine Zufallsvariable Z folgendermaßen definiert ist:

$$\kappa(Z) = \frac{E(Z - EZ)^4}{(\text{Var } Z)^2}$$

Angewendet auf bedingt heteroskedastische Prozesse mit Erwartungswert null erhält man

$$EX_t^4 = E\varepsilon_t^4 E\sigma_t^4 \geq E\varepsilon_t^4 (E\sigma_t^2)^2$$

Die \geq -Relation wird mit der Jensen'schen Ungleichung (siehe Seite 29) erklärt. Man erweitert auf der rechten Seite mit $\frac{(E\varepsilon_t^2)^2}{(E\varepsilon_t^2)^2}$ und, da $(\varepsilon_t) \sim IID(0, 1)$ ist, folgt

$$EX_t^4 \geq \kappa_{\varepsilon_0} (EX_t^2)^2$$

Daraus ergibt sich

$$\kappa_{X_t} \geq \kappa_{\varepsilon_0} \quad \forall t \in \mathbb{Z} \quad (4.18)$$

Da meist $(\varepsilon_t) \sim NID(0, 1)$ vorausgesetzt wird, bedeutet 4.18, dass die unbedingte Verteilung des GARCH(p, q)-Prozesses leptokurtisch ist, also „dickere Schwänze“ als eine Normalverteilung besitzt.

5 Skalare ARMA–Prozesse und ihre Identifizierbarkeit

Definition 5.0.2 (ARMA(p, q)–Prozess) Ein Prozess $(X_t | t \in \mathbb{Z})$ heißt ARMA(p, q) (autoregressive moving average)–Prozess, wenn (X_t) (schwach) stationär ist und $\forall t \in \mathbb{Z}$ die lineare Differenzgleichung

$$a_0 X_t + a_1 X_{t-1} + \dots + a_{p-1} X_{t-p+1} + a_p X_{t-p} = b_0 \varepsilon_t + b_1 \varepsilon_{t-1} + \dots + b_{q-1} \varepsilon_{t-q+1} + b_q \varepsilon_{t-q} \quad (5.1)$$

erfüllt. Dabei sind $a_0, a_1, \dots, a_{p-1}, a_p, b_0, b_1, \dots, b_{q-1}, b_q \in \mathbb{R}$, $a_p \neq 0$, $b_q \neq 0$, $p, q \in \mathbb{N}_0$, (ε_t) ist ein skalares Weißes Rauschen.

$(X_t | t \in \mathbb{Z})$ heißt ein ARMA(p, q)–Prozess mit Erwartungswert μ , wenn der Prozess $(X_t - \mu)$ ein ARMA(p, q)–Prozess ist.

Es wird der Verschiebeoperator (Backward–Shift–Operator) B eingeführt, der auf \mathbb{Z} durch

$$B(X_t | t \in \mathbb{Z}) = (X_{t-1} | t \in \mathbb{Z}) \quad (5.2)$$

definiert wird. Der Verschiebeoperator B ist linear und bijektiv.

Mit Hilfe des Verschiebeoperators B kann 5.1 kompakter durch

$$a(B)X_t = b(B)\varepsilon_t \quad \text{mit} \quad \begin{cases} a(B) = a_0 + a_1 B + \dots + a_{p-1} B^{p-1} + a_p B^p \\ b(B) = b_0 + b_1 B + \dots + b_{q-1} B^{q-1} + b_q B^q \end{cases} \quad (5.3)$$

dargestellt werden. ($a(B)X_t$ wird kurz für $a(B)(X_t | t \in \mathbb{Z})$ geschrieben.) $a(B)$ und $b(B)$ sind Polynome vom Grad p beziehungsweise q mit dem Verschiebeoperator B .

Eine Lösung von 5.3 ist also durch

$$X_t = a^{-1}(B) b(B) \varepsilon_t \quad (5.4)$$

gegeben, wenn das inverse Element von $a(B)$ existiert.

Zwei Fragen stellen sich jetzt. Unter welchen Bedingungen existiert $a^{-1}(B)$? Wie schaut $a^{-1}(B)$ aus?

Zuerst betrachtet man einmal nur die linke Seite von 5.1.

$$\sum_{j=0}^p a_j X_{t-j} = \sum_{j=0}^p a_j \int_{-\pi}^{\pi} e^{i\lambda(t-j)} dz_X(\lambda) = \int_{-\pi}^{\pi} \underbrace{\sum_{j=0}^p a_j e^{-i\lambda j}}_{=a(\lambda)} e^{i\lambda t} dz_X(\lambda)$$

$z_X(\lambda)$ ist der zu (X_t) gehörige Prozess mit orthogonalen Zuwächsen.

Diese Gleichung stellt eine lineare Transformation eines (schwach) stationären Prozesses mit der Transferfunktion $a(\lambda) = \sum_{j=0}^p a_j e^{-i\lambda j}$ dar.

Folgende Überlegungen kann man jetzt anstellen:

$e^{-i\lambda} = \cos \lambda - i \sin \lambda$ ist eine komplexe Zahl mit $|e^{-i\lambda}| = 1$. Daraus folgt

$$a(\lambda) = \sum_{j=0}^p a_j (e^{-i\lambda})^j = \sum_{j=0}^p a_j z^j \quad \text{mit} \quad |z| = 1$$

Das heißt, $a(\lambda)$ ist ein Polynom, das auf dem Einheitskreis $\{z \mid |z| = 1\}$ definiert ist.

Hier wird ein Satz aus der Analysis (ohne Beweis) angewendet, der besagt, dass rationale Funktionen, wie Polynome, *eindeutig* von der Menge $\{z \mid |z| = 1\}$ auf ganz \mathbb{C} erweitert werden können, man betrachtet also $a(z)$ mit $z \in \mathbb{C}$.

Die Polynome $a(B)$ und $a(z)$ sind von gleicher Bauart, und das kann man zur Bestimmung von $a^{-1}(B)$ ausnützen.

Denn die Mengen $\left\{ \sum_{j=-\infty}^{\infty} k_j z^j \mid z \in \mathbb{C} \right\}$ und $\left\{ \sum_{j=-\infty}^{\infty} k_j B^j \mid B \dots \text{Verschiebeoperator} \right\}$ sind *isomorph* bezüglich der Multiplikation. Reihen von der Bauart $\sum_{j=-\infty}^{\infty} k_j z^j$ werden *Laurentreihen* (verallgemeinerte Potenzreihen) genannt.

Da dieser Isomorphismus existiert, können folgende Überlegungen für Polynome $a(z)$ mit $z \in \mathbb{C}$ durchgeführt werden, und die Resultate kann man dann eindeutig auf Polynome mit dem Verschiebeoperator $a(B)$ übertragen.

Unter der Annahme $a(z) \neq 0$ für alle $z \in \mathbb{C}$ mit $|z| = 1$ existiert das inverse Element $a^{-1}(z)$. Damit also $a^{-1}(z)$ existiert, darf das Polynom $a(z)$ keine Nullstellen am Einheitskreis $\{z \mid |z| = 1\}$ haben. Mit dem Argument, dass $a(z)$ stetig ist, kann man aus $a(z) \neq 0, \forall |z| = 1$, folgern, dass zwei Konstanten $0 < r_1 < 1 < r_2$ existieren, sodass $a(z) \neq 0$ für alle z , die innerhalb des Kreisrings $\{z \mid r_1 < |z| < r_2\}$ liegen, gilt. Das Polynom hat also außerdem keine Nullstellen innerhalb dieses Kreisrings. Da ein Polynom eine analytische¹ Funktion ist, kann $a(z)$ innerhalb dieses Kreisrings (mit Mittelpunkt $z_0 = 0$) in eine Laurentreihe $\sum_{j=-\infty}^{\infty} a_j (z-z_0)^j$ entwickelt werden, die dort konvergent ist.

Diese Laurentreihe kann folgendermaßen bestimmt werden. Das Polynom $a(z)$ kann

¹Eine Funktion $f(x)$ heißt in einem Gebiet G *analytisch*, wenn sie in allen Punkten von G differenzierbar ist.

gleich dem Ausdruck $c(z - z_1) \dots (z - z_p)$ gesetzt werden, dabei sind $z_1, \dots, z_p \in \mathbb{C}$ die Nullstellen von $a(z)$, die laut Annahme außerhalb von $\{z \mid r_1 < |z| < r_2\}$ liegen, c ist eine Konstante.

$$\Rightarrow \frac{1}{a(z)} = \frac{1}{c} \cdot \prod_{i=1}^p \frac{1}{z - z_i}$$

- Fall $|z_i| > 1$ (Nullstelle liegt außerhalb des Einheitskreises):

$$\frac{1}{z - z_i} = -\frac{1}{z_i} \cdot \frac{1}{1 - \frac{z}{z_i}} = -\frac{1}{z_i} \sum_{j=0}^{\infty} \left(\frac{z}{z_i}\right)^j = -\frac{1}{z_i} \sum_{j=0}^{\infty} z_i^{-j} z^j$$

denn $|z| < |z_i| \Rightarrow \left|\frac{z}{z_i}\right| < 1$ (Entwicklung in eine geometrische Reihe)

- Fall $|z_i| < 1$ (Nullstelle liegt innerhalb des Einheitskreises):

$$\frac{1}{z - z_i} = \frac{1}{z} \cdot \frac{1}{1 - \frac{z_i}{z}} = \frac{1}{z} \sum_{j=0}^{\infty} \left(\frac{z_i}{z}\right)^j = \sum_{j=0}^{\infty} z_i^j z^{-(j+1)} = \frac{1}{z_i} \sum_{j=-\infty}^{-1} z_i^{-j} z^j$$

denn $|z| > |z_i| \Rightarrow \left|\frac{z_i}{z}\right| < 1$

$$\Rightarrow \frac{1}{z - z_i} = \begin{cases} -\frac{1}{z_i} \sum_{j=0}^{\infty} z_i^{-j} z^j & \forall |z| < |z_i| \text{ im Fall } |z_i| > 1 \\ \frac{1}{z_i} \sum_{j=-\infty}^{-1} z_i^{-j} z^j & \forall |z| > |z_i| \text{ im Fall } |z_i| < 1 \end{cases}$$

$a^{-1}(z)$ erhält man aus dem Produkt dieser Laurentreihen multipliziert mit $\frac{1}{c}$. Die einzelnen Faktoren sind konvergente geometrische Reihen, und das endliche Produkt dieser Reihen ist wieder konvergent. $a^{-1}(z)$ lässt sich also durch eine Laurentreihe $\sum_{j=-\infty}^{\infty} h_j z^j$ darstellen, deren Koeffizienten h_j geometrisch gegen 0 für $j \rightarrow \pm\infty$ konvergieren. Die Laurentreihe von $a^{-1}(z)$ muss noch mit $b(z)$ multipliziert werden, was an der Konvergenz nichts ändert. Man erhält

$$l(z) = a^{-1}(z)b(z) = \left(\sum_{j=-\infty}^{\infty} h_j z^j \right) b(z) = \sum_{j=-\infty}^{\infty} l_j z^j$$

wobei $\sum_{j=-\infty}^{\infty} |l_j| < \infty$. Durch den zuvor erwähnten Isomorphismus erhält man $a^{-1}(B)$ und für die Lösung von 5.3

$$X_t = l(B)\varepsilon_t = \sum_{j=-\infty}^{\infty} l_j \varepsilon_{t-j} \quad \text{mit} \quad l(B) = a^{-1}(B)b(B)$$

Unter der Annahme $a(z) \neq 0, \forall |z| \leq 1$, ist leicht zu erkennen, dass $h_j = 0, \forall j < 0$. (Dann existieren keine Nullstellen innerhalb des Einheitskreises.)

Zusammenfassend wurde gezeigt:

Satz 5.0.8 • *Unter der Annahme $a(z) \neq 0, \forall |z| = 1$, existiert eine eingeschwingene und stabile Lösung des ARMA-Systems 5.3, und sie hat die Gestalt*

$$X_t = l(B)\varepsilon_t = \sum_{j=-\infty}^{\infty} l_j \varepsilon_{t-j} \quad \text{mit } l(B) = a^{-1}(B)b(B) = \left(\sum_{j=-\infty}^{\infty} h_j B^j \right) b(B) \quad (5.5)$$

• *Unter der Annahme $a(z) \neq 0, \forall |z| \leq 1$ existiert eine eingeschwingene, stabile und kausale Lösung des ARMA-Systems 5.3, und sie hat die Gestalt*

$$X_t = l(B)\varepsilon_t = \sum_{j=0}^{\infty} l_j \varepsilon_{t-j} \quad \text{mit } l(B) = a^{-1}(B)b(B) = \left(\sum_{j=0}^{\infty} h_j B^j \right) b(B) \quad (5.6)$$

In beiden Fällen gilt $\sum_{j=-\infty}^{\infty} |l_j| < \infty$, wodurch die Existenz der Summen gesichert ist.

Die Bedingung $a(z) \neq 0, \forall |z| \leq 1$, wird *Stabilitätsbedingung* genannt.

Bemerkung:

Diese beiden Lösungen sind partikuläre Lösungen von 5.3. Die gesamte Lösungsmenge des ARMA(p, q)-Systems besteht aus den partikulären Lösungen plus der Menge aller homogenen Lösungen von $a(B)X_t = 0$.

In dieser Arbeit werden nur kausale ARMA-Prozesse betrachtet. Im kausalen Fall werden die Koeffizienten $l_j, j = 0, 1, \dots$, von $l(B)$ durch den Koeffizientenvergleich von

$$a^{-1}(z)b(z) = l(z) \quad \Leftrightarrow \quad a(z)l(z) = b(z)$$

berechnet. Man erhält

$$z^0 \quad a_0 l_0 = b_0 \quad \Rightarrow \quad l_0 = a_0^{-1} b_0$$

$$z \quad a_0 l_1 + a_1 l_0 = b_1 \quad \Rightarrow \quad l_1 = a_0^{-1} (b_1 - a_1 a_0^{-1} b_0)$$

⋮

5.1 Das Spektrum eines (kausalen) ARMA-Prozesses

Zuerst wird das Spektrum von Weißem Rauschen (ε_t) berechnet. Aus den Eigenschaften der spektralen Dichte von Seite 16 folgt $\int_{-\pi}^{\pi} f_{\varepsilon}(\lambda) d\lambda = \gamma(0) = E \varepsilon_t^2 = \sigma^2$.

$$\Rightarrow \quad f_{\varepsilon}(\lambda) = \frac{\sigma^2}{2\pi} \quad (5.7)$$

f_ε ist also konstant.

Das Spektrum eines (kausalen) ARMA-Prozesses erhält man folgendermaßen:

Ein ARMA-Prozess $X_t = \sum_{j=0}^{\infty} l_j \varepsilon_{t-j}$ ist eine lineare Transformation von Weißem Rauschen mit der Transferfunktion $l(e^{-i\lambda}) = \sum_{j=0}^{\infty} l_j e^{-i\lambda j} = a^{-1}(e^{-i\lambda})b(e^{-i\lambda})$.

Aus 2.15 folgt dann

$$f_X(\lambda) = \frac{\sigma^2}{2\pi} a^{-1}(e^{-i\lambda})b(e^{-i\lambda})\overline{b(e^{-i\lambda})} \overline{a^{-1}(e^{-i\lambda})} \quad (5.8)$$

für $\lambda \in [-\pi, \pi]$.

5.2 Identifizierbarkeit von skalaren ARMA-Prozessen

Unter „Identifizierbarkeit“ versteht man das Problem, ob es möglich ist, aus den gegebenen zweiten Momenten der Daten (Beobachtungen) f_X beziehungsweise γ_X , eindeutig auf die Parameter des zugrundeliegenden Prozesses zu schließen.

Definition 5.2.1 (Beobachtungsäquivalenz von ARMA-Systemen) *Zwei ARMA-Systeme $a(B)X_t = b(B)\varepsilon_t$ und $\tilde{a}(B)X_t = \tilde{b}(B)\varepsilon_t$ heißen beobachtungsäquivalent, wenn für ein gegebenes σ^2 ein $\tilde{\sigma}^2$ existiert, sodass (a, b, σ^2) und $(\tilde{a}, \tilde{b}, \tilde{\sigma}^2)$ das gleiche f_X erzeugen. Dies muss für alle zulässigen σ^2 gelten.*

Wenn nur f_X betrachtet wird, sind beobachtungsäquivalente Systeme von außen nicht zu unterscheiden.

Definition 5.2.2 (Identifizierbarkeit) *Eine Klasse von ARMA-Systemen heißt identifizierbar, wenn sie keine unterschiedlichen, beobachtungsäquivalenten Systeme enthält.*

Mit anderen Worten, um Identifizierbarkeit zu gewährleisten, muss Beobachtungsäquivalenz ausgeschlossen werden.

Die Identifizierbarkeit von ARMA-Prozessen ist leichter zu beschreiben, wenn man das Problem sozusagen „teilt“. Zuerst wird die Identifizierbarkeit von MA(q)-Prozessen behandelt, dann die Identifizierbarkeit von AR(p)-Prozessen. Die rechte und die linke „Seite“ von ARMA-Prozessen wird also „getrennt“ durchgemacht.

1. MA(q)-Prozesse: $X_t = b(B)\varepsilon_t = \sum_{j=0}^q b_j \varepsilon_{t-j}$

Das Spektrum eines MA(q)-Prozesses ist

$$f_X(\lambda) = \frac{\sigma^2}{2\pi} b(e^{i\lambda})\overline{b(e^{i\lambda})} \quad \lambda \in [-\pi, \pi] \quad (5.9)$$

Mit dem Argument von Seite 36 kann die auf $\{z \mid |z| = 1\}$ definierte spektrale Dichte eindeutig auf ganz \mathbb{C} erweitert werden:

$$f_X(z) = \frac{\sigma^2}{2\pi} b(z)b(1/z) \quad z \in \mathbb{C} \quad (5.10)$$

Da das Weiße Rauschen (ε_t) nicht beobachtet werden kann, können die Prozesse (ε_t) und $B^l(\varepsilon_t) = (\varepsilon_{t-l})$, $l \in \mathbb{Z}$, allein mit der Information aus f_X nicht unterschieden werden. Anders ausgedrückt, erzeugen $b(B)$ und $\tilde{b}(B) = B^l(b)$ das gleiche f_x . Die zugehörigen MA-Systeme sind beobachtungsäquivalent. Daraus ist einfach zu sehen, dass $b_0 = 0$ vorausgesetzt werden kann.

Außerdem sind die MA-Systeme $b(B)$ mit der Varianz σ^2 sowie $\tilde{b}(B) = cb(B)$, c konstant, mit der Varianz $\tilde{\sigma}^2 = \sigma^2/c$ beobachtungsäquivalent. Denn Weißes Rauschen bleibt auch nach der Multiplikation mit einer Konstanten Weißes Rauschen, es ändert sich nur die Varianz. Deshalb kann man $b_0 = 1$ voraussetzen.

Um die restlichen Koeffizienten sowie den Grad des Polynoms $b(z)$ zu bestimmen, werden die Nullstellen der spektralen Dichte f_X betrachtet. Da das Polynom $b(z)$ laut Voraussetzung reelle Koeffizienten hat, treten seine Nullstellen in konjugiert komplexen Paaren auf, das heißt, wenn z_j eine Nullstelle ist, dann ist auch \bar{z}_j eine Nullstelle.

Andererseits sind dann wegen der Bauart von f_X auch $\frac{1}{z_j}$ sowie $\frac{1}{\bar{z}_j}$ Nullstellen von f_X . Die Hälfte der Nullstellen von f_X gehört also zu $b(z)$.

Da sich $b(z)$ durch

$$b(z) = c(z - z_1)(z - z_2) \dots (z - z_q)$$

darstellen lässt, entspricht q der halben Anzahl der Nullstellen von f_X .

Nun bleibt noch zu klären, welche Nullstellen dem Polynom $b(z)$ und welche dem Polynom $b(1/z)$ zugeordnet werden. Dieses Problem heißt *spektrales Faktorisierungsproblem*.

Logischerweise gilt, wenn z_j außerhalb des Einheitskreises liegt, dann liegt $\frac{1}{z_j}$ innerhalb des Einheitskreises und umgekehrt. Deshalb liegt die eine Hälfte der Nullstellen von f_X innerhalb, die andere Hälfte außerhalb des Einheitskreises.

Unter der Bedingung $b(z) \neq 0, \forall |z| \leq 1$, wird die Entscheidung, welche Nullstellen zur Bestimmung von $b(z)$ verwendet werden, eindeutig. Dann sind die Nullstellen von $b(z)$ nämlich jene Nullstellen z_j von f_X mit $|z_j| > 1$, denn $b(z)$ hat dann keine Nullstellen innerhalb oder am Einheitskreis.

Die Konstante c wird mit Hilfe der Bedingung $b_0 = 1$ bestimmt, und es gilt

$$c = \frac{1}{(-1)^q \prod_{j=1}^q z_j} \quad (5.11)$$

σ^2 wird aus 5.10 berechnet.

So wurde gezeigt, dass die Klasse der skalaren MA-Systeme mit $b_0 = 1$ und $b(z) \neq 0, \forall |z| \leq 1$, identifizierbar ist.

2. AR(p)-Prozesse: $a(B)X_t = \sum_{j=0}^p a_j X_{t-j} = \varepsilon_t$

Unter der Stabilitätsbedingung $a(z) \neq 0, \forall |z| \leq 1$, hat dieses AR(p)-System eine eingeschwungene, stabile und kausale Lösung.

Das Spektrum eines AR(p)-Prozesses kann wieder eindeutig auf \mathbb{C} erweitert werden und hat die Form

$$f_X(z) = \frac{\sigma^2}{2\pi} \frac{1}{a(z)a(1/z)} \quad (5.12)$$

Die Bestimmung von $a(z)$ erfolgt analog der Bestimmung von $b(z)$ bei MA(q)-Prozessen, nur dass an Stelle der Nullstellen von f_X die Pole von f_X verwendet werden müssen. Für die Pole von f_X kann man die gleichen Überlegungen wie für die Nullstellen vorher anstellen. Jene Pole, die außerhalb des Einheitskreises liegen, sind die Nullstellen von $a(z)$. p entspricht also der halben Anzahl der Pole von f_X . Das ist durch die Stabilitätsbedingung vorausgesetzt. $a(z)$ kann aus den Polen von f_X bis auf einen konstanten Faktor eindeutig bestimmt werden.

Zusätzlich setzt man $a_0 = 1$ voraus. Daraus kann dann die Konstante berechnet werden. σ^2 erhält man eindeutig aus 5.12.

Daher ist auch die Klasse der stabilen und kausalen skalaren AR(p)-Systeme mit $a_0 = 1$ identifizierbar.

Mit diesen Ergebnissen lässt sich die Identifizierbarkeit von skalaren ARMA-Prozessen einfach beschreiben.

5.8 kann eindeutig auf ganz \mathbb{C} erweitert werden:

$$f_X(z) = \frac{\sigma^2}{2\pi} \frac{b(z)b(1/z)}{a(z)a(1/z)} \quad (5.13)$$

$b(z)$ kann analog der Darstellung bei den MA(q)-Prozessen, $a(z)$ analog der Beschreibung bei den AR(p)-Prozessen bestimmt werden. Man sucht also alle Nullstellen und Pole von f_X mit einem Betrag, der größer als 1 ist, und bestimmt damit $a(z)$ und $b(z)$ bis auf einen konstanten Faktor eindeutig.

Folgende Schwierigkeit muss aber noch beachtet werden. Die Bedingung, dass $a(z)$ und $b(z)$ keine gemeinsamen (nichttrivialen) Teiler haben dürfen, muss noch eingeführt werden.

Die Konstanten von $a(z)$ und $b(z)$ werden über die Bedingung $a_0 = b_0 = 1$ bestimmt, und σ^2 wird aus 5.13 berechnet.

Zusammenfassend gilt folgender Satz:

Satz 5.2.1 (Identifizierbarkeit von skalaren ARMA(p, q)-Systemen) *Eine Klasse von skalaren ARMA(p, q)-Systemen ist identifizierbar, wenn folgende Bedingungen gelten:*

1. $a(z) \neq 0 \quad \forall |z| \leq 1$ (Stabilitätsbedingung)
2. $b(z) \neq 0 \quad \forall |z| \leq 1$
3. $a_0 = b_0 = 1$
4. $a(z)$ und $b(z)$ haben keine gemeinsamen Nullstellen.

6 Der quadrierte GARCH(p, q)–Prozess

Ausgangspunkt der folgenden Darstellungen sind wieder die Gleichungen 4.3 und 4.4. Es wird angenommen, dass eine stark stationäre und damit auch schwach stationäre Lösung (X_t) des GARCH(p, q)–Systems existiert.

Nun definiert man:

$$v_t := X_t^2 - \sigma_t^2 \quad (6.1)$$

$$\begin{aligned} \Rightarrow X_t^2 &= \sigma_t^2 + v_t \\ &= \alpha_0 + \sum_{i=1}^q \alpha_i X_{t-i}^2 + \sum_{j=1}^p \beta_j \sigma_{t-j}^2 + v_t = \\ &= \alpha_0 + \sum_{i=1}^{\max(p,q)} \alpha_i X_{t-i}^2 + \sum_{i=1}^{\max(p,q)} \beta_i X_{t-i}^2 - \sum_{i=1}^{\max(p,q)} \beta_i X_{t-i}^2 + \sum_{j=1}^p \beta_j \sigma_{t-j}^2 + v_t \end{aligned}$$

$$\text{wobei } \begin{cases} \alpha_{q+1}, \dots, \alpha_{\max(p,q)} = 0 & \text{falls } q < p \\ \beta_{p+1}, \dots, \beta_{\max(p,q)} = 0 & \text{falls } q > p \end{cases}$$

$$\begin{aligned} \Rightarrow X_t^2 &= \alpha_0 + \sum_{i=1}^{\max(p,q)} (\alpha_i + \beta_i) X_{t-i}^2 - \underbrace{\sum_{j=1}^p \beta_j (X_{t-j}^2 - \sigma_{t-j}^2)}_{=v_{t-j}} + v_t \\ X_t^2 - \sum_{i=1}^{\max(p,q)} (\alpha_i + \beta_i) X_{t-i}^2 &= \alpha_0 + v_t - \sum_{j=1}^p \beta_j v_{t-j} \end{aligned} \quad (6.2)$$

Diese Differenzgleichung zeigt, wenn (v_t) ein Weißes Rauschen und (X_t^2) schwach stationär ist, dann ist (X_t^2) ein ARMA($\max(p, q), p$)–Prozess.

Es gilt:

- $E(v_t | \mathcal{F}_{t-1}) = E(X_t^2 | \mathcal{F}_{t-1}) - E(\sigma_t^2 | \mathcal{F}_{t-1}) = \sigma_t^2 - \sigma_t^2 = 0$
 - $E v_t = E(E(v_t | \mathcal{F}_{t-1})) = 0$
 - $Var v_t = E(v_t^2) = E(E(v_t^2 | \mathcal{F}_{t-1})) = E(\sigma_t^4)(E(\varepsilon_t^4) - 1)^1 < \infty$
 - $Cov(v_t, v_{t-k}) = E(Cov(v_t, v_{t-k} | \mathcal{F}_{t-1})) = 0 \quad \forall k > 0$
- $\Rightarrow (v_t)$ ist ein Weißes Rauschen.

¹Bemerkung: Falls $(\varepsilon_t) \sim NID(0, 1)$, dann ist $Var v_t = 2E(\sigma_t^4)$.

6.1 Schwache Stationarität von (X_t^2)

Der Prozess (X_t^2) ist genau dann schwach stationär, wenn folgende Bedingungen gelten:

1. $E(X_t^4) < \infty, \quad \forall t \in \mathbb{Z}$
2. $E(X_t^2) = \text{konstant}, \quad \forall t \in \mathbb{Z}$
3. $Cov(X_t^2 X_{t-k}^2)$ hängt nur von $k > 0$ ab, $\forall t \in \mathbb{Z}$

6.1.1 Bestimmung von $E(X_t^4)$

Eine Möglichkeit $E(X_t^4)$ zu bestimmen wurde von Karanasos [5] hergeleitet.

Satz 6.1.1 (Karanasos [5]) (X_t, σ_t) sei ein (schwach) stationärer GARCH(p, q)-Prozess, der durch die Gleichungen 4.3 und 4.4 definiert ist. Das vierte Moment des GARCH(p, q)-Prozesses (X_t, σ_t) ist gegeben durch

$$E(X_t^4) = \frac{\alpha_0^2 \Phi}{(1 - \sum_{i=1}^p \alpha_i - \sum_{i=1}^q \beta_i)(1 - \delta)} \cdot E(\varepsilon_t^4) \quad (6.3)$$

Φ und δ sind folgendermaßen definiert:

$$\Phi = 1 + \alpha_q + \beta_q +$$

$$+ \sum_{i=1}^{q-1} (\alpha_i + \alpha_q(\alpha_{q-i} + \beta_{q-i})) \sum_{k=1}^{p+q-2} \gamma_{p+i-1, k} +$$

$$+ \sum_{i=1}^{p-1} (\beta_i + \beta_p(\alpha_{p-i} + \beta_{p-i})) \sum_{k=1}^{p+q-2} \gamma_{i, k} +$$

$$+ \alpha_q \sum_{i=1}^{p-q} \beta_{q+i} \sum_{k=1}^{p+q-2} \gamma_{i, k} + \beta_p \sum_{i=1}^{q-p} \alpha_{p+i} \sum_{k=1}^{p+q-2} \gamma_{p+i-1, k}$$

$$\delta = v\alpha_q^2 + \alpha_q\beta_q + \alpha_p\beta_p + \beta_p^2 +$$

$$+ \sum_{i=1}^{q-1} (\alpha_i + \alpha_q(\alpha_{q-i} + \beta_{q-i})) \sum_{k=1}^{p+q-2} (\rho_k \alpha_{k-\delta_k} + \beta_{k-\delta_k}) \gamma_{p+i-1, k} +$$

$$+ \sum_{i=1}^{p-1} (\beta_i + \beta_p(\alpha_{p-i} + \beta_{p-i})) \sum_{k=1}^{p+q-2} (\rho_k \alpha_{k-\delta_k} + \beta_{k-\delta_k}) \gamma_{i, k} +$$

$$+ \alpha_q \sum_{i=1}^{p-q} \beta_{q+i} \sum_{k=1}^{p+q-2} (\rho_k \alpha_{k-\delta_k} + \beta_{k-\delta_k}) \gamma_{i, k} +$$

$$+ \beta_p \sum_{i=1}^{q-p} \alpha_{p+i} \sum_{k=1}^{p+q-2} (\rho_k \alpha_{k-\delta_k} + \beta_{k-\delta_k}) \gamma_{p+i-1, k}$$

γ_{ij} ist das ij . Element der Matrix $\mathbf{\Gamma} = \mathbf{A}^{-1}$. \mathbf{A} ist eine $(p+q-2) \times (p+q-2)$ -dimensionale Matrix, die aus vier Teilmatrizen

$$\mathbf{A} = \begin{pmatrix} \mathbf{A}_1 & \mathbf{A}_2 \\ \mathbf{A}_3 & \mathbf{A}_4 \end{pmatrix}$$

besteht.

$\mathbf{A}_1 = (a_{ij}^{(1)})$ ist eine $(p-1) \times (p-1)$ -dimensionale Matrix mit

$$a_{ij}^{(1)} = \begin{cases} -\beta_{i+j} & i < j \\ 1 - \beta_{i+j} & i = j \\ -(\alpha_{i-j} + \beta_{i-j} + \beta_{i+j}) & i > j \end{cases}$$

wobei in allen drei Fällen $\begin{cases} \alpha_k = 0 & k > q \\ \beta_k = 0 & k > p \end{cases}$ gilt.

$\mathbf{A}_2 = (a_{ij}^{(2)})$ ist eine $(p-1) \times (q-1)$ -dimensionale Matrix mit

$$a_{ij}^{(2)} = \begin{cases} -\alpha_{i+j} & i+j \leq q \\ 0 & \text{sonst} \end{cases}$$

$\mathbf{A}_3 = (a_{ij}^{(3)})$ ist eine $(q-1) \times (p-1)$ -dimensionale Matrix mit

$$a_{ij}^{(3)} = \begin{cases} -\beta_{i+j} & i+j \leq p \\ 0 & \text{sonst} \end{cases}$$

$\mathbf{A}_4 = (a_{ij}^{(4)})$ ist eine $(q-1) \times (q-1)$ -dimensionale Matrix mit

$$a_{ij}^{(4)} = \begin{cases} -\alpha_{i+j} & i < j \\ 1 - \alpha_{i+j} & i = j \\ -(\alpha_{i-j} + \beta_{i-j} + \beta_{i+j}) & i > j \end{cases}$$

wobei in allen drei Fällen $\begin{cases} \alpha_k = 0 & k > q \\ \beta_k = 0 & k > p \end{cases}$ gilt.

$$\rho_k = \begin{cases} E(\varepsilon_t^4) & k \geq p \\ 1 & \text{sonst} \end{cases}$$

$$\delta_k = \begin{cases} p-1 & k \geq p \\ 0 & \text{sonst} \end{cases}$$

Bedingungen für die Existenz von $E(X_t^4)$ sind $\det(\mathbf{A}) \neq 0$ und $\delta < 1$.

Beweis:

Da (σ_t) und (ε_t) unabhängig sind, gilt $E(X_t^4) = E(\sigma_t^4)E|\varepsilon_t|^4$.

Es werden folgende Ausdrücke zur Vereinfachung der Schreibweise eingeführt:

$$v := E(\varepsilon_0^4) = E(\varepsilon_t^4), \text{ da } (\varepsilon_t) \sim IID(0, 1)$$

$$f := E(\sigma_0^4) = E(\sigma_t^4), \text{ da } (X_t, \sigma_t) \text{ laut Voraussetzung stationär ist}$$

$$g := E(\sigma_0^2) = E(\sigma_t^2) = \frac{\alpha_0}{1 - \sum_{i=1}^q \alpha_i - \sum_{j=1}^p \beta_j}$$

v und g sind bekannt, f wird gesucht.

$$i \in \mathbb{N} : \quad \lambda_i := E(\sigma_t^2 X_{t-i}^2) \quad c_i := E(\sigma_t^2 \sigma_{t-i}^2)$$

$$\Rightarrow E(X_t^2 X_{t-i}^2) = \lambda_i \quad E(X_t^2 \sigma_{t-i}^2) = c_i \quad E(X_t^2 \sigma_t^2) = f \quad E(X_t^4) = fv$$

Ausgehend von den Gleichungen

$$X_t^2 \sigma_t^2 = X_t^2 (\alpha_0 + \sum_{i=1}^q \alpha_i X_{t-i}^2 + \sum_{i=1}^p \beta_i \sigma_{t-i}^2)$$

$$\sigma_t^2 \sigma_{t-j}^2 = \sigma_{t-j}^2 (\alpha_0 + \sum_{i=1}^q \alpha_i X_{t-i}^2 + \sum_{i=1}^p \beta_i \sigma_{t-i}^2) \quad j = 1, \dots, q$$

$$\sigma_t^2 X_{t-j}^2 = X_{t-j}^2 (\alpha_0 + \sum_{i=1}^q \alpha_i X_{t-i}^2 + \sum_{i=1}^p \beta_i \sigma_{t-i}^2) \quad j = 1, \dots, p$$

erhält man folgende Gleichungen, indem man von beiden Seiten die Erwartungswerte bestimmt:

1. $f = \alpha_0 g + \sum_{i=1}^q \alpha_i \lambda_i + \sum_{i=1}^p \beta_i c_i$
2. $c_j = \alpha_0 g + (\alpha_j + \beta_j) f + \sum_{i=1}^{j-1} (\alpha_{j-i} + \beta_{j-i}) c_i + \sum_{i=1}^{q-j} \alpha_{j+i} \lambda_i + \sum_{i=1}^{p-j} \beta_{j+i} c_i \quad j = 1, \dots, p$
3. $\lambda_j = \alpha_0 g + (v\alpha_j + \beta_j) f + \sum_{i=1}^{j-1} (\alpha_{j-i} + \beta_{j-i}) \lambda_i + \sum_{i=1}^{q-j} \alpha_{j+i} \lambda_i + \sum_{i=1}^{p-j} \beta_{j+i} c_i$
 $j = 1, \dots, q$

Dabei ist $\alpha_i = 0$ für $i > q$ und $\beta_i = 0$ für $i > p$.

$$\Rightarrow c_p = \alpha_0 g + (\alpha_p + \beta_p) f + \sum_{i=1}^{p-1} (\alpha_{p-i} + \beta_{p-i}) c_i + \sum_{i=1}^{q-p} \alpha_{p+i} \lambda_i$$

$$\lambda_q = \alpha_0 g + (v\alpha_q + \beta_q) f + \sum_{i=1}^{q-1} (\alpha_{q-i} + \beta_{q-i}) \lambda_i + \sum_{i=1}^{p-q} \beta_{q+i} c_i$$

c_p und λ_q werden in die 1. Gleichung eingesetzt:

$$\begin{aligned}
\Rightarrow f &= \alpha_0 g(1 + \alpha_q + \beta_p) + f(v\alpha_q^2 + \alpha_q\beta_q + \alpha_p\beta_p + \beta_p^2) + \\
&+ \sum_{i=1}^{q-1} (\alpha_i + \alpha_q\alpha_{q-i} + \alpha_q\beta_{q-i})\lambda_i + \sum_{i=1}^{p-1} (\beta_i + \alpha_{p-i}\beta_p + \beta_{p-i}\beta_p)c_i + \\
&+ \alpha_q \sum_{i=1}^{p-q} \beta_{q+i}c_i + \beta_p \sum_{i=1}^{q-p} \alpha_{p+i}\lambda_i
\end{aligned}$$

Aus den Gleichungen für c_1, \dots, c_{p-1} und $\lambda_1, \dots, \lambda_{q-1}$ erhält man das $(p+q-2) \times (p+q-2)$ -dimensionale Gleichungssystem

$$\mathbf{Ac} = \mathbf{B} \quad (6.4)$$

mit

$$\mathbf{c} = \begin{pmatrix} c_1 \\ \vdots \\ c_{p-1} \\ \lambda_1 \\ \vdots \\ \lambda_{q-1} \end{pmatrix} \quad (6.5)$$

$$\mathbf{B} = \begin{pmatrix} \alpha_0 g + (\alpha_1 + \beta_1)f \\ \vdots \\ \alpha_0 g + (\alpha_{p-1} + \beta_{p-1})f \\ \alpha_0 g + (v\alpha_1 + \beta_1)f \\ \vdots \\ \alpha_0 g + (v\alpha_{q-1} + \beta_{q-1})f \end{pmatrix} \quad \text{wobei} \quad \begin{cases} \alpha_{q+1}, \dots, \alpha_{p-1} = 0 & \text{falls } p > q \\ \beta_{p+1}, \dots, \beta_{q-1} = 0 & \text{falls } p < q \end{cases} \quad (6.6)$$

$\mathbf{A} = \begin{pmatrix} \mathbf{A}_1 & \mathbf{A}_2 \\ \mathbf{A}_3 & \mathbf{A}_4 \end{pmatrix}$ ist die $(p+q-2) \times (p+q-2)$ -dimensionale Matrix, die im Satz definiert wurde.

Wenn man das Gleichungssystem 6.4 löst, erhält man \mathbf{c} als Funktion von f

$$\mathbf{c} = \mathbf{A}^{-1}\mathbf{B}$$

Sei $\mathbf{\Gamma} = (\gamma_{ij}) = \mathbf{A}^{-1} \in \mathbb{R}^{(p+q-2) \times (p+q-2)}$

$$\Rightarrow \mathbf{c} = \mathbf{\Gamma}\mathbf{B}$$

$$c_j = \alpha_0 g \sum_{k=1}^{p+q-2} \gamma_{jk} + f \left[\sum_{k=1}^{p-1} \gamma_{jk}(\alpha_k + \beta_k) + \sum_{k=1}^{q-1} \gamma_{j,p+k-1}(v\alpha_k + \beta_k) \right] \quad j = 1, \dots, p-1 \quad (6.7)$$

$$\lambda_j = \alpha_0 g \sum_{k=1}^{p+q-2} \gamma_{p+j-1,k} + f \left[\sum_{k=1}^{p-1} \gamma_{p+j-1,k}(\alpha_k + \beta_k) + \sum_{k=1}^{q-1} \gamma_{p+j-1,p+k-1}(v\alpha_k + \beta_k) \right] \quad j = 1, \dots, q-1 \quad (6.8)$$

c_j und λ_j werden in f eingesetzt, und diese Gleichung wird nach f aufgelöst. Mit Hilfe der im Satz eingeführten Variablen δ und Φ erhält man

$$f = \frac{\alpha_0 g \Phi}{1 - \delta} \quad (6.9)$$

$$\Rightarrow E(X_t^2) = f v = \frac{v \alpha_0 g \Phi}{1 - \delta} \quad (6.10)$$

□

6.1.2 Bedingungen für die schwache Stationarität von (X_t^2)

Nach Satz 4.1.1 ist ein GARCH(p, q)-Prozess, für den die Bedingung $\sum_{i=1}^q \alpha_i + \sum_{j=1}^p \beta_j < 1$ gilt, schwach stationär mit $E X_t = 0$, $Var X_t = E \sigma_t^2 = E X_t^2 = \frac{\alpha_0}{1 - \sum_{i=1}^q \alpha_i - \sum_{j=1}^p \beta_j}$, $\forall t \in \mathbb{Z}$, und $Cov(X_t, X_s) = 0$ für $t \neq s$.

$E X_t^2$ ist also unter der Bedingung $\sum_{i=1}^q \alpha_i + \sum_{j=1}^p \beta_j < 1$ konstant, $\forall t \in \mathbb{Z}$.

Die dritte Bedingung für die schwache Stationarität, dass die Kovarianz zwischen X_t^2 und X_{t-k}^2 , $k > 0$, nur von k abhängt, wird folgendermaßen gezeigt.

Die Darstellung 6.2, die den quadrierten GARCH(p, q)-Prozess in der Differenzgleichung der Form nach als ARMA($\max(p, q), p$)-Prozess zeigt, kann mit Hilfe des in Kapitel 5 eingeführten Verschiebeoperators B vereinfacht geschrieben werden:

$$\tilde{A}(B) X_t^2 = \alpha_0 + \tilde{B}(B) v_t \quad (6.11)$$

Dabei ist

$$\tilde{A}(B) = 1 - (\alpha_1 + \beta_1)B - (\alpha_2 + \beta_2)B^2 - \dots - (\alpha_{\max(p,q)} + \beta_{\max(p,q)})B^{\max(p,q)}$$

mit $\alpha_{q+1}, \dots, \alpha_{\max(p,q)} = 0$ falls $\max(p, q) = p$, beziehungsweise $\beta_{p+1}, \dots, \beta_{\max(p,q)} = 0$ falls $\max(p, q) = q$.

$$\tilde{B}(B) = 1 - \beta_1 B - \beta_2 B^2 - \dots - \beta_p B^p$$

$\tilde{A}(B)$ und $\tilde{B}(B)$ sind Polynome vom Grad $\max(p, q)$ beziehungsweise q mit dem Verschiebeoperator B .

Wenn der quadrierte GARCH(p, q)-Prozess (X_t^2) existiert, dann kann er durch

$$X_t^2 = \tilde{A}^{-1}(B)[\alpha_0 + \tilde{B}(B) v_t]$$

dargestellt werden.

Das inverse Element $\tilde{A}^{-1}(B)$ des Polynoms $\tilde{A}(B)$ kann nach der in Kapitel 5 beschriebenen Methode berechnet werden.

Nach Satz 5.0.8 existiert unter der Annahme $\tilde{A}(z) \neq 0, \forall |z| \leq 1, z \in \mathbb{C}$, eine eingeschwungene, stabile und kausale Lösung des ARMA-Systems 6.11.

$\tilde{A}^{-1}(B)$ lässt sich durch eine Potenzreihe darstellen:

$$\tilde{A}^{-1}(B) = \sum_{j=0}^{\infty} l_j B^j \quad \text{mit} \quad \sum_{j=0}^{\infty} |l_j| < \infty \quad (6.12)$$

Die Lösung von 6.11 hat die Darstellung

$$X_t^2 = \tilde{\alpha}_0 + \sum_{j=0}^{\infty} \tilde{a}_j v_{t-j} \quad (6.13)$$

mit $\tilde{\alpha}_0 = \alpha_0 \sum_{j=0}^{\infty} l_j$

$$\tilde{a}_j = \begin{cases} -\sum_{i=0}^j \beta_i l_{j-i} & \text{falls } 0 \leq j \leq p \\ -\sum_{i=0}^p \beta_i l_{j-i} & \text{falls } j > p \end{cases}$$

$$\beta_0 := -1$$

Da $\sum_{j=0}^{\infty} l_j$ beschränkt ist, gilt auch

$$\sum_{j=0}^{\infty} |\tilde{a}_j| \leq \sum_{j=0}^{\infty} |l_j| \sum_{i=0}^p |\beta_i| < \infty \quad (6.14)$$

Die Kovarianzen des quadrierten Prozesses sind durch

$$\text{Cov}(X_t^2, X_{t-k}^2) = E(X_t^2 X_{t-k}^2) - EX_t^2 EX_{t-k}^2 \quad \forall k > 0$$

gegeben.

Da EX_t^2 und $EX_{t-k}^2 \forall t \in \mathbb{Z}$ und $\forall k > 0$ konstant sind, genügt es zu zeigen, dass $E(X_t^2 X_{t-k}^2)$ nur von k abhängt.

$$\begin{aligned} E(X_t^2 X_{t-k}^2) &= E[(\tilde{\alpha}_0 + \sum_{j=0}^{\infty} \tilde{a}_j v_{t-j})(\tilde{\alpha}_0 + \sum_{i=0}^{\infty} \tilde{a}_i v_{t-k-i})] = \\ &= \tilde{\alpha}_0^2 + \tilde{\alpha}_0 \sum_{j=0}^{\infty} \tilde{a}_j \underbrace{E v_{t-j}}_{=0} + \tilde{\alpha}_0 \sum_{i=0}^{\infty} \tilde{a}_i \underbrace{E v_{t-k-i}}_{=0} + \sum_{j=0}^{\infty} \sum_{i=0}^{\infty} \tilde{a}_j \tilde{a}_i E v_{t-j} v_{t-k-i} = \end{aligned}$$

$$\begin{aligned}
&= \tilde{\alpha}_0^2 + \sum_{j=0}^{\infty} \tilde{a}_j \tilde{a}_{j-k} E v_{t-j} = \tilde{\alpha}_0^2 + \sum_{j=k}^{\infty} \tilde{a}_j \tilde{a}_{j-k} \text{Var} v_{t-j} = \\
&= \tilde{\alpha}_0^2 + E(\sigma_t^4)(E(\varepsilon_t^4) - 1) \sum_{j=k}^{\infty} \tilde{a}_j \tilde{a}_{j-k} \\
&\Rightarrow \text{Cov}(X_t^2, X_{t-k}^2) \text{ ist nur von } k > 0 \text{ abhängig.}
\end{aligned}$$

Folgende Bedingungen müssen also erfüllt sein, damit der quadrierte GARCH(p, q)-Prozess (X_t^2) ein schwach stationärer ARMA($\max(p, q), p$)-Prozess ist, wenn der GARCH(p, q)-Prozess stark stationär ist:

- Das auf $z \in \mathbb{C}$ definierte Polynom $\tilde{A}(z) = 1 - (\alpha_1 + \beta_1)z - (\alpha_2 + \beta_2)z^2 - \dots - (\alpha_{\max(p,q)} + \beta_{\max(p,q)})z^{\max(p,q)}$ darf innerhalb des Einheitskreises keine Nullstellen haben. Es muss

$$\tilde{A}(z) = 1 - (\alpha_1 + \beta_1)z - (\alpha_2 + \beta_2)z^2 - \dots - (\alpha_{\max(p,q)} + \beta_{\max(p,q)})z^{\max(p,q)} \neq 0 \quad \forall |z| \leq 1 \quad (6.15)$$

gelten (Stabilitätsbedingung).

- Auch die Bedingung für die Existenz des vierten Moments von X_t muss erfüllt sein:

$$\begin{aligned}
\delta &= v\alpha_q^2 + \alpha_q\beta_q + \alpha_p\beta_p + \beta_p^2 + \\
&+ \sum_{i=1}^{q-1} (\alpha_i + \alpha_q(\alpha_{q-i} + \beta_{q-i})) \sum_{k=1}^{p+q-2} (\rho_k \alpha_{k-\delta_k} + \beta_{k-\delta_k}) \gamma_{p+i-1,k} + \\
&+ \sum_{i=1}^{p-1} (\beta_i + \beta_p(\alpha_{p-i} + \beta_{p-i})) \sum_{k=1}^{p+q-2} (\rho_k \alpha_{k-\delta_k} + \beta_{k-\delta_k}) \gamma_{i,k} + \\
&+ \alpha_q \sum_{i=1}^{p-q} \beta_{q+i} \sum_{k=1}^{p+q-2} (\rho_k \alpha_{k-\delta_k} + \beta_{k-\delta_k}) \gamma_{i,k} + \\
&+ \beta_p \sum_{i=1}^{q-p} \alpha_{p+i} \sum_{k=1}^{p+q-2} (\rho_k \alpha_{k-\delta_k} + \beta_{k-\delta_k}) \gamma_{p+i-1,k} < 1
\end{aligned}$$

(siehe Satz 6.1.1)

7 Identifizierbarkeit von univariaten GARCH(p, q)–Prozessen

Zu Beginn dieses Kapitels werden noch einmal kurz die wichtigsten „Werkzeuge“ für die Identifizierbarkeit zusammengefasst.

Nach den Sätzen 4.1.1 und 4.1.4 ist eine Lösung (X_t) der Gleichungen 4.3 und 4.4 ein stark und damit auch schwach stationärer GARCH(p, q)–Prozess, wenn die Parameter folgende Bedingungen erfüllen:

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

Es gilt: $E X_t = 0 \quad \forall t \in \mathbb{Z}$

$$\text{Var } X_t = E X_t^2 = E \sigma_t^2 = \frac{\alpha_0}{1 - \sum_{i=1}^q \alpha_i - \sum_{j=1}^p \beta_j} = \textit{konstant} \quad \forall t \in \mathbb{Z}$$

$$\text{Cov}(X_t, X_s) = 0 \quad t \neq s$$

Der quadrierte GARCH(p, q)–Prozess lässt sich als Lösung eines ARMA($\max(p, q), p$)–Systems darstellen:

$$X_t^2 - \sum_{i=1}^{\max(p,q)} (\alpha_i + \beta_i) X_{t-i}^2 = \alpha_0 + v_t - \sum_{j=1}^p \beta_j v_{t-j}$$

Dabei ist $v_t = X_t^2 - \sigma_t^2$ ein Weißes Rauschen.

Unter den Bedingungen von Seite 50

$$1 - (\alpha_1 + \beta_1)z - (\alpha_2 + \beta_2)z^2 - \dots - (\alpha_{\max(p,q)} + \beta_{\max(p,q)})z^{\max(p,q)} \neq 0 \quad \forall |z| \leq 1, z \in \mathbb{C}$$

und $\delta < 1$ hat dieses ARMA–System einen schwach stationären ARMA–Prozess als Lösung. Nach 6.13 kann man diesen durch

$$X_t^2 = \tilde{\alpha}_0 + \sum_{j=0}^{\infty} \tilde{\alpha}_j v_{t-j}$$

mit $\tilde{\alpha}_0 = \alpha_0 \sum_{j=0}^{\infty} l_j$

$$\tilde{\alpha}_j = \begin{cases} -\sum_{i=0}^j \beta_i l_{j-i} & \text{falls } 0 \leq j \leq p \\ -\sum_{i=0}^p \beta_i l_{j-i} & \text{falls } j > p \end{cases}$$

$$\beta_0 := -1$$

darstellen.

Für die Konstante $\tilde{\alpha}_0$ gilt:

$$\begin{aligned} \tilde{\alpha}_0 &= \alpha_0 \sum_{j=0}^{\infty} l_j = \alpha_0 \cdot \tilde{A}^{-1}(1) = \frac{\alpha_0}{1 - (\alpha_1 + \beta_1) - (\alpha_2 + \beta_2) - \dots - (\alpha_{\max(p,q)} + \beta_{\max(p,q)})} = \\ &= \frac{\alpha_0}{1 - \sum_{i=1}^q \alpha_i - \sum_{j=1}^p \beta_j} = E X_t^2 \quad \forall t \in \mathbb{Z} \end{aligned}$$

Aus 6.13 folgt dann

$$X_t^2 - E X_t^2 = \sum_{j=0}^{\infty} \tilde{\alpha}_j v_{t-j}$$

Mit Satz 2.3.3 und Satz 2.3.4 folgt weiter

$$\begin{aligned} Cov(X_t^2, X_0^2) &= E (X_t^2 - E X_t^2) (X_0^2 - E X_0^2) = \\ &= E (\sum_{j=0}^{\infty} \tilde{\alpha}_j v_{t-j}) (\sum_{k=0}^{\infty} \tilde{\alpha}_k v_{-k}) = \\ &= \sum_{j=0}^{\infty} \sum_{k=0}^{\infty} \tilde{\alpha}_j \tilde{\alpha}_k \underbrace{E v_{t-j} v_{-k}}_{\gamma_v(t-j+k)} = \\ &= \sum_{j=0}^{\infty} \sum_{k=0}^{\infty} \tilde{\alpha}_j \tilde{\alpha}_k \int_{-\pi}^{\pi} e^{i\lambda(t-j+k)} f_v(\lambda) d\lambda = \\ &= \int_{-\pi}^{\pi} \underbrace{\sum_{j=0}^{\infty} \tilde{\alpha}_j e^{-i\lambda j}}_{=\tilde{A}^{-1}(e^{-i\lambda})\tilde{B}(e^{-i\lambda})} e^{i\lambda t} f_v(\lambda) \underbrace{\sum_{k=0}^{\infty} \tilde{\alpha}_k e^{-i\lambda k}}_{=\overline{\tilde{A}^{-1}(e^{-i\lambda})\tilde{B}(e^{-i\lambda})}} d\lambda \end{aligned}$$

Mit Satz 2.3.3 folgt außerdem

$$\begin{aligned} Cov(X_t^2, X_0^2) &= \gamma_{X^2}(\lambda) = \int_{-\pi}^{\pi} e^{i\lambda t} f_{X^2}(\lambda) d\lambda \\ \Rightarrow \int_{-\pi}^{\pi} e^{i\lambda t} f_{X^2}(\lambda) d\lambda &= \int_{-\pi}^{\pi} e^{i\lambda t} \tilde{A}^{-1}(e^{-i\lambda}) \tilde{B}(e^{-i\lambda}) f_v(\lambda) \overline{\tilde{A}^{-1}(e^{-i\lambda}) \tilde{B}(e^{-i\lambda})} d\lambda \end{aligned}$$

Mit dem gleichen Argument wie bei Satz 2.3.4 folgt daraus für das Spektrum von (X_t^2)

$$f_{X^2}(\lambda) = \tilde{A}^{-1}(e^{-i\lambda}) \tilde{B}(e^{-i\lambda}) f_v(\lambda) \overline{\tilde{A}^{-1}(e^{-i\lambda}) \tilde{B}(e^{-i\lambda})} \quad (7.1)$$

$$f_v(\lambda) = \frac{1}{2\pi} \text{Var}v_t = \frac{1}{2\pi} E\sigma_0^2(E\varepsilon_0^4 - 1) =: \frac{\sigma_v^2}{2\pi} \quad (7.2)$$

f_{X^2} kann wieder eindeutig von der Menge $\{z \in \mathbb{C} \mid |z| = 1\}$ auf ganz \mathbb{C} erweitert werden.

$$\Rightarrow f_{X^2}(z) = \frac{\sigma_v^2 \tilde{B}(z)\tilde{B}(\frac{1}{z})}{2\pi \tilde{A}(z)\tilde{A}(\frac{1}{z})} \quad z \in \mathbb{C} \quad (7.3)$$

Nun kann man analog der Beschreibung in Abschnitt 5.2 weiter vorgehen.

Um das Polynom $\tilde{B}(z) = 1 - \beta_1 z - \beta_2 z^2 - \dots - \beta_p z^p$ zu bestimmen, werden alle Nullstellen z_j von $f_{X^2}(z)$ berechnet. Die Hälfte der Nullstellen zählt zum Polynom $\tilde{B}(z)$ und die andere Hälfte der Nullstellen zu $\tilde{B}(1/z)$.

Der Grad p des Polynoms $\tilde{B}(z)$ entspricht der halben Anzahl der Nullstellen von $f_{X^2}(z)$.

Man führt wieder die Bedingung $\tilde{B}(z) \neq 0, \forall |z| \leq 1$, ein, durch die alle Nullstellen z_j von $f_{X^2}(z)$ mit $|z_j| > 1$ eindeutig dem Polynom $\tilde{B}(z)$ zugeordnet werden. $\tilde{B}(z)$ wird durch

$$\tilde{B}(z) = c_B(z - z_1)(z - z_2) \cdots (z - z_p)$$

berechnet. c_B ist eine Konstante.

Bei dem quadrierten GARCH(p, q)-Prozess (X_t^2) ist beim Polynom $\tilde{B}(z)$ als Koeffizient von z^0 1 vorgegeben. Dies ist nach 5.2 sowieso eine Bedingung für die Identifizierbarkeit von ARMA-Prozessen.

Nach 5.11 ist

$$c_B = \frac{1}{(-1)^p \prod_{j=1}^p z_j}$$

Mit einem Koeffizientenvergleich zwischen den beiden Darstellungen $1 - \beta_1 z - \beta_2 z^2 - \dots - \beta_p z^p$ und $c_B(z - z_1)(z - z_2) \cdots (z - z_p)$ erhält man die Parameter β_1, \dots, β_p eindeutig.

Für die Bestimmung von $\tilde{A}(z)$ werden die Pole von $f_{X^2}(z)$ betrachtet. Da die Anzahl der Pole von f_{X^2} der Anzahl der Nullstellen von $\tilde{A}(z)\tilde{A}(1/z)$ entspricht, ist der Grad von $\tilde{A}(z)$ gleich der halben Anzahl der Pole von f_{X^2} .

Unter der Voraussetzung der Stabilitätsbedingung für den quadrierten GARCH(p, q)-Prozess

$$1 - (\alpha_1 + \beta_1)z - (\alpha_2 + \beta_2)z^2 - \dots - (\alpha_{\max(p,q)} + \beta_{\max(p,q)})z^{\max(p,q)} \neq 0 \quad \forall |z| \leq 1$$

sind jene Pole von $f_{X^2}(z)$, die betragsmäßig größer als 1 sind, das ist genau die Hälfte aller Pole, die Nullstellen des Polynoms $\tilde{A}(z)$.

Damit kann $\tilde{A}(z)$ durch

$$\tilde{A}(z) = c_A(z - p_1)(z - p_2) \cdots (z - p_{\max(p,q)})$$

dargestellt werden, wobei $p_j, j = 1, \dots, \max(p, q)$, jetzt die Pole von $f_{X^2}(z)$ mit $|p_j| > 1$ sind.

Der Koeffizient von z^0 ist wieder 1, und daraus erhält man c_A mit 5.11:

$$c_A = \frac{1}{(-1)^{\max(p,q)} \prod_{j=1}^{\max(p,q)} p_j}$$

Damit berechnet man die Koeffizienten $\alpha_i + \beta_i, i = 1, \dots, \max(p, q)$, des Polynoms $\tilde{A}(z)$ mittels Koeffizientenvergleich. (Achtung auf die Vorzeichen!)

Da nach den Voraussetzungen von Seite 25 für die Parameter eines GARCH(p, q)-Prozesses $\alpha_1, \dots, \alpha_{q-1} \geq 0, \alpha_q > 0, \beta_1, \dots, \beta_{p-1} \geq 0, \beta_p > 0$ gelten muss, können aus den Koeffizienten von $\tilde{A}(z)$ die Parameter $\alpha_1, \dots, \alpha_q$ berechnet werden.

Die zusätzliche Bedingung, dass die Polynome $1 - (\alpha_1 + \beta_1)z - (\alpha_2 + \beta_2)z^2 - \dots - (\alpha_{\max(p,q)} + \beta_{\max(p,q)})z^{\max(p,q)}$ und $1 - \beta_1z - \beta_2z^2 - \dots - \beta_pz^p$ keine gemeinsamen Nullstellen haben, muss noch eingeführt werden.

Aus der Gleichung 7.3 kann die Varianz σ_v^2 des Weißen Rauschen (v_t) berechnet werden. Aus 7.2 folgt

$$E\sigma_0^2 = E\sigma_t^2 = \frac{\sigma_v^2}{E\varepsilon_0^4 - 1}$$

da (σ_t) schwach stationär ist.

$$\Rightarrow E X_t^2 = E\varepsilon_t^2 E\sigma_t^2$$

mit $(\varepsilon_t) \sim IID(0, 1)$

Daraus kann der Parameter α_0 berechnet werden (siehe Seite 52):

$$\alpha_0 = E X_t^2 \left[1 - \sum_{i=1}^q \alpha_i - \sum_{j=1}^p \beta_j \right] = \frac{\sigma_v^2}{(E\varepsilon_0^4 - 1)} \left[1 - \sum_{i=1}^q \alpha_i - \sum_{j=1}^p \beta_j \right] \quad (7.4)$$

Abschließend kann man also sagen:

Univariate GARCH(p, q)-Prozesse, die die Bedingungen

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

2. $\tilde{A}(z) = 1 - (\alpha_1 + \beta_1)z - (\alpha_2 + \beta_2)z^2 - \dots - (\alpha_{\max(p,q)} + \beta_{\max(p,q)})z^{\max(p,q)} \neq 0 \quad \forall |z| \leq 1, z \in \mathbb{C}$
3. $\tilde{B}(z) = 1 - \beta_1z - \beta_2z^2 - \dots - \beta_pz^p \neq 0 \quad \forall |z| \leq 1, z \in \mathbb{C}$
4. $\tilde{A}(z)$ und $\tilde{B}(z)$ haben keine gemeinsamen Nullstellen.
5. $\delta < 1$

erfüllen, sind identifizierbar.

Literaturverzeichnis

- [1] Bollerslev, Tim, 1986. Generalized Autoregressive Conditional Heteroskedasticity. *Journal of Econometrics* 31 (1986), 307–327.
- [2] Brockwell, Peter J., Davis, Richard A., 1991. *Time Series: Theory and Methods*. Second Edition. Springer Series in Statistics, Springer, New York.
- [3] Deistler, Manfred, Scherrer, Wolfgang, 1994. *Econometrics II - The Prague Lectures*. TU Wien.
- [4] Engle, Robert F., 2003. *Risk and Volatility. Econometric Models and Financial Practice*. Nobel Lecture, 2003.
- [5] Karanasos, Menelaos, 1999. The second moment and the autocovariance function of the squared errors of the GARCH model. *Journal of Econometrics* 90 (1999), 63–76.
- [6] Lindner, Alexander M., 2007. Stationarity, Mixing, Distributional Properties and Moments of $\text{GARCH}(p, q)$. <http://www-ma4.ma.tum.de/Papers/index.en.html>.
- [7] Straumann, Daniel, 2005. *Estimation in Conditionally Heteroscedastic Time Series Models*. Lecture Notes in Statistics 1881, Springer, Berlin.