Statistik der Finanzmärkte

Thorsten Schmidt [1]

Version vom 25. Mai 2007

[1]Fakultät für Mathematik und Informatik, Universität Leipzig, Augustusplatz 10/11 04109 Leipzig Germany. Email: thorsten.schmidt@math.uni-leipzig.de

Inhaltsverzeichnis

1 Zeitreihenmodelle für Finanzdaten

In diesem Kapitel werden Zeitreihenmodelle vorgestellt, die sich als besonders geeignet zur Modellierung von Finanzdaten erwiesen haben. Das in Engle (1982) vorgestellte ARCH (auto regressive conditional heteroscedastic) Modell ist eine Erweiterung des bis dato favorisierten AR (auto regressive) Modells. Zu dieser Zeit wurden viele Versuche gemacht, die Defizite des AR Modells zu beheben. Das von Engle vorgestellte Modell glänzt durch seine Einfachheit und erfasst viele Symptome die Finanzdaten aufweisen.

1.1 Einführung

Betrachte einen Wahrscheinlichkeitsraum $(\Omega, \mathcal{F}, \mathbb{P})$. Zeitreihen sind spezielle *stochastische Prozesse*, eben solche die in diskreter Zeit ablaufen. Eine Zeitreihe ist definiert als eine folge von Zufallsvariablen $(Y_t)_{t\in\mathcal{T}}$, wobei $\mathcal{T}$ eine höchstens abzählbare Menge ist; typischerweise $\{1, \dots, n\}$, $\mathbb{Z}$ oder $\mathbb{N}$. Hierbei nimmt Y Werte im sogenannten *Zustandsraum* $\mathcal{Y}$ an. Ist $\mathcal{Y} \subset \mathbb{R}$, so nennen wir Y *univariat*. Im folgenden sei $\mathcal{T}$ stets $\mathbb{N}$ oder $\mathbb{Z}$ und wir schreiben kurz Y oder (Y_t) für den stochastischen Prozess $(Y_t)_{t\in\mathcal{T}}$.

Bekanntermaßen ist eine Zufallsvariable eine Abbildung von Ω in ihren Zustandsraum. Für festes $\omega \in \Omega$ nennen wir $(Y(\omega)_t)_{t\in\mathcal{T}}$ einen *Pfad* des Prozesses Y. Ein berühmter Satz von Kolmogorov besagt nun, dass die Verteilung des Prozesses Y bereits durch seine endlich-dimensionalen Randverteilungen (fidi's), d.h. die Verteilung aller $(Y_{t_1}, \dots, Y_{t_n})$ mit $(t_1, \dots, t_n) \subset \mathcal{Y}$, festgelegt ist.

Es wird sich als nützlich erweisen, hierbei auch bedingte Verteilungen zu betrachten. Im Falle stetiger Verteilungen kann man die Verteilung jeweils durch die Dichte beschreiben[1] und bedingte Verteilungen eben durch bedingte Dichten. Die fidi's sind durch die Dichten der $(Y_{t_1}, \dots, Y_{t_n})$, im folgenden mit $f(y_{t_1}, \dots, y_{t_n})$ bezeichnet, beschrieben. Schreiben wir für die bedingte Dichte von Y_t gegeben $Y_{t-1}, \dots, Y_{t-k}$

$$f_{t|t-1,\dots,t-k}(y_t \,|\, y_{t-1}, \dots, y_{t-k})$$

so ist die Dichte des Vektors $(Y_t, Y_{t-1}, \dots, Y_{t-k})$ gegeben durch

$$f_{t|t-1,\dots,t-k}(y_t|y_{t-1}, \dots, y_{t-k}) \cdots f_{t-k+1|t-k}(y_{t-k+1}|y_{t-k})\, f_{t-k}(y_{t-k}).$$

[1]Die *Verteilungsfunktion* einer (univariaten) Zufallsvariablen ξ, $F(x) := \mathbb{P}(\xi \le x)$, ergibt durch Ableiten die *Dichte*: $f(x) = \frac{\partial}{\partial x} F(x)$, falls sie existiert. Alle Verteilungsgrößen können dann mit der Dichte berechnet werden, wie etwa Erwartungswerte oder m-te Momente:

$$\mathbb{E}(\xi^m) = \int_\Omega \xi(\omega)^m\, \mathbb{P}(d\omega) = \int_{\mathbb{R}} x^m\, f(x)\, dx.$$

Mit $f_{t-k}(\cdot)$ haben wir hierbei die Dichte von Y_{t-k} bezeichnet.

Mitunter ist es nützlich, nur bestimmte Merkmale einer Verteilung zu betrachten, etwa Erwartungswert oder Varianz/Kovarianz. Im Gaußschen Fall beschreiben diese Kenngrößen bereits die gesamte Verteilung. Dazu definieren wir, falls sie existiert, durch

$$m(t) := \mathbb{E}(Y_t), \qquad t \in \mathcal{T}$$

die **Erwartungswertfunktion** sowie durch

$$\gamma(t,h) := \operatorname{Cov}(Y_t, Y_{t+h}) = \mathbb{E}\big[(Y_t - m_t)\,(Y_{t+h} - m_{t+h})\big], \qquad t, t+h \in \mathcal{T}$$

die **Autokovarianzfunktion** (AKF) für den Abstand h. Dabei beschreibt $\gamma(t,h)$ die lineare Abhängigkeit des Prozesses zu Zeitpunkten t und $t+h$. Prozesse, für die m und γ stets existieren heissen *2-ter Ordnung*. Manchmal benutzt man die **Autokorrelationsfunktion** an Stelle von γ, definiert durch

$$\rho(t,h) := \frac{\gamma(t,h)}{\gamma(t,0)} = \operatorname{Corr}(Y_{t+h}, Y_t).$$

Insbesondere für das Schätzen von stochastischen Prozessen stellt sich *Stationarität* als wichtige Eigenschaft heraus.

Definition 1.1.1. (Stationarität)

(i) Ein stochastischer Prozess Y heisst **strikt stationär***, falls die Verteilungen der Vektoren $(Y_{t+1}, \dots, Y_{t+n})$ und $(Y_1, \dots, Y_n)$ für alle t identisch sind.*

(ii) Ein Prozess 2-ter Ordnung heisst **schwach stationär***, falls Erwartungswert- und Autokovarianzfunktion nicht von t abhängen, genauer:*

$$m_t = m_0, \quad \textit{und} \quad \gamma(t,h) = \gamma(0,h), \qquad \forall t, h \in \mathcal{T}.$$

Man beachte, dass für stationäre Prozesse m und γ nicht mehr von t abhängen. Dann erweist sich vor allem γ als nützliches Hilfsmittel, um etwa verschieden Zeitreihentypen voneinander zu unterscheiden.

Beispiel 1.1.2.

(i) *(i.i.d.)* Eine Folge von Zufallsvariablen (Y_i) heißt **i.i.d.** (unabhängig und identisch verteilt), falls Y_i und Y_j für alle $i \neq j$ unabhängig sind und alle Y_i die gleiche Verteilung haben. Die Zeitreihe (Y_i) ist stationär. Existiert weiterhin $\operatorname{Var}(Y_i) := \sigma^2$, so ist

$$\gamma(h) = \begin{cases} \sigma^2 & \text{für } h = 0 \\ 0 & \text{sonst.} \end{cases}$$

(ii) *(Random Walk)* Für $(X_i)_{i \geq 0}$ i.i.d. sei der Prozess $(Y_i)_{i \geq 0}$ definiert durch

$$Y_i := X_1 + \cdots + X_i$$

und $Y_0 := 0$. Dann bezeichnet man Y als "Random Walk". Für $X_i = \pm 1$ erhält man einen Binomialbaum. Für die AKF erhält man leicht $\gamma(t,h) = t\operatorname{Var}(X_1)$, diese Zeitreihe ist also nicht stationär. Was ist $\rho(t)$?

Des weiteren muss man sich Gedanken machen, wie stark ein Prozess von seiner Vergangenheit abhängt. Es wäre möglich, dass sich die gesamte Information aus der Vergangenheit in einem einzigen Wert, dem aktuellen Wert des Prozesses, kumuliert.

Definition 1.1.3. *Ein stochastischer Prozess Y heisst Markovsch der Ordnung k, falls die bedingte Verteilung nur von den letzten k Werten abhängt, formal geschrieben: Für alle $K \geq k$ und beschränkte und stetige f ist*

$$\mathbb{E}\big(f(Y_t)|Y_{t-1},\ldots,Y_{t-k},\ldots,Y_{t-K}\big) = \mathbb{E}\big(f(Y_t)|Y_{t-1},\ldots,Y_{t-k}\big).$$

Existieren die bedingten Dichten, so läßt sich die Markov-Eigenschaft auch hierdurch ausdrücken:

$$f_{t\,|\,t-1,\ldots,t-K}(y_t\,|\,y_{t-1},\ldots,y_{t-K}) = f_{t\,|\,t-1,\ldots,t-k}(y_t\,|\,y_{t-1},\ldots,y_{t-k}).$$

Diese Idee inspiriert gerade Autoregressive Prozesse: Solche Prozesse, die nur von einer festen Anzahl vergangener Werte abhängen. Die Kovarianzfunktion wird für diese Prozesse eine besondere Rolle spielen, so dass wir a priori ihre Existenz fordern.

Definition 1.1.4. *Wir definieren*[2] $\mathcal{F}_t^Y := \sigma(Y_t, Y_{t-1},\ldots)$ *die Y-Vergangenheit ab dem Zeitpunkt t. Ein stochastischer Prozess $(Y_t)_{t\in\mathcal{T}}$ 2-ter Ordnung heißt autoregressiv der Ordnung k, falls*

$$\mathbb{E}(Y_t|\mathcal{F}_{t-1}^Y) = \mathbb{E}(Y_t|Y_{t-1},\ldots,Y_{t-k}).$$

Falls die Funktion auf der rechten Seite affin ist, bezeichnen wir den Prozess als linear autoregressiv oder AR(k). Es gilt die Darstellung

$$\mathbb{E}(Y_t|\mathcal{F}_{t-1}^Y) = a_0 + a_1 Y_{t-1} + \cdots + a_k Y_{t-k}.$$

1.2 Lineare und nicht-lineare Prozesse

Die klassische Zeitreihenanalyse befasst sich im wesentlichen mit sogenannten ARMA Prozessen (auto regressive moving average). Einen gleitenden Durchschnitt könnte man sich folgendermaßen vorstellen. Wähle k fest und setze

$$Y_t = \frac{1}{k}\sum_{i=1}^{k} Z_{t-i},$$

wobei $Z_t : t \in \mathcal{T}$ zentrierte[3] und beispielsweise unabhängige Zufallsvariablen sein könnten. Man erhält sofort die folgende, allgemeinere Formulierung

[2] Durch $\sigma(\xi_1,\xi_2,\ldots)$ bezeichnen wir die σ-Algebra, die durch die Zufallsvariablen $\xi_1,\xi_2,\ldots$ erzeugt wird. Sie repräsentiert die Information, die in diesen Zufallsvariablen enthalten ist.

[3] Mit *zentriert* bezeichnen wir eine Zufallsvariable, die den Erwartungswert (oder bedingten Erwartungswert, je nach Kontext) Null hat.

Definition 1.2.1. *Eine Folge von Zufallsvariablen* $(Z_t)_{t\in\mathcal{T}}$ *bezeichnen wir als* **weißes Rauschen**, *geschrieben* $(Z_t) \sim WN(0, \sigma^2)$, *falls alle* Z_t *zentriert mit Varianz* σ^2 *paarweise unkorreliert sind, d.h.* $\mathbb{E}(Z_t Z_{t+h}) = 0$ *für alle* $h \neq 0$.

Ein stochastischer Prozess Y *heißt* **ARMA(p,q)**, *falls er strikt stationär ist und, für alle* $t \in \mathcal{T}$ *folgende Darstellung besitzt:*

$$Y_t - \phi_1 Y_{t-1} - \cdots - \phi_p Y_{t-p} = c + Z_t + \theta_1 Z_{t-1} + \cdots + \theta_q Z_{t-q}.$$

Dabei seien $(Z_t) \sim WN(0, \sigma^2)$, $c \in \mathbb{R}$ *und die Polynome* $\phi(z) := (1 - \phi_1 z - \cdots - \phi_p z^p)$, $\theta(z) := (1 + \theta_1 z + \cdots + \theta_q z^q)$ *haben keine gemeinsame Faktoren.*

Um obige Gleichung etwas eleganter darstellen zu können, führt man den Shift-Operator L (L wie "lag"), definiert durch $LY_t := Y_{t-1}$, ein. Mit den obigen Polynomen erhält man formal $\phi(L)$ und $\theta(L)$ und die kompakte Darstellung

$$\phi(L) Y_t = c + \theta(L) Z_t.$$

Beispiel 1.2.2. MA(1) und AR(1)

(i) *MA(1)* In diesem Fall ist

$$Y_t = c + Z_t + \theta_1 Z_{t-1}.$$

Für die AKF erhält man leicht (beachte: $\mathbb{E}(Z_t^2) = \sigma^2$) für $t \geq 1$

$$\gamma(t,h) = \begin{cases} \sigma^2 & h = 0 \\ \theta_1 \sigma^2 & h = 1 \\ 0 & \text{sonst.} \end{cases}$$

Ist Y_0 also verteilt wie beispielsweise Z_1 und unabhängig von allen Z, so ist Y stationär.

(ii) *AR(1)* Dieser Prozess hat die Darstellung

$$Y_t = c + \phi_1 Y_{t-1} + Z_t.$$

Für eine einfachere Darstellung setzen wir $c = 0$. Damit kann man statt der Varianz direkt zweite Momente betrachten:

$$\gamma(t,h) = \begin{cases} \gamma(t,0) & h = 0 \\ \mathbb{E}(Y_t Y_{t+1}) = \mathbb{E}(Y_t^2 \phi_1) = \phi_1 \gamma(t,0) & h = 1 \\ \phi_1^h \gamma(t,0) & h \text{beliebig.} \end{cases}$$

Betrachten wir die Zeitreihe $(Y_t)_{t\in\mathbb{Z}}$. Gibt es eine Zeitreihe, die obige Gleichung erfüllt, also ein AR(1) Prozess ist ? (Antwort: ja! Diese Zeitreihe ist stationär). Für eine stationäre Zeitreihe folgt auch $\gamma(t,0) = \gamma(0)$ und

$$\gamma(0) = \text{Cov}(Y_t, Y_t) = \text{Cov}(\phi_1 Y_{t-1} + Z_t, \phi_1 Y_{t-1} + Z_t) = \phi_1^2 \gamma(0) + \sigma^2.$$

Damit folgt

$$\gamma(0) = \frac{\sigma^2}{1 - \phi_1^2}.$$

Wie in obigem Beispiel gesehen, spielt Stationarität einer Zeitreihe eine wichtige Rolle. Im Folgenden werden wir die Stationarität von ARMA Zeitreihen genauer untersuchen.

Im AR(1) Fall hat man eine Darstellung

$$\begin{aligned} Y_t &= \phi_1 Y_{t-1} + Z_t \\ &= Z_t + \phi_1 Z_{t-1} + \cdots + \phi_1^k Z_{t-k} + \phi_1^{k+1} Y_{t-k-1} \\ &\stackrel{?}{=} \sum_{j=0}^{\infty} \phi_1^j Z_{t-j}. \end{aligned} \tag{1.1}$$

Wann gilt letztes Gleichheitszeichen und in welchem Sinne hat man die Konvergenz der Reihe zu verstehen ? Die Antwort darauf gründet im wesentlichen auf folgenden Satz:

Satz 1.2.3. *Sei $(Z_t)_{t\in\mathbb{Z}}$ eine Folge von Zufallsvariablen mit $\sup_{t\in\mathbb{Z}} \mathbb{E}(Z_t^2) < \infty$ und $(\psi_i)_{i\in\mathbb{Z}}$ eine Folge mit $\sum_{i\in\mathbb{Z}} |\psi_i| < \infty$. Dann konvergiert die Reihe*

$$\psi(L)Z_t := \sum_{i\in\mathbb{Z}} \psi_i L^i Z_t$$

absolut mit Wahrscheinlichkeit 1. Ebenso konvergiert die Reihe im quadratischen Mittel.

Für den Beweis verweisen wir auf Brockwell und Davis (1991, p. 83).

Mit obigen Satz erhalten wir sofort: Ist $|\phi_1| < 1$, so konvergiert (1.1) mit Wahrscheinlichkeit 1. Diese Reihe ist also eine Lösung der AR(1)-Gleichung. Weiterhin ist sie stationär. Man zeigt leicht, dass es auch die einzige Lösung ist.

Satz 1.2.4. *Ist $(X_t)_{t\in\mathbb{Z}}$ ein schwach stationärer Prozess mit AKF $\gamma(\cdot)$ und gilt $\sum |\psi_i| < \infty$, so konvergiert*

$$\sum_{i\in\mathbb{Z}} \psi_i L^i X_t =: \psi(L)X_t$$

für alle $t \in \mathbb{Z}$ mit Wahrscheinlichkeit 1. Setzen wir

$$Y_t := \psi(L)X_t,$$

so ist Y_t wieder schwach stationär mit AKF

$$\gamma_y(h) = \sum_{j,k\in\mathbb{Z}} \psi_j \psi_k \gamma(h - j + k)$$

Beweis. Ist (X_t) schwach stationär, so ist $\mathbb{E}|X_t| \le (\mathbb{E}(X_t^2))^{\frac{1}{2}} = \sigma^2 < \infty$. $\rightarrow$ Anwendung von Satz 1.2.3. Für die Stationarität von Y nutzen wir

$$\begin{aligned} \mathbb{E}(Y_t) &= \lim_{n\to\infty} \sum_{j=-n}^{n} \psi_j \mathbb{E}(X_{t-j}) = \mathbb{E}X_t \cdot \sum_{j\in\mathbb{Z}} \psi_j \\ \mathbb{E}(Y_{t+h}Y_t) &= \lim_{n\to\infty} \mathbb{E}\Big[\Big(\sum_{j=-n}^{n} \psi_j X_{t+h-j} \Big) \cdot \Big(\sum_{k=-n}^{n} \psi_k X_{t-k} \Big) \Big] \\ &= \sum_{j,k\in\mathbb{Z}} \psi_j \psi_k [\gamma(h - j + k) + \underbrace{\mathbb{E}(X_t)^2}_{=\sigma^2}]. \end{aligned}$$

Aus der letzten Zeile liest man leicht die AKF ab. ∎

Betrachtet man die Darstellung

$$\phi(L)Y_t = c + \theta(L)Z_t$$

so erhält man, falls ϕ invertierbar ist,

$$Y_t = \phi^{-1}(L)c + \phi^{-1}(L)\theta(L)Z_t$$

also eine Darstellung des ARMA Prozesses als MA(∞)-Prozess. ARMA Prozesse bilden also eine Teilklasse/Approximation von MA(∞) Prozessen. Die Bedeutung von MA(∞) Prozessen stammt von folgendem Satz:

Theorem 1.2.5. (Wold-Zerlegung) *Ist (Y_t) eine nicht-deterministische, schwach stationäre Zeitreihe, so hat Y die Darstellung*

$$Y_t = \sum_{j=0}^{\infty} \psi_j Z_{t-j} + V_t,$$

mit deterministischem (V_t) und $\psi_0 = 1, \sum_{j=0}^{\infty} \psi_j^2 < \infty, Z_t \sim WN(0,\sigma^2)$. Z und V sind jeweils unkorreliert.

Beispiel 1.2.6. Im folgenden werden einige Beispiele aus der Anwendung vorgestellt. Insbesondere das letzte Beispiel ist in der Finanzmathematik von Interesse.

(i) **Bilineare Modelle**. Sie sind Linear in Y, Z aber nicht in beiden Variablen:

$$Y_Z = Z_t + \frac{1}{2} Y_{t-2} Z_{t-1}, \qquad \epsilon_t \text{ iid } \mathcal{N}(0,\sigma^2)$$

$\Rightarrow \mathbb{E}(Y_t) = 0$ und

$$\mathbb{E}(Y_t Y_{t-h}) = \begin{cases} h \geqslant 2: & 0 \\ h = 1: & 0 \\ h = 0: & \mathbb{E}(Y_t^2) = \sigma^2 + \left(\frac{1}{2}\sigma\right)^2 \mathbb{E}(Y_{t-2}^2) \end{cases}$$

Man erhält Konvergenz für $\frac{1}{2}\sigma < 1$.

(ii) **Stochastische Autoregression**.

$$Y_t = Z_t + \begin{cases} B_1 Y_{t-1} & \text{m.W. } p_1 \\ B_2 Y_{t-2} & p_2 \\ 0 & 1 - p_1 - p_2 \end{cases}$$

Ist $0 < \beta_1, \beta_2 < 1$, so kann man die Vert. v. Z_t so wählen, dass $Y_t \sim \text{Exp}(1)$.

(iii) **Zeittransformierte Prozesse**.

$(Y_n)_{n \in \mathbb{N}}$ ist der Preis an Tick n (Trading time) und $(Y_{N_t})_{t \geq 0}$ ist der Prozess in Kalenderzeit. Wir nehmen an, dass Y, N unabhängig sind und weiterhin N iid Zuwächse hat. Dann gilt

$$Y_n = \phi Y_{n-1} + \sigma Z_n, \qquad Z_n \text{ iid } \mathcal{N}(0,1).$$

Hieraus folgt,

$$Y_{N_t} = \phi^{N_t - N_{t-1}} Y_{N_{t-1}} + \sigma Z_{N_t} + \phi\sigma Z_{N_{t-1}} + \phi^{N_t - N_{t-1} - 1}\sigma Z_{N_{t-1}}$$

wobei $\sigma Z_{N_t} + \phi\sigma Z_{N_{t-1}} + \phi^{N_t - N_{t-1} - 1}\sigma Z_{N_{t-1}}$ normalverteilt, mit Erwartungswert 0 und Varianz

$$\sigma^2 \frac{1 - \phi^{2(N_t - N_{t-1})}}{1 - \phi^2}.$$

Betrachten wir Erwartungswert und Varianz. Wir setzen $N_t - N_{t-1} := \mu_t$. Dann

$$\begin{aligned} \mathbb{E}(Y_t|Y_{t-1}, N_t - N_{t-1}) &= \phi^{\mu_t}\, Y_{t-1} \\ \mathrm{Var}(Y_t|Y_{t-1}, N_t - N_{t-1}) &= \sigma^2 \frac{1 - \phi^{2\mu_t}}{1 - \phi^2} \end{aligned}$$

und somit $\mathbb{E}(Y_t|Y_{t-1}) = \mathbb{E}(\phi^{\mu_t}) \cdot Y_{t-1}$. Wir setzen $r := \mathbb{E}(\phi^{\mu_t})$. Dann

$$\begin{aligned} \mathrm{Var}(Y_t|Y_{t-1}) &= \mathbb{E}(Y_t^2|Y_{t-1}) - \mathbb{E}(Y_t|Y_{t-1})^2 \\ &= \mathbb{E}\Big(\mathbb{E}(Y_t^2|Y_{t-1}, \mu_t)\big|Y_{t-1}\Big) - \mathbb{E}\Big(\mathbb{E}(Y_t|Y_{t-1}, \mu_t)\big|Y_{t-1}\Big)^2 \\ &= \sigma^2 \frac{1 - \mathbb{E}(\phi^{2\mu_t})}{1 - \phi^2} + Y_{t-1}^2 \Big[\underbrace{\mathbb{E}\big(\phi^{2\mu_t}\big) - r^2}_{=\mathrm{Var}(\phi^{\mu_t})}\Big] \end{aligned}$$

Die Eigenschaft, dass die bedingte Varianz von vorherigen Werten der Zeitreihe (durch Y_{t-1}) abhängt, nennt man *Heteroskedastizität.* Somit hat dieser Prozess ebenso wie ARCH-Prozesse diese Eigenschaft. Aus einem Autoregressiven Prozess entstand durch die Zeittransformation ein Prozess mit völlig anderen Eigenschaften!

Definition 1.2.7. *Ein ARMA(p,q) Prozess* (Y_t) *heißt kausal, falls es eine Folge von Konstanten* Ψ_j *mit* $\sigma|\Psi_j| < \infty$ *gibt, und*

$$Y_t = \sum_{j=0}^{\infty} \Psi_j Z_{t-j}$$

D.h. Der Prozess Y hängt nur von vergangenen Z_t (und nicht etwa von zukünftigen "nichtkausal" "überirdischen" Werten ab!)

Theorem 1.2.8. *Ein ARMA (p,q) Prozess, dessen Polynome* ϕ *und* θ *keine gemeinsamen Nullen besitzen, ist kausal genau dann, wenn* $\phi(z) \neq 0 \quad \forall z \in C$ *mit* $|z| < 1$. Ψ_j *ergeben sich nach*

$$\Psi(z) = \sum_{j=0}^{\infty} \Psi_j z^j = \frac{\theta(z)}{\phi(z)}.$$

Beweis. Mit $\phi(z) \neq 0 \quad \forall |z| \leq 1$ erhält man die Reihendarstellung

$$\frac{l}{\phi(z)} = \sum_{j=0}^{\infty} \zeta_j z^j =: \zeta(z) \qquad |z| \leq 1$$

Man erhält die Darstellung

$$Y_t = \zeta(L)\theta(L) Z_t$$

und Y ist kausal. Nehmen wir umgekehrt an, dass Y kausal ist, so liefert ein Koeffizientenvergleich genau diese Darstellung. ∎

1.2.1 Schätzen schwach stationärer Prozesse

Ein schwach stationärer Prozess ist beschrieben durch m und γ. Schätzen dieser beiden Objekte ist also äußerst wichtig.

m wird hierbei durch den arithmetischen Mittelwert geschätzt:

$$\hat{m}_n := \frac{1}{n}\sum_{i=1}^{n} Y_i.$$

Lemma 1.2.9. *$\hat{m}_n$ ist unverzerrt und*

$$\operatorname{Var}(\hat{m}_n) = \frac{1}{n}\sum_{h=-n}^{n}\Big(1 - \frac{|h|}{n}\gamma(h)\Big)$$

Beweis. Zunächst ist

$$\mathbb{E}(\hat{m}_n) = \frac{1}{n}\sum_{i=1}^{n}\mathbb{E}(Y_i) = m.$$

Weiterhin ist

$$\operatorname{Var}(\hat{m}_n) = \frac{1}{n^2}\operatorname{Var}\left(\sum_{i=1}^{n} Y_i\right) = \frac{1}{n^2}\sum_{i,j}\operatorname{Cov}(Y_i, Y_j)$$

Bei einem stationären Prozess ist gerade $\operatorname{Cov}(Y_i, Y_j) = \gamma(i-j)$. Hierbei gibt es das Paar $i = j$ n-Mal, $i - j = 1$ gerade $n - 1$ und $i - j = n$ gerade $(n - (n-1)) = 2$ Mal, und somit

$$\operatorname{Var}(\hat{m}_n) = \frac{1}{n^2}\sum_{h=-n}^{n}(n - |h|)\,\gamma(h). \qquad \blacksquare$$

Der folgende Satz gibt eine erste Beschreibung des asymptotischen Verhaltens von $\hat{m}_n$.

Satz 1.2.10. *Für eine schwach stationäre Zeitreihe $(Y_t)_{t\in\mathbb{Z}}$ mit $\gamma(n) \to 0$ gilt*

$$\operatorname{Var}(\hat{m} - n) \to 0.$$

Ist weiterhin $\sum|\gamma(h)| < \infty$, so gilt auch

$$\operatorname{Var}(\sqrt{n}\,\hat{m}_n) \to \sum_{k\in\mathbb{Z}}\gamma(h)$$

Beweis.

(i) Mit obigem Lemma ist

$$\operatorname{Var}(\hat{m}_n) \le \frac{1}{n}\sum_{k=-n}^{n}\gamma(h).$$

Mit dem Grenzwertsatz v. Cauchy folgt aus $\gamma(n) \to 0$, dass auch die rechte Seite gegen Null konvergiert.

(ii) Wir werden erwarten, dass $\sqrt{n}\,\hat{m}_n$ gegen eine Zufallsvariable konvergiert! Aus dem vorherigen Lemma kennen wir bereits den EW (m) und es ist allso nützlich, auch die Varianz zu kennen!

Nach dem vorherigen Lemma ist

$$\mathrm{Var}(\sqrt{n}\hat{m}_n) = \sum_{h=-n}^{n} \gamma(h) - \sum_{h=-n}^{n} \frac{|h|}{n}\gamma(h).$$

Es bleibt also zu zeigen, dass der letzte Ausdruck gegen Null konvergiert. Dies folgte wieder aus dem Cauchy'schen GWS, falls nur $|h|\gamma(h) \to \sigma$. Da $\sum |\gamma(h)| < \infty$, muss dies aber gelten (Die Reihe über $\frac{1}{n}$ divergiert!) ∎

Um einen zentralen Grenzwertsatz anwenden zu können, ist es immer nützlich, *iid* Variablen zu haben! Nicht verwunderlich, dass dies im folgenden Theorem auftaucht:

Theorem 1.2.11. *Ist (Y_t) gegeben durch den stationären Prozess*

$$Y_t = \mu + \sum_{j=-\infty}^{\infty} \Psi_j\, Z_{t-j}, \quad Z_t \sim iid \text{ mit } \mathbb{E}(Z_t) = \sigma, \mathrm{Var}(Z_t) = \sigma^2$$

und gilt $\sum |\Psi_j| < \infty$ sowie $\sum \Psi_j \neq \sigma$, so ist

$$\sqrt{n}\,\hat{m}_n \xrightarrow[n\to\infty]{\mathscr{L}} \mathcal{N}\Big(\mu, \sum_{h\in\mathbb{Z}} \gamma(h)\Big).$$

Des weiteren ist

$$\sum_{n\in\mathbb{Z}} \gamma(h) = \sigma^2 \left[\sum_{j\in\mathbb{Z}} \Psi_j\right]^2.$$

Beweis. B/D sec. 7.3 ∎

Neben der Schätzung von m muss man sich natürlich auch mit der Schätzung von γ befassen! Allerdings hat die Stichproben AKF nicht so vorteilhafte Eigenschaften, weswegen wir sie nur kurz behandeln werden. Sie ist definiert durch:

$$\hat{\gamma}_n(h) := \frac{1}{n} \sum_{t=1}^{n-|h|} (Y_{t+|h|} - \hat{m}_n)(Y_t - \hat{m}_n)$$

Die obige Formel macht bereits von der Symmetrie der AKF bei stationären Zeitreihen Gebrauch!

Wie die Stichprobenkovarianz ist obiger Schätzer *verzerrt.*

$$\mathbb{E}(\hat{\gamma}_n) = \frac{1}{n} \sum_{t=1}^{n} \mathbb{E}\big(\underbrace{(Y_{t+|h|} - \hat{m}_n)(Y_t - \hat{m}_n)}_{Y_{t+|h|}Y_t - Y_{t+|h|}\hat{m}_n - Y_t\hat{m}_n + (\hat{m}_n)^2} \big)$$

Natürlich ist

$$\begin{aligned}\mathbb{E}(Y_{t+|h|}Y_t) &= \mathbb{E}\left((Y_{t+|h|} - m + m)(Y_t - m + m)\right) \\ &= \gamma(h) + m^2\end{aligned}$$

und

$$\mathbb{E}(Y_t \frac{1}{n}\sum_{j=1}^{n} Y_i) = \left[\frac{1}{n}\sum_{j=1}^{n}\gamma(t-i)\right] + m^2.$$

Aus Lemma folgt $\mathbb{E}(\hat{m}_1^2) = \frac{l}{n}\sum_{|h|\leq n}(1 - \frac{|h|}{n})\gamma(n) + m^2$ und somit

$$\begin{aligned}\mathbb{E}(\hat{\gamma}_n) &= \gamma(h) - \frac{1}{n^2}\sum_{t,j}\left[\gamma(t - i|h| - j) + \gamma(t-j)\right] + \frac{1}{n}\sum_{|h|\leq n}(1 - \frac{|h|}{n})\gamma(h) \\ &= \gamma(h) - \frac{1}{n^2}\sum_{t,i}\left[\gamma(t + |h| - i) + \gamma(t-i)\right] + \frac{1}{n}\sum_{|h|} \leq n(1 - \frac{|h|}{n})\gamma(h)\end{aligned}$$

Erinnern wir uns an den Beweis von Lemma 1.2.9. Danach ist $\frac{1}{n}\sum_{t,i}\gamma(t-i)$ gerade gleich $\sum_{|h|\leq n}(1 - \frac{|n|}{n}\gamma(h))$ und wir erhalten

$$\mathbb{E}(\hat{\gamma}_n) = \gamma(h) - \frac{1}{n^2}\sum_{t,i}\gamma(t + |h| - i)$$

Man erkennt das Problem: Wäre $h = 0$, so wäre obige Gleichung gerade 0. h ist so eine Art Verschiebung in der Stichprobe $1, \ldots, n$. Spielt diese Verschiebung keine große Rolle, ist also n sehr großen Vergleich zu n, so ist $\hat{\gamma}_n$ *fast* unverzerrt.

Ebenso sind Schätzungen für $h \sim n$ natürlich nicht sehr sinnvoll, schließlich stehen nur wenige Daten zur Verfügung. Man sagt eine Schätzung ist sinnvoll (Box, Jenkins) falls $n \geq 50$ und $h \leq \frac{n}{t}$.

Immerhin kann man zeigen, dass $\hat{\gamma}_n \to \gamma$ und ebenso asymptotische Normalität. Siehe Brockwell und Davis (1991).

1.2.2 Schätzung von ARMA-Modellen

In diesem Abschnitt beschrieben wir (aber nur kurz) die Schätzung von ARMA-Modellen.

Zunächst nehmen wir einmal an, p und q seien bekannt. Des weiteren arbeiten wir mit zentrierten Daten und $c = 0$.

Yule-Walker Gleichungen

Für einen Autoregressiven Prozess AR(p) kann man wir folgt vorgehen:

$$Y_t = \phi_1 Y_{t-1} + \cdots + \phi_p Y_{t-p} + Z_t,$$

Z_t White Noise $(0, \sigma^2)$

Damit ist auch

$$\begin{aligned}\mathbb{E}\left(Y_t Y_{t+j}\right) &= \phi_1 \mathbb{E}\left(Y_t t - 1 Y_{t+j}\right) + \cdots + \phi_p \mathbb{E}\left(Y_{t-p} Y_{t+j}\right) + 0 \\ \gamma(j) &= \phi_1(j-1) + \cdots + \phi_p \gamma(j-p)\end{aligned}$$

Man erhält die sog. **Yule-Walker Gleichungen**

$$\begin{pmatrix}\gamma(1)\\ \vdots \\ \gamma(p)\end{pmatrix} = \begin{pmatrix}\gamma(0) & \cdots & \gamma(p)\\ \vdots & \ddots & \vdots \\ \gamma(p) & \cdots & \gamma(0)\end{pmatrix}\begin{pmatrix}\phi_1\\ \vdots \\ \phi_p\end{pmatrix}$$

Kurz kann man das durch $\gamma_p = \Gamma_p \phi$ schreiben, wobei $\Gamma_p = \big(\gamma(i-j)\big)_{i,j=1}^p$. Weiterhin ist $\sigma^2 = \gamma(0) - \langle \phi, \gamma_p \rangle$.

Diese Gleichungen kann man nutzen, um die AKF γ aus σ und ϕ zu berechnen.

Andererseits, schätzen wir σ und γ durch $\hat{\sigma}$ bzw. $\hat{\gamma}$, so erhält man

$$\hat{\Gamma_p}\,\hat{\phi} = \hat{\gamma}_p \quad \text{und } \hat{\sigma}^2 = \hat{\gamma}(0) - \langle \hat{\phi}, \hat{\gamma}_p \rangle,$$

wobei natürlich $\hat{\Gamma_p} = \big(\hat{\gamma}(i-j)\big)_{i,j=1}^p$.

Ist nun $\hat{\gamma}(0) > 0$, so ist $\hat{\Gamma_m}$ nicht singulär für alle m, und wir können nach ϕ auflösen und erhalten die **Stichproben-Yule-Walker Gleichungen**:

$$\hat{\phi} = \hat{\Gamma_p}^{-1} \hat{\gamma}_p$$

und

$$\hat{\sigma}^2 = \hat{\gamma}(0) - \underbrace{\left(\hat{\Gamma_p}^{-1} \hat{\gamma}_p\right)' \hat{\gamma}_p}_{\hat{\gamma}_p' \hat{\Gamma_p}^{-1} \hat{\gamma}_p}.$$

Ebenso haben wir asymptotische Normalität und Erwartungstreue:

$$\hat{\phi} \xrightarrow[n\to\infty]{\mathscr{L}} \mathcal{N}\Big(\phi, \frac{1}{n}\sigma^2 \Gamma_p^{-1}\Big).$$

Der Yule-Walker Schätzer ergibt die beste lineare Vorhersage von X_{t+1} auf der Basis von $\{X_1, \ldots, X_t\}$ unter der Annahme, dass die AKF der X gleich der geschätzten AKF ist.

1.2.3 Univariate ARCH Modelle

Eine statistische Analyse der zweiten Modelle von Finanzdaten zeigt, dass die bedingte Varianz (od. das bedingte zweite Moment) nicht konstant ist. Im Gegenteil, man erhält eine AKF ähnlich wie für AR Modelle. Damit ist bereits klar, dass sich ARMA Modelle nicht besonders für Finanzdaten eignen. In diesem Abschnitt stellen wir geeignete Verallgemeinerungen dar.

Das ARCH (1) Modell.

Das von Engle 1982 vorgestellte ARCH-Modell sieht im einfachsten Fall wie folgt aus:

Sei S(Z_t) iid mit $\mathbb{E}(Z_t) = 0, \mathrm{Var}(Z_t) = 1$. Wir schreiben $Z_t \sim SWN(0,1)$.

Definition 1.2.12. *Ein Prozess $(Y_t)_{t\in\mathbb{Z}}$ heißt ARCH(1)-Prozess, falls für alle $t \in \mathbb{Z}$ die folgenden Eigenschaften erfüllt sind:*

$$\begin{array}{ll} (i) & Y_t = \sigma_t Z_t \\ (ii) & \sigma_t^2 = \alpha_0 + \alpha_1 Y_{t-1}^2, \quad \alpha_0, \alpha_1 > 0 \end{array} .$$

Es folgt direkt

$$\mathbb{E}(Y_t|Y_{t-1}) = 0 \qquad \text{und } \mathrm{Var}(Y_t|Y_{t-1}) = \sigma_t^2.$$

In Engle (1982) waren die Z_t sogar iid $\mathcal{N}(0,1)$. Dann ist die *bedingte* Verteilung von Y_t gegeben Y_{t-1} normal. Die unbedingte Verteilung ist eine sogenannte Mischung, gegeben durch

$$\begin{aligned} &\mathbb{E}\left(\mathbb{P}(Y_t \le x|Y_{t-1})\right) \\ &= \mathbb{E}\left(\Phi\Big(\frac{x}{\sigma_t}\Big)\right) = \int \Phi\Big(\frac{x}{y}\Big) f_{\sigma_t}(y)dy, \end{aligned}$$

wobei f_{σ_t} die unbedingte Verteilung von σ_t ist. Diese Verteilung hat viel dickere tails als die Normalverteilung.

Bemerkung: Für einen ARCH(1) Prozess gelten folgende Eigenschaften:

Ist Y schwach stationär, so ist $\mathrm{Var}(Y_t) = \frac{\lambda_0}{1-\lambda_1}$. Des weiteren ist

$$\begin{aligned} \mathrm{Cov}(Y_t; Y_{t-h}) &= \mathbb{E}(Y_t Y_{t-h}) \\ &= \mathbb{E}\left(\mathbb{E}(Y_t|Y_{t-1,\dots,}Y_{t-1})Y_{t-h}\right) = 0 \end{aligned}$$

und somit die AKF des Prozesses für alle $h \neq 0$ gerade 0. Interessant ist insbesondere ii) in obiger Definition. Wir erhalten aus $\mathbb{E}(Y_t|Y_{t-1}) = 0$

$$\mathrm{Var}(Y_t|Y_{t-1}) = \mathbb{E}(Y_t^2|Y_{t-1}) = \lambda_0 + \lambda_1 Y_{t-1}^2$$

also

$$\begin{aligned} Y_t^2 &= \mathbb{E}(Y_t^2|Y_{t-1}) + \underbrace{Y_t^2 - \mathbb{E}(Y_t|Y_{t-1})}_{=:\upsilon_t} \\ &= \lambda_0 + \lambda_1 Y_{t-1}^2 + \upsilon_t. \end{aligned}$$

Hierbei ist $\mathbb{E}(\upsilon_t|\upsilon_{t-1}) = 0$. Man kann zeigen, dass unter $\mathbb{E}(Y_t^4) < \infty$ gilt: $\mathrm{Var}(\upsilon_t) = \text{const.}$ Da ebenso $\mathrm{Cov}(\upsilon_t, \upsilon_{t-1}) = 0$ ist in diesem Falle also $(\upsilon_t)_{t\in\mathbb{Z}}$ ein weißes Random und der Prozess $(Y_t^2)_{t\in\mathbb{Z}}$ ein AR(1) Prozess.

Für die strikte Stationarität eines ARCH(1)-Modells ist eine hinreichende und notwendige Bedingung, das $\mathbb{E}(\ln(\alpha_1 Z_t^2)) < 0$, also $\ln\alpha_1 + 2\mathbb{E}(\ln|Z_t|) < 0$. Eine Bedingung an α_1 ist demnach mit der Verteilung von Z eng verbunden. Ist $Z \sim \mathcal{N}(0,1)$, so ist die s gerade $\alpha_1 < 3.562$.

Im Gegensatz dazu ist schwache Stationarität immer (unabhängig von der Verteilung der Z's) an $\alpha_1 < 1$ geknüpft.

Satz 1.2.13. *Der ARCH(1)-Prozess ist schwach stationär dann und nur dann, wenn* $\alpha_1 < 1$.

Beweis. Wie bereits erwähnt, ist im Falle schwacher Stationarität $\mathrm{Var}(Y_t) = \frac{\alpha_0}{1\alpha_1}$, welche existiert für $\alpha_1 < 1$. Umgekehrt folgt aus $Y_t = \sigma_t Z_t$

$$\begin{aligned} Y_t^2 &= Z_t^2(\alpha_0 + \alpha_1 Y_{t-1}^2) && (1.2)\\ &= Z_t^2(\alpha_0 + \alpha_1 Z_{t-1}^2(\alpha_0 + \alpha_1 Y_{t-2}^2)) && (1.3)\\ &= \alpha_0 Z_t^2 + \alpha_1 Z_t^2 Z_{t-1}^2(\alpha_0 + \alpha_1 Y_{t-2}^2) = \alpha_0 Z_t^2 + \alpha_0\alpha_1 Z_t^2 Z_{t-1}^2 + \dots \\ &= \alpha_0 \sum_{i=0}^{\infty} \alpha_1^i \prod_{j=0}^{i} Z_{t-j}^2 + \lim_{i\to\infty} \alpha_1^i \prod_{j=0}^{i} Z_{t-j}^2 Y_{t-j}^2. \end{aligned}$$

Wie bereits bei den ARMA-Prozessen kann man (aber mit mehr technischem Aufwand) zeigen, dass der letzte Term verschwindet (unter $\mathbb{E}(\ln\alpha_1, Z_t^2) < 0$) und die Summe absolut konvergiert, und zwar mit Wahrscheinlichkeit 1.

Aus $\alpha_1 < 1$ folgt auch

$$\mathbb{E}\left(\ln(\alpha_1 Z_t^2)\right) \le \ln \mathbb{E}(\alpha_1 Z_t^2) = \ln\alpha_1 < 0,$$

und somit

$$\mathbb{E}(Y_t^2) = \mathbb{E}\left[(\alpha_0 \sum_{i=0}^{\infty} \alpha_1^i \prod_{j=0}^{i} Z_{t-j}^2)\right] = \alpha_0 \sum_{i=0}^{\infty} \alpha_1^i = \frac{\alpha_0}{1-\alpha_1}.$$

Des weiteren ist $\mathrm{Cov}(Y_t; Y_{t+n}) = 0$, also ist Y stationär.

∎

Bemerkung 1.2.14. Darüber hinaus ist Y sogar weißes Rauschen! Also geeignet als Fehlerverteilung für ARMA-Prozesse. Ist die Varianz von Y gleich ∞, so ist Y **nicht** schwach stationär, aber für $\mathbb{E}(\ln\alpha_1 Z_t^2) < 0$ möglicherweise strikt stationär!!

Wie sieht es mit Momenten von Y aus?

Satz 1.2.15. *Betrachte* $m \ge 1$ *und* $\alpha_1 < 1$. *Dann hat der ARCH(1)-Prozess endliche Momente der Ordnung* $2m$ *genau dann, wenn folgende Bedingungen erfüllt sind:*

(i) $\mathbb{E}(Z_t^{2m}) < \infty$

(ii) $\alpha_1 < \frac{1}{\sqrt[m]{\mathbb{E}(Z_t^{2m})}}$

Beweis. Wieder nutzen wir (1.2). Dabei ist

$$\mathbb{E}(Y_t^{2m}) = \mathbb{E}[(\alpha^0 \sum_{i=0}^{\infty} \alpha_1^i \prod_{j=0}^{i} Z_{t-j}^2)^m] = \mathbb{E}(Z_t^2)\mathbb{E}\left((\sum_{i=0}^{\infty} \alpha_0\alpha_1^i \prod_{j=1}^{i} Z_{t-j}^2)^m\right)$$

wobei wir $\prod_\emptyset = 1$ setzen. Aus der Minkowski-Ungleichung folgt

$$\left(\mathbb{E}(\sum_{i=0}^{\infty} |\zeta_i|^m)\right)^{\frac{1}{m}} \le \sum_{i=0}^{\infty} \mathbb{E}(|\zeta_i|^m)^{\frac{1}{m}},$$

also

$$\mathbb{E}(Y_t^{2m}) \leq \mathbb{E}(Z_t^2)\Bigg(\sum_{i=0}^{\infty} \underbrace{\mathbb{E}(\alpha_0^m \alpha_1^{im} \prod_{j=1}^{i} Z_{t-j}^{2m})^{\frac{1}{m}}}_{\alpha_0^m \alpha_1^{im}}\Bigg)^m = \mathbb{E}(Z_t^2)\Bigg(\alpha_0 \sum_{i=0}^{\infty} \alpha_1^i \mathbb{E}(Z_t^{2m})^{\frac{i}{m}}\Bigg)^m$$

also muss gelten $\alpha_1^i \mathbb{E}(Z_t^{2m})^{\frac{i}{m}} < 1 \Leftrightarrow \alpha_1 < \frac{1}{\sqrt[m]{\mathbb{E}(Z_t^{2m})}}$. ■

Als Beispiel erhalten wir für $Z \sim \mathcal{N}(0,1)$ und $m = 2$ und $\alpha_1 < \frac{1}{\sqrt{3}}$ vierte Momente.

Existiert das vierte Moment, so ist

$$\mathbb{E}(Y_t^4) = \frac{1+\alpha_1}{1-\alpha_1} \cdot \frac{\alpha_0^2 \mathbb{E}(Z_t^4)}{1-\alpha_1^2 E(Z_t^4)}$$

1.2.4 GARCH-Prozesse

Definition 1.2.16. *Sei $(Z_t)_{t\in\mathbb{Z}}$ striktes weißes Rauschen. Dann heißt ein Prozess $(Y_t)_{t\in\mathbb{Z}}$* **GARCH(p,q)***-Prozess, falls folgende Bedingungen erfüllt sind:*

(i) Y ist strikt stationär

(ii) $\forall\, t \in \mathbb{Z}$ und geeignete $\alpha_0 > 0, \alpha_i, \beta_j \geq 0 \quad i = 1,\dots,p\ j = 1,\dots,q$ gilt

$$\begin{aligned} Y_t &= \sigma_t Z_t \\ \sigma_t^2 &= \alpha_0 + \sum_{i=1}^{p} \alpha_i Y_{t-i}^2 + \sum_{j=1}^{q} \beta_i \sigma_{t-i}^2. \end{aligned}$$

Dabei ist ein GARCH(p,0)-Prozess gerade ein ARCH(p)-Prozess. Die Verallgemeinerung ist ähnlich wie von AR zu ARMA, aber **nicht identisch!** Wie erwartet, werden die Y_t^2 gerade einen ARMA-Prozess bilden. So ist

$$\begin{aligned} Y_t^2 &= \mathbb{E}(Y_t^2|Y_{t-1}, Y_{t-2}, \dots) + \underbrace{Y_t - \mathbb{E}(Y_t|Y_{t-1}, Y_{t-2}, \dots)}_{=:V_t} \\ &= \alpha_0 + \sum_{i=1}^{p} \alpha_i Y_{t-i}^2 + \sum_{j=1}^{q} \beta_i \sigma_{t-i}^2. \end{aligned}$$

Dabei ist $V_t = \sigma_t^2(Z_t^2 - 1) = Y_t^2 - \sigma_t^2$. Wir erhalten

$$Y_t^2 = \alpha_0 + \sum_{i=1}^{max(p,q)} (\alpha_i + \beta_i) Y_{t-i}^2 - \sum_{j=1}^{q} \beta_i V_{t-i}^2 + V_t,$$

wobei wir $\alpha_i = 0$ und $\beta_j = 0$ für $i > p$ bzw. $j > q$ setzen. Das GARCH-Modell führt also zu einer ARMA-Darstellung in den Quadraten!

Man sieht bereits, dass große σ's nun aufgrund zweier Ursachen entstehen: Einerseits - wie bei ARCH - große Y_{t-i}^2, andererseits kann dies aber auch schlicht durch große, σ_{t-i}^2 ausgelöst werden. Man erhält bei GARCH also längere Zyklen höherer Volatilität als bei ARCH.

Im folgenden wollen wir zunächst die theoretischen Eigenschaften genauer betrachten.

Ähnlich wie bei ARCH kann man die Frage nach Stationarität mit Stochastischen Rekurrenz Gleichungen lösen. Wir werden nur kurz die entsprechenden Ergebnisse zitieren.

Satz 1.2.17. *Nelson (1990). Ein GARCH (1,1)-Prozess ist strikt stationär genau dann, wenn*

$$\mathbb{E}\left(\ln(\alpha_1 Z_0^2 + \beta_1)\right) < 0. \tag{1.4}$$

Der allgemeine GARCH(p,q)-Fall wurde 1992 von Bougerol und Picard geklärt. Lange Zeit vorher hatte man stets $\sum \alpha_i + \sum \beta_i < 1$ angenommen, da unter dieser Voraussetzung die ARMA-Darstellung der Y^2 stationär ist.

Unter obiger Bedingung (1.4) erhält man die Konvergenz von

$$\sigma_t^2 = \alpha_0 + \alpha_0 \sum_{i=1}^{\infty} \prod_{j=1}^{i} (\alpha_1 Z_{t-j}^2 + \beta_1)$$

und damit eine Darstellung der Lösung der GARCH (1,1)-Gleichung

$$Y_t = Z_t \cdot \sqrt{\alpha_0 + \alpha_0 \sum_{i=1}^{\infty} \prod_{j=1}^{i} (\alpha_1 Z_{t-j}^2 + \beta_1)}. \tag{1.5}$$

Unter schwacher Stationarität können wir wieder leicht die Varianz von Y ausrechnen: (beachte $\mathbb{E}(Y_t) = 0$)

$$\begin{aligned} \operatorname{Var}(Y_t) &= \mathbb{E}\left((1 + \alpha_1 Y_{t-1}^2 + \beta_1 \sigma_{t-1}) Z_t^2\right) \\ &= 1 + \alpha_1 \operatorname{Var}(Y_{t-1}^2) + \beta_1 \cdot \operatorname{Var}(Y_{t-1}^2), \end{aligned}$$

also

$$\operatorname{Var}(Y_t) = \frac{\alpha_0}{1 - \alpha_1 - \beta_1}.$$

Ebenso erhalten wir

Satz 1.2.18. *Ein GARCH (1,1)-Prozess ist genau dann stationär, wenn $\alpha_1 + \beta_1 < 1$.*

Ebenso kann Darstellung (1.5) benutzen, um höhere Momente auszurechnen bzw. Bedingungen für deren Existenz herzuleiten.

Wieder erhalten wir

$$\operatorname{Cov}(Y_t; Y_{t-1}) = 0.$$

Beispiele / Erweiterungen

Empirische Untersuchungen von Hoch-frequenz-Daten (täglich oder häufiger) ergeben oft GARCH-Modelle mit $\sum \lambda_i + \beta_i$ in der Nähe von 1. Eine Reihe von Arbeiten beschäftigt sich deswegen mit dem Fall, indem diese Summe $= 1$ ist.

Wir sprechen von einem Integrierten GARCH-Prozess (IGARCH), falls $\sum \alpha_i + \beta_i = 1$. Schauen wir uns die Quadrate dieses Prozesses genauer an:

$$Y_t^2 = \alpha_0 + Y_{t-1}^2 - \beta_1 V_{t-1}^2 + V_t,$$

also

$$\delta Y_t^2 := Y_t^2 - Y_{t-1}^2 = \alpha_0 - (1 - \alpha_1) V_{t-1}^2 + V_t \tag{1.6}$$

Natürlich ist $Y_t^2 = \sum_{j=0}^{\infty} \delta Y_{t-j}^2$, weswegen man einen Prozess wie (1.6) auch einen Integrierten ARMA oder ARIMA-Prozess nennt. Allerdings ist wegen $\mathbb{E} Y_t^2 = \infty$ auch $\mathbb{E}|V_t| = \infty$ und somit V_t kein weißes Rauschen!

ARMA mit GARCH-Fehlern

Unter schwacher Staitonarität verhält sich ein GARCH-Prozess wie schwaches weißes Rauschen. Es bietet sich also an, einen ARMA-Prozess mit GARCH-Fehlern zu betrachten.

Definition 1.2.19. *Sei $(Z_t)_{t\in\mathbb{Z}} SWN(0, \sigma^2)$. Ein Prozess Y heißt ARMA (p_1, q_1)-Prozess mit GARCH (p_2, q_2)-Fehlern, falls*

$$\begin{aligned}
Y_t &= \mu_t + \sigma_t Z_t \\
\mu_t &= \mu + \sum_{i=1}^{p_1} \phi_i (Y_{t-i} - \mu) + \sum_{j=1}^{q_1} \beta_j \sigma_{t-j} Z_{t-j} \\
\sigma_t^2 &= \alpha_0 + \sum_{i=1}^{p_2} \alpha_i (Y_{t-i} - \mu)^2 + \sum_{j=1}^{q_2} \beta_j \sigma_{t-j}^2.
\end{aligned}$$

Hierbei ist $\alpha_0 > 0, \alpha_i, \beta_i \geq 0$ und $\sum \alpha_i + \beta_j < 1$.

Unter geeigneten Bedingungen sind μ_t und σ_t $\mathcal{F}_{t-1} = \sigma(Y_{t-1}, \dots)$-messbar, und man erhält

$$\mathbb{E}(Y_t|\mathcal{F}_{t-1}) = \mu_t, \ \mathrm{Var}(Y_t|\mathcal{F}_{t-1}) = \sigma_t^2.$$

Der Fortschritt im Mittel wird demnach durch den ARMA-Port bestimmt, und die Varianzen durch den GARCH-Teil. Auf diesem Prozess kann man die Zeitreihentechniken für ARMA-Prozess anwenden.

Asymmetrische GARCH-Modelle

Typischerweise beobachtet man in Finanzzeitreihen ein unterschiedliches Verhalten, je nachdem, ob die Aktie eben steigt oder fällt. Man kann sich gut vorstellen, dass Investoren immer nervöser werden, wenn eine Aktie fällt und somit die Volatilität steigt, wenn eine Aktie fällt.

Definition 1.2.20. *Sei* $(Z_t)_{t\in\mathbb{Z}}$ *SWN* $(0,\sigma^2)$*. Ein Prozess* $(Y_t)_{t\in\mathbb{Z}}$ *heißt AGARCH (1,1)-Prozess, falls er strikt stationär ist und für ein* $\delta \in [-1,1]$

$$\begin{aligned} Y_t &= \sigma_t Z_t \\ \sigma_t^2 &= \alpha_0 + \alpha_1 (Y_{t-1} + \delta |Y_{t-1}|)^2 + \beta_1 \sigma_{t-1}^2. \end{aligned}$$

Man beachte, dass

$$Y_{t-1} + \delta|Y_{t-1}| = Y_{t-1} \begin{cases} 1+\delta & \text{falls } Y \geq 0 \\ 1-\delta & \text{sonst.} \end{cases}$$

Dieses Modell ist also in der Lage, negative und positive Renditen zu unterscheiden.

Ähnlich motiviert sind sogenannte Threshold (TGARCH)Modelle. Hier ist

$$\sigma_t^2 = \alpha_0 + Y_{t-1}^2(\alpha_0 + \tilde{\alpha}_0 1_{\{Y_{t-1}<0\}}) + \beta_1 \sigma_{t-1}^2.$$

Andere Möglichkeiten, Asymetrien herbeizuführen, besteht auch durch Nutzung einer asymetrischen Fehlerverteilung (etwa verallgemeinert hyperbolisch ...)

Das ARCH-M Modell

In seiner Arbeit stellte Duan 1995 das ARCH-M Modell vor. Im Unterschied zu vielen anderen, statistich motivierten ARCH-Modelen, gelingt es im ARCH-M Modell ein äquivalentes Martingalmaß anzugeben. Dies bedeutet, dass das betrachtete Modell frei von Arbitrage ist.

In einem Arbitrage-freien Modell müssen diskontierte Aktienpreise unter dem risikoneutralen Maß Q Martingale sein, d.h. es muss

$$\mathbb{E}^Q(e^{-rT} S_T | \mathcal{F}_t) = e^{-rt} S_t$$

gelten. Der Einfachheit halber betrachten wir das Modell direkt unter Q und lassen das Q am Erwartungswert im folgenden weg. Obige Gleichung ist äquivalent zu

$$e^{-r(T-t)} \mathbb{E}\Big({}^{S_T}/_{S_t} | \mathcal{F}_t\Big) = 1.$$

Man muss also in der Lage sein, derartige Erwartungswerte zu berechnen. Ein ARCH-M(p) Modell ist definiert durch

$$\begin{aligned} Y_t &= -\frac{1}{2}\sigma_t^2 + \sigma_t Z_t, \qquad (1.7) \\ \sigma_t^2 &= \alpha_0 + \alpha_1 \sigma_{t-1}^2 Z_{t-1}^2 + \cdots + \sigma_{t-p}^2 Z_{t-p}^2. \end{aligned}$$

Satz 1.2.21. *Gilt* (1.7)*, wobei σ_t messbar bezüglich $\mathcal{F}_{t-1} := \sigma(Y_{t-1}, Y_{t-2}, \dots)$ ist und sind die Z_i i.i.d. standardnormalverteilt, so ist $\mathbb{E}^Q(\exp(Y_T)|\mathcal{F}_t) = \exp(Y_t)$.*

Beweis. Es genügt, die Behauptung für $T = t+1$ zu zeigen. Für die Laplacetransformierte einer Normalverteilung gilt

$$\begin{aligned}\mathbb{E}\Big(\exp(\lambda\mathbb{Z})\Big) &= \int e^{\lambda z}\frac{1}{\sqrt{2\pi}}e^{-\frac{z^2}{2}}dz \\ &= \int \frac{1}{\sqrt{2\pi}}\exp\Big(-\frac{(z-\lambda)^2-\lambda^2}{2}dz = \exp\Big(\frac{\lambda^2}{2}\Big).\end{aligned}$$

Damit folgt insbesondere

$$\mathbb{E}\Big(\exp(-\frac{1}{2}\sigma_t^2 + \sigma_t Z_t)|\mathcal{F}_{t-1}\Big) = 1,$$

und die Behauptung folgt. ■

Setzt man als Modell für den Aktienkurs

$$S_t := S_{t-1}e^{r+Y_t}, \qquad t = 1, \dots$$

mit einem festen S_0 so erhält man $\mathbb{E}(\exp(-rT)S_T|\mathcal{F}_t) = \exp(-rt)S_t$, was notwendige Bedingung (unter Q) für einen Arbitragefreien Markt war.

Für die bedingte Varianz im ARCH-M(1) Modell gilt

$$\sigma_t^2 = \alpha_0 + \alpha_1(Y_{t-1} + \frac{1}{2}\sigma_{t-1}^2) = \alpha_0 + \alpha_1\left[Y_{t-1}^2 + Y_{t-1}\sigma_{t-1}^2 + \frac{1}{4}\sigma_{t-1}^4\right]$$

man hat also eine modifizierte GARCH-Struktur.

1.2.5 Maximum Likelihood Schätzung

Der genauen Analyse sei eine kurze Motivation vorangestellt. Eine ausführliche Behandlung der Maximum-Likelihood Schätzer findet sich in meinem Skript zu Statistik 1.

Motivation:

Betrachte n unabhängige Realisierungen $X_1, \dots, X_n$ einer Normalverteilten $\mathcal{N}(\mu, \sigma^2)$ ZV. Dabei sei σ bekannt, und μ soll geschätzt werden. Aufgrund der Beobachtungen könnte man versuchen, dass μ zu wählen, wofür die Beobachtung $X_1, \dots, X_n$ am wahrscheinlichsten wird. Da ein jedes Ereignis die W.o. hat, betrachtet man statt dessen das Ereignis $\{x_1 \pm \epsilon, X_2 \pm \epsilon, \dots\}$ und erhält

$$\mathbb{P}(X_1 \in [x_1-\varepsilon, x_1+\varepsilon], \dots, X_n \in [x_n-\varepsilon, x_n+\varepsilon]) = \prod_{i=1}^n \mathbb{P}(X_i \in [x_i-\varepsilon, x_i+\varepsilon]).$$

Nun ist $\mathbb{P}(X_i \in [x_i-\varepsilon, x_i+\varepsilon]) = F(x_i+\varepsilon) - F(x_i-\varepsilon)$. Für das Maximierungsproblem kann man mit einer positiven Zahl multiplizieren und erhält

$$\frac{1}{(2\epsilon)^n}\prod_{i=1}^n [F(x_i+\epsilon) - F(x_i-\epsilon)] \xrightarrow[\epsilon\to 0]{} \prod_{i=1}^n f(x_i),$$

also das Produkt der Dichten.

Definition 1.2.22. *Betrachte eine Parametrische Zufallsverteilung mit Dichte $f_\theta(x)$. Der Maximum-Likelihood-Schätzer von θ ist definiert durch*

$$\hat{\theta}_n(X_1, \ldots, X_n) := \operatorname{argmax}_{\theta \in A} \prod_{i=1}^{n} f_\theta(X_i).$$

Beispiel 1.2.23. Oft ist es günstig, die sogenannte log-likelihood-funktion

$$l_\theta(x_1, \ldots, x_n) = \sum_{i=1}^{n} \ln f_\theta(x_i)$$

zu betrachten. Für die Normalverteilung ist beispielsweise

$$l_\mu(x_1, \ldots, x_n) = \sum_{i=1}^{n} -\frac{(x_i - \mu)^2}{2} + \text{const.}$$

und $\frac{\partial}{\partial_\mu} l_\mu = 0$ ergibt $\sum_{i=1}^{n} x_i = n \cdot \mu$, also $\hat{\mu} = \frac{1}{n} \sum_{i=1}^{n} X_i$.

Bemerkenswert an dem oben abgeleiteten ML-Schätzer ist, dass er robust gegenüber Änderungen an der Verteilung ist! Von dem starken Gesetz der großen Zahl bzw. dem zentralen Grenzwertsatz weiß man, dass dieser Schätzer konsistent und asymptotisch normal unter wesentlich schwächeren Voraussetzungen ist. Dies macht man sich bei sogenannten Quasi-Maximum-Likelihood Schätzern (QMLE) zunutze.

1.3 Quasi-Maximum-Likelihood Schätzung

Wir bezeichnen mit $f_t(y; \theta)$ die bedingte Dichte von Y_t gegeben $Y_0, \ldots, Y_{t-1}$. $\theta \in \mathbb{R}^d$ ist hierbei der zu schätzende Parameter. Da

$$\begin{aligned}
\mathbb{P}(X \leqslant x, Y \leqslant y) &= \mathbb{E}\left(\mathbb{P}(Y \leqslant y | X) 1_{\{X \leqslant x\}}\right) \\
&= \int_0^x \int_0^y f_{y|x}(z) dz f_x dx
\end{aligned}$$

kann man das Konzept des ML-Schätzers (welches ja Unabhängigkeit voraussetzte) auch auf den allgemeineren Fall ausweiten, in welchem man bedingte Dichten betrachtet. Eine Schwierigkeit bereitet allerdings die Dichte von y_0. Da sie aber asymptotisch keine Rolle spielt, bedingt man einfach auf Y_0, was heißt, das man Y_0 praktisch wie eine (bekannte) Konstante behandelt.

Im folgenden schreiben wir kurz:

$$f_{Y_t | Y_{t-1}, \ldots, Y_0}(y_t | y_{t-1}, \ldots, y_0; \theta) =: f_t(y_t | y_{t-1}, \ldots, y_0; \theta).$$

Wir definieren die *Likelihoodfunktion* (bedingt auf Y_0) durch

$$L(\mathbf{y}; \theta) = \prod_{t=1}^{T} f_t(y_t | y_{t-1}, \ldots, y_0; \theta).$$

Eine präzisere Notation sieht so aus:

$$L_{Y_1,\dots,Y_T|Y_0}(y_1,\dots,y_T|y_0;\theta) = \prod_{t=1}^{T} f_t(y_t|y_{t-1},\dots,y_0;\theta).$$

Die obige Notation verdeutlicht noch einmal, dass die Likelihood-Funktion bedingt Y_0 betrachtet wird.

Der M-L Schätzer für θ ist definiert durch

$$\hat{\theta}_T := \operatorname{argmax}_\theta L(\mathbf{Y};\theta) = \operatorname{argmax}_\theta \log L(\mathbf{Y};\theta).$$

Wird die wahre bedingte Dichte f nun durch die Normalverteilungsdichte ersetzt, so spricht man von einem *Quasi-Maximum-Likelihood Schätzer*. Genauer: Man ersetzt die bedingte Dichte von $Y_t|Y_{t-1},\dots$ durch die Dichte von $\mathcal{N}\big(\mathbb{E}(Y_t|Y_{t-1},\dots),\operatorname{Var}(Y_t|Y_{t-1},\dots)\big)$. Wir bezeichnen diese Dichte mit $\tilde{f}_t(y_t|y_{t-1},\dots;\theta)$ und definieren den QMLE mit

$$\tilde{L}(y;\theta) := \prod_{t=1}^{T} \tilde{f}_t(y_t|y_{t-1},\dots,y_0;\theta)$$

durch

$$\tilde{\theta}_T := \operatorname{argmax}_\theta \tilde{L}(\mathbf{Y};\theta).$$

Unter geeigneten Regularitätsannahmen an die wahre Verteilung der Y ist dieser Schätzer konsistent (auch, wenn die Y nicht normalverteilt sind) und asymptotisch normal.

Seine Kovarianzmatrix ist gegeben durch

$$\begin{aligned}
V_{QS}\left[\sqrt{T}\,\tilde{\theta}_T\right] &= J^{-1} I J^{-1},\\
J &= \mathbb{E}\left[-\frac{\partial^2 \ln \tilde{f}_t(Y_t|Y_{t-1},\dots,Y_0;\theta)}{\partial\theta\partial\theta'}\right]\\
I &= \mathbb{E}\left[\frac{\partial \ln \tilde{f}_t(Y_t|Y_{t-1},\dots,Y_0;\theta)}{\partial\theta}\cdot\frac{\partial \ln \tilde{f}_t(Y_t|Y_{t-1},\dots,Y_0;\theta)}{\partial\theta'}\right],
\end{aligned}$$

wobei wir mit θ' den transponierten Vektor bezeichnen. Hierbei ist allerdings zu beachten, dass die obigen Erwartungswerte bezüglich der wahren zugrunde liegenden Verteilung zu bilden sind! Wenn die Verteilung der Y sozusagen kompatibel mit der Normalverteilung ist, so ist $I = J$ und V_{QS} schlicht I.

I bezeichnet man auch als *Fisher Information*. Sie gibt die asymptotische Varianz des Schätzers und damit etwas über seinen Informationsgehalt an. Man kann sich darunter die Information, die eine Zufalllsvariable über den unbekannten Parameter θ trägt, vorstellen. Man kann leicht zeigen, dass $\mathbb{E}\left(\frac{\partial}{\partial\theta}\ln \tilde{f}_t(Y_t|Y_{t-1},\dots,Y_0;\theta)\right) = 0$ also

$$I = \operatorname{Var}[\frac{\partial}{\partial\theta}\ln \tilde{f}_t(Y_t|Y_{t-1},\dots,Y_0;\theta)]$$

Die Fisher Information ist also die Varianz des sogenannten Scores.

Für den arithmetischen Mittelwert im Normalverteilungsfall ist

$$\frac{\partial}{\partial\mu}\ln f_i(Y;\mu) = \frac{\partial}{\partial\mu}\ln\frac{1}{\sqrt{2\pi\sigma^2}}e^{-(Y-\mu)^2/2\sigma^2} = \frac{\partial}{\partial\mu} - \frac{(Y-\mu)}{2\sigma^2} = \frac{Y-\mu}{\sigma^2}$$

Die Score Funktion stellt den Einfluss eines Datums auf den Schätzer dar. Je weiter entfernt das Datum von μ ist, desto größer der Einfluss. Das arithmetische Mittel ist somit sensitiv gegenüber Ausreißern. Die Fisher Information ist

$$\mathbb{E}\left(\frac{(Y-\mu)^2}{\sigma^2}\right) = \frac{1}{\sigma^4}\cdot\sigma^2 = \frac{1}{\sigma^2}$$

Je geringer die Varianz, desto höher also die Information! Betrachtet man $\theta = \binom{\mu}{\sigma}^2$ so erhält man

$$J^{-1}IJ^{-1} = \begin{pmatrix} \frac{1}{\sigma^2} & 0 \\ 0 & 2\sigma^4 \end{pmatrix} \quad \text{und somit asymptotisch unkorrelliertheit von } \hat{\mu} \text{ und } \hat{\sigma}$$

1.4 QMLE für GARCH

Wir betrachten ein GARCH(1,1)-Modell. Die Dichten der Fehler Z bezeichnen wir mit f_z.

$$\mathbb{E}(Y_t|Y_{t-1}) = 0 \text{ und } \operatorname{Var}(Y_t|Y_{t-1}) = \sigma_t^2 = \alpha_0 + \alpha_1 Y_{t-1}^2 + \beta_1\sigma_{t-1}^2 \tag{1.8}$$

ist

$$l_t(Y;\theta) = f_{Y_t,\ldots}(\theta) = \phi(\frac{Y}{\sigma_t})\cdot\frac{1}{\sigma_t} = \frac{1}{\alpha_0 + \alpha_1 Y_{t-1}^2 + \beta_1\sigma_{t-1}^2}\phi\left(\frac{Y}{\alpha_0 + \alpha_1 Y_{t-1}^2 + \beta_1\sigma_{t-1}^2}\right) \tag{1.9}$$

Für den ARCH(1)-Fall erhalten wir sofort den QMLE. Für den GARCH(1,1)-Fall ist allerdings σ_0 nicht beobachtbar. Man wählt einen Startwert, etwa die Stichprobenvarianz von $Y_1,\ldots,Y_T$ oder sogar 0.

Die Likelihood Funktion muss man nun numerisch maximieren.

1.5 Momentenschätzer im MARCH-Modell

Literatur: Schmidt und Stute (2006).

Im MARCH-Modell ist

$$y_i = r - \frac{1}{2}\sigma_i^2 + \underbrace{\sigma_i Z_i}_{\varepsilon_i}$$
$$\sigma_i^2 = \alpha_0 + \alpha_1\sigma_{i-1}^2 Z_{i-1}^2$$

r ist hierbei bekannt, $\alpha_0, \alpha_1 > 0$ sind zu schätzen. Die Z_i sind $i.i.d. \sim \mathcal{N}(0,1)$.

Setze $\varepsilon_i = \sigma_i Z_i$.

Untersuchen wir die ε_i genauer

$$\begin{aligned}
\varepsilon_i^2 = \sigma_i^2 Z_i^2 &= Z_i^2(\alpha_0 + \alpha_1 \varepsilon_{i-1}^2) \\
&= Z_i^2(\alpha_0 + \alpha_1 Z_{i-1}^2(\alpha_0 + \alpha_1 \varepsilon_{i-2}^2) \\
&= Z_i^2(\alpha_0 + \alpha_0\alpha_1 Z_{i-1}^2 + \alpha_1^2 Z_{i-2}^2 Z_{i-1}^2(\alpha_0 + \alpha_1 \varepsilon_{i-3}^2)) \\
&= \alpha_0 Z_i^2(1 + \sum_{k=1}^{n}\prod_{j=1}^{k}(\alpha_1 Z_{i-j}^2)) + Z_i^2(\prod_{i=1}^{n}\alpha_1 Z_{i-j}^2 \varepsilon_{i-n}^2)
\end{aligned}$$

Betrachten wir den Erwartungswert, so erhalten wir

$$\begin{aligned}
E(\varepsilon_i^2) &= \alpha_0(1 + \sum_{k=1}^{n}\alpha_1^k) + E(\varepsilon_{i-n}^2)\alpha_1^h \\
&\xrightarrow[n\to\infty]{\alpha_1<1} \frac{\alpha_0}{1-\alpha_1}
\end{aligned}$$

Damit gilt ebenso:

$$E(g_i) = r - \frac{1}{2}\underbrace{E(\sigma_{i-1}^2)}_{E(\varepsilon_{i-1}^2)} - \underbrace{E(\sigma_i Z_i)}_{0=} \to r - \frac{1}{2}\frac{\alpha_0}{1-\alpha_1}$$

Haben wir schwache Stationarität, so lässt sich der Erwartungswert der Y_i durch das arithmetische Mittel schätzen, bezeichnet durch $\overline{Y}_n$.

$$\Rightarrow \overline{Y}_n \sim r - \frac{1}{2}\frac{\alpha_0}{1-\alpha_1}$$

Wäre beispielsweise α_0 bekannt, so erhielten wir direkt einen Schätzer für α_1. Man kann ebenso zeigen, dass

$$\operatorname{Var}(Y_i) \to \frac{\alpha_0}{1-\alpha_1} + \frac{\alpha_0^2\alpha_1^2}{2(1-\alpha_1)^2(1-3\alpha_1^2)}$$

Man erhält durch Schätzer für α_0, α_1 die konsistent und asymptotisch normalverteilt sind. Weiteres: Schmidt und Stute (2006)

1.5.1 Der "News Impact" oder M-GARCH Modellierung mit Asymmetrien

Literatur: Hafner und Härdle (2000) "Discrete time option pricing with flexible volatility estimation".

Diese Arbeit greift die Ideen von Duan (1995) auf, um in ARCH-Modellen Optionen zu bewerten.

Dabei betrachten sie die diskrete Rendite des Aktienkurses

$$Y_t := \frac{S_t - S_{t-1}}{S_{t-1}}$$

und stellen hierfür ein entsprechendes M-GARCH Modell unter P und Q vor. Der Einfachheit halber betrachten wir ein GARCH(1,1) Modell:

Unter P:

$$
\begin{aligned}
Y_t &= \mu_t + \sigma_t Z_t, \quad \text{mit } Z_t \sim \text{ i.i.d. } \mathcal{N}(0,1) \\
\sigma_t^2 &= \alpha_0 + \alpha_1 \sigma_{t-1}^2 Z_{t-1}^2 + \beta_1 \sigma_{t-1}^2
\end{aligned}
$$

Die Autoren machen die Annahme eines konstanten market price of risk:

$$
\lambda = \frac{\mu_t - \sigma}{\sigma_t}
$$

$\Rightarrow \mu_t = \sigma + \lambda \sigma_t$

Dadurch taucht σ_t in Y_t auf und dieses Modell heißt M-GARCH (GARCH in mean). Für den Übergang zu Q machen die Autoren analog zu Duan noch folgende Annahme

$$
\mathrm{Var}^P(Y_t|\mathcal{F}_{t-1}) = \mathrm{Var}^Q(Y_t|\mathcal{F}_{t-1})
$$

Unter Q muss gelten:

$$
E^Q(Y_t|\mathcal{F}_{t-1}) = r
$$

Das ist gleichbedeutend mit:

$$
Z_t^* := Z_t + \lambda \sim_Q \mathcal{N}(0,1)
$$

Damit sieht unser Modell unter Q folgendermaßen aus:

$$
\begin{aligned}
Y_t &= r + \sigma_t Z_t^*, \quad Z_t^* \text{ iid } \mathcal{N}(0,1) \\
\sigma_t^2 &= \alpha_0 + \alpha_1 \sigma_t t - 1^2 (Z_{t-1}^* - \lambda)^2 + \beta_1 \sigma_{t-1}^2
\end{aligned}
$$

dabei ist

$$
\begin{aligned}
\sigma_{t-1}^2 (Z_t^* - \lambda)^2 &= [\sigma_{t-1}] \\
&= [Y_{t-1} - \underbrace{(r + \lambda \sigma_{t-1})}_{\mu_{t-1}}]^2 = (Y_{t-1} - \mu_{t-1})^2
\end{aligned}
$$

also ist σ_{t-1}^2 eine Quadratische Funktion von Y_{t-1}(falls $\beta_1 = 0$).

Man beobachtet, dass $Y_t - \mu_t$ einen symmetrischen impact (Einfluss) auf die zukünftige Volatilität hat. Dass dies in der Realität nicht zutrifft, belegen die Autoren mit einer eingehenden satistischen Analyse des DAX. Hierbei betrachten sie das Modell

$$
Y_t = g(Y_{t-1}) Z_t
$$

und schätzen g.

Dieser Graph zeigt eine deutliche Asymmetrie. Eine Modellklasse, die in der Lage ist, Asymetrien zu modellieren, sind die bereits vorgestellten TGARCH-Modelle. Die Autoren wählen

$$
\sigma_t^2 = \alpha_0 + \alpha_1 \underbrace{\sigma_{t-1}^2 Z_{t-1}^2}_{=:\varepsilon_{t-1}^2} 1_{\{\varepsilon_{t-1} > 0\}} + \alpha_2 \varepsilon_{t-1}^2 1_{\{\varepsilon_{t-1} \leq 0\}} + \beta_1 \sigma_{t-1}^2
$$

Man kann sich allerdings fragen, inwiefern dieses Modell den geschätzten impact im obigen Graph korrekt wiedergibt. Insbesondere scheint dies für beweglich große y (Renditen) nicht der Fall.

Zunächst einmal ist $\varepsilon_t^2 = (Y_t - \mu_t)^2$, also

$$\sigma_t^2 = \alpha_0 + \alpha_1(Y_{t-1} - \mu_{t-1})^2 1_{\{Y_{t-1} < \mu_{t-1}\}} + \alpha_2(Y_{t-1} - \mu_{t-1})^2 1_{\{Y_{t-1} \geq \mu_{t-1}\}} + \beta_1 \sigma_{t-1}^2$$

Wie sieht es mit der Stationarität des Modells aus?

Satz 1.5.1. *Für einen schwach stationären M-TGARCH Prozess ist die unbedingte Varianz gerade:*

$$\mathrm{Var}^Q(\sigma_t Z_t^*) = \frac{\alpha_0}{1 - \psi(\lambda)(\alpha_1 - \alpha_2) - \alpha_2(1 + \lambda^2) - \beta_1}$$

mit $\psi(\lambda) = \lambda\phi(\lambda) + (1 + \lambda^2)\Phi(\lambda)$. Φ *ist hierbei die Verteilungsfunktion der Standardnormalverteilung.*

Für $\lambda > 0$ erhalten wir $\Psi(\lambda) > \frac{1}{2}$.

Beweis. Siehe Vorlesung. ∎

1.5.2 Praktische Anwendung des Modells in Bezug auf DAX-Calls

Die Autoren schätzen das Modell mit DAX-Daten aus den Jahren 1988- 1991. Sie verwenden QMLE für das M-GARCH und M-TGARCH Modell. Für das TGARCH-Modell erhalten sie:

α_0	α_1	α_2	β_1	λ
$1.9 \cdot 10^{-5}$	0.2	0.05	0.77	0.0385

Die Likelihood ist besser als für das GARCH-Modell. Außerdem $\alpha_1 > \alpha_2$.

Vergleicht man die Optionspreise von Calls, die unter dem TGARCH, GARCH und Black-Scholes-Modell bewertet wurden, so erhält man den besten Fit für das TGARCH-Modell; insbesondere für "out-of-the-money" Optionen. Der Schluss der Autoren ist, dass Optionspreise bereits Asymmetrien berücksichtigen.

2 Verschiedenes

In diesem Abschnitt diskutieren wir noch verschieden Verfahren, die für Statistik auf Finanzmärkten von Nutzen sein können. Zunächst besprechen wir die Methode der Kurvenschätzung am Beispiel von Dichteschätzung durch Kernschätzer. Diese Methode fand gerade im vorigen Abschnitt Anwendung, als im TGARCH-Modell die Funktion g aus Daten geschätzt werden sollte. Kurvenschätzungen stellen also Hilfsmittel bereit, neben Dichten auch allgemeiner beliebige Funktionen aus Daten schätzen zu können.

Im anschließenden Kapitel stellen wir das Central-Asset-Pricing Modell (CAPM) vor und analysieren die Methoden, dieses Modell aus Daten zu schätzen.

2.1 Kurvenschätzung am Beispiel von Dichteschätzungen

2.1.1 Histogramm

Ziel: Beispielsweise Schätzung der Renditeverteilung zur Bestimmung von Risikokennzahlen, wie etwa VaR (Value-at-Risk).

Ziel dieses Abschnittes wird es sein, nichtparametrische Verfahren zur Schätzung einer Dichte vorzustellen. eine nichtparametrische Schätzung der Verteilungsfunktion erhält man durch die empirische Verteilungsfunktion. $F_n(x) = \frac{1}{n} \sum_{i=1}^{n} 1_{\{X_i \leq X\}}$

Da die Ableitung von F_n nicht existiert, ist $(F_n)'$ nicht als Dichteschätzer geeignet.

Die älteste Methode Dichten zu schätzen, ist das Histogramm. Ein Histogramm gruppiert die Daten in Intervalle, typischerweise mit der gleichen Länge. Für ein X_0 und ein $h > 0$ sei

$$I = (x_0 + jh, x_0 + (j+1)h], \qquad j \in Z \text{ (Partition)}$$

Die Dichte in einem Intervall wird bei dem Histogramm wie folgt geschätzt:

für $x \in I_j$

$$\begin{aligned} f_n(x) &= \frac{\#\text{ Daten in } I_j}{n \cdot h} \\ &= \frac{F_n(x_0 + (j+1)h) - F_n(x_0 + j - h)}{h} \end{aligned}$$

Frage: Wie kann man das x_0 an den Daten fest machen? Die Wahl von h hat schon einen großen Einfluss, aber auch x_0 ist wichtig. Da das Histogramm bei unterschiedlicher Wahl des x_0 unterschiedlich aussieht. Die erste Schwierigkeit besteht in der guten Wahl von h.

Ist h zu klein oder zu groß, erhält eine schlechte, wenig informative Schätzung. Das zweite Problem ist die Wahl von x_0.

Dieses Problem kann man umgehen, indem man die Intervalle direkt um die Daten legt.

Wir schätzen für einen Punkt x und feste Bandbreite h

$$\begin{aligned} f_n(x) &= \frac{F_n(x+\frac{n}{2}) - F_n(x-\frac{n}{2})}{h} \\ &= \frac{\#\text{ Daten in}(x-\frac{n}{2}, x+\frac{n}{2}]}{n \cdot h} \end{aligned}$$

d.h. man legt um die Datenpunkte Intervalle, und die Funktion zählt die Überschneidungen der Intervalle.

Leider stören immer noch die Sprünge. Wie erhält man einen glatten Schätzer für die Dichte?

Wir schreiben für obigen Schätzer

$$f_n(x) = \frac{1}{n} \int K\Big(\frac{x-y}{n}\Big) F_n(dy)$$

mit $K(x) = 1_{(-\frac{1}{2},\frac{1}{2}]}(x)$

Nun hat man die Möglichkeit K, den "Kern"durch andere Funktionen zu ersetzen, etwa

$$K(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$$

Damit kann man den Einfluss des Datenpunktes steuern. Direkt an dem Datenpunkt ist er sehr groß und mit zunehmender Entfernung wird er immer geringer.

Definition 2.1.1. *Für iid Daten $X_1, \ldots, X_n$ mit Dichte f heißt der Schätzer*

$$f_n(x) := \frac{1}{h_n} \int K\Big(\frac{x-y}{h_n}\Big) F_n(dy)$$

mit Bandbreite $h_n > 0$ und Kern K **Kernschätzer** *von f, falls K folgende Bedingungen erfüllt:*

(i) $K(x) \geq 0$

(ii) $\int K(u)du = 1$

Wir berechnen den Erwartungswert des Kernschätzers.

Lemma 2.1.2. Für einen Kernschätzer f_n gilt:

$$\mathbb{E}(f_n(x)) = \frac{1}{h_n} \int K(\frac{x-y}{h_n}) f(y)dy.$$

Beweis: ÜA

Dieser Erwartungswert ist typischerweise verschieden von $f(x)$. Der Kernschätzer ist verzerrt, man sagt, er hat einen Bias.

Merke: Ähnlich wie bei dem Histogramm darf die Bandbreite nicht zu groß und nicht zu klein gewählt werden. A priori gibt es keine optimale Bandbreitenwahl (obwohl das einige Verfahren suggerieren).

Wir schauen uns die Verzerrung noch genauer an.

$$\begin{aligned}
\text{Bias} :&= E(f_n(x)) - f(x)\\
&= \frac{1}{h_n}\int K\Big(\underbrace{\frac{x-y}{h_n}}_{=:z}\Big) f(y)dy - f(x)\\
&= \int K(z) f(x-h_n z)dz - f(x)\cdot\int K(z)dz\\
&= \int K(z)[f(x-h_n z) - f(x)]dz
\end{aligned}$$

Lemma 2.1.3. *Ist f beschränkt und stetig, so gilt*

$$\mathbb{E}(f_n(x)) \xrightarrow[n\to\infty]{} f(x)$$

für jede Folge h_n mit $h_n \to 0$.

Mit einer Taylorentwicklung erhält man für den Mittleren quadratischen Fehler (MSE)

$$\begin{aligned}
\mathbb{E}((f_n(x)-f(x))^2) &= \operatorname{Var}(f_n(x)) + Bias^2\\
&\sim \frac{f(x)\int K^2(u)du}{n\cdot h_n} + h_n^4\Big[\frac{f''(x)}{2}\int u^2K(u)du\Big]^2
\end{aligned}$$

Ableiten, gleich Null setzen liefert ein optimales h_n :

$$h_n = n^{-1/5}\left[\frac{f(x)}{4}\int K^2(u)du\right]^{\frac{1}{5}}\left[\frac{f''(x)}{2}\int u^2K(u)du\right]^{\frac{2}{5}} = cn^{-1/5}$$

Der optimale Kern ist der Epanecknikov-Kern:

$$K_0(x) = \frac{3}{4}(1-\frac{1}{5}x^2)/\sqrt{5}\cdot 1_{\{|x|<\sqrt{5}\}}$$

Er ist symmetrisch und hat einen kompakten Träger. Allerdings ist der Effizienzverlust im Vergleich zum Gaußkern nicht sehr groß.

3 Das CAPM

3.1 Portfolio Optimierung

Das CAPM ermöglicht den einfachsten Zugang zu einer statischen Portfoliooptimierung. Wir betrachten einen Markt mit n Wertpapieren $S_1, \dots, S_n$. Ziel wird es sein, eine optimale Rendite bei geringer Varianz zu erzielen.

Betrachten wir zunächst einen Markt mit zwei Aktien S_1, S_2. Ein Portfolio besteht aus Linearkombinationen

$$\phi_1 S_1 + \phi_2 S_2$$

mit, sagen wir, $\phi_1 + \phi_2 = 1$. Dabei müssen die ϕ_i nicht unbedingt positiv sein (Leerverkäufe erlaubt).

Die Rendite des Portfolios ist

$$\phi_1 \frac{S_1(t) - S_1(o)}{S_1(0)} + \phi_2 \frac{S_2(t) - S_2(0)}{S_2(0)} = \phi_1 R_1 + \phi_2 R_2.$$

Es genügt also, im Folgenden sich auf Renditen zurückzuziehen.

Die erwartete Rendite ist natürlich ϕ_i (zur Zeit 0 bekannt, bzw. messbar)

$$\mathbb{E}(\phi' R) = \phi' \mu \quad \text{mit} \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix} = \mathbb{E} \begin{pmatrix} R_1 \\ R_2 \end{pmatrix} \text{ und } \phi = \begin{pmatrix} \phi_1 \\ \phi_2 \end{pmatrix}.$$

Die erwartete Varianz ist

$$\begin{aligned} \operatorname{Var}(\phi' R) &= \operatorname{Var}(\phi_1 R_1 + \phi_2 R_2) \\ &= \phi_1^2 \operatorname{Var}(R_1) + \phi_2^2 \operatorname{Var}(R_2) + 2\phi_1\phi_2 \operatorname{Cov}(R_1, R_2) \\ &= (\phi_1 \phi_2) \begin{pmatrix} \operatorname{Var}(R_1) & \operatorname{Cov}(R_1, R_2) \\ \operatorname{Cov}(R_1, R_2) & \operatorname{Var}(R_2) \end{pmatrix} \begin{pmatrix} \phi_1 \\ \phi_2 \end{pmatrix} \\ &= \phi' \Sigma \phi \end{aligned}$$

mit $\Sigma = \operatorname{Cov}(R)$, die Varianz-Kovarianz-Matrix von R. Nehmen wir einmal an, $\operatorname{Var}(R_i) > 0$, so erhält durch auf unsere obige Fragestellung ein Optimierungsproblem. Dazu nehmen wir zunächst an, dass

$$\operatorname{Cov}(R_1, R_2) = 0.$$

Wie sieht für vorgegebenes $\phi_1 (\phi_2 = 1 - \phi_1)$ die Rendite/Volaritätskurve aus?

$$\phi_1 \to \frac{\phi_1 \mu_1 + (1 - \phi_1)\mu_2}{\sqrt{\phi_1^2 \sigma_1^2 + (1 - \phi_1)^2 \sigma_2^2}}$$

Beispiel 3.1.1.

$$\mu_1 = 0.1\ \mu_2 = 0.05\ \sigma_1 = 0.4,\ \sigma_2 = 0.3$$

ϕ	erw.Rendite	erw.Var.d.R.	Vola
-0.5	1.15	0.4	0.5
-0.25	0.975	0.15	0.4
0	0.8	0.09	0.3
0.25	0.625	0.06	0.25
0.5	0.45	0.06	0.25
0.75	0.275	0.1	0.3
1	0.1	0.16	0.4

Wir erhalten (bis auf Reskalierung) eine nach rechts gekippte Parabel. Welches Portfolio wäre für uns optimal?

Antwort: Es kommt drauf an! Klar, $\phi_1 = 1$ ist schlecht, da $\phi_1 = -0.25$ eine viel bessere Rendite bei gleichem Risiko bietet. Man sucht sich also am besten ein Portfolio am oberen Randoms. Je nach persönlicher Risikoaversion weiter oben bzw. unten.

Für unterschiedliche Korrelationen $\in (-1, 1)$ erhält man einen anderen Öffnungswinkel der Parabel. Für $\delta = \pm 1$ lässt sich das eine Instrument aus dem anderen berechnen und wir erhalten einen linearen Zusammenhang.

3.1.1 Das Minimierungsproblem - Allgemeiner Fall ohne risikoloses Asset

Für ein allgemeines Modell mit n-Assets erhalten wir eine Portfolio-Rendite

$$\mathbb{E}(\phi' R) = \phi' \mu$$

und Varianz

$$\phi' \Sigma \phi.$$

Dabei gilt $\phi' \cdot 1 = \phi_1 + \cdots + \phi_n = 1$. Für eine Optimierung fixieren wir vorab eine Rendite $\overline{\mu}$ und erhalten die zweite Nebenbedingung $\phi' \mu = \overline{\mu}$.

Zu minimieren ist natürlich die Varianz. Das Minimierungsproblem löst man mit Hilfe der Lagrangemultiplikatoren:

$$\min L(\phi, \lambda_1, \lambda_2) = \phi' \Sigma \phi + \lambda_1(\overline{\mu} - \phi' \mu) + \lambda_2(1 - \phi' 1).$$

Lemma 3.1.2. *Das Varianz-optimale Portfolio mit erwarteter Rendite $\overline{\mu}$ ist gegeben durch*

$$\phi = g + h \cdot \overline{\mu} \quad \textit{mit}$$

$$\begin{aligned} A &= 1' \Sigma^{-1} \mu & \qquad & g = \frac{1}{A^2 - CB}\left[A \Sigma^{-1} \mu - B \Sigma^{-1} 1\right] \\ B &= \mu' \Sigma^{-1} \mu & \textit{und} & \\ C &= 1' \Sigma^{-1} 1 & & h = \frac{1}{A^2 - CB}\left[A \Sigma^{-1} 1 - C \Sigma^{-1} \mu\right] \end{aligned}$$

Beweis. Wir müssen die Lagrangefunktionen in $\phi, \lambda_1, \lambda_2$ minimieren. D.h. wir erhalten

$$0 = \frac{\partial L}{\partial \phi} \quad = 2\Sigma\phi - \lambda_1\mu - \lambda_2 1 = 0 \tag{3.1}$$

$$0 = \frac{\partial L}{\partial \lambda_1} \quad \Leftrightarrow \overline{\mu} = \phi'\mu \tag{3.2}$$

$$0 = \frac{\partial L}{\partial \lambda_2} \quad \Leftrightarrow 1 = \phi' 1 \tag{3.3}$$

Aus (3.1) folgt direkt

$$2\phi = \Sigma^{-1}(\lambda_1\mu + \lambda_2 1) \quad \text{und} \tag{3.4}$$

$$\phi = \frac{1}{2}\Sigma^{-1}(\lambda_1\mu + \lambda_2 1).$$

Wir nutzen (3.2) und erhalten aus $\mu' \cdot$ (3.4)

$$2\mu'\phi = 2\underbrace{\phi'\mu}_{=\overline{\mu}} = \mu'\Sigma^{-1}(\lambda_1\mu + \lambda_2 1)$$

bzw. aus (3.3) durch $1' \cdot$ (3.4)

$$2 \cdot 1'\phi = 2\underbrace{\phi' 1}_{=1} = 1'\Sigma^{-1}(\lambda_1\mu + \lambda_2 1)$$

das GLS

$$2\overline{\mu} = \lambda_1 \underbrace{\mu'\Sigma^{-1}\mu}_{=:\beta} + \lambda_2 \underbrace{\mu'\Sigma^{-1}1}_{=:A}$$

$$2 = \lambda_1 \underbrace{1'\Sigma^{-1}\mu}_{=:A} + \lambda_2 \underbrace{1'\Sigma^{-1}1}_{=:C}$$

also

$$\lambda_2 = \frac{2\overline{\mu} \cdot A - 2B}{A^2 - CB} \quad \text{und } \lambda_1 = \frac{2A - 2\overline{\mu}C}{A^2 - CB}$$

Einsetzen liefert die Behauptung. ■

Welche Varianz hat dieses Portfolio?

Lemma 3.1.3. *Das Varianz-optimale Portfolio hat zur Rendite $\overline{\mu}$ die Varianz*

$$(\sigma^*_{(\overline{\mu})})^2 = \frac{C}{CB - A^2}\left[\left(\overline{\mu} - \frac{A}{C}\right)^2\right] + \frac{1}{C}.$$

Beweis. Für zwei Portfolios ϕ und $\tilde{\phi}$ ist die Covarianz

$$\begin{aligned}
\mathrm{Cov}(\phi'R, \underbrace{\tilde{\phi}'R}) &= \mathbb{E}\Big([\phi'R - \mathbb{E}(\dots)][R'\tilde{\phi} - \mathbb{E}(\dots)]\Big) \\
&= \mathbb{E}\Big(\phi'R\, R'\tilde{\phi} - \phi'R\mathbb{E}(R'\tilde{\phi}) - \mathbb{E}(\phi'R)R'\tilde{\phi} + \mathbb{E}(\phi'R)\mathbb{E}(R'\tilde{\phi})\Big)
\end{aligned}$$

Der Erwartungswert ist ein linearer Operator und deswegen gilt

$$\begin{aligned}\mathrm{Cov}(\phi' R, \tilde{\phi}' R) &= \mathbb{E}\left[(\phi' R\, R'\tilde{\phi}) - \mathbb{E}(\phi' R)\mathbb{E}(R'\tilde{\phi})\right] \\ &= \phi' \underbrace{\mathrm{Cov}(R)}_{=\Sigma} \tilde{\phi}\end{aligned}$$

Wir nehmen an, dass ϕ und $\tilde{\phi}$ Varianz-optimale Portfolios sind mit Erwartungswert μ und $\tilde{\mu}$. Dann gilt $\phi = g + h\mu$ und $\tilde{\phi} = g + h\tilde{\mu}$ mit den Bezeichnungen des vorigen Lemmas. Es folgt

$$(g + h\mu)'\Sigma(g + h\tilde{\mu}) = g'\Sigma g + \mu h'\Sigma g + \tilde{\mu} g'\Sigma h + \mu\tilde{\mu} h'\Sigma h$$

Dabei ist

$$\begin{aligned}g'\Sigma g &= \frac{1}{A^2 - CB}\left[A\Sigma^{-1}\mu - B\Sigma^{-1}\mathbf{1}\right]' \Sigma \frac{1}{A^2 - CB}\left[A\Sigma^{-1}\mu - B\Sigma^{-1}\mathbf{1}\right] \\ &= \frac{1}{(A^2 - CB)^2}\left[A^2\mu'\Sigma^{-1}\Sigma\Sigma^{-1}\mu - AB\mu'\Sigma^{-1}\Sigma\Sigma^{-1}\mathbf{1} - AB\mathbf{1}'\Sigma^{-1}\Sigma\Sigma^{-1}\mu + B^2\mathbf{1}'\Sigma^{-1}\Sigma\Sigma^{-1}\mathbf{1}\right] \\ &= \frac{1}{(A^2 - CB)^2}\left[A^2\underbrace{\mu^1\Sigma^{-1}\mu}_{B} - 2AB\underbrace{\mu'\Sigma^{-1}\mathbf{1}}_{A} + B^2\underbrace{\mathbf{1}'\Sigma^{-1}\mathbf{1}}_{C}\right] \\ &= \frac{1}{(A^2 - CB^2)}\left[A^2B - 2A^2B + B^2C\right] = \frac{-B}{A^2 - CB^2}\end{aligned}$$

Ebenso

$$g'\Sigma h = \frac{A}{A^2 - CB} \text{ und } h'\Sigma h = \frac{-C}{A^2 - CB}$$

und so erhalten wir

$$\begin{aligned}\mathrm{Cov}(\phi' R, \tilde{\phi}' R) &= \frac{1}{A^2 - CB}\left[-B + (\mu + \tilde{\mu})A - \mu\tilde{\mu}C\right] \cdot C \cdot \left[-(\mu - \frac{A}{C})(\hat{\mu} - \frac{A}{C}) + \frac{A^2}{C^2} + \frac{B}{C}\right] \\ &= \frac{C}{A^2 - CB}\left[(\mu - \frac{A}{C})(\hat{\mu} - \frac{A}{C})\right] + \frac{1}{C}\end{aligned}$$

Für $\mu = \tilde{\mu}$ folgt die Behauptung. ∎

Wir erkennen, dass für $\mu = \frac{A}{C}$ das Portfolio mit insgesamt der kleinsten Varianz erreicht wird. Das entsprechende ϕ bezeichnen wir mit ϕ_{min}. Man beachte, dass die Kovarianz mit einem beliebigen Varianz-optimalen Portfolio immer $\frac{1}{C}$ ist. Man möchte die Renditen der zugrunde liegenden Aktien mit der Portfolio-Rendite vergleichen. Dabei erhält man

Lemma 3.1.4. *Für ein Varianz-optimales Portfolio ϕ mit Rendite μ und $\mu_0 := \frac{A\mu - B}{C\mu - A}$ erhält man*

$$\mu_i - \mu_0 = \beta_i \cdot (\mu - \mu_0) \text{ mit } \beta = \frac{\mathrm{Cov}(R_i, \phi' R)}{(\sigma^*_{(\mu)})^2}$$

Hierbei ist β der Anteil der Aktie i an der Rendite des Varianz-optimierten Portfolios. Ein hohes β entspricht dabei einer hohen mittleren Rendite; β kann allerdings auch negativ sein! Wir erhalten die Darstellung

$$\mu_i = \mu_0 + (\mu - \mu_0)\frac{\mathrm{Cov}(R_i, \phi' R)}{(\sigma^*_{(\mu)})^2}.$$

3.1.2 Das Minimierungsproblem mit risikoloser Anlage

Enthält der betrachtete Markt eine risikolose Anlage, so wird, wie bereits gesehen, das optimale Portfolio eine Kombination aus risikoloser Anlage und Marktportfolio sein.

Zunächst ergänzen wir den Markt um das Asset 0, mit Rendite r und Varianz 0.

Ein Portfolio hat also die Rendite

$$\phi_0 r + \sum_{i=1}^{n} \phi_i R_i.$$

Die Normierungsbedingung ist $\sum_{i=0}^{n} \phi_i = 1$, man kann also obiges Portfolio auch als Konvexkombination auffassen. Wir erhalten für den Erwartungswert

$$\phi_0 r + (1 - \phi_0)\tilde{\mu}$$

und für die Varianz

$$(1 - \phi_0)^2 \tilde{\sigma}^2.$$

Im Mean-Variance Diagramm stellt sich dies also durch

$$S \mapsto \frac{S}{\tilde{\sigma}} \cdot \tilde{\mu} + r(1 - \frac{s}{\sigma}) = r + S \cdot \frac{\tilde{\mu} - r}{\sigma}$$

dar. **Plot** Nun wollen wir noch versuchen, das Taugentialportfolio zu berechnen. Es ist gekennzeichnet durch die größte Steigung $\frac{\mu^* - r}{\sigma^*}$

Theorem 3.1.5. *(Mutual Fund Theorem) Das optimale Portfolio besteht aus folgender Kombination*

$$\phi_0 r + (1 - \phi_0)\phi_M' R$$

mit

$$\phi_M' = \frac{1}{\mathbf{1}'\Sigma^{-1}(\mu - \mathbf{1}r)} \Sigma^{-1}(\mu - \mathbf{1}r)$$

Beweis. Wir werden zeigen, dass das optimale Portfolio durch den Vektor ϕ^* beschrieben ist. Dann folgt

$$\phi_0^* r + \phi^{*'} R \quad \text{und } \phi_0^* + \phi^{*\prime} 1 = 1,$$

was äquivalent ist zu

$$\phi_0^* r + (1 - \phi_0^*) \cdot \left(\frac{\phi^{*\prime}}{\phi^{*\prime}\mathbf{1}} \right)' R,$$

wobei wir

$$\phi_M := \frac{\phi^{*\prime}}{\phi^{*\prime}\mathbf{1}}$$

setzen. Wir benötigen also ϕ^*. Unsere Überlegungen zeigen, dass das optimale Portfolio das Varianz-minimale Portfolio zu vorgegebener Rendite unter Nebenbedingung $\phi_0 + \phi'\mathbf{1} = 1$ ist bzw.

$$\phi_0 r + \phi'\mu = \overline{\mu} = (1 - \phi'\mathbf{1})r + \phi'\mu = \overline{\mu}.$$

Die Lagrangegleichung ist

$$\phi'\Sigma\phi + \lambda\left((1-\phi'\mathbf{1})r + \phi'\mu - \overline{\mu}\right).$$

Setzen wir die Ableitung nach ϕ gleich Null, so erhalten wir

$$2\Sigma\phi - \lambda 1 r + \lambda\mu \stackrel{!}{=} 0$$

und

$$(1-\phi'\mathbf{1})r + \phi'\mu - \overline{\mu} = 0.$$

Da $\phi = \frac{1}{2}\Sigma^{-1}1(1^r - \mu)$ und $\phi'(\mu - 1r) = \overline{\mu} - r$ gilt also

$$\begin{aligned} & \frac{1}{2}\lambda(1r-\mu)'\Sigma^{-1}(\mu-1r) = \overline{\mu} - r \\ \Rightarrow \qquad & \lambda = -\frac{\overline{\mu}-r}{2(\mu-\mathbf{1}r)'\Sigma^{-1}(\mu-\mathbf{1}r)} \\ \Rightarrow \qquad & \phi^{*\prime} = \frac{\overline{\mu}-r}{(\mu-\mathbf{1}r)'\Sigma^{-1}(\mu-\mathbf{1}r)} \cdot \Sigma^{-1}(\mu-\mathbf{1}r) \end{aligned}$$

und $\phi_0 = 1 - \phi^{*\prime}\mathbf{1}$

Wir erhalten

$$\phi_M = \frac{\phi^*}{\mathbf{1}'\phi^*} = \frac{\Sigma^{-1}(\mu-\mathbf{1}r)}{\mathbf{1}'\Sigma^{-1}(\mu-\mathbf{1}r)}. \qquad \blacksquare$$

In der Notation "Marktportfolio" spielt also die vorher festgelegte Rendite $\overline{\mu}$ keine Rolle mehr. Man wählt schlicht ein ϕ_0, womit man die erwartete Rendite optimal erzielt. Wie erhält man nun in der Praxis ein optimales Portfolio? Man orientiert sich an einem gewissen, relativ großen Portfolio etwa dem DAX (oder S & P 500).

Die entsprechenden μ und Σ ersetzt man durch die empirische Schätzer

$$\begin{aligned} \hat{\mu}_N &= \frac{1}{N}\sum_{i=1}^{N} R_i \\ \hat{\Sigma}_N &= (\sigma_{ij}^r)_{i\sigma=1}^n \ \text{ mit } \sigma_{ij} = \frac{1}{n-1}\sum_{i-1}^{N}(R_i - \mu_N^i)(R_i - \mu_N^j) \end{aligned}$$

Betrachtet man beispielsweise ein Portfolio aus Allianz, BASF, Bayer, BMW, Siemens und VW erhält man (Feb. 95 - Dez. 2004)

$$\hat{\mu} = \begin{pmatrix} 0.54 \\ 1.29 \\ 0,70 \\ 1.24 \\ 1.58 \\ 0.93 \end{pmatrix} \qquad \hat{\Sigma} = \begin{pmatrix} 126 & & & & & \\ 48 & 61 & & & & \\ 71 & 55 & 89 & & & \\ 50.5 & 44 & 50 & 85 & & \\ 75 & 44 & 52 & 46 & 144 & \\ 55.7 & 43 & 46 & 60 & 50.7 & 103 \end{pmatrix}$$

mit $r = 0.035 \Rightarrow \phi(0.4, 1.3, -0.7, 0.5, 0.5, -0.17)'$

Wie hoch ist der Beitrag einer Einzelaktie zur gesamten Rendite?

Diagramm

Dazu machen wir folgende interessante Beobachtung:

Varianz und EW des Marktportfolios bezeichnen wir mit σ_M^2 bzw. μ. Dann gilt

$$\sigma^2 = \phi_M' \Sigma \phi_M = \frac{\phi_M'(\mu - \mathbf{1}r)}{1'\Sigma^{-1}(\mu - \mathbf{1}r)}.$$

Erinnern wir uns:

$$\phi_0 r + \phi_M' \mu = \mu_M \Rightarrow (1 - \phi_M' \mathbf{1})r + \phi_M' \mu = \mu_M$$

also

$$\phi_M'(\mu - 1r) = \mu_m - r.$$

Damit ist

$$\sigma_M^2 = \frac{\mu_m - r}{\mathbf{1}'\Sigma^{-1}(\mu - \mathbf{1}r)}$$

Wir erhalten

$$\frac{\mu_m - r}{\sigma_2} = \mathbf{1}'\Sigma^{-1}(\mu - \mathbf{1}r)$$

$$\Rightarrow \Sigma\phi_M = \frac{\mu - \mathbf{1}r}{\mathbf{1}'\Sigma^{-1}(\mu - \mathbf{1}r)} = (\mu - \mathbf{1}r) \cdot \frac{\sigma^2}{\mu_M - r}$$

Schauen wir uns die i-te Zeile dieses GLS an:

$$\underbrace{\phi_M^1 \operatorname{Cov}(R_i, R_1) + \cdots + \phi_M^n \operatorname{Cov}(R_i, R_n)}_{\operatorname{Cov}(R_i, \phi_M' R)} = (\mu_i - r) \cdot \frac{\sigma^2}{\mu_M - r}$$

und wir erhalten

$$\mu_i - r = (\mu_M - r) \cdot \underbrace{\frac{\operatorname{Cov}(R_i, R_M)}{\sigma_M^2}}_{=:\beta_i}.$$

Der Faktor, um welchen der Renditeüberschuss der Aktie i vom Renditeüberschuss des Marktportfolios abweicht, nennt man β.

Obige Gleichung bezeichnet man auch als Wertpapiertmarktlinie.

Diesen Zusammenhang kann man auch unmittelbar zur Bewertung einer Aktie nutzen:

Beachte: $\mu_i = \mathbb{E}(R_i) = \mathbb{E}\left(\frac{S_1^i - S_0^i}{S_0^i}\right) = \frac{\mathbb{E}(S_1^i}{S_0^i} - 1$ und deswegen $S_0^i = \frac{\mathbb{E}(S_1^i)}{1+\mu_i} = \frac{\mathbb{E}(S_1^i)}{1+r+\beta_i-(\mu-r)}$

Wir erhalten also eine Bewertungsregel, in der der Diskontierungsfaktor gemäß der Rendite der Aktie adjustiert wird, und zwar über den Faktor β.

3.1.3 Schätzung des $\beta's$

Eine Schätzung des $\beta's$ benutzt eine zeitliche Entwicklung der Renditen. Wir betrachten das Mehrperiodenmodell und setzen

$$Z_t^i := R_t^i - r_t.$$

Für das Marktportfolio schreiben wir Z_t^M. Als empirisches Modell nehmen wir an, dass

$$Z_t^i = \beta_i Z_t^M + \epsilon_{it}$$

mit $i = 1, \ldots, n$ und $t = 1, \ldots, T$, wobei $\epsilon_{it} \sim$ i.i.d. $\mathcal{N}(0, \sigma^2)$.

Der Kleinst-Quadrate-Schätzer minimiert

$$\sum_{t=1}^{T} (Z_t - \beta Z_t^M)^2 \to \min.$$

Hinreichend ist

$$\sum_t \left[Z_t - \hat{\beta} Z_t^M \right] \cdot (-Z_t^M) = 0$$

$$\Leftrightarrow \hat{\beta} = \frac{\sum_{t=1}^T Z_t Z_t^M}{\sum_{t=1}^T (Z_t^M)^2}.$$

Für Tests nutzen wir die Erwartungstreue, und dass

$$\frac{\hat{\beta} - \beta}{\sqrt{\sigma^2 (\sum_t (Z_t^M)^2)^{-1}}} \Big|_{(Z_t^M)_{t=1,\ldots,T}} \sim \mathcal{N}(0, 1),$$

wobei σ natürlich leicht aus den Daten geschätzt werden kann.

Man könnte z.B. auf $\beta = 0$ testen. Ebenso interessant ist der Test auf einen Achsenabschnitt $\alpha \neq 0$

Vergleiche auch Schmidt und Trede (2006).

A Die goldenen Regeln

In seinem Skriptum zur Stochastik III an der Universität Gießen stellte Herr Prof. Dr. Stute die "Goldenen Regeln" vor. Sie sind das alltägliche Handwerkszeug für den Umgang mit bedingten Erwartungen und werden in diesem Abschnitt, allerdings ohne Beweise, zitiert. Es ist durchaus eine lohnenswerte Übung, die Aussagen zu beweisen.

Sei $(\Omega, \mathcal{A}, \mathbb{P})$ der zugrundeliegende Wahrscheinlichkeitsraum. Wir bezeichnen mit $L^1(\Omega, \mathcal{A}, \mathbb{P})$ die Menge aller $\mathbb{P}$-integrierbaren Funktionen $\xi : \Omega \mapsto \mathbb{R}$.

Definition A.1.6. *Sei $\xi \in L^1(\Omega, \mathcal{A}, \mathbb{P})$ und $\mathcal{B} \subset \mathcal{A}$. Eine Zufallsvariable $z \in L^1(\Omega, \mathcal{A}, \mathbb{P})$ heisst* **bedingte Erwartung** *(von ξ gegeben $\mathcal{B}$) genau dann, wenn*

(i) z ist $\mathcal{B}$-messbar

(ii) für alle $B \in \mathcal{B}$ gilt

$$\int_B z \, d\mathbb{P} = \int_B \xi \, d\mathbb{P}.$$

Wir setzen

$$\mathbb{E}(\xi|\mathcal{B}) := \{z : z \text{ erfüllt (i) und (ii) von A.1.6}\}.$$

Es ist üblich auch für jedes z auch $\mathbb{E}(\xi|\mathcal{B})$ zu schreiben. Immerhin gilt für $z_1, z_2 \in \mathbb{E}(\xi|\mathcal{B})$, dass $z_1 = z_2$ $\mathbb{P}$-f.s. Die bedingte Wahrscheinlichkeit ist definiert durch $\mathbb{P}(A|\mathcal{B}) := \mathbb{E}(1_A|\mathcal{B})$.

Einige Beispiele mögen das Konzept erläutern.

(i) Ist $\mathcal{B} = \{\emptyset, \Omega\}$, so ist $\mathbb{E}(\xi|\mathcal{B}) = \mathbb{E}(\xi)$.

(ii) Wird $\mathcal{B}$ durch paarweise disjunkte $A_i \in \mathcal{A}$, erzeugt, so gilt

$$\mathbb{E}\Big(\xi|\mathcal{B}\Big) = \sum_{i=1}^{\infty} 1_{A_i} \frac{\int_{A_i} \xi \, d\mathbb{P}}{\mathbb{P}(A_i)},$$

wobei man den Bruch gleich Null setzt, wenn $\mathbb{P}(A_i) = 0$.

Der nun folgende erste Satz von Goldenen Regeln bildet die Grundausstattung. Vor allem Punkt (i) haben wir bereits wiederholt in der Vorlesung eingesetzt. (wo?)

(1. Goldene Regel) Sei $\xi, \xi_1, \xi_2 \in L^1(\Omega, \mathcal{A}, \mathbb{P})$ und $a, b \in \mathbb{R}$. Dann gilt

(i) $\int_B \mathbb{E}(\xi|\mathcal{B})\, d\mathbb{P} = \int_B \xi\, d\mathbb{P}$ für alle $B \in \mathcal{B}$

(ii) $\mathbb{E}\Big(\mathbb{E}(\xi|\mathcal{B})\Big) = \mathbb{E}(\xi)$

(iii) $\xi = a$ $\mathbb{P}$-f.s. $\Rightarrow \mathbb{E}(\xi|\mathcal{B}) = a$ $\mathbb{P}$-f.s.

(iv) $0 \leq \xi$ $\mathbb{P}$-f.s. $\Rightarrow 0 \leq \mathbb{E}(\xi|\mathcal{B})$ $\mathbb{P}$-f.s.

(v) $\mathbb{E}(a_1\xi_1 + a_2\xi_2|\mathcal{B}) = a\mathbb{E}(\xi_1|\mathcal{B}) + b\mathbb{E}(\xi_2|\mathcal{B})$ $\mathbb{P}$-f.s.

(vi) $\xi_1 \leq \xi_2$ $\mathbb{P}$-f.s. $\Rightarrow \mathbb{E}(\xi_1|\mathcal{B}) \leq \mathbb{E}(\xi_2|\mathcal{B})$ $\mathbb{P}$-f.s.

(vii) ξ ist $\mathcal{B}$-messbar $\Rightarrow \mathbb{E}(\xi|\mathcal{B}) = \xi$.

(2. Goldene Regel) Sei $\xi \in L^1(\Omega, \mathcal{A}, \mathbb{P})$ und η $\mathcal{B}$-messbar (!), so dass $\eta \cdot \xi \in L^1(\Omega, \mathcal{A}, \mathbb{P})$. Dann gilt

$$\mathbb{E}\Big(\eta \cdot \xi|\mathcal{B}\Big) = \eta\, \mathbb{E}\Big(\xi|\mathcal{B}\Big) \qquad \mathbb{P}\text{-f.s..}$$

Diese Regel erlaubt es, messbare Faktoren aus dem bedingten Erwartungswert herauszuziehen. In der dritten Goldenen Regel behandeln wir zwei unterschiedliche σ-Algebren.

(3. Goldene Regel) Für $\xi \in L^1(\Omega, \mathcal{A}, \mathbb{P})$ und $\mathcal{B}_1 \subset \mathcal{B}_2 \subset \mathcal{A}$ gilt

$$\mathbb{E}\Big[\mathbb{E}(\xi|\mathcal{B}_1)|\mathcal{B}_2\Big] = \mathbb{E}(\xi|\mathcal{B}_1) = \mathbb{E}\Big[\mathbb{E}(\xi|\mathcal{B}_2)|\mathcal{B}_1\Big] \qquad \mathbb{P}\text{-f.s..}$$

Was hat man in diesem neuen Kontext unter Unabhängigkeit zu verstehen ? Zunächst können σ-Algebren unabhängig voneinander sein.

Definition A.1.7. *$\mathcal{B}_1$ und $\mathcal{B}_2$ heißen ($\mathbb{P}$-)unabhängig genau dann, wenn*

$$\mathbb{P}(B_1 \cap B_2) = \mathbb{P}(B_1)\,\mathbb{P}(B_2) \qquad \textit{für alle } B_i \in \mathcal{B}_i, i = 1, 2.$$

Für ξ_i B_i-messbar erhalten wir sofort $\mathbb{E}(\xi_1\xi_2) = \mathbb{E}(\xi_1)\mathbb{E}(\xi_2)$. Was aber sind unabhängige Zufallsvariablen? Dies kann man auf σ-Algebren zurückführen. Dazu definiert man die von ξ **generierte σ-Algebra** durch

$$\sigma(\xi) := \Big\{\xi^{-1}(D) : D \subset \mathcal{B}(\mathbb{R})\Big\},$$

wobei $\mathcal{B}(\mathbb{R})$ die Borel σ-Algebra auf $\mathbb{R}$ bezeichnet. Damit erhalten wir

(4. Goldene Regel) Sei $\xi \in L^1(\Omega, \mathcal{A}, \mathbb{P})$ und $\sigma(\xi)$ unabhängig von $\mathcal{B}$. Dann ist

$$\mathbb{E}(\xi|\mathcal{B}) = \mathbb{E}(\xi) \qquad \mathbb{P}\text{-f.s..}$$

Wenn man den bedingten Erwartungswert als Erwartungswert auf Basis von gewissen Informationen (z.B. wenn $\mathcal{B} = \mathcal{F}_t$ ist) interpretiert, so bedeutet die 4. Goldene Regel, dass unter Unabhängigkeit diese Information keinen Zusatznutzen bringt, der bedingte Erwartungswert ist gleich dem unbedingten Erwartungswert.

Wir schreiben

$$\mathbb{E}(\xi|\eta) := \mathbb{E}\Big(\xi \,\Big|\, \sigma(\eta)\Big)$$

für die **bedingte Erwartung bezüglich** η.

In diesem Fall hat die bedingte Erwartung eine besonders schöne Gestalt.

Lemma A.1.8. *Es existiert eine messbare Funktion* $f : \mathbb{R} \mapsto \mathbb{R}$*, so dass*

$$\mathbb{E}(\xi|\eta) = f(\eta).$$

Diese Funktion f nennt man die **Faktorisierung** von ξ bezgl. η. Mann schreibt dann

$$f(x) =: \mathbb{E}(\xi|\eta = x).$$

In der Statistik nennt man die Funktion f auch die **Regressionsfunktion**. Man kann dann die bedingte Dichte ausrechnen. Sei $f(x, y)$ die gemeinsame Dichte von (ξ, η) und f die Dichte von η. Dann gilt auf $\{x : f(x) > 0\}$

$$\mathbb{E}(\xi|\eta = x) = \frac{\int y\, f(x, y)\, dy}{f(x)} =: \int y\, f(y|x)\, dy.$$

Die bedingte Dichte $f(y|\eta = x)$ ist also gerade $f(x, y)/f(x)$.

In der Finanzmathematik erweist sich die nächste Regel als äußerst nützlich.

(5. Goldene Regel) ξ und η seien unabhängig und $T : \mathbb{R}^2 \mapsto \mathbb{R}$ so dass $T(\xi, \eta) \in L^1(\Omega, \mathcal{A}, \mathbb{P})$. Dann haben wir f.s. (bezgl. der Verteilung von η), dass

$$\mathbb{E}\Big[T(\xi, \eta)|\eta = x\Big] = \mathbb{E}\Big[T(\xi, x)\Big].$$

Für den bedingten Erwartungswert $\mathbb{E}(T(\xi, \eta)|\eta)$ müssen wir also den unbedingten Erwartungswert rechts ausrechnen und dann wieder für x η einsetzen.

Ein Beispiel: ist $T(x, y) = xy$, so schließt man, dass $\mathbb{E}(\xi\eta|\eta) = \eta\mathbb{E}(\xi)$, was wir natürlich schon von der 2. Goldenen Regel wissen.

Betrachten wir eine diskrete Filtration $\{\mathcal{F}_n\}_{n\in T}$ und diesbezüglich diskrete Stoppzeiten, so definieren wir die σ-Algebra der bis τ anfallenden Information durch

$$\mathcal{F}_\tau := \{A \in \mathcal{A} : A \cap \{\tau \le n\} \in \mathcal{F}_n \text{ für alle } n \in T\}.$$

(6. Goldene Regel) Ist $\xi \in L^1(\Omega, \mathcal{A}, \mathbb{P})$ und τ eine Stoppzeit bezüglich der diskreten Filtration $\{\mathcal{F}_n\}_{n\in T}$, so gilt

$$\mathbb{E}(\xi|\mathcal{F}_\tau) = \mathbb{E}(\xi|\mathcal{F}_n) \qquad \mathbb{P}\text{-f.s. auf} \quad \{\tau = n\}.$$

Das Konzept der Unabhängigkeit kann man im Rahmen mit bedingten Erwartungen neu fassen.

Definition A.1.9. *Sei $\mathcal{F} \subset \mathcal{A}$ eine σ-Algebra. $A_1, \ldots, A_n$ heißen bedingt unabhängig bezgl. $\mathcal{F}$ genau dann, wenn*

$$\mathbb{P}\Big(\bigcap_{i=1}^{n} A_i|\mathcal{F}\Big) = \prod_{i=1}^{n} \mathbb{P}(A_i|\mathcal{F}) \qquad \mathbb{P}\text{-}f.s..$$

Sind weiterhin $\mathcal{F}_1, \mathcal{F}_1, \mathcal{F}_3 \subset \mathcal{A}$ σ-Algebren, so nennen wir $\mathcal{F}_1$ und $\mathcal{F}_2$ unabhängig bedingt $\mathcal{F}_3$, falls für alle $F_1 \in \mathcal{F}_1, F_2 \in \mathcal{F}_2$

$$\mathbb{P}(F_1 \cap F_2|\mathcal{F}_3) = \mathbb{P}(F_1|\mathcal{F}_3)\, \mathbb{P}(F_2|\mathcal{F}_3).$$

(7. Goldene Regel) $\mathcal{F}_1$ und $\mathcal{F}_2$ sind unabhängig bedingt $\mathcal{F}_3$ genau dann, wenn für jedes $\xi \in L^1(\Omega, \mathcal{F}_1, \mathbb{P})$

$$\mathbb{E}\Big[\xi|\sigma(\mathcal{F}_2 \cup \mathcal{F}_3)\Big] = \mathbb{E}(\xi|\mathcal{F}_3) \qquad \mathbb{P}\text{-f.s. .}$$

Literatur

Brockwell, P. J. und R. A. Davis (1991). *Time Series: Theory and Methods* (2nd ed.). Springer Verlag. Berlin Heidelberg New York.

Duan, J. C. (1995). The garch option pricing model. *Mathematical Finance*, 13–32.

Engle, R. F. (1982). Autoregressive conditional heteroscedasticity with estimates of the variance of U. K. inflation. *Econometrica 50*, 987–1008.

Hafner, C. und W. Härdle (2000). Discrete time option pricing with flexible volatility estimation. *Finance and Stochastics*, 189–207.

Schmidt, F. und M. Trede (2006). *Finanzmarktstatistik*. Springer Verlag. Berlin Heidelberg New York.

Schmidt, T. und W. Stute (2006). A moment estimator for the march- modell. *Working paper*.