Das GARCH Modell zur Modellierung von Finanzmarktzeitreihen

Seminararbeit von Frauke Heuermann

Juni 2010

Inhaltsverzeichnis

0 Einleitung

In den bereits kennengelernten ARMA-Prozessen war die Varianz der zu untersuchenden Zeitreihe immer als konstant vorausgesetzt. Dagegen nimmt der ARCH-und später der GARCH-Prozess eine sich verändernde Varianz an, was besser die Realität widerspiegelt. Dies sieht man z.B. daran, dass Aktienkurse während Krisenperioden stärker schwanken als zu ruhigeren Zeiten. Deshalb sind GARCH-Modelle besonders geeignet Finanzmarktzeitreihen zu modellieren.

1 Der ARCH-Prozess

Wir wollen uns zunächst mit dem ARCH-Modell befassen.

Definition 1.1 (ARCH-Prozess)**.** *Es sei $(Z_t)_{t\in\mathbb{Z}}$ ein iid Weißes Rauschen. $(X_t)_{t\in\mathbb{Z}}$ heißt dann autoregressive conditional heteroskedasticity (kurz: ARCH(p)) Prozess der Ordnung p, wenn er strikt stationär ist und die folgenden Gleichungen für alle $t \in \mathbb{Z}$ erfüllt*

$$X_t = \sigma_t Z_t \tag{1}$$

$$\sigma_t^2 = \alpha_0 + \sum_{i=1}^{p} \alpha_i X_{t-i}^2 \tag{2}$$

mit $\alpha_0 > 0$, $\alpha_i \geq 0$ $\forall i = 1, \ldots, p$ und einen strikt positiv-wertigen Prozess $(\sigma_t)_{t\in\mathbb{Z}}$.

Die Begriff ARCH erklärt sich anhand der folgenden Definitionen:
−autoregressiv bedeutet, dass X_t von seiner Vergangenheit abhängt, also von X_{t-1} abhängig ist
−bedingt heteroskedastisch bedeutet, dass die bedingte Varianz nicht konstant ist.

Bemerkung 1.2. *Bezeichne $\mathcal{F}_t = \sigma(X_s : s \leq t)$. Dann ist $(\mathcal{F}_t)_{t\in\mathbb{Z}}$ die natürliche Filtration. Offenbar ist σ_t messbar bezüglich $\mathcal{F}_{t-1}$. Ist dann $E(|X_t|) < \infty$, so folgt, da zusätzlich Z_t unabhängig von $\mathcal{F}_{t-1}$ ist:*

$$E(X_t|\mathcal{F}_{t-1}) = E(\sigma_t Z_t|\mathcal{F}_{t-1}) = \sigma_t \underbrace{E(Z_t)}_{=0} = 0$$

D.h. der ARCH-Prozess ist eine Martingaldifferenz bezüglich $(\mathcal{F}_t)_{t\in\mathbb{Z}}$.

Die Unabhängigkeit von Z_t und $\mathcal{F}_{t-1}$ folgt, da der ARCH-Prozess kausal ist, d.h. die Gleichungen (1) und (2) müssen eine Lösung der Form $X_t = f(Z_t, Z_{t-1}, \ldots)$ haben, so dass Z_t unabhängig von seinen früheren Werten ist. Dies ist jedoch gegeben, da Z_t als iid vorausgesetzt ist. Der ARCH-Prozess ist kausal, weil er nur von der Vergangenheit abhängt.

1.1 Das ARCH(1)-Modell

Für die weitere Diskussion der Eigenschaften betrachten wir nun das ARCH(1)-Modell.
Im Fall $p = 1$ folgt aus (1) $X_t^2 = \sigma_t^2 Z_t^2$ und damit aus (2):

$$X_t^2 = \sigma_t^2 Z_t^2 = (\alpha_0 + \alpha_1 X_{t-1}^2) Z_t^2 = \alpha_0 Z_t^2 + \alpha_1 X_{t-1}^2 Z_t^2 \tag{3}$$

Um das ARCH(1)-Modell genauer zu verstehen, müssen wir herausfinden, wann diese Gleichung stationäre Lösungen, die wir mithilfe von (Z_t) ausdrücken wollen, besitzt. Wir suchen somit Lösungen der Form $X_t^2 = f(Z_t, Z_{t-1}, \ldots)$ für eine Funktion f. Dabei müssen wir zwischen schwach und stark stationären Lösungen unterscheiden.

Satz 1.3. *Sei*

$$Y_t = A_{t-1} Y_{t-1} + B_t \tag{4}$$

mit Folgen von unabhängig, identisch verteilten Zufallsgrößen $(A_t)_{t\in\mathbb{Z}}$ und $(B_t)_{t\in\mathbb{Z}}$. Dann ist eine hinreichende Bedingung für die stationäre Lösung von dieser Gleichung

$$E(ln^+(|B_t|)) < \infty$$

und

$$E(ln|A_t|) < 0$$

Dabei bezeichnet $ln^+(x) = max(0, lnx)$.
Die eindeutige Lösung ist dann gegeben durch

$$Y_t = B_t + \sum_{i=1}^{\infty} B_{t-i} \prod_{j=0}^{i-1} A_{t-j} \tag{5}$$

und die Summe konvergiert dabei fast sicher absolut.

Durch k-maliges Einsetzen obiger Gleichung erhält man

$$\begin{aligned} Y_t &= A_t Y_{t-1} + B_t && (6)\\ &= A_t(A_{t-1}Y_{t-2} + B_{t-1}) + B_t && (7)\\ &= B_t + \sum_{i=1}^{k} B_{t-1} \prod_{j=0}^{i-1} A_{t-j} + Y_{t-k-1} \prod_{i=0}^{k} A_{t-i} && (8) \end{aligned}$$

1.2 Stationarität des ARCH(1)-Prozesses

Wir wollen nun das ARCH(1)-Modell auf strikte Stationarität überprüfen. Dazu stellen wir zunächst fest, dass das ARCH(1)-Modell von der Form (4) mit $A_t = \alpha_1 Z_t^2$ und $B_t = \alpha_0 Z_t^2$ ist. Diese sind offenbar unabhängig, identisch verteilt, da $(Z_t) \sim SWN(0, 1)$. Die Bedingungen für Stationarität lassen sich dann also umformen zu

$$E(\ln^+(|\alpha_0 Z_t^2|) < \infty$$

Man kann zeigen, dass dies eine notwendige und hinreichende Bedingung für die strikte Statinarität ist (vgl. Bougerol und Picard 1992). Nach (8) ist dann

$$X_t^2 = \alpha_0 \sum_{i=0}^{\infty} \alpha_1^i \prod_{j=0}^{i} Z_{t-j}^2 \tag{9}$$

eine Lösung von (3).
Strikte Stationarität hängt also von der Verteilung der Innovationen (Z_t) ab. So ist zum Beispiel das Kriterium für strikte Stationarität, falls Z_t standard normalverteilt ist :$\alpha_1 < 3.562$, oder falls Z_t t-verteilt mit vier Freiheitsgraden und Varianz 1 ist :$\alpha_1 < 5.437$.

Wir wollen nun ein Kriterium für die schwache Stationarität für das ARCH(1)-Modell kennenlernen:

Satz 1.4. *Der ARCH(1)-Prozess $(X_t)_{t\in\mathbb{Z}}$ ist genau dann ein schwach-stationäres weißes Rauschen, wenn $\alpha_1 < 1$ gilt. Dann ist die Varianz gegeben durch $\frac{\alpha_0}{1-\alpha_1}$.*

Beweis: „$\Longrightarrow$“Sei also $(X_t)_{t\in\mathbb{Z}}$ ein schwach stationäres Weißes Rauschen. Dann gilt aufgrund der Gleichung (**??**) und $E(Z_t^2) = 1$ für die Varianz von X_t:

$$\begin{aligned}
Var(X_t) &= E(X_t^2) + \underbrace{(E(X_t))^2}_{=0, E(X_t)=\underbrace{E(Z_t)}_{=0} E(\sigma_t)} \\
&= \alpha_0 + \alpha_1 E(X_{t-1}^2) \\
&= \alpha_0 + \alpha_1 Var(X_t)
\end{aligned}$$

$\Rightarrow Var(X^2) = \frac{\alpha_0}{1-\alpha_1}$
Aus dieser Gleichung folgt dann offenbar auch $\alpha_1 < 1$, da die Varianz positiv ist und $\alpha_0 > 0$ gilt.
„$\Longleftarrow$“Sei nun $\alpha_1 < 1$. Aufgrund der Jensenschen Ungleichung und wieder wegen $E(Z_t^2) = 1$ gilt:

$$E(\ln(\alpha_1 Z_t^2)) \leq \ln(E(\alpha_1 Z_t^2)) = \ln(\alpha_1) < 0 \tag{10}$$

da $\alpha_1 < 1$. (X_t) ist eine Martingaldifferenz mit endlichem, zeitlich unabhängigem zweiten Moment. Somit ist (X_t) ein Weißes Rauschen. □

Satz 1.5. *Für $m \geq 1$ hat der strikt stationäre ARCH(1)-Prozess genau dann endliche Momente der Ordnung 2m, wenn $E(Z_t^{2m}) < \infty$ und $\alpha_1 < (E(Z_t^{2m}))^{-\frac{1}{m}}$*

Beweis: Wir können zunächst (9) umformen zu $X_t^2 = Z_t^2 \sum_{i=0}^{\infty} Y_{t,i}$ für positive Zufallsvariablen $Y_{t,i} = \alpha_0 \alpha_1 \prod_{j=1}^{i} Z_{t-j}^2$, $i \geq 1$ und $Y_{t,0} := \alpha_0$. Sei $m \geq 1$, dann gilt:

$$E(Y_{t,1}^m) + E(Y_{t,2}^m) \leq E((Y_{t,1}^m + Y_{t,2}^m)^m) \leq ((E(Y_{t,1}^m))^{\frac{1}{m}} + (E(Y_{t,2}^m))^{\frac{1}{m}})^m \tag{11}$$

unter Benutzung der Minkowskischen Ungleichung. Wegen

$$E(X_t^{2m}) = E(Z_t^{2m})E((\sum_{i=0}^{\infty} Y_{t,i})^m)$$

gilt

$$E(Z_t^{2m})\sum_{i=0}^{\infty} E(Y_{t,i}^m) \leq E(X_t^{2m}) \leq E(Z_t^{2m})(\sum_{i=0}^{\infty}(E(Y_{t,i}^m))\frac{1}{m})^m$$

Aus $E(Y_{t,j}^m) = \alpha_0^m \alpha_1^m (E(Z_t^{2m}))i$ folgt nun, dass die drei Mengen genau dann endlich sind, wenn $E(Z_t^{2m}) < \infty$ und $\alpha_1 < (E(Z_t^{2m}))^{-\frac{1}{m}}$ gilt. □

1.3 Das vierte Moment und die Wölbung

Man nennt κ_Z die Kurtosis oder Wölbung von Z, falls gilt

$$\kappa_Z = \frac{E(Z^4)}{(Var(Z^2))^2}.$$

Damit $E(X^4) < \infty$ gilt, müssen wir fordern, dass $\alpha_1 < \frac{1}{\sqrt{3}}$ für normal verteilte Innovationen und $\alpha_1 < \frac{1}{\sqrt{6}}$ im Fall von t-verteilten Innovationen mit mindestens sechs Freiheitsgraden, da das vierte Moment für nur vier Freiheitsgrade nicht definiert ist. Dafür kann man obigen Satz für $m = 2$ anwenden. Wenn man nun die Existenz des vierten Moments voraussetzt und den Erwartungswert von dem Quadrat von (3) bildet, erhält man

$$EX_t^4 = \frac{\alpha_0^2 EZ_t^4(1-\alpha_1^2)}{(1-\alpha_1)^2(1-\alpha_1^2 EZ_t^4)}$$

und für die Wölbung von X_t gilt:

$$\kappa_X = \frac{EX_t^4}{E(X_t^2)^2} = \frac{\kappa_Z(1-\alpha_1^2)}{(1-\alpha_1^2\kappa_Z)}$$

wobei $\kappa_Z = EZ_t^4$ die Wölbung der Innovationen ist. Für normalverteilte und t Innovationen ist $\kappa_X > 3$, d.h. die Wölbung ist leptokurtisch oder steilgipflig.

1.4 Parallelen zwischen dem AR(1)-Prozess und dem ARCH(1)-Prozess

Wie der Name schon sagt, existiert ein Zusammenhang zwischen dem ARCH- und dem AR-Prozess. Sei $(X_t)_{t\in\mathbb{Z}}$ ein schwach stationärer ARCH(1)-Prozess, also gilt $\alpha_1 < 1$. Wir schreiben dann

$$X_t^2 = \sigma_t^2 Z_t^2 = \alpha_t^2 + V_t = \alpha_0 + \alpha_1 X_{t-1}^2 + V_t \tag{12}$$

Hierbei bezeichnet $V_t = \sigma_t^2(Z_t^2 - 1)$ und es gilt $(V_t)_{t\in\mathbb{Z}}$ ist eine Folge von Martingaldifferenzen, wenn $E|V_t| < \infty$ und $E(V_t|\mathcal{F}_{t-1}) = \sigma_t^2 E(Z_t^2 - 1) = 0$ gilt. Dabei bemerkt man schnell die Ähnlichkeit zu einem AR(1)-Prozess, der sich nur darin unterscheidet, dass V_t nicht notwendigerweise ein Weißes Rauschen seien muss.

Das kann man aber dadurch erreichen, dass man $E(X_t^4) < \infty$ fordern. Dann gilt nämlich:

$$E(\sigma_t^2(Z_t^2 - 1)) = E\sigma_t^2 E(Z_t^2 - 1) = 0$$

$$\begin{aligned}
Var(\sigma_t^2(Z_t^2 - 1)) &= E\sigma_t^4 E((Z_t^2 - 1)^2) \\
&= 2E(\alpha_0 + \alpha_1 X_{t-1}^2)^2 \\
&= 2\alpha_0^2 + \alpha_1\alpha_0 \underbrace{EX_{t-1}^2}_{<\infty} + \alpha_1^2 \underbrace{EX_{t-1}^4}_{<\infty} \\
&= KONST
\end{aligned}$$

Und somit ist V_t ein Weißes Rauschen.

Für diesen Fall folgt dann also, dass X_t^2 ein AR(1)-Prozess von der Form

$$\left(X_t^2 - \frac{\alpha_0}{1-\alpha_1}\right) = \alpha_1 \left(X_{t-1}^2 - \frac{\alpha_0}{1-\alpha_1}\right) + V_t$$

mit Erwartungswert $\frac{\alpha_0}{1-\alpha_1}$ und ACF $\rho(h) = \alpha_1^{|h|}$, $h \in \mathbb{Z}$ ist. D.h. die Quadrate eines ARCH(1)-Prozesses (X_t^2) folgen einem AR(1)-Prozess.

2 Der GARCH-Prozess

Definition 2.1 (GARCH-Prozess). *Sei $(Z_t)_{t\in\mathbb{Z}} \sim$ SWN(0,1). $(X_t)_{t\in\mathbb{Z}}$ heißt dann verallgemeinerter ARCH-Prozess der Ordnung (p,q) (kurz: GARCH(p,q)-Prozess),wenn die folgenden Gleichungen erfüllt sind:*

$$\begin{aligned}
X_t &= \sigma_t Z_t \\
\sigma_t^2 &= \alpha_0 + \sum_{i=1}^{p} \alpha_i X_{t-i}^2 + \sum_{j=1}^{q} \beta_j \sigma_{t-j}^2
\end{aligned}$$

für $\alpha_0 > 0$, $\alpha_i \geq 0$, $\forall i = 1, \ldots, p$ und $\beta_j \geq 0, \forall j = 1, \ldots, q$ und einen strikt positiv-wertigen Prozess $(\sigma_t)_{t\in\mathbb{Z}}$.

Für $q = 0$ entspricht das GARCH-Modell einem ARCH-Modell, was die Bezeichnung des GARCH-Prozesses erklärt. Jedoch hängt die Varianz eines GARCH-Modells noch von den früheren Varianzen und Werten des Prozesses ab.

2.1 Der GARCH(1,1)-Prozess

In der Praxis werden häufig GARCH(1,1)-Modelle verwendet, da die Anzahl der unbekannten Parameter sehr gering ist. Im GARCH(1,1)-Modell tendieren Perioden mit hoher Volatilität persistent zu werden, da $|X_t|$ groß sein kann, falls entweder $|X_{t-1}|$ oder σ_{t-1} groß ist . Den gleichen Effekt kann man bei einem ARCH-Modell beobachten, jedoch wird der Zustand bei einem GARCH-Modell schon wesentlich früher erreicht.

Wie schon bei dem ARCH(1)-Modell stellt sich die Frage, wann eine schwach stationäre Lösung (X_t) existiert. Diese wird durch den folgenden Satz beantwortet:

Satz 2.2. *Der GARCH(1,1)-Prozess hat genau dann eine schwach stationäre Lösung (X_t), wenn*

$$\alpha_1 + \beta_1 < 1$$

gilt. Dann folgt $Var(X_t) = \frac{\alpha_0}{1-\alpha_1-\beta_1}$.

Beweis: Der Beweis verläuft analog zu dem Beweis für das ARCH-Modell. □

Bemerkung 2.3. *Für das GARCH(1,1)-Modell gilt*

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

und besitzt somit die Form (4) für $Y_t = \sigma_t^2$, $A_t = \alpha_1 Z_{t-1}^2 + \beta_1$ und $B_t = \alpha_0$. Die Bedingung für strikte Stationarität lässt sich dann umformen zu $E(ln(\alpha_1 Z_{t-1}^2 + \beta_1)) < 0$ und die allgemeine Lösung lautet dann

$$\sigma_t^2 = \alpha_0 + \alpha_0 \sum_{i=0}^{\infty} \prod_{j=1}^{i} (\alpha_1 Z_{t-j}^2 + \beta_1)$$

Ist dann $(\sigma_t^2)_t$ strikt stationär, so gilt dies auch für X_t, falls $X_t = \sigma_t Z_t$ und $Z_t \sim SWN$ gilt. Dann ist die strikt stationäre Lösung für X_t:

$$X_t = Z_t(\alpha_0(1 + \sum_{i=0}^{\infty} \prod_{j=1}^{i} (\alpha_1 Z_{t-j}^2 + \beta_1)))^{\frac{1}{2}}$$

2.2 Das vierte Moment und die Wölbung

Beh.: Eine notwendige und hinreichende Bedingung für die Existenz des 4ten Moments ist $E((\alpha_1 Z_t^2 + \beta_1)^2) < 1$ oder äquivalent dazu $(\alpha_1 + \beta_1)^2 < 1 - (\kappa_Z - 1)\alpha_1^2$

Beweis: Es folgt aus der obigen Gleichung für σ_t:

$$\begin{aligned} E(\sigma_t^4) &= (\alpha_0 + (\alpha_1 Z_{t-1}^2 + \beta_1)\sigma_{t-1}^2) \\ &= \alpha_0^2 + 2\alpha_0(\alpha_1 E(Z_{t-1}^2) + \beta_1)E(\sigma_{t-1}^2) + (\alpha_1 E(Z_{t-1}^2) + \beta_1)^2 E(\sigma_{t-1}^4) \\ &= \alpha_0^2 + 2\alpha_0(\alpha_1 + \beta_1)E(\sigma_t^2) + (\alpha_1^2 \kappa_Z + \beta_1\alpha_1 + \beta_1^2)E(\sigma_t^4) \end{aligned}$$

Wenn wir nun diese Gleichung nach $E(\sigma_t^4)$ auflösen und benutzen, dass $E(\sigma_t^2) = E(Xt^2) = \alpha_0/(1 - \alpha_1 - \beta_1)$ und $E(X_t^4) = \kappa_Z E(\sigma_t^4)$ gilt, folgt:

$$E(X_t^4) = \frac{\alpha_0^2 \kappa_Z (1 - (\alpha_1 + \beta_1)^2)}{(1 - \alpha_1 - \beta_1)^2 (1 - \alpha_1^2 \kappa_z - \beta_1^2 - 2\alpha_1\beta_1)}.$$

□

Mit dieser Gleichung erhalten wir dann für die Wölbung von X_t:

$$\kappa_X = \frac{\kappa_Z (1 - (\alpha_1 + \beta_1)^2)}{(1 - \alpha_1 - \beta_1)^2 - (\kappa_Z - 1)\alpha_1^2}$$

Daran erkennt man, dass die Wölbung von X_t größer ist als die von Z_t, falls $\kappa_Z > 1$ gilt. Dies ist insbesondere der Fall für die Normal- und die t-Verteilung.

2.3 Parallelen zwischen dem ARMA- und dem GARCH-Prozess

Auch hier lässt sich der GARCH-Prozess umschreiben zu

$$X_t^2 = \alpha_0 + \alpha_1 X_{t-1}^2 + \beta_1 \sigma_{t-1}^2 + V_t$$

mit einer Martingaldifferenz $V_t := \sigma_t^2(Z_t^2 - 1)$. Definiert man $\sigma_{t-1}^2 := X_{t-1}^2 + V_t$, so lässt sich dies umformen zu

$$X_t^2 = \alpha_0 + (\alpha_1 + \beta_1) X_{t-1}^2 - \beta_1 V_{t-1} - V_{t-1} \tag{13}$$

Dies erinnert stark an einen ARMA(1,1)-Prozess. Wenn wir nun zusätzlich noch $E(X_t^4) < \infty$ fordern, erhält man aufgrund von $\alpha_1 + \beta_1 < 1$, dass

$$\left(X_t^2 - \frac{\alpha_0}{1 - \alpha_1 - \beta_1} \right) = (\alpha_1 + \beta_1) \left(X_{t-1}^2 - \frac{\alpha_0}{1 - \alpha_1 - \beta_1} \right) - \beta_1 V_{t-1} + V_t$$

ein ARMA(1,1)-Prozess ist.

2.4 Das GARCH(p,q)-Modell

Die ARCH- und GARCH-Modelle höherer Ordnung verhalten sich genauso wie die Modelle mit $p = q = 1$. Die hinreichende und notwendige Bedingung für schwache Stationarität ist dann

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

Der quadratische GARCH(p,q)-Prozess hat die Form

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

mit $\alpha_i = 0$, $i = p+1, \ldots, q$ falls $q > p$ und $\beta_j = 0$, $j = q+1, \ldots, p$ falls $q < p$. Dies ist dann ein ARMA(max $\{p, q\}$,q)-Prozess, falls $E(X_t^4) < \infty$ gilt.

2.5 Der IGARCH-Prozess

Als IGARCH(p,q)-Prozess bezeichnet man einen GARCH(p,q)-Prozess für den $\sum\limits_{i=1}^{p} \alpha_i + \sum\limits_{j=1}^{q} \beta_j = 1$ gilt. Bei einem einfachen GARCH-Modell kann es vorkommen, dass $\sum\limits_{i=1}^{p} \alpha_i + \sum\limits_{j=1}^{q} \beta_j > 1$ gilt, was jedoch zur Folge hat, dass der Prozess eine unendliche Varianz hat und somit nicht mehr schwach stationär sein kann.
Wir betrachten also das IGARCH(1,1)-Modell. Dann folgt mithilfe von (13):

$$\Delta X_t^2 = X_t^2 - X_{t-1}^2 = \alpha_0 - (1 - \alpha_1) V_{t-1} + V_t$$

mit V_t und σ_t wie oben. Dies erinnert an einen ARIMA(0,1,1)-Modell, außer dass V_t nicht unbedingt ein Weißes Rauschen sein muss oder eine Martingaldifferenz (da $E(V_t | \mathcal{F}_{t-1})$ nicht definiert ist, solange $E\sigma_t^2 = EX_t^2 = \infty$). Dann ist auch $E(|V_t|)$ nicht definiert.

2.6 Erweiterte GARCH-Modelle

Wir wollen nun ein paar Beispiele für Erweiterungen des GARCH-Modells kennenlernen:

2.6.1 Das ARMA-Modell mit GARCH-Fehlern

Definition 2.4 (ARMA-Prozess mit GARCH-Fehlern)**.** *Sei* $(Z_t)_{t\in\mathbb{Z}}$ $SWN(0,1)$*. Wir nennen einen Prozess* $(X_t)_{t\in Z_t^2}$ $ARMA(p_1,q_1)$ *Prozess mit* $GARCH(p_2,q_2)$*-Fehlern, falls er schwach stationär ist und den folgenden Gleichungen genügt:*

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

für $\alpha_0 > 0$*,* $\alpha_i \geq 0$ *für* $i = 1,\dots,p_2$*,* $\beta_j \geq 0$ *für* $j = 1,\dots,q_2$ *und* $\sum_{i=1}^{p_2} \alpha_i + \sum_{i=1}^{q_2} \beta_j < 1$*.*

Der ARMA-Prozess ist kausal und invertierbarer linearer Prozess, falls die Polynome $\tilde{\phi}(z) = 1 - \phi_1 z - \dots - \phi_{p_1} z^{p_1}$ und $\tilde{\theta}(z) = 1 + \theta_1 z + \dots + \theta_{q_1} z^{q_1}$ keine Nullstellen auf oder innerhalb des Einheitskreises haben.

2.6.2 GARCH-Modell mit Hebelwirkung

In dem nun kennengelernten GARCH-Modell wird der leverage Effekt (=Hebelwirkung) nicht berücksichtigt. Dieser sagt aus, dass bei negativer Rendite -wie zum Beispiel fallenden Aktienkursen- eine höhere Volatilität zu beobachten ist. Marktinformationen sollten also einen asymmetrischen Effekt auf die Volatilität haben. Dabei gilt, dass schlechte Nachrichten die Volatilität erhöht.
Man möchte diesen Effekt auch bei dem GARCH(1,1)-Modell simulieren. Dazu addiert man einen Parameter zu der Volatilitätsgleichung des GARCH-Modells und erhält:

$$\sigma_t^2 = \alpha_0 + \alpha_1 (X_{t-1} + \delta |X_{t-1}|)^2 + \beta_1 \sigma_{t-1}^2 \tag{14}$$

Dabei setzen wir voraus, dass $\delta \in [-1,1]$ und $\alpha_1 \geq 0$ gilt.
σ_t^2 lässt sich dann wie folgt ausdrücken:

$$\sigma_t^2 = \begin{cases} \alpha_0 + \alpha_1 (1+\delta)^2 X_{t-1}^2 + \beta_1 \sigma_{t-1}^2, & \text{falls } X_{t-1}^2 \geq 0 \\ \alpha_0 + \alpha_1 (1-\delta)^2 X_{t-1}^2 + \beta_1 \sigma_{t-1}^2, & \text{falls } X_{t-1}^2 < 0 \end{cases}$$

und deshalb gilt

$$\frac{\partial \sigma_t^2}{\partial X_{t-1}^2} = \begin{cases} \alpha_1 (1+\delta)^2, & \text{falls } X_{t-1}^2 \geq 0 \\ \alpha_1 (1-\delta)^2, & \text{falls } X_{t-1}^2 < 0 \end{cases}$$

Wir setzen generell $\delta < 0$ voraus, so dass schlechte Nachrichten eine größeren Effekt auf die Volatilität beinhaltet.

2.6.3 Threshold GARCH

Das TGARCH-Modell ist eine Verallgemeinerung des GARCH-Modells mit leverage Effekt. Bei diesem Modell möchte man eine Auftrennung von σ_t^2 erreichen. Diese Trennung wird hier bei 0 gemacht. Allgemein kann man sie aber auch woanders setzten. Das ist auch der Grund für den Namen Threshold (=Schwelle). Wenn wir (14) umformulieren, erhält wir

$$\sigma_t^2 = \alpha_0 + \tilde{\alpha_1} X_{t-1}^2 + \tilde{\delta}\mathbf{I}_{\{X_{t-1}<0\}} X_{t-1}^2 + \beta_1 \sigma_{t-1}^2$$

mit $\tilde{\alpha_1} = \alpha_1(1+\delta)^2$ und $\tilde{\delta} = -4\delta\alpha_1$. Dies ist die allgemeinste Form eines TGARCH Modells.

Bemerkung 2.5. *Eine weitere Möglichkeit eine Asymmetrie des GARCH-Modells zu erreichen besteht darin, dass man eine asymmetrische Verteilung der Innovationen wählt. Diese sollte jedoch so normiert werden, dass der Erwartungswert Null und die Varianz Eins ist.*

2.7 Maximum Likelihood Schätzung

2.7.1 Bilden des MLS

Wir betrachten nun wieder das ARCH(1)- und das GARCH(1,1)-Modell, wobei die Aussagen für die allgemeinen Modelle daraus direkt folgen. Unser Ziel ist es die Parameter dieser Modelle zu schätzen. Seien $X_0, \dots, X_n$ Zufallsvariablen. Die gemeinsame Dichte von $X_0, \dots, X_n$ ist dann durch

$$f_{X_0,\dots,X_n}(x_0,\dots,x_n) = f_{X_0}(x_0)\prod_{t=1}^{n} f_{X_t|X_{t-1},\dots,X_0}(x_t|x_{t-1},\dots,x_0)$$

gegeben. Das Problem beim Erstellen der Maximum Likelihood Funktion ist, dass wir die Randdichte f_{X_0} für das ARCH- und GARCH- Modell nicht kennen. Deshalb führt man die bedingte Maximum Likelihood Funktion (CML) gegeben X_0 ein, die durch

$$f_{X_1,\dots,X_n|X_0}(x_1,\dots,x_n|x_0) = \prod_{t=1}^{n} f_{X_t|X_{t-1},\dots,X_0}(x_t|x_{t-1},\dots,x_0)$$

berechnet wird.
Sei g(z) die Dichte der Innovationen $(Z_t)_{t\in\mathbb{Z}}$ und $\sigma_t = (\alpha_0 + \alpha_1 x_{t-1}^2)^{\frac{1}{2}}$. Dann gilt für die Dichte des ARCH(1)-Prozesses

$$f_{X_t|X_{t-1},\dots,X_0}(x_t|x_{t-1},\dots,x_0) = f_{X_t|X_{t-1}}(x_t|x_{t-1}) = \frac{1}{\sigma_t} g\left(\frac{x_t}{\sigma_t}\right),$$

da die bedingte Dichte für einen ARCH(1)-Prozess nur von X_{t-1} abhängt.
Dann folgt für die CML mit $\sigma_t = (\alpha_0 + \alpha_1 X_{t-1}^2)^{\frac{1}{2}}$

$$\begin{aligned} L(\alpha_0,\alpha_1;X) &= f_{X_1,\dots,X_n|X_0}(x_1,\dots,x_n|x_0) \\ &= \prod_{t=1}^{n} f_{X_t|X_{t-1},\dots,X_0}(x_t|x_{t-1},\dots,x_0) \\ &= \prod_{t=1}^{n} \frac{1}{\sigma_t} g\left(\frac{X_t}{\sigma_t}\right). \end{aligned}$$

Für das GARCH(1,1) Modell konstruiert man den CML wie oben, nur macht man dies in Abhängigkeit von X_0 und σ_0, so dass man folgende Gleichung erhält

$$\begin{aligned} L(\alpha_0, \alpha_1, \beta; X) &= f_{X_1,\ldots,X_n|X_0}(x_1, \ldots, x_n|x_0) \\ &= \prod_{t=1}^{n} f_{X_t|X_{t-1},\ldots,X_0,\sigma_0}(x_t|x_{t-1}, \ldots, x_0, \sigma_0) \end{aligned}$$

Wie schon bei dem ARCH(1) Modell hängt die bedingte Dichte hier nur von σ_t ab. Diese kann man rekursiv darstellen mithilfe von $\sigma_0, X_0, \ldots, X_{t-1}$, wenn man die Eigenschaft des GARCH(1,1) Modells ausnutzt, dass $\sigma_t^2 = \alpha_0 + \alpha_1 X_{t-1}^2 + \beta_1 \sigma_{t-1}^2$ gilt. Dann erhält man eine modifizierte Version des CML

$$L(\alpha_0, \alpha_1, \beta; X) == \prod_{t=1}^{n} \frac{1}{\sigma_t} g\left(\frac{X_t}{\sigma_t}\right)$$

mit $\sigma_t = (\alpha_0 + \alpha_1 X_{t-1}^2 + \beta_1 \sigma_{t-1}^2)^{\frac{1}{2}}$.
Dabei löst man das Problem, dass man den Wert von σ_0^2 nicht kennt, indem man sich einen Startwert vorgibt, z.B. Null oder die Stichprobenvarianz von $X_1, \ldots, X_t$ (da $E\sigma_t^2 = Var X_t^2$ gilt).
Für ein GARCH(p,q) Modell benötigen wir $p + n$ Werte $X_{-p+1}, \ldots, X_0$, $X_1, \ldots, X_n$. Dann wollen wir die CML für die beobachteten Werte $X_{-p+1}, \ldots, X_0$ genauso wie für die nicht beobachteten Werte $\sigma_{-p+1}, \ldots, \sigma_0$ -die wir uns wieder vorgeben müssen- anpassen. Z.B. geben wir uns für ein GARCH(1,3)-Modell Startwerte $\sigma_0, \sigma_{-1}, \sigma_{-2}$ vor.
Man kann statt dessen auch das ARMA-Modell mit GARCH-Fehlern wie in Definition 2.4 betrachten und dafür eine likelihood Funktion bestimmen. Dann erhält man die CML

$$L(\theta, X) = \prod_{t=1}^{n} \frac{1}{\sigma_t} g\left(\frac{X_t - \mu_t}{\sigma_t}\right)$$

wobei σ_t einem GARCH-Prozess und μ_t einem ARMA-Prozess folgen und alle nicht bekannten Parameter werden in θ zusammengefasst.
Wir definieren dann die bedingte log-likelihood Funktion durch

$$\ln L(\theta; X) = \sum_{t=1}^{n} l_t(\theta)$$

wobei l_t die log-likelihood Anteil entstehend aus der t-ten Beobachtung beschreibt. Damit wir dann das Maximum bestimmen können, müssen wir die bedingte log-likelihood Funktion ableiten und Null setzen.

$$\frac{\partial}{\partial \theta} \ln L(\theta; X) = \sum_{t=1}^{n} \frac{\partial l_t(\theta)}{\partial \theta} = 0.$$

Man nennt hierbei die linke Seite den score vector der bedingten log-likelihood Funktion. Diese Gleichung wird häufig auf numerische Weise durch das Newton-Raphson Verfahren bestimmt oder man bestimmt sie mithilfe des BHHH(=Berndt, Hall, Hall, Hausmann)-Algorithmus.

Wir werden als nächstes das Verhalten der Schätzparameter diskutieren. Dabei unterscheiden wir in zwei Situationen. Die erste Situation setzt ein Modell voraus, dass die Daten durch eine Zeitreihenmodell generiert werden, das korrekt spezifiziert ist. Es werden dann die asymptotischen Eigenschaften des Maximum Likelihood Schätzers (MLS)unter diesen idealisierten Voraussetzungen beschrieben.
Bei der zweiten Situation wird irrtümlich angenommen, dass die Innovationen einer Gaußverteilung folgen. Diese Situation nennt man dann die Quasi-Maximum Likelihood Schätzung (QML) und die erhaltenen Schätzer heißen QMLS.

2.7.2 Verhalten des MLS

Wir wollen nun auf die asymptotische Eigenschaft des MLS für ein GARCH Modell ohne ARMA-Komponenten oder leverage Struktur eingehen, aber die Ergebnisse gelten auch im allgemeineren Fall.
Unter den oben angegebenen Voraussetzungen an das GARCH(p,q) Modell (also dass das Modell korrekt spezifiziert ist) kann man zeigen, dass folgendes gilt

$$\sqrt{n}(\hat{\theta_n} - \theta) \xrightarrow{d} N_{p+q+1}(0, I(\theta)^{-1}),$$

wobei

$$I(\theta) = E\left(\frac{\partial l_t(\theta)}{\partial \theta}\frac{\partial l_t(\theta)}{\partial \theta'}\right) = -E\left(\frac{\partial^2 l_t(\theta)}{\partial \theta \partial \theta'}\right)$$

die Matrix der Fischer-Information ist, die aus jeder einzelnen Beobachtung entsteht. Dann haben wir einen konsistenten und asymptotischen Schätzer für die Parameter des GARCH-Modells. Allgemein wird die erwartete Informationsmatrix $I(\theta)$ durch eine beobachtete Informationsmatrix angenähert. Die beobachtete Informationsmatrix kann man in einer der beiden äquivalenten Formen für die erwartete Informationsmatrix darstellen

$$\begin{aligned}\bar{I}(\theta) &= \frac{1}{n}\sum_{t=1}^{n}\left(\frac{\partial l_t(\theta)}{\partial \theta}\frac{\partial l_t(\theta)}{\partial \theta'}\right)\\ \bar{J}(\theta) &= -\frac{1}{n}\sum_{t=1}^{n}\left(\frac{\partial^2 l_t(\theta)}{\partial \theta \partial \theta'}\right),\end{aligned}$$

dabei sagt man, dass die erste Matrix die outer-product Form und die zweite die Hessian Form besitzt. Diese Matrizen untersucht man mit den MLS um $\bar{I}(\hat{\theta})$ und $\bar{J}(\hat{\theta})$ zu erhalten. Im Allgemeinen macht man dies mithilfe der ersten und zweiten Ableitung von der log-likelihood Funktion und die notwendigen Matrizen erhält man als Nebenprodukt von dem BHHH Algorithmus.
Wenn das Modell gut gewählt ist, sollten die Schätzer für $\bar{I}(\hat{\theta})$ und $\bar{J}(\hat{\theta})$ weitestgehend gleich sein, da sie auf zwei verschiedenen Methoden, die Fisher-Information darzustellen, beruht. In der Praxis wird häufig $\bar{I}(\hat{\theta})$ durch den sogenannten Sandwich-Schätzer $\bar{J}(\hat{\theta})\bar{I}(\hat{\theta})\bar{J}(\hat{\theta})^{-1}$ ermittelt. Dieser wird bei der QML Methode verwendet, die wir nun untersuchen werden.

2.7.3 Verhalten von QMLS

Wir haben nun also ein GARCH(p,q)-Modell mit Innovationen, die nicht normalverteilt sind, aber wir versuchen die Parameter unter Annahme normalverteilter Innovationen zu schätzen, indem wir die likelihood Funktion maximieren. Dennoch erhalten wir konsistente Schätzer und falls zusätzlich das fünfte Moment der wirklichen Verteilung der Innovationen endlich ist, erhalten wir einen normal asymptotischen Schätzer. Die Form der asymptotischen Kovarianzmatrix ändert sich also.
Wir müssen nun zwischen den beiden folgenden Matrizen unterscheiden

$$\bar{I}(\theta) = (\frac{\partial l_t(\theta)}{\partial \theta}\frac{\partial l_t(\theta)}{\partial \theta'})$$
$$\bar{J}(\theta) = -E(\frac{\partial^2 l_t(\theta)}{\partial \theta \partial \theta'}),$$

wobei der Erwartungswert bezüglich der wahren Verteilung bestimmt wird. Solange kein Gauß Modell (also die Innovationen nicht normalverteilt sind) vorliegt unterscheiden sich diese beiden Matrizen. Man kann zeigen, dass

$$\sqrt{n}(\hat{\theta_n} - \theta) \xrightarrow{d} N_{p+q+1}(0, (\bar{J}(\theta))^{-1}\bar{I}(\theta)(\bar{J}(\theta))^{-1})$$

gilt und man sagt, dass die Kovarianzmatrix die Sandwich-Form besitzt. Diese kann man durch $\bar{J}(\hat{\theta})\bar{I}(\hat{\theta})\bar{J}(\hat{\theta})^{-1}$ schätzen. Dabei gelten die Definitionen wie oben. Wenn die Modelldiagnose - die in dem nächsten Abschnitt beschrieben ist - besagt, dass das Modell gut gewählt ist, aber die Annahme der Normalverteilung zweifelhaft erscheint, dann sollten die Standard Fehler bei dem Kovarianz-Matrix Schätzer liegen.

2.7.4 Modelldiagnose

Wie schon bei dem ARMA Modell ist es nach der Schätzung der Parameter wichtig zu überprüfen, ob das Modell gut angepasst ist. Man stellt also fest, ob das Modell mit den geschätzten Parametern passend ist. Dafür setzen wir ein ARMA-GARCH Modell der Form $X_t - \mu_t = \varepsilon_t = \sigma_t Z_t$ mit μ_t und σ_t wie in Definition 2.4 voraus und unterscheiden zwischen standardisierten und nicht-standardisierten Residuen.

Zunächst betrachten wir die nicht-standardisierten Residuen $\hat{\varepsilon}_1, \ldots, \hat{\varepsilon}_t$ von dem ARMA- Anteil des Modells. Diese sollten sich nun unter dem gefundenen Modell genauso verhalten wie ein GARCH-Prozess.

$$\hat{Z}_t = \frac{\hat{\varepsilon}_t}{\hat{\sigma}_t}, \tag{15}$$

$$\hat{\sigma}_t^2 = \hat{\alpha}_0 + \sum_{i=1}^{p_2} \hat{\alpha}_i \hat{\varepsilon}_{t-i}^2 + \sum_{j=1}^{q_2} \hat{\beta}_j \hat{\sigma}_{t-j}^2 \tag{16}$$

Die standardisierten Residuen verhalten sich wie ein SWN. Dies kann man mithilfe eines sogenannten Portmanteau-Test (dieser testet für mehrere Autokorrelationskoeffizienten die Signifikanz zu Null) feststellen. Falls die Vermutung, dass es sich um ein SWN handelt, nicht abgelehnt wurde, kann man

die Gültigkeit der im ML benutzten Verteilung mithilfe eines QQplots oder goodness-of-fit Tests -für die normal oder t-Verteilungen untersucht- werden. Wenn sich dann die Residuen nicht wie eine iid Standard-Normalverteilung bei einer einfachen Likelihood Schätzung verhalten, so bestimmt man den QMLS mithilfe des Sandwich-Schätzers.
Man erhält somit ein zwei-Schritte Programm, wobei man in der ersten Phase einen QMLS bestimmt und als zweites anhand der gefundenen Schätzer ein Modell bildet und diese Schätzer überprüft.

3 Volatilitätsprognose

Wir wollen nun eine Methode , die zukünftige Volatilität zu prognostizieren anhand eines GARCH-Modells, angeben. Folgen also $X_{t-n-1}, \ldots, X_t$ einem bestimmten GARCH-Modell, so wollen wir σ_{t+h} für ein $h \geq 1$ vorhersagen können. Wir setzen voraus, dass wir die Werte bis zum Zeitpunkt t kennen, also $\mathcal{F} = \sigma(\{X_s : s \geq t\})$ und, dass das GARCH-Modell angepasst und die Parameter geschätzt sind.
Betrachten wir nun also eine Volatilitäsprognosen für das einfache GARCH-Modell

Beispiel 3.1. *Als Erstes betrachten wir ein einfaches GARCH(1,1)-Modell wie in Definition 2.1 und setzen noch voraus, dass das Modell schwach stationär (d.h. $E(X_t^2) = E(\sigma_t^2) < \infty$) ist. Falls $(X_t)_{t\in\mathbb{Z}}$ eine Martingaldifferenz ist, so ist die optimale Vorhersage für X_{t+h} Null. Eine Vorhersage für X_{t+1}^2 basierend auf $\mathcal{F}_t$ ist der bedingte Mittelwert σ_{t+1}^2, der gegeben ist durch*

$$E(X_{t+1}^2|\mathcal{F}_t) = \sigma_{t+1}^2 = \alpha_0 + \alpha_1 X_t^2 + \beta_1 \sigma_t^2.$$

Falls man noch $E(X_t^4) < \infty$ fordert, so ist dies die optimale quadratische Fehlervorhersage.
Die allgemein übliche Herangehensweise an das Problem, dass man die vergangenen Werte, die man für die Schätzung von σ_t^2 benötigt, normalerweise nicht kennt, ist indem man die Volatilität mithilfe der Gleichung (16) schätzt.

$$\hat{\sigma}_{t+1}^2 = \hat{E}(X_{t+1}^2|\mathcal{F}_t) = \alpha_0 + \alpha_1 X_t^2 + \beta_1 \hat{\sigma}_t^2$$

Falls $h > 1$, so gilt für den Zeitpunkt t, dass X_{t+h}^2 und σ_{t+h}^2 Zufallsvariablen sind und ihre Vorhersage stimmt überein und lautet

$$\begin{aligned} E(X_{t+h}^2|\mathcal{F}_t) &= E(\sigma_{t+h}^2|\mathcal{F}_t) \\ &= \alpha_0 + \alpha_1 E(X_{t+h-1}^2|\mathcal{F}_t) + \beta_1 E(\sigma_{t+h-1}^2|\mathcal{F}_t) \\ &= \alpha_0 + (\alpha_1 + \beta_1) E(X_{t+h-1}^2|\mathcal{F}_t) \end{aligned}$$

Durch h-maliges Einsetzen erhält man dann also:

$$E(X_{t+h}^2|\mathcal{F}_t) = \alpha_0 \sum_{i=0}^{h-1} (\alpha_1 + \beta_1)^i + (\alpha_1 + \beta_1)^{h-1} (\alpha_1 X_t^2 + \beta_1 \sigma_t^2)$$

Wir bemerken, dass für $h \to \infty$ *gilt:* $E(X_{t+h}^2|\mathcal{F}_t) \to \alpha_0/(1-\alpha_1-\beta_1)$ *fast sicher. Das heißt, die Vorhersage der quadrierten Varianz konvergiert gegen die unbedingte Varianz des Prozesses!*