6 Processes with Autoregressive Conditional Heteroskedasticity (ARCH)

6.1 Summary

In particular in the case of financial time series one often observes a highly fluctuating volatility (or variance) of a series: Agitated periods with extreme amplitudes alternate with rather quiet periods being characterized by moderate observations. After some short preliminary considerations concerning models with time-dependent heteroskedasticity, we will discuss the model of autoregressive conditional heteroskedasticity (ARCH), for which Robert F. Engle was awarded the Nobel prize in the year 2003. After a generalization (GARCH), there will be a discussion on extensions relevant for practice. Throughout this chapter, the innovations or shocks $\{\varepsilon_t\}$ stand for a pure random process as defined in Example 2.7.

6.2 Time-Dependent Heteroskedasticity

The heteroskedasticity allowed for here is modeled as time-dependent **volatility** by[1]

$$x_t = \sigma_t \, \varepsilon_t \,, \quad \varepsilon_t \sim \text{iid}(0, 1) \,, \tag{6.1}$$

[1]The following equation could be extended by a mean function, e.g. of a regression-type,

$$x_t = \alpha + \beta z_t + \sigma_t \, \varepsilon_t \,,$$

or

$$x_t = a_1 x_{t-1} + \cdots + a_p x_{t-p} + \sigma_t \, \varepsilon_t \,.$$

We restrict our exposition and concentrate on modeling volatility exclusively, although in practice time-dependent heteroskedasticity is often found with regression errors.

U. Hassler, *Stochastic Processes and Calculus*, Springer Texts in Business and Economics, DOI 10.1007/978-3-319-23428-1_6

where $\{\sigma_t\}$ is the volatility process being stochastically independent of $\{\varepsilon_t\}$ by assumption, and $\{\varepsilon_t\}$ is a pure random process with unit variance. There exist two routes to model the volatility process. The first one, often labeled as processes with stochastic volatility, assumes an unobserved, or latent, process $\{h_t\}$ behind the volatility: $\sigma_t = \exp(h_t/2)$. This implies for the squared data

$$x_t^2 = e^{h_t}\varepsilon_t^2 \quad \text{or} \quad \log(x_t^2) = h_t + \log(\varepsilon_t^2)\,.$$

For an early survey on stochastic volatility processes see Taylor (1994). A second strand in the literature assumes that σ_t depends on observed data, in particular on past observations x_{t-j}. This class of models has been called autoregressive conditional heteroskedasticity (ARCH). ARCH processes are widely spread and successful in practice and will be the focus of attention in the present chapter.

Heteroskedasticity as a Function of the Past

In this chapter the variance function is modeled by the observed past of the process itself:

$$\sigma_t^2 = f\left(x_{t-1}, x_{t-2}, \cdots\right)\,. \tag{6.2}$$

By plugging in x_{t-j} from (6.1) one obtains:

$$\sigma_t^2 = f\left(\sigma_{t-1}\varepsilon_{t-1}, \sigma_{t-2}\varepsilon_{t-2}, \cdots\right)\,.$$

We will show that the process from (6.1) is a martingale difference sequence. Remember the definition of the information set $\mathcal{I}_{t-1}$ generated by the past of x_t up to x_{t-1}. Then it holds that[2]

$$\begin{aligned}
\mathrm{E}\left(x_t \,|\, \mathcal{I}_{t-1}\right) &= \mathrm{E}\left(x_t \,|\, x_{t-1}, x_{t-2}, \cdots\right)\\
&= \mathrm{E}\left(\sigma_t\varepsilon_t \,|\, x_{t-1}, x_{t-2}, \cdots\right)\\
&= \sigma_t \mathrm{E}\left(\varepsilon_t \,|\, x_{t-1}, x_{t-2}, \cdots\right)\\
&= \sigma_t \mathrm{E}\left(\varepsilon_t\right)\,,
\end{aligned}$$

as ε_t is independent of x_{t-j} for $j > 0$ by construction. With ε_t being zero on average, it follows that

$$\mathrm{E}\left(x_t \,|\, \mathcal{I}_{t-1}\right) = 0\,,$$

[2] When conditioning on $\mathcal{I}_{t-1}$, one often writes $\mathrm{E}\left(\cdot \,|\, x_{t-1}, x_{t-2}, \cdots\right)$ instead of $\mathrm{E}\left(\cdot \,|\, \mathcal{I}_{t-1}\right)$.

which proves (integrability of $\{x_t\}$ assumed) that $\{x_t\}$ is in fact a martingale difference sequence, see below (2.11). The variance of the martingale difference is determined in the following way (as x_t is zero on average,

$$\begin{aligned}\text{Var}(x_t) &= \text{E}\left(x_t^2\right) \\ &= \text{E}\left(\sigma_t^2 \varepsilon_t^2\right) \\ &= \text{E}\left(\sigma_t^2\right) \text{E}\left(\varepsilon_t^2\right) \\ &= \text{E}\left(\sigma_t^2\right) ,\end{aligned}$$

as σ_t^2 from (6.2) and ε_t (with variance 1 from (6.1)) are stochastically independent. Hence, the following proposition is verified.

Proposition 6.1 (Heteroskedastic Martingale Differences) *Let $\{x_t\}$ be from (6.1) and $\{\sigma_t^2\}$ from (6.2) with $E(\sigma_t^2) < \infty$ independent of $\{\varepsilon_t\}$. Then $\{x_t\}$ is a martingale difference sequence with variance*

$$Var(x_t) = E(x_t^2) = E(\sigma_t^2) \,.$$

Let us remember Proposition 2.2. Due to the martingale difference property it holds that

$$\text{E}(x_t) = 0 \quad \text{and} \quad \gamma(h) = \text{E}(x_t x_{t+h}) = 0 \,, \; h \neq 0 \,.$$

Hence, the process is serially uncorrelated with expectation zero which would be supposed e.g. for returns. However, the process is generally not independent over time. The (weak) stationarity of $\{x_t\}$ depends on the possibly variable variance; if the variance $\text{Var}(x_t)$ is constant, then the entire process is stationary.

Heuristics

Now, the question is how the functional dependence in (6.2) should be specified and parameterized. Heteroskedasticity as an empirical phenomenon has been known to observers on financial markets for a long time. Before ARCH models were introduced, it had been measured by moving a window of width B through the data and averaging over the squares:

$$s_t^2 = \frac{1}{B} \sum_{i=1}^{B} x_{t-i}^2 \,.$$

For every point in time t one averages over the past preceding B values in order to determine the variance in t. In doing so, we do not center x_{t-i} around the arithmetic mean as we think of returns with $\text{E}(x_t) = 0$ when applying the procedure, cf.

Proposition 6.1. With daily observations (with five trading days a week) one chooses e.g. $B = 20$ which approximately corresponds to a time window of a month. A first improvement of the time-dependent volatility measurement is obtained by using a weighted average where the weights, $g_i \geq 0$, are not negative:

$$s_t^2 = \sum_{i=1}^{B} g_i\, x_{t-i}^2 \quad \text{with} \quad \sum_{i=1}^{B} g_i = 1. \tag{6.3}$$

Example 6.1 (Exponential Smoothing) For the weighting function g_i one often uses an exponential decay:

$$g_i = \frac{\lambda^{i-1}}{1 + \lambda + \ldots + \lambda^{B-1}} \quad \text{with} \quad 0 < \lambda < 1.$$

Note that the denominator is just defined such that it holds that $\sum_{i=1}^{B} g_i = 1$. With growing B one furthermore obtains

$$1 + \lambda + \ldots + \lambda^{B-1} = \frac{1-\lambda^B}{1-\lambda} \to \frac{1}{1-\lambda}.$$

Inserting the exponentially decaying weights in (6.3), we get the following result for $B \to \infty$:

$$s_t^2(\lambda) = (1-\lambda) \sum_{i=1}^{\infty} \lambda^{i-1}\, x_{t-i}^2.$$

Now it is an easy exercise to verify the following recursive relation:

$$s_t^2(\lambda) = (1-\lambda)\, x_{t-1}^2 + \lambda\, s_{t-1}^2(\lambda). \tag{6.4}$$

We will call $s_t^2(\lambda)$ the exponentially smoothed volatility or variance. In order to be able to calculate it for $t = 2, \ldots, n$, we need a starting value. Typically, $s_1^2(\lambda) = x_1^2$ is chosen which leads to $s_2^2(\lambda) = x_1^2$. ■

The ARCH and GARCH processes which are subsequently introduced are models leading to volatility specifications which generalize s_t^2 and $s_t^2(\lambda)$ from (6.3) and (6.4), respectively.

6.3 ARCH Models

So-called autoregressive conditional heteroskedasticity models can be traced back to Engle (1982). We consider the case of the order q and specify σ_t^2 from (6.2) as follows:

$$\sigma_t^2 = \alpha_0 + \alpha_1\, x_{t-1}^2 + \ldots + \alpha_q\, x_{t-q}^2\,, \tag{6.5}$$

where it is assumed that

$$\alpha_0 > 0 \quad \text{and} \quad \alpha_i \geq 0\,,\; i = 1, \ldots, q\,, \tag{6.6}$$

in order to guarantee $\sigma_t^2 > 0$. Note that this variance function corresponds to s_t^2 from (6.3). For $\alpha_1 = \cdots = \alpha_q = 0$, however, the case of homoskedasticity is modeled.

Conditional Moments

Given $x_{t-1}, \ldots, x_{t-q}$, one naturally obtains zero for the conditional expectation as ARCH processes are martingale differences, see Proposition 6.1. For the conditional variance it holds that

$$\begin{aligned}\operatorname{Var}\left(x_t \,|\, x_{t-1}, \ldots, x_{t-q}\right) &= \operatorname{E}\left(x_t^2 \,|\, x_{t-1}, \ldots, x_{t-q}\right) \\ &= \sigma_t^2 \operatorname{E}\left(\varepsilon_t^2 \,|\, x_{t-1}, \cdots, x_{t-q}\right) \\ &= \sigma_t^2\,,\end{aligned}$$

as ε_t is again independent of x_{t-j} and has a unit variance. Hence, for the variance it conditionally holds that:

$$\operatorname{Var}(x_t|x_{t-1}, \ldots, x_{t-q}) = \alpha_0 + \alpha_1\, x_{t-1}^2 + \ldots + \alpha_q\, x_{t-q}^2\,,$$

which explains the name of the models: The conditional variance is modeled autoregressively (where "autoregressive" means in this case: dependent on the past of the process). Thus, extreme amplitudes in the previous period are followed by high volatility in the present period resulting in so-called volatility clusters. If the assumption of normality of the innovations is added, then the conditional distribution of x_t given the past is normal as well:

$$x_t \,|\, x_{t-1}, \ldots x_{t-q} \sim \mathcal{N}\left(0, \alpha_0 + \alpha_1\, x_{t-1}^2 + \ldots + \alpha_q\, x_{t-q}^2\right)\,.$$

But, although the original work by Engle (1982) assumed $\varepsilon_t \sim \text{ii}\,\mathcal{N}(0, 1)$, the assumption of normality is not crucial for ARCH effects.

Stationarity

By Proposition 6.1 it holds for the variance that

$$\operatorname{Var}(x_t) = \operatorname{E}(\sigma_t^2) = \alpha_0 + \alpha_1 \operatorname{E}(x_{t-1}^2) + \cdots + \alpha_q \operatorname{E}(x_{t-q}^2)\,.$$

If the process is stationary, $\operatorname{Var}(x_t) = \operatorname{Var}(x_{t-j})$, $j = 1, \ldots, q$, then its variance results as

$$\operatorname{Var}(x_t) = \frac{\alpha_0}{1 - \alpha_1 - \ldots - \alpha_q}\,.$$

For a positive variance expression, this requires necessarily (due to $\alpha_0 > 0$):

$$1 - \alpha_1 - \ldots - \alpha_q > 0\,.$$

In fact, this condition is sufficient for stationarity as well. In Problem 6.1 we therefore show the following result.

Proposition 6.2 (Stationary ARCH)
Let $\{x_t\}$ be from (6.1) and $\{\sigma_t^2\}$ from (6.5) with (6.6). The process is weakly stationary if and only if it holds that

$$\sum_{j=1}^{q} \alpha_j < 1\,.$$

Correlation of the Squares

We define $e_t = x_t^2 - \sigma_t^2$. Due to Proposition 6.1 the expected value is zero: $\mathrm{E}(e_t) = 0$. Adding x_t^2 to both sides of (6.5), one immediately obtains

$$x_t^2 = \alpha_0 + \alpha_1\, x_{t-1}^2 + \ldots + \alpha_q\, x_{t-q}^2 + e_t\,.$$

From this we learn that an ARCH(q) process implies an autoregressive structure for the squares $\{x_t^2\}$. The serial dependence of an ARCH process originates from the squares of the process. Because of $\alpha_i \geq 0$, x_t^2 and x_{t-i}^2 are positively correlated which again allows to capture volatility clusters.

In Figs. 6.1 and 6.2 ARCH(1) time series of the length 500 are simulated. For this purpose pseudo-random numbers $\{\varepsilon_t\}$ are generated as normally distributed and $\alpha_0 = 1$ is chosen. The effect of α_1 is now quite obvious. The larger the value of this parameter, the more obvious are the volatility clusters. Long periods with little movement are followed by shorter periods of vehement, extreme amplitudes which can be negative as well as positive. These volatility clusters become even more obvious in the respective lower panel of the figures, in which the squared observations $\{x_t^2\}$ are depicted. Because of the positive autocorrelation of the squares, small amplitudes tend again to be followed by small ones while extreme observations appear to follow each other.

Skewness and Kurtosis

In the first section of the chapter on basic concepts from probability theory we have defined the kurtosis by means of the fourth moment of a random variable and we have denoted the corresponding coefficient by γ_2. For $\gamma_2 > 3$ the density function is more "peaked" than the one of the normal distribution: On the one hand the

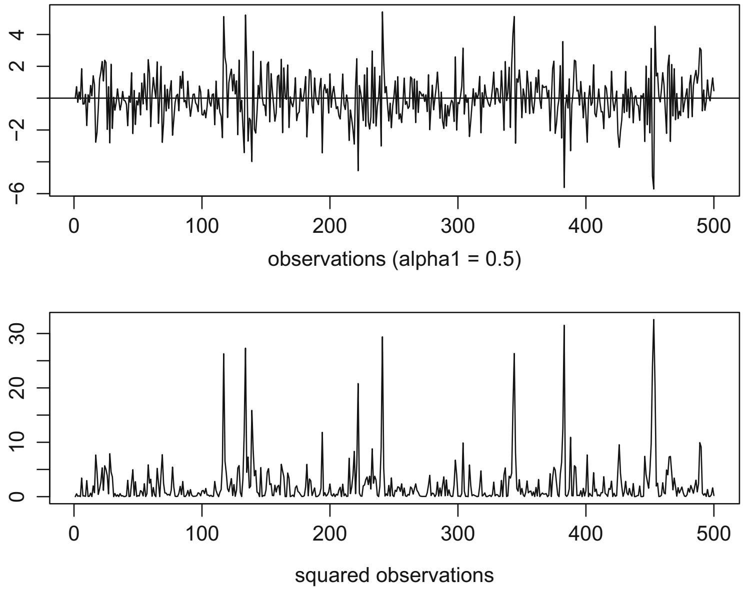

Fig. 6.1 ARCH(1) with $\alpha_0 = 1$ and $\alpha_1 = 0.5$

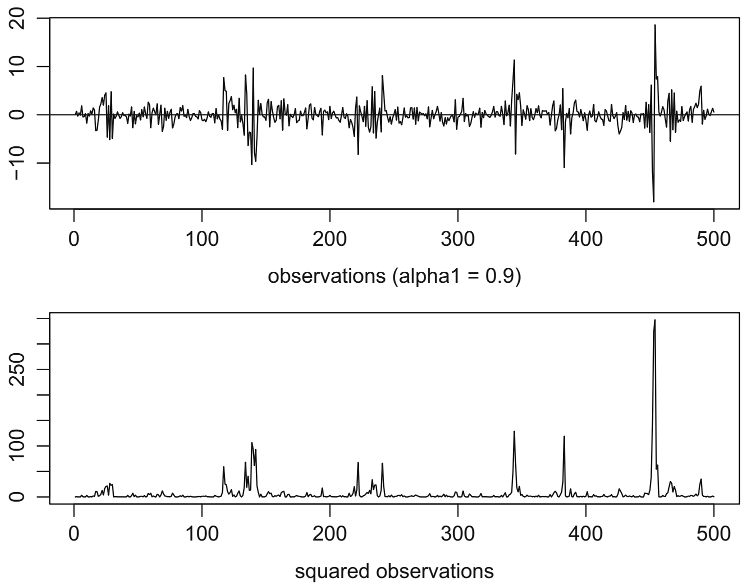

Fig. 6.2 ARCH(1) with $\alpha_0 = 1$ and $\alpha_1 = 0.9$

values are more concentrated around the expected value, on the other hand there occur extreme observations in the tail of the distribution with higher probability ("fat-tailed and highly peaked"). For stationary ARCH processes with Gaussian innovations ($\varepsilon_t \sim \text{ii}\,\mathcal{N}(0,1)$) it holds that the kurtosis exceeds 3 (provided it exists at all):

$$\gamma_2 > 3\,.$$

The corresponding derivation can be found in Problem 6.2. Due to this excess kurtosis, ARCH is generally incompatible with the assumption of an *un*conditional Gaussian distribution.

We define the skewness coefficient γ_1 similarly to the kurtosis by the third moment of a standardized random variable. The skewness coefficient of ARCH models depends on the symmetry of ε_t. If this innovation is symmetric, then it follows that $\mathrm{E}(\varepsilon_t^3) = 0$. Hence, $\gamma_1 = 0$ follows for the corresponding ARCH process (due to independence of σ_t and ε_t):

$$\begin{aligned}\mathrm{E}(x_t^3) &= \mathrm{E}(\sigma_t^3)\cdot\mathrm{E}(\varepsilon_t^3)\\ &= \mathrm{E}(\sigma_t^3)\cdot 0 = 0.\end{aligned}$$

Thereby it was only used that ε_t is symmetrically distributed.

Example 6.2 (ARCH(1)) In particular for a stationary ARCH(1) process with $\alpha_1^2 < \frac{1}{3}$ and Gaussian innovations ε_t the kurtosis is finite and it results as (see Problem 6.3):

$$\gamma_2 = 3\,\frac{1-\alpha_1^2}{1-3\,\alpha_1^2} > 3\,.$$

For $\alpha_1^2 \geq \frac{1}{3}$ there occur extreme observations with a high probability such that the kurtosis is no longer constant. Consider a stationary ARCH(1) process ($\alpha_1 < 1$) with $\alpha_1^2 = 1/3$. Under this condition one has for $\mu_{4,t} = \mathrm{E}(x_t^4)$ with $\mathrm{E}(x_{t-1}^2) = \alpha_0/(1-\alpha_1)$ assuming Gaussianity:

$$\begin{aligned}\mu_{4,t} &= \mathrm{E}(\varepsilon_t^4)\,\mathrm{E}(\sigma_t^4) = 3\,\mathrm{E}\left((\alpha_0+\alpha_1 x_{t-1}^2)^2\right)\\ &= 3\left(\alpha_0^2 + 2\,\frac{\alpha_0^2\alpha_1}{1-\alpha_1} + \alpha_1^2\mu_{4,t-1}\right) = c + 3\,\alpha_1^2\mu_{4,t-1} = c + \mu_{4,t-1}\,,\end{aligned}$$

where the constant c is appropriately defined. Continued substitution yields thus

$$\mu_{4,t} = c\,t + \mu_{4,0}\,.$$

Hence, we observe that the kurtosis grows linearly over time if $3\,\alpha_1^2 = 1$. ■

6.4 Generalizations

Some extensions of the ARCH model having originated from empirical features of financial data will be covered in the following.

GARCH

In practice, with many financial series it can be observed that the correlation of the squares reaches far into the past. Therefore, for an adequate modeling a large q is needed, i.e. a large number of parameters. A very economical parametrization, however, is allowed for by the GARCH model.

Generalized ARCH processes of the order p and q were introduced by Bollerslev (1986) and are defined by their volatility function

$$\sigma_t^2 = \alpha_0 + \alpha_1 x_{t-1}^2 + \cdots + \alpha_q x_{t-q}^2 + \beta_1 \sigma_{t-1}^2 + \cdots + \beta_p \sigma_{t-p}^2 . \tag{6.7}$$

The result process is abbreviated as GARCH(p, q). In addition to the parameter restrictions from (6.6) it is required that

$$\beta_i \geq 0 \, , \; i = 1, \ldots, p \, . \tag{6.8}$$

Jointly, these restrictions are clearly sufficient for $\sigma_t^2 > 0$ but stricter than necessary. Substantially weaker assumptions were derived by Nelson and Cao (1992).

We adopt the stationarity conditions for GARCH models from Bollerslev (1986, Theorem 1). The resulting variance will be determined in Problem 6.4. Thus, we obtain the following results.

Proposition 6.3 (Stationary GARCH)
Let $\{x_t\}$ be from (6.1) and $\{\sigma_t^2\}$ from (6.7) with (6.6) and (6.8). The process is weakly stationary if and only if

$$\sum_{j=1}^{q} \alpha_j + \sum_{j=1}^{p} \beta_j < 1 \, .$$

Then it holds for the variance that

$$Var(x_t) = \frac{\alpha_0}{1 - \sum_{j=1}^{q} \alpha_j - \sum_{j=1}^{p} \beta_j} \, .$$

It can be shown that the stationary GARCH process can be considered as an ARCH(∞) process. Under the conditions from Proposition 6.3 it holds that (see Problem 6.5):

$$\sigma_t^2 = \gamma_0 + \sum_{i=1}^{\infty} \gamma_i \, x_{t-i}^2 \quad \text{with } \gamma_i \geq 0 \text{ and } \sum_{i=1}^{\infty} |\gamma_i| < \infty \,. \tag{6.9}$$

Thus, the GARCH process allows for modeling an infinitely long dependence of the volatility on the past of the process itself with only $p + q$ parameters although this dependence decays with time (i.e. $\gamma_i \to 0$ for $i \to \infty$). The fact that GARCH can be considered as ARCH(∞) has the nice consequence that results for stationary ARCH processes also hold for GARCH models. In particular, GARCH models are again special cases of processes with volatility (6.2) and therefore examples of martingale differences, i.e. Proposition 6.1 holds true. If we assume a Gaussian distribution of $\{\varepsilon_t\}$, it follows, just as for the ARCH(q) process of finite order, that the skewness is zero and that the kurtosis exceeds the value 3.

Example 6.3 (GARCH(1,1)) Consider the GARCH(1,1) case more explicitly. It is by far the most frequently used GARCH specification in practice. Continued substitution shows under the assumption of stationarity that ($\alpha_1 + \beta_1 < 1$):

$$\sigma_t^2 = \frac{\alpha_0}{1 - \beta_1} + \alpha_1 \sum_{i=1}^{\infty} \beta_1^{i-1} \, x_{t-i}^2 \,.$$

Hence, we have an explicit ARCH(∞) representation of GARCH(1,1). Assuming that

$$1 - (\alpha_1 + \beta_1)^2 - 2\alpha_1^2 \;>\; 0,$$

the kurtosis is defined. In Problem 6.6 we show (with Gaussian distribution of $\{\varepsilon_t\}$):

$$\gamma_2 = 3 \, \frac{1 - (\alpha_1 + \beta_1)^2}{1 - (\alpha_1 + \beta_1)^2 - 2\alpha_1^2} \;>\; 3.$$

Furthermore one shows by

$$\begin{aligned} \sigma_t^2 &= \alpha_0 + \alpha_1 \, x_{t-1}^2 + \beta_1 \, \sigma_{t-1}^2 \\ &\Longleftrightarrow \\ x_t^2 &= \alpha_0 + (\alpha_1 + \beta_1) \, x_{t-1}^2 - \beta_1 \left(x_{t-1}^2 - \sigma_{t-1}^2 \right) + x_t^2 - \sigma_t^2 \end{aligned}$$

the equation

$$x_t^2 = \alpha_0 + (\alpha_1 + \beta_1) \, x_{t-1}^2 + e_t - \beta_1 \, e_{t-1}$$

with

$$e_t = x_t^2 - \sigma_t^2\,, \quad \mathrm{E}(e_t) = 0\,.$$

The GARCH(1,1) process $\{x_t\}$ therefore corresponds to an ARMA(1,1) structure of the squares $\{x_t^2\}$. ■

In Figs. 6.3 and 6.4 the influence of the sum of the parameters $\alpha_1 + \beta_1$ is illustrated by means of simulated GARCH(1,1) observations. We therefore fix $\alpha_0 = 1$ and $\alpha_1 = 0.3$ and vary β_1 in such a way that stationarity is ensured. The larger β_1 (and therefore the sum of $\alpha_1 + \beta_1$), the more pronounced is the change from quiet periods with little or, in absolute value, moderate amplitudes to excited periods in which extreme amplitudes follow each other. Again, this pattern of volatility becomes particularly apparent with the serially correlated squares in the lower panel, respectively.

IGARCH

Considering the volatility of GARCH(1,1),

$$\sigma_t^2 = \alpha_0 + \alpha_1\, x_{t-1}^2 + \beta_1\, \sigma_{t-1}^2,$$

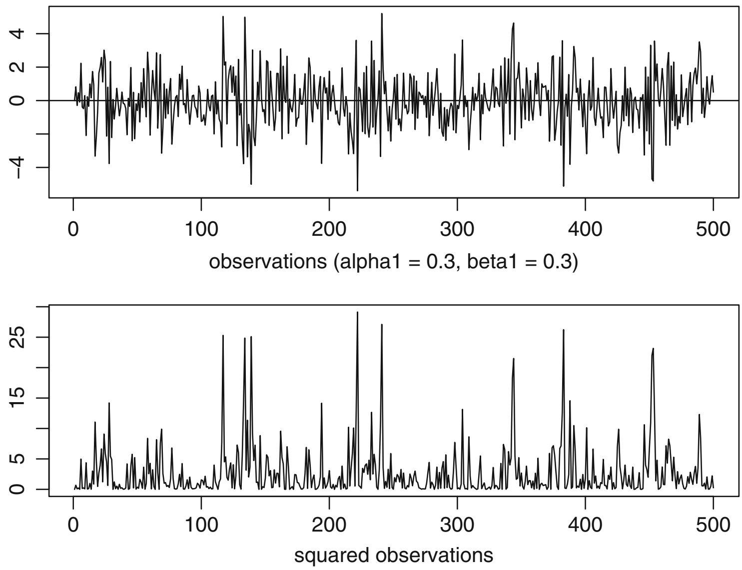

Fig. 6.3 GARCH(1,1) with $\alpha_0 = 1$, $\alpha_1 = 0.3$ and $\beta_1 = 0.3$

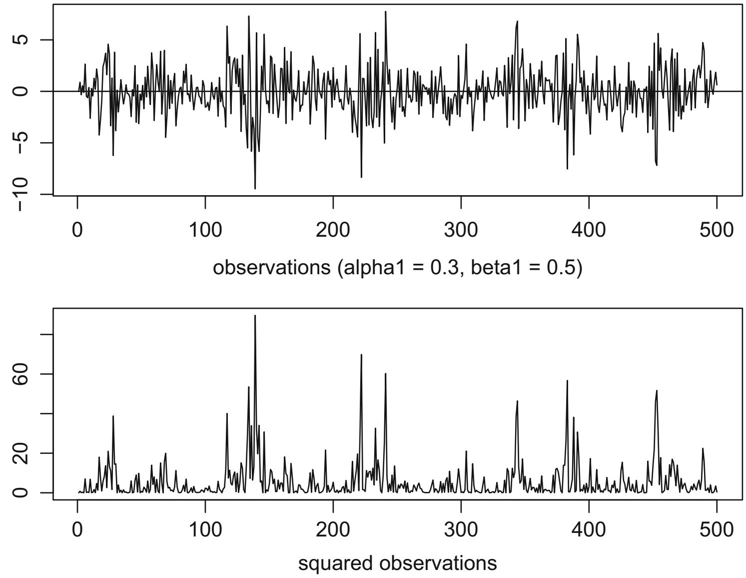

Fig. 6.4 GARCH(1,1) with $\alpha_0 = 1$, $\alpha_1 = 0.3$ and $\beta_1 = 0.5$

we are reminded of $s_t^2(\alpha)$ from (6.4). The difference being that $\alpha_0 = 0$, and it holds that $\alpha_1 + \beta_1 = 1$ (i.e. $\alpha_1 = 1 - \lambda$ and $\beta_1 = \lambda$, respectively). Models with such a restriction violate the stationarity condition ($\alpha_1 + \beta_1 < 1$). This can be shown when forming the expected value of σ_t^2 with $x_{t-1}^2 = \sigma_{t-1}^2\, \varepsilon_{t-1}^2$:

$$\begin{aligned} \mathrm{E}(\sigma_t^2) &= \alpha_0 + \alpha_1\, \mathrm{E}(\sigma_{t-1}^2)\, \mathrm{E}(\varepsilon_{t-1}^2) + \beta_1 \mathrm{E}(\sigma_{t-1}^2) \\ &= \alpha_0 + (\alpha_1 + \beta_1)\, \mathrm{E}(\sigma_{t-1}^2). \end{aligned}$$

With $\alpha_1 + \beta_1 = 1$ one obtains

$$\mathrm{E}(\sigma_t^2 - \sigma_{t-1}^2) = \alpha_0 > 0.$$

In other words: The expectations for the increments of the volatility are positive for every point in time, modeling a volatility expectation which tends to infinity with t. This idea was generalized in literature. With

$$\sum_{j=1}^{q} \alpha_j + \sum_{j=1}^{p} \beta_j = 1$$

one talks about integrated GARCH processes (IGARCH) since Engle and Bollerslev (1986). This is a naming which becomes more understandable in the chapter on integrated processes.

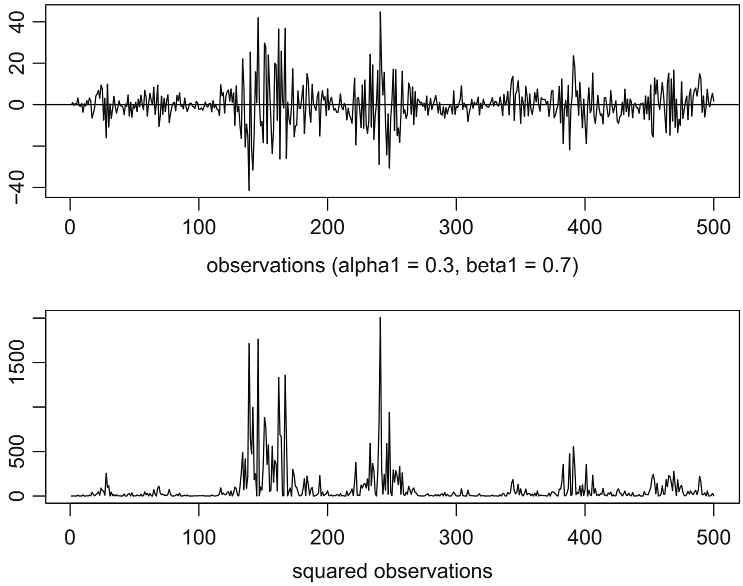

Fig. 6.5 IGARCH(1,1) with $\alpha_0 = 1$, $\alpha_1 = 0.3$ and $\beta_1 = 0.7$

In Fig. 6.5, an IGARCH(1,1) process ($\alpha_1 + \beta_1 = 1$) was simulated according to the scheme from Figs. 6.3 and 6.4. In comparison to the previous figures, in this case we find considerably more extreme volatility clusters which, however, are not exaggerated. The kind of depicted dynamics in Fig. 6.5 can be frequently observed in financial practice.

GARCH-M

We talk about "GARCH in mean"[3] (GARCH-M) if the volatility term influences the (mean) level of the process. In order to explain this with regard to contents, we think of risk premia: For a high volatility of an investment (high-risk), a higher return is expected, on average. In equation form we write this down as follows:

$$x_t = \theta\, \sigma_t + u_t\,, \tag{6.10}$$

where $\{u_t\}$ is a GARCH process:

$$u_t = \sigma_t\, \varepsilon_t\,, \quad \sigma_t^2 = \alpha_0 + \alpha_1\, u_{t-1}^2 + \cdots + \alpha_q\, u_{t-q}^2 + \beta_1\, \sigma_{t-1}^2 + \ldots + \beta_p\, \sigma_{t-p}^2\,. \tag{6.11}$$

[3]Originally, the ARCH-M model was proposed by Engle, Lilien, and Robins (1987).

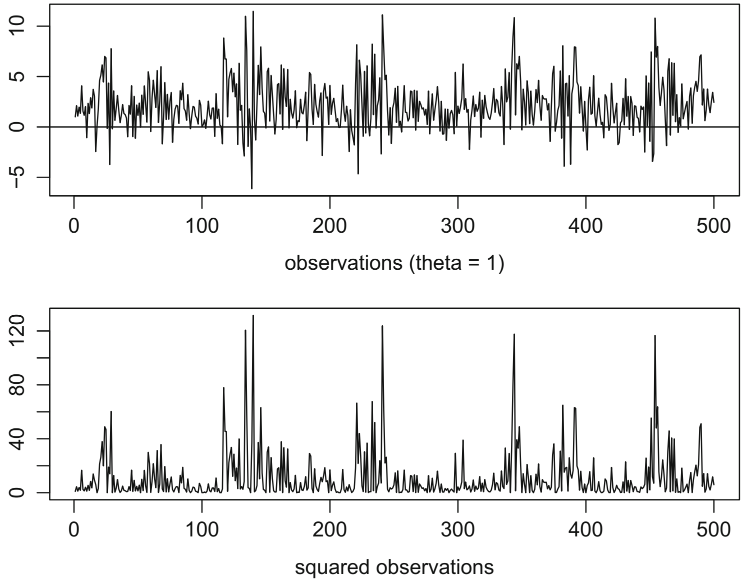

Fig. 6.6 GARCH(1,1)-M from (6.11) with $\alpha_0 = 1$, $\alpha_1 = 0.3$ and $\beta_1 = 0.5$

Therefore, in this case the mean function μ_t is set to $\theta\,\sigma_t$. In some applications it has proved successful to model the risk premium as a multiple of the variance instead of modeling it by the standard deviation ($\theta\,\sigma_t$):

$$x_t = \theta\,\sigma_t^2 + u_t\,.$$

For both mean functions the GARCH-M process $\{x_t\}$ is no longer free from serial correlation for $\theta > 0$; it is no longer a martingale difference sequence.

In Fig. 6.6 a GARCH-M series was generated as in (6.11). The volatility cluster can be well identified in the lower panel of the squares.The effect of $\theta = 1$ becomes apparent in the upper panel: In the series of x_t local, reversing trends can be spotted. Upward trends involve a high volatility, whereas quiet periods are marked by a decreasing or lower level.

EGARCH

We talk about exponential GARCH when the volatility is modeled as an exponential function of the past squares x_{t-i}^2. This suggestion originates from Nelson (1991) and was made in order to capture the asymmetries in the volatility.[4] It is observed that decreasing stock prices (negative returns) tend to involve higher volatilities than

[4]We do not exactly present Nelson's model but a slightly modified implementation which is used in the software package *EViews*.

increasing ones. This so-called **leverage effect** is not captured by ordinary GARCH models.

In *EViews* the variance for EGARCH is calculated as follows:

$$\log \sigma_t^2 = \omega + \sum_{j=1}^{p} \beta_j \log \sigma_{t-j}^2 + \sum_{j=1}^{q} \left(\alpha_j |\varepsilon_{t-j}| + \gamma_j \, \varepsilon_{t-j} \right),$$

where ε_{t-j} is defined as

$$\varepsilon_{t-j} = \frac{x_{t-j}}{\sigma_{t-j}}.$$

For $\gamma_j = 0$ the sign is not an issue. However, for $\gamma_j < 0$ in the negative case

$$\alpha_j |\varepsilon_{t-j}| + \gamma_j \, \varepsilon_{t-j} = (\alpha_j - \gamma_j)|\varepsilon_{t-j}| \quad \text{for} \quad \varepsilon_{t-j} < 0$$

has a stronger effect on $\log \sigma_t^2$ than in the positive case

$$\alpha_j |\varepsilon_{t-j}| + \gamma_j \, \varepsilon_{t-j} = (\alpha_j + \gamma_j)\varepsilon_{t-j} \quad \text{for} \quad \varepsilon_{t-j} > 0.$$

Note that for EGARCH the expression σ_t^2 is without parameter restrictions always positive by construction. Applying the exponential function, it results that

$$\sigma_t^2 = \exp \left[\omega + \sum_{j=1}^{p} \beta_j \log \sigma_{t-j}^2 + \sum_{j=1}^{q} \left(\alpha_j |\varepsilon_{t-j}| + \gamma_j \, \varepsilon_{t-j} \right) \right].$$

Example 6.4 (EGARCH (1,1)) Again, as special case we treat the situation with $p = q = 1$,

$$\log \sigma_t^2 = \omega + \beta_1 \, \log \sigma_{t-1}^2 + \alpha_1 |\varepsilon_{t-1}| + \gamma_1 \, \varepsilon_{t-1} \,,$$

or after applying the exponential function:

$$\begin{aligned} \sigma_t^2 &= e^{\omega} \sigma_{t-1}^{2\beta_1} \cdot \exp\left(\alpha_1 |\varepsilon_{t-1}| + \gamma_1 \, \varepsilon_{t-1}\right) \\ &= e^{\omega} \sigma_{t-1}^{2\beta_1} \cdot \begin{cases} \exp\left(|\varepsilon_{t-1}|(\alpha_1 - \gamma_1)\right), & \varepsilon_{t-1} \le 0 \\ \exp\left(\varepsilon_{t-1}(\alpha_1 + \gamma_1)\right), & \varepsilon_{t-1} \ge 0 \end{cases}. \end{aligned}$$

In this case it is again shown that for $\gamma_1 < 0$ the leverage effect is modeled in such a way that negative observations have a larger volatility effect than positive observations of the same absolute value. ■

In Fig. 6.7 a realization of a simulated EGARCH(1,1) process is depicted. The "leverage parameter" is $\gamma_1 = -0.5$. When the graphs of the squared and the original observations are compared, it can be detected that the most extreme amplitudes are

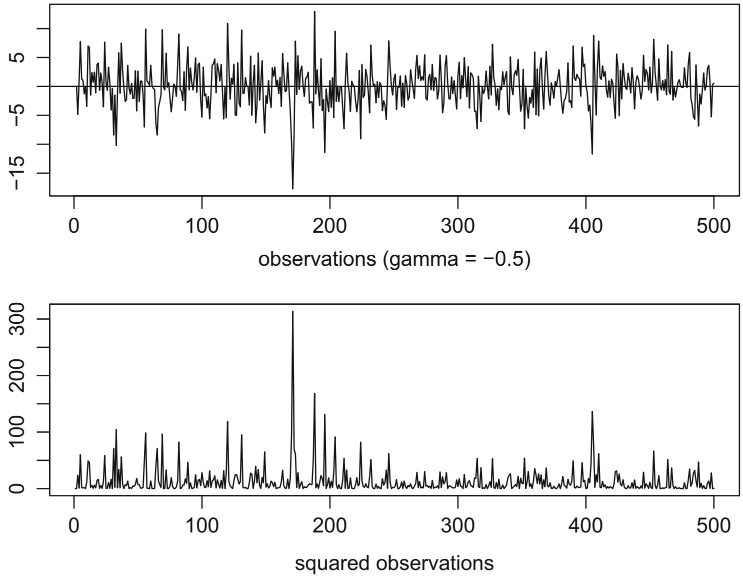

Fig. 6.7 EGARCH(1,1) with $\omega = 1$, $\alpha_1 = 0.3$, $\beta_1 = 0.5$ and $\gamma_1 = -0.5$

in fact negative. Furthermore, it can be observed that periods with predominantly negative values are characterized by a high volatility.

YAARCH

The works of Engle (1982) and Bollerslev (1986) have set the stage for a downright ARCH industry. A large number of generalizations and extensions has been published and applied in practice. Most of these versions were published under more or less appealing acronyms. When Engle (2002) balanced the books after 20 years of ARCH, he added with some irony another acronym: YAARCH standing for Yet Another ARCH. There is no end in sight for this literature.

6.5 Problems and Solutions

Problems

6.1 Prove Proposition 6.2.
Hint: According to Engle (1982, Theorem 2) the process is stationary if and only if it holds that

$$\alpha(z) = 0 \Rightarrow |z| > 1\,, \tag{6.12}$$

with $\alpha(z) := 1 - \alpha_1 z - \ldots - \alpha_q z^q$.

6.2 Show that the kurtosis γ_2 of an ARCH process exceeds the value 3. Assume a Gaussian distribution of $\{\varepsilon_t\}$ and $\mathrm{E}(\sigma_t^4) < \infty$.

6.3 Calculate the kurtosis of a stationary ARCH(1) process as given in Example 6.2 for the case that it exists. Assume a Gaussian distribution of $\{\varepsilon_t\}$.

6.4 Assume $\{x_t\}$ to be a stationary GARCH process. Determine the variance expression from Proposition 6.3.

6.5 Assume $\{x_t\}$ to be a stationary GARCH process with (6.6) and (6.8). Determine the ARCH(∞) representation from (6.9).

6.6 Calculate the kurtosis of a stationary GARCH(1,1) process as given in Example 6.3 for the case that it exists. Assume a Gaussian distribution of $\{\varepsilon_t\}$.

Solutions

6.1 We have to show the equivalence of (6.12) and the condition from Proposition 6.2, given (6.6). This condition can also be written as $\alpha(1) > 0$. Hence, we have to prove the equivalence:

$$(6.12) \iff \alpha(1) > 0.$$

We proceed in two steps.
"$\Rightarrow$": Under the condition by Engle (1982) it holds that

$$\mathrm{Var}(x_t) = \frac{\alpha_0}{1 - \alpha_1 - \ldots - \alpha_q} = \frac{\alpha_0}{\alpha(1)} \geq 0\,.$$

Due to $\alpha_0 > 0$ it immediately follows that $\alpha(1) \geq 0$. The case $\alpha(1) = 0$, however, is due to (6.12) excluded, such that $\alpha(1) > 0$ can be concluded.
"$\Leftarrow$": For a root z of $\alpha(z)$ it holds that:

$$1 = \sum_{j=1}^{q} \alpha_j z^j\,.$$

By the triangle inequality and applying (6.6) it follows that:

$$\begin{aligned}1 \leq \sum_{j=1}^{q} \left|\alpha_j z^j\right| = \sum_{j=1}^{q} \alpha_j \left|z^j\right| \leq \max_j \left|z^j\right| \sum_{j=1}^{q} \alpha_j \\ = \max_j \left|z^j\right| (1 - \alpha(1)) \\ < \max_j \left|z^j\right|,\end{aligned}$$

where the assumption was used for the last inequality. Therefore, we have shown for a root z of $\alpha(z)$ that it holds that $\max_j |z^j| > 1$ and thus $|z| > 1$.
Hence, the proof is completed.

6.2 From Example 2.4 we adopt due to the Gaussian distribution

$$\mathrm{E}(\varepsilon_t^4) = \mathrm{E}\left(\left(\frac{\varepsilon_t - 0}{1}\right)^4\right) = 3\,.$$

Therefore, in a first step it follows due to the independence of σ_t and ε_t:

$$\mathrm{E}(x_t^4) = \mathrm{E}\left(\sigma_t^4\,\varepsilon_t^4\right) = \mathrm{E}\left(\sigma_t^4\right)\,\mathrm{E}\left(\varepsilon_t^4\right) = 3\,\mathrm{E}\left(\sigma_t^4\right).$$

Hence, because of $\mathrm{E}(x_t) = 0$ and Proposition 6.1 the kurtosis of x_t results as:

$$\gamma_2 = \frac{\mathrm{E}(x_t^4)}{(\mathrm{Var}(x_t))^2} = \frac{3\,\mathrm{E}\left(\sigma_t^4\right)}{\left(\mathrm{E}\left(\sigma_t^2\right)\right)^2}\,.$$

The usual variance decomposition, see Eq. (2.1),

$$\mathrm{Var}\left(\sigma_t^2\right) = \mathrm{E}\left(\sigma_t^4\right) - \left(\mathrm{E}\left(\sigma_t^2\right)\right)^2 \geq 0\,,$$

yields

$$\frac{\mathrm{E}\left(\sigma_t^4\right)}{\left(\mathrm{E}\left(\sigma_t^2\right)\right)^2} \geq 1\,.$$

Hence, the claim is verified: $\gamma_2 \geq 3$.

6.3 As σ_t and ε_t are stochastically independent, it holds that

$$\mathrm{E}(x_t^k) = \mathrm{E}(\sigma_t^k)\mathrm{E}(\varepsilon_t^k),$$

whereby the k-th central moment is given, as $\{x_t\}$ is a martingale difference sequence with zero expectation. On the assumption of a standard normally distributed random process one obtains

$$\mathrm{E}(\varepsilon_t^2) = 1, \quad \mathrm{E}(\varepsilon_t^3) = 0 \quad \text{and} \quad \mathrm{E}(\varepsilon_t^4) = 3.$$

This implies for the ARCH(1) process that the skewness is zero due to $\mathrm{E}(x_t^3) = 0$.

In order to determine the kurtosis, we first observe that the fourth moment is constant under the condition $3\alpha_1^2 < 1$. To that end define $\mu_{4,t}$ and use $\mu_2 = \mathrm{Var}(x_t) = \frac{\alpha_0}{1-\alpha_1}$:

$$\mu_{4,t} = \mathrm{E}(x_t^4) = \mathrm{E}(\varepsilon_t^4)\,\mathrm{E}(\sigma_t^4) = 3\,\mathrm{E}\left((\alpha_0 + \alpha_1 x_{t-1}^2)^2\right)$$
$$= 3\left(\alpha_0^2 + 2\,\frac{\alpha_0^2\alpha_1}{1-\alpha_1} + \alpha_1^2\mu_{4,t-1}\right) = c + 3\,\alpha_1^2\mu_{4,t-1}\,,$$

where the constant c is defined appropriately. Infinite substitution yields

$$\mu_{4,t} = c\left(1 + 3\alpha_1^2 + (3\alpha_1^2)^2 + \cdots\right) = \frac{c}{1-3\alpha_1^2} = \mu_4\,.$$

With a constant μ_4 (and μ_2) one obtains

$$\mu_4 = \mathrm{E}(x_t^4) = 3\,\mathrm{E}(\sigma_t^4) = 3\,\mathrm{E}(\alpha_0^2 + 2\alpha_0\,\alpha_1\,x_{t-1}^2 + \alpha_1^2\,x_{t-1}^4)$$
$$= 3[\alpha_0^2 + 2\alpha_0\,\alpha_1\,\mu_2 + \alpha_1^2\,\mu_4]$$

or

$$\mu_4 = \frac{3}{1-3\,\alpha_1^2}\left[\alpha_0^2 + \frac{2\alpha_0^2\,\alpha_1}{1-\alpha_1}\right]$$
$$= \frac{3}{1-3\,\alpha_1^2}\,\frac{\alpha_0^2(1+\alpha_1)}{1-\alpha_1}.$$

From this it follows that

$$\gamma_2 = \frac{\mu_4}{(\mathrm{Var}(x_t))^2} = \frac{3}{1-3\,\alpha_1^2}(1-\alpha_1)(1+\alpha_1)$$
$$= 3\,\frac{1-\alpha_1^2}{1-3\,\alpha_1^2}.$$

Of course, these transformations were only possible for $1-3\alpha_1^2 > 0$. Hence, this is the condition for a finite, constant kurtosis.

6.4 We use the fact that σ_t from (6.7) is again independent of ε_t. This can be shown by substitution of σ_{t-j}^2 and x_{t-i}^2 according to (6.1). Thus, as in Proposition 6.1, for stationarity and for arbitrary points in time, it holds that:

$$\mathrm{Var}(x_t) = \mathrm{E}\left(\sigma_t^2\right) = \gamma(0)\,.$$

Hence, by forming the expected value we obtain from (6.7):

$$\gamma(0) = \alpha_0 + \alpha_1\,\gamma(0) + \ldots + \alpha_q\,\gamma(0) + \beta_1\,\gamma(0) + \ldots + \beta_p\,\gamma(0)\,.$$

Therefore, we can solve

$$\gamma(0) = \frac{\alpha_0}{1 - \sum_{j=1}^{q} \alpha_j - \sum_{j=1}^{p} \beta_j}\,,$$

as claimed.

6.5 We define the lag polynomial $\beta(L)$ with

$$\beta(L) = 1 - \beta_1\,L - \ldots - \beta_p\,L^p\,.$$

Hence, it holds that

$$\beta(L)\,\sigma_t^2 = \alpha_0 + \alpha_1\,x_{t-1}^2 + \ldots + \alpha_q\,x_{t-q}^2\,.$$

By assumption

$$\beta_j \geq 0\,, \quad j = 1, \ldots, p\,, \quad \text{and} \quad \beta(1) > 0\,.$$

In Problem 6.1 we have shown that this is equivalent to

$$\beta(z) = 0 \quad \Rightarrow \quad |z| > 1\,.$$

This is in turn the condition of invertibility known from Proposition 3.3 which guarantees a causal, absolutely summable series expansion with coefficients $\{c_j\}$:

$$\frac{1}{\beta(L)} = \sum_{j=0}^{\infty} c_j\,L^j\,, \quad \sum_{j=0}^{\infty} |c_j| < \infty\,.$$

By comparison of coefficients one obtains from

$$1 = \left(1 - \beta_1 L - \ \ldots\ - \beta_p L^p\right) \sum_{j=0}^{\infty} c_j L^j$$

as usual

$$\begin{aligned} c_0 &= 1 \\ c_1 &= \beta_1\,c_0 = \beta_1 \geq 0 \end{aligned}$$

$$c_2 = \beta_1 c_1 + \beta_2 c_0 = \beta_1^2 + \beta_2 \geq 0$$

$$\vdots$$

$$c_j = \beta_1 c_{j-1} + \ldots + \beta_p c_{j-p} \geq 0\,, \quad j \geq p\,.$$

Thus, the inversion of $\beta(L)$ yields:

$$\begin{aligned}\sigma_t^2 &= \frac{\alpha_0}{\beta(1)} + \frac{\alpha_1 x_{t-1}^2 + \ldots + \alpha_q x_{t-q}^2}{\beta(L)} \\ &= \gamma_0 + \sum_{i=1}^{\infty} \gamma_i x_{t-i}^2\,,\end{aligned}$$

where γ_i, $i > 0$, results by convolution of

$$\frac{\alpha_1 L + \ldots + \alpha_q L^q}{\beta(L)} = \sum_{k=1}^{q} \alpha_k L^k \sum_{j=1}^{\infty} c_j L^j.$$

The non-negativity and summability of $\{\alpha_k\}$ and $\{c_j\}$ is conveyed to the series $\{\gamma_i\}$. Hence, the proof is complete.

6.6 As for the ARCH(1) case it holds that

$$\mu_4 = \mathrm{E}\left(x_t^4\right) = 3\,\mathrm{E}\left(\sigma_t^4\right)\,.$$

Applying $\mathrm{E}\left(x_t^2\right) = \mathrm{E}\left(\sigma_t^2\right) = \frac{\alpha_0}{1-\alpha_1-\beta_1}$ yields:

$$\begin{aligned}\mathrm{E}\left(\sigma_t^4\right) &= \mathrm{E}\left(\left[\alpha_0 + \alpha_1 x_{t-1}^2 + \beta_1 \sigma_{t-1}^2\right]^2\right) \\ &= \mathrm{E}\left(\alpha_0^2 + \alpha_1^2 x_{t-1}^4 + \beta_1^2 \sigma_{t-1}^4\right) \\ &\quad + \mathrm{E}\left(2\,\alpha_0 \alpha_1 x_{t-1}^2 + 2\,\alpha_0 \beta_1 \sigma_{t-1}^2 + 2\,\alpha_1 \beta_1 x_{t-1}^2 \sigma_{t-1}^2\right) \\ &= \alpha_0^2 + 3\,\alpha_1^2\,\mathrm{E}\left(\sigma_{t-1}^4\right) + \beta_1^2\,\mathrm{E}\left(\sigma_{t-1}^4\right) \\ &\quad + \frac{2\,\alpha_0^2 \alpha_1}{1-\alpha_1-\beta_1} + \frac{2\,\alpha_0^2 \beta_1}{1-\alpha_1-\beta_1} + 2\,\alpha_1 \beta_1\,\mathrm{E}\left(\sigma_{t-1}^4\right)\,\mathrm{E}\left(\varepsilon_{t-1}^2\right)\,.\end{aligned}$$

As for the ARCH(1) case one has to show that $\mathrm{E}\left(\sigma_t^4\right)$ turns out to be constant under stationarity and the condition $1-2\,\alpha_1^2-(\alpha_1+\beta_1)^2 > 0$. We omit this step here and take $\mathrm{E}\left(\sigma_t^4\right) = \mathrm{E}\left(\sigma_{t-1}^4\right)$ for granted. It then holds that:

$$\mathrm{E}\left(\sigma_t^4\right) = \frac{\alpha_0^2 + \frac{2\,\alpha_0^2(\alpha_1+\beta_1)}{1-\alpha_1-\beta_1}}{1-2\,\alpha_1^2-(\alpha_1+\beta_1)^2}\,.$$

From this it follows that

$$\begin{aligned}\gamma_2 &= 3\,\mathrm{E}\left(\sigma_t^4\right)\,\frac{1}{\left(\mathrm{Var}\left(x_t^2\right)\right)^2}\\ &= 3\,\frac{\alpha_0^2\,(1+\alpha_1+\beta_1)}{(1-\alpha_1-\beta_1)(1-2\,\alpha_1^2-(\alpha_1+\beta_1)^2)}\;\frac{(1-\alpha_1-\beta_1)^2}{\alpha_0^2}\\ &= 3\,\frac{(1+(\alpha_1+\beta_1))(1-(\alpha_1+\beta_1))}{1-2\,\alpha_1^2-(\alpha_1+\beta_1)^2}\\ &= 3\,\frac{1-(\alpha_1+\beta_1)^2}{1-2\,\alpha_1^2-(\alpha_1+\beta_1)^2}\,.\end{aligned}$$

This is in accordance with the claim.

References

Bollerslev, T. (1986). Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics, 31*, 307–327.

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

Engle, R. F. (2002). New frontiers for ARCH models. *Journal of Applied Econometrics, 17*, 425–446.

Engle, R. F., & Bollerslev T. (1986). Modelling the persistence of conditional variances. *Econometric Reviews, 5*, 1–50.

Engle, R. F., Lilien, D. M., & Robins, R. P. (1987). Estimating time-varying risk premia in the term structure: the ARCH-M model. *Econometrica, 55*, 391–407.

Nelson, D. B. (1991). Conditional heteroskedasticity in asset returns: A new approach. *Econometrica, 59*, 347–370.

Nelson, D. B., & Cao, Ch. Q. (1992). Inequality constraints in the univariate GARCH model. *Journal of Business & Economic Statistics, 10*, 229–235.

Taylor, S.J. (1994). Modeling stochastic volatlity: A review and comparative study. *Mathematical Finance, 4*, 183–204.