Models of Volatility

8

The prices of financial market securities are often shaken by large and time-varying shocks. The amplitudes of these price movements are not constant. There are periods of high volatility and periods of low volatility. Within these periods volatility seems to be positively autocorrelated: high amplitudes are likely to be followed by high amplitudes and low amplitudes by low amplitudes. This observation which is particularly relevant for high frequency data such as, for example, daily stock market returns implies that the conditional variance of the one-period forecast error is no longer constant (homoskedastic), but time-varying (heteroskedastic). This insight motivated Engle (1982) and Bollerslev (1986) to model the time-varying variance thereby triggering a huge and still growing literature.[1] The importance of volatility models stems from the fact that the price of an option crucially depends on the variance of the underlying security price. Thus with the surge of derivative markets in the last decades the application of such models has seen a tremendous rise. Another use of volatility models is to assess the risk of an investment. In the computation of the so-called value at risk (VaR), these models have become an indispensable tool. In the banking industry, due to the regulations of the Basel accords, such assessments are in particular relevant for the computation of the required equity capital backing-up assets of different risk categories.

The following exposition focuses on the class of autoregressive conditional heteroskedasticity models (ARCH models) and their generalization the generalized autoregressive conditional heteroskedasticity models (GARCH models). These

[1]Robert F. Engle III was awarded the Nobel prize in 2003 for his work on time-varying volatility. His Nobel lecture (Engle 2004) is a nice and readable introduction to this literature.

K. Neusser, *Time Series Econometrics*, Springer Texts in Business and Economics,
DOI 10.1007/978-3-319-32862-1_8

models form the basis for even more generalized models (see Bollerslev et al. (1994) or Gouriéroux (1997)). Campbell et al. (1997) provide a broader economically motivated approach to the econometric analysis of financial market data.

8.1 Specification and Interpretation

8.1.1 Forecasting Properties of AR(1)-Models

Models of volatility play an important role in explaining the behavior of financial market data. They start from the observation that periods of high (low) volatility are clustered in specific time intervals. In these intervals high (low) volatility periods are typically followed by high (low) volatility periods. Thus volatility is usually positively autocorrelated as can be observed in Fig. 8.3. In order to understand this phenomenon we recapitulate the forecasting properties of the AR(1) model.[2] Starting from the model

$$X_t = c + \phi X_{t-1} + Z_t, \qquad Z_t \sim \text{IID}(0, \sigma^2) \text{ and } |\phi| < 1,$$

the best linear forecast in the mean-squared-error sense of X_{t+1} conditional on $\{X_t, X_{t-1}, \ldots\}$, denoted by $\mathbb{P}_t X_{t+1}$, is given by (see Chap. 3)

$$\mathbb{P}_t X_{t+1} = c + \phi X_t.$$

In practice the parameters c and ϕ are replaced by an estimate.

The *conditional* variance of the forecast error then becomes:

$$\mathbb{E}_t \left(X_{t+1} - \mathbb{P}_t X_{t+1} \right)^2 = \mathbb{E}_t Z_{t+1}^2 = \sigma^2,$$

where $\mathbb{E}_t$ denotes the conditional expectation operator based on information $X_t, X_{t-1}, \ldots$. The conditional variance of the forecast error is therefore constant, irrespective of the current state.

The *unconditional* forecast is simply the expected value of $\mathbb{E} X_{t+1} = \mu = \frac{c}{1-\phi}$ with forecast error variance:

$$\mathbb{E}\left(X_{t+1} - \frac{c}{1-\phi} \right)^2 = \mathbb{E}\left(Z_{t+1} + \phi Z_t + \phi^2 Z_{t-1} + \ldots \right)^2 = \frac{\sigma^2}{1-\phi^2} > \sigma^2.$$

Thus the conditional as well as the unconditional variance of the forecast error are constant. In addition, the conditional variance is smaller and thus more precise because it uses more information. Similar arguments can be made for ARMA models in general.

[2] Instead of assuming $Z_t \sim \text{WN}(0, \sigma^2)$, we make for convenience the stronger assumption that $Z_t \sim \text{IID}(0, \sigma^2)$.

8.1.2 The ARCH(1) Model

The volatility of financial market prices exhibit a systematic behavior so that the conditional forecast error variance is no longer constant. This observation led Engle (1982) to consider the following simple model for heteroskedasticity (non-constant variance).

Definition 8.1 (ARCH(1) Model). *A stochastic process $\{Z_t\}$, $t \in \mathbb{Z}$, is called an autoregressive conditional heteroskedastic process of order one, ARCH(1) process, if it is the solution of the following stochastic difference equation:*

$$Z_t = \nu_t \sqrt{\alpha_0 + \alpha_1 Z_{t-1}^2} \qquad \text{with } \alpha_0 > 0 \text{ and } 0 < \alpha_1 < 1, \tag{8.1}$$

where $\nu_t \sim \text{IID}\,\text{N}(0,1)$ and where ν_t and Z_{t-1} are independent from each other for all $t \in \mathbb{Z}$.

We will discuss the implications of this simple model below and consider generalizations in the next sections. First we prove the following theorem.

Theorem 8.1. *Under conditions stated in the definition of the ARCH(1) process, the difference equation (8.1) possesses a unique and strictly stationary solution with $\mathbb{E}Z_t^2 < \infty$. This solution is given by*

$$Z_t = \nu_t \sqrt{\alpha_0 \left(1 + \sum_{j=1}^{\infty} \alpha_1^j \nu_{t-1}^2 \nu_{t-2}^2 \ldots \nu_{t-j}^2\right)}. \tag{8.2}$$

Proof. Define the process

$$Y_t = Z_t^2 = \nu_t^2(\alpha_0 + \alpha_1 Y_{t-1}) \tag{8.3}$$

Iterating backwards k times we get:

$$\begin{aligned} Y_t &= \alpha_0 \nu_t^2 + \alpha_1 \nu_t^2 Y_{t-1} = \alpha_0 \nu_t^2 + \alpha_1 \nu_t^2 \nu_{t-1}^2 (\alpha_0 + \alpha_1 Y_{t-2}) \\ &= \alpha_0 \nu_t^2 + \alpha_0 \alpha_1 \nu_t^2 \nu_{t-1}^2 + \alpha_1^2 \nu_t^2 \nu_{t-1}^2 Y_{t-2} \\ &\qquad \ldots \\ &= \alpha_0 \nu_t^2 + \alpha_0 \alpha_1 \nu_t^2 \nu_{t-1}^2 + \ldots + \alpha_0 \alpha_1^k \nu_t^2 \nu_{t-1}^2 \ldots \nu_{t-k}^2 \\ &\quad + \alpha_1^{k+1} \nu_t^2 \nu_{t-1}^2 \ldots \nu_{t-k}^2 Y_{t-k-1}. \end{aligned}$$

Define the process $\{Y_t'\}$ as

$$Y_t' = \alpha_0 \nu_t^2 + \alpha_0 \sum_{j=1}^{\infty} \alpha_1^j \nu_t^2 \nu_{t-1}^2 \ldots \nu_{t-j}^2.$$

The right-hand side of the above expression just contains nonnegative terms. Moreover, making use of the IID N(0, 1) assumption of $\{\nu_t\}$,

$$\begin{aligned}\mathbb{E}Y_t' &= \mathbb{E}(\alpha_0 \nu_t^2) + \alpha_0 \mathbb{E}\left(\sum_{j=1}^{\infty} \alpha_1^j \nu_t^2 \nu_{t-1}^2 \ldots \nu_{t-j}^2\right) \\ &= \alpha_0 \sum_{j=0}^{\infty} \alpha_1^j = \frac{\alpha_0}{1-\alpha_1}.\end{aligned}$$

Thus, $0 \leq Y_t' < \infty$ a.s. Therefore, $\{Y_t'\}$ is strictly stationary and satisfies the difference equation (8.3). This implies that $Z_t = \sqrt{Y_t'}$ is also strictly stationary and satisfies the difference equation (8.1).

To prove uniqueness, we follow Giraitis et al. (2000). For any fixed t, it follows from the definitions of Y_t and Y_t' that for any $k \geq 1$

$$|Y_t - Y_t'| \leq \alpha_1^{k+1} \nu_t^2 \nu_{t-1}^2 \ldots \nu_{t-k}^2 |Y_{t-k-1}| + \alpha_0 \sum_{j=k+1}^{\infty} \alpha_1^j \nu_t^2 \nu_{t-1}^2 \ldots \nu_{t-j}^2.$$

The expectation of the right-hand side is bounded by

$$\left(\mathbb{E}|Y_1| + \frac{\alpha_0}{1-\alpha_1}\right) \alpha_1^{k+1}.$$

Define the event A_k by $A_k = \{|Y_t - Y_t'| > 1/k\}$. Then,

$$\mathbf{P}(A_k) \leq k\mathbb{E}|Y_t - Y_t'| \leq k \left(\mathbb{E}|Y_1| + \frac{\alpha_0}{1-\alpha_1}\right) \alpha_1^{k+1}$$

where the first inequality follows from Chebyschev's inequality setting $r = 1$ (see Theorem C.3). Thus, $\sum_{k=1}^{\infty} \mathbf{P}(A_k) < \infty$. The Borel-Cantelli lemma (see Theorem C.4) then implies that $\mathbf{P}\{A_k \; i.o.\} = 0$. However, as $A_k \subset A_{k+1}$, $\mathbf{P}(A_k) = 0$ for any k. Thus, $Y_t = Y_t'$ a.s. □

Remark 8.1. Note that the normality assumption is not necessary for the proof. The assumption $\nu_t \sim \text{IID}(0, 1)$ would be sufficient. Indeed, in practice it has been proven useful to adopt distributions with fatter tail than the normal, like the t-distribution (see the discussion in Sect. 8.1.3).

Given the assumptions made above $\{Z_t\}$ has the following properties:

(i) The expected value of Z_t is:

$$\mathbb{E}Z_t = \mathbb{E}\nu_t \mathbb{E}\sqrt{\alpha_0 + \alpha_1 Z_{t-1}^2} = 0.$$

This follows from the assumption that ν_t and Z_{t-1} are independent.

(ii) The covariances between Z_t and Z_{t-h}, $\mathbb{E}Z_t Z_{t-h}$, for $h \neq 0$ are given by:

$$\begin{aligned}\mathbb{E}Z_t Z_{t-h} &= \mathbb{E}\left(\nu_t \sqrt{\alpha_0 + \alpha_1 Z_{t-1}^2}\, \nu_{t-h}\sqrt{\alpha_0 + \alpha_1 Z_{t-h-1}^2}\right) \\ &= \mathbb{E}\nu_t\nu_{t-h}\, \mathbb{E}\sqrt{\alpha_0 + \alpha_1 Z_{t-1}^2}\, \sqrt{\alpha_0 + \alpha_1 Z_{t-h-1}^2} = 0.\end{aligned}$$

This is also a consequence of the independence assumption between ν_t and Z_{t-1}, respectively between ν_{t-h} and Z_{t-h-1}.

(iii) The variance of Z_t is:

$$\begin{aligned}\mathbb{V}Z_t &= \mathbb{E}Z_t^2 = \mathbb{E}\nu_t^2\left(\alpha_0 + \alpha_1 Z_{t-1}^2\right) \\ &= \mathbb{E}\nu_t^2\; \mathbb{E}\left(\alpha_0 + \alpha_1 Z_{t-1}^2\right) = \frac{\alpha_0}{1-\alpha_1} < \infty.\end{aligned}$$

This follows from the independence assumption between ν_t and Z_{t-1} and from the stationarity of $\{Z_t\}$. Because $\alpha_0 > 0$ and $0 < \alpha_1 < 1$, the variance is always strictly positive and finite.

(iv) As ν_t is normally distributed, its skewness, $\mathbb{E}\nu_t^3$, equals zero. The independence assumption between ν_t and Z_{t-1}^2 then implies that the skewness of Z_t is also zero, i.e.

$$\mathbb{E}Z_t^3 = 0.$$

Z_t therefore has a symmetric distribution.

The properties (i), (ii) and (iii) show that $\{Z_t\}$ is a white noise process. According to Theorem 8.1 it is not only stationary but even strictly stationary. Thus $\{Z_t\}$ is *uncorrelated* with $Z_{t-1}, Z_{t-2}, \ldots$, but not *independent* from its past! In particular we have:

$$\begin{aligned}\mathbb{E}\left(Z_t | Z_{t-1}, Z_{t-2}, \ldots\right) &= \mathbb{E}_t \nu_t \sqrt{\alpha_0 + \alpha_1 Z_{t-1}^2} = 0 \\ \mathbb{V}(Z_t | Z_{t-1}, Z_{t-2}, \ldots) &= \mathbb{E}\left(Z_t^2 | Z_{t-1}, Z_{t-2}, \ldots\right) \\ &= \mathbb{E}_t \nu_t^2\left(\alpha_0 + \alpha_1 Z_{t-1}^2\right) = \alpha_0 + \alpha_1 Z_{t-1}^2.\end{aligned}$$

The conditional variance of Z_t therefore depends on Z_{t-1}. Note that this dependence is positive because $\alpha_1 > 0$.

In order to guarantee that this conditional variance is always positive, we must postulate that $\alpha_0 > 0$ and $\alpha_1 > 0$. The stability of the difference equation requires in addition that $\alpha_1 < 1$.[3] Thus high volatility in the past, a large realization of Z_{t-1}, is followed by high volatility in the future. The precision of the forecast, measured by the conditional variance of the forecast error, thus depends on the history of the process. This feature is not compatible with linear models and thus underlines the *non-linear* character of the ARCH model and its generalizations.

Despite the fact that ν_t was assumed to be normally distributed, Z_t is not normally distributed. Its distribution deviates from the normal distribution in that extreme realizations are more probable. This property is called the *heavy-tail* property. In particular we have[4]:

$$\begin{aligned}\mathbb{E}Z_t^4 &= \mathbb{E}\nu_t^4 \left(\alpha_0 + \alpha_1 Z_{t-1}^2\right)^2 = \mathbb{E}\nu_t^4 \left(\alpha_0^2 + 2\alpha_0\alpha_1 Z_{t-1}^2 + \alpha_1^2 Z_{t-1}^4\right) \\ &= 3\alpha_0^2 + \frac{6\alpha_0^2\alpha_1}{1-\alpha_1} + 3\alpha_1^2 \mathbb{E}Z_{t-1}^4.\end{aligned}$$

The strict stationarity of $\{Z_t\}$ implies $\mathbb{E}Z_t^4 = \mathbb{E}Z_{t-1}^4$ so that

$$\begin{aligned}(1-3\alpha_1^2)\mathbb{E}Z_t^4 &= \frac{3\alpha_0^2(1+\alpha_1)}{1-\alpha_1} \qquad \Longrightarrow \\ \mathbb{E}Z_t^4 &= \frac{1}{1-3\alpha_1^2} \times \frac{3\alpha_0^2(1+\alpha_1)}{1-\alpha_1}.\end{aligned}$$

$\mathbb{E}Z_t^4$ is therefore positive and finite if and only if $3\alpha_1^2 < 1$, respectively if $0 < \alpha_1 < 1/\sqrt{3} = 0.5774$. For high correlation of the conditional variance, i.e. high $\alpha_1 > 1/\sqrt{3}$, the fourth moment and therefore also all higher even moments will no longer exist. The kurtosis κ is

$$\kappa = \frac{\mathbb{E}Z_t^4}{[\mathbb{E}Z_t^2]^2} = 3 \times \frac{1-\alpha_1^2}{1-3\alpha_1^2} > 3,$$

if $\mathbb{E}Z_t^4$ exists. The heavy-tail property manifests itself by a kurtosis greater than 3 which is the kurtosis of the normal distribution. The distribution of Z_t is therefore leptokurtic and thus more vaulted than the normal distribution.

Finally, we want to examine the autocorrelation function of Z_t^2. This will lead to a test for ARCH effects, i.e. for time varying volatility (see Sect. 8.2 below).

[3] The case $\alpha_1 = 1$ is treated in Sect. 8.1.4.

[4] As $\nu_t \sim \mathrm{N}(0,1)$ its even moments, $m_{2k} = \mathbb{E}\nu_t^{2k}$, $k = 1, 2, \ldots$, are given by $m_{2k} = \prod_{j=1}^{k}(2j-1)$. Thus we get $m_4 = 3$, $m_6 = 15$, etc. As the normal distribution is symmetric, all odd moments are equal to zero.

Theorem 8.2. *Assuming that* $\mathbb{E}Z_t^4$ *exists,* $Y_t = \frac{Z_t^2}{\alpha_0}$ *has the same autocorrelation function as the AR(1) process* $W_t = \alpha_1 W_{t-1} + U_t$ *with* $U_t \sim \mathrm{WN}(0,1)$. *In addition, under the assumption* $0 < \alpha_1 < 1$, *the process* $\{W_t\}$ *is also causal with respect to* $\{U_t\}$.

Proof. From $Y_t = \nu_t^2(1 + \alpha_1 Y_{t-1})$ we get:

$$\begin{aligned}
\gamma_Y(h) &= \mathbb{E}Y_tY_{t-h} - \mathbb{E}Y_t\mathbb{E}Y_{t-h} = \mathbb{E}Y_tY_{t-h} - \frac{1}{(1-\alpha_1)^2} \\
&= \mathbb{E}\nu_t^2\,(1 + \alpha_1 Y_{t-1})\,Y_{t-h} - \frac{1}{(1-\alpha_1)^2} \\
&= \mathbb{E}Y_{t-h} + \alpha_1\mathbb{E}Y_{t-1}Y_{t-h} - \frac{1}{(1-\alpha_1)^2} \\
&= \frac{1}{1-\alpha_1} + \alpha_1\left(\gamma_Y(h-1) + \frac{1}{(1-\alpha_1)^2}\right) - \frac{1}{(1-\alpha_1)^2} \\
&= \alpha_1\gamma_Y(h-1) + \frac{1-\alpha_1+\alpha_1-1}{(1-\alpha_1)^2} = \alpha_1\gamma_Y(h-1).
\end{aligned}$$

Therefore, $\gamma_Y(h) = \alpha_1^h\gamma_Y(0) \Rightarrow \rho(h) = \alpha_1^h$. □

The unconditional variance of X_t is:

$$\mathbb{V}X_t = \mathbb{V}\left(\frac{c}{1-\phi} + \sum_{j=0}^{\infty}\phi^j Z_{t-j}\right) = \frac{1}{1-\phi^2}\,\mathbb{V}Z_t = \frac{\alpha_0}{1-\alpha_1} \times \frac{1}{1-\phi^2}.$$

The unconditional variance of X_t involves all parameters of the model. Thus modeling the variance of X_t induces a trade-off between ϕ, α_0 and α_1.

Figure 8.1 plots the realizations of two AR(1)-ARCH(1) processes.Both processes have been generated with the same realization of $\{\nu_t\}$ and the same parameters $\phi = 0.9$ and $\alpha_0 = 1$. Whereas the first process (shown on the left panel of the figure) was generated with a value of $\alpha_1 = 0.9$, the second one had a value of $\alpha_1 = 0.5$. In both cases the stability condition, $\alpha_1 < 1$, is fulfilled, but for the first process $3\alpha_1^2 > 1$, so that the fourth moment does not exist. One can clearly discern the large fluctuations, in particular for the first process.

8.1.3 General Models of Volatility

The simple ARCH(1) model can be and has been generalized in several directions. A straightforward generalization proposed by Engle (1982) consists by allowing further lags to enter the ARCH equation (8.1). This leads to the ARCH(p) model:

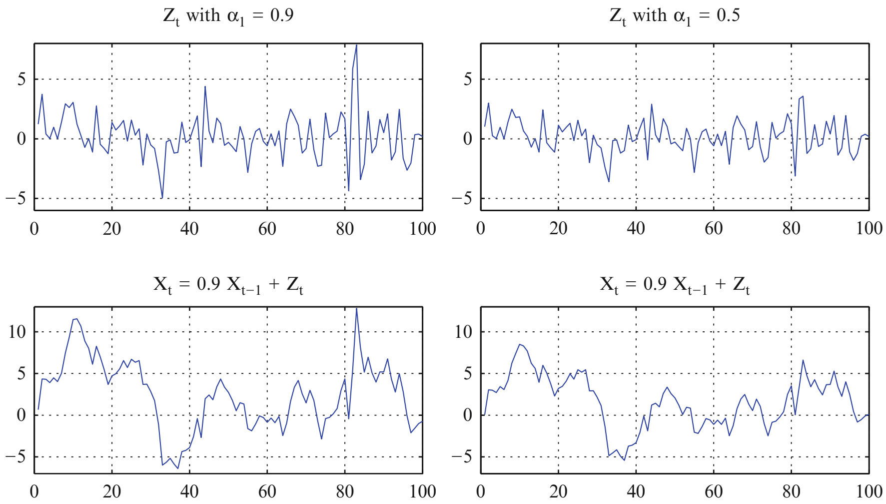

Fig. 8.1 Simulation of two ARCH(1) processes ($\alpha_1 = 0.9$ and $\alpha_1 = 0.5$)

$$\text{ARCH(p)}: \qquad Z_t = \nu_t \sigma_t \quad \text{with } \sigma_t^2 = \alpha_0 + \sum_{j=1}^{p} \alpha_j Z_{t-j}^2 \tag{8.4}$$

where $\alpha_0 \geq 0$, $\alpha_j \geq 0$ and $\nu_t \sim \text{IID N}(0, 1)$ with ν_t independent from Z_{t-j}, $j \geq 1$. A further popular generalization was proposed by Bollerslev (1986):

$$\text{GARCH(p, q)}: \qquad Z_t = \nu_t \sigma_t \quad \text{with } \sigma_t^2 = \alpha_0 + \sum_{j=1}^{p} \alpha_j Z_{t-j}^2 + \sum_{j=1}^{q} \beta_j \sigma_{t-j}^2 \tag{8.5}$$

where we assume $\alpha_0 \geq 0$, $\alpha_j \geq 0$, $\beta_j \geq 0$ and $\nu_t \sim \text{IID N}(0, 1)$ with ν_t independent from Z_{t-j}, $j \geq 1$, as before. This model is analogous the ordinary ARMA model and allows for a parsimonious specification of the volatility process. All coefficients should be positive to guarantee that the variance is always positive. In addition it can be shown (see for example Fan and Yao (2003, 150) and the literature cited therein) that $\{Z_t\}$ is (strictly) stationary with finite variance if and only if $\sum_{j=1}^{p} \alpha_j + \sum_{j=1}^{q} \beta_j < 1$.[5] Under this condition $\{Z_t\} \sim \text{WN}(0, \sigma_Z^2)$ with

$$\sigma_Z^2 = \mathbb{V}(Z_t) = \frac{\alpha_0}{1 - \sum_{j=1}^{p} \alpha_j - \sum_{j=1}^{q} \beta_j}.$$

[5] A detailed exposition of the GARCH(1,1) model is given in Sect. 8.1.4.

As ν_t is still normally distributed the uneven moments of the distribution of Z_t are zero and the distribution is thus symmetric. The fourth moment of Z_t, $\mathbb{E}Z_t^4$, exists if

$$\sqrt{3}\,\frac{\sum_{j=1}^{p}\alpha_j}{1-\sum_{j=1}^{q}\beta_j} < 1.$$

This condition is sufficient, but not necessary.[6] Furthermore, $\{Z_t\}$ is a white noise process with heavy-tail property if $\{Z_t\}$ is strictly stationary with finite fourth moment.

In addition, $\{Z_t^2\}$ is a causal and invertible ARMA$(\max\{p,q\},q)$ process satisfying the following difference equation:

$$\begin{aligned} Z_t^2 &= \alpha_0 + \sum_{j=1}^{p}\alpha_j Z_{t-j}^2 + \sum_{j=1}^{q}\beta_j\sigma_{t-j}^2 + e_t \\ &= \alpha_0 + \sum_{j=1}^{\max\{p,q\}}(\alpha_j+\beta_j)Z_{t-j}^2 + e_t - \sum_{j=1}^{q}\beta_j e_{t-j}, \end{aligned}$$

where $\alpha_{p+j} = \beta_{q+j} = 0$ for $j \geq 1$ and "error term"

$$e_t = Z_t^2 - \sigma_t^2 = (\nu_t^2 - 1)\left(\alpha_0 + \sum_{j=1}^{p}\alpha_j Z_{t-j}^2 + \sum_{j=1}^{q}\beta_j\sigma_{t-j}^2\right).$$

Note, however, there is a circularity here because the noise process $\{e_t\}$ is defined in terms of Z_t^2 and is therefore not an exogenous process driving Z_t^2. Thus, one has to be precautious in the interpretation of $\{Z_t^2\}$ as an ARMA process.

Further generalizations of the GARCH(p,q) model can be obtained by allowing deviations from the normal distribution for ν_t. In particular, distributions such as the t-distribution which put more weight on extreme values have become popular. This seems warranted as prices on financial markets exhibit large and sudden fluctuations.[7]

The Threshold GARCH Model

Assuming a symmetric distribution for ν_t and specifying a linear relationship between σ_t^2 and Z_{t-j}^2 bzw. σ_{t-j}^2, $j > 0$, leads to a symmetric distribution for Z_t. It has, however, been observed that downward movements seem to be different from

[6] Zadrozny (2005) derives a necessary and sufficient condition for the existence of the fourth moment.

[7] A thorough treatment of the probabilistic properties of GARCH processes can be found in Nelson (1990), Bougerol and Picard (1992a), Giraitis et al. (2000), Klüppelberg et al. (2004, theorem 2.1), and Lindner (2009).

upward movements. This asymmetric behavior is accounted for by the asymmetric GARCH(1,1) model or threshold GARCH(1,1) model (TGARCH(1,1) model). This model was proposed by Glosten et al. (1993) and Zakoïan (1994):

$$\text{asymmetric GARCH}(1,1): \; Z_t = \nu_t \sigma_t \quad \text{with}$$
$$\sigma_t^2 = \alpha_0 + \alpha_1 Z_{t-1}^2 + \beta \sigma_{t-1}^2 + \gamma \mathbf{1}_{\{Z_{t-1}<0\}} Z_{t-1}^2.$$

$\mathbf{1}_{\{Z_{t-1}<0\}}$ denotes the indicator function which takes on the value one if Z_{t-1} is negative and the value zero otherwise. Assuming, as before, that all parameters α_0, α_1, β and γ are greater than zero, this specification postulates a leverage effect because negative realizations have a greater impact than positive ones. In order to obtain a stationary process the condition $\alpha_1 + \beta + \gamma/2 < 1$ must hold. This model can be generalized in an obvious way by allowing additional lags Z_{t-j}^2 and $\sigma_{t-j}^2, j > 1$ to enter the above specification.

The Exponential GARCH Model

Another interesting and popular class of volatility models was introduced by Nelson (1991). The so-called exponential GARCH models or EGARCH models are defined as follows:

$$\log \sigma_t^2 = \alpha_0 + \beta \log \sigma_{t-1}^2 + \gamma \left| \frac{Z_{t-1}}{\sigma_{t-1}} \right| + \delta \frac{Z_{t-1}}{\sigma_{t-1}}$$
$$= \alpha_0 + \beta \log \sigma_{t-1}^2 + \gamma |\nu_{t-1}| + \delta \nu_{t-1}.$$

Note that, in contrast to the previous specifications, the dependent variable is the logarithm of σ_t^2 and not σ_t^2 itself. This has the advantage that the variance is always positive irrespective of the values of the coefficients. Furthermore, the leverage effect is exponential rather than quadratic because $Z_t = \nu_t \exp(\sigma_t/2)$. The EGARCH model is also less recursive than the GARCH model as the volatility is specified directly in terms of the noise process $\{\nu_t\}$. Thus, the above EGARCH model can be treated as an AR(1) model of $\log \sigma_t^2$ with noise process $\gamma|\nu_{t-1}| + \delta\nu_{t-1}$. It is obvious that the model can be generalized to allow for additional lags both in σ_t^2 and ν_t. This results in an ARMA process for $\{\log \sigma_t^2\}$ for which the usual conditions for the existence of a causal and invertible solution can be applied (see Sect. 2.3). A detailed analysis and further properties of this model class can be found in Bollerslev et al. (1994), Gouriéroux (1997) and Fan and Yao (2003, 143–180).

The ARCH-in-Mean Model

The ARCH-in-mean model or ARCH-M model was introduced by Engle et al. (1987) to allow for a feedback of volatility into the mean equation. More specifically, assume for the sake of simplicity that the variance equation is just represented the ARCH(1) model

$$Z_t = \nu_t \sigma_t \quad \text{with } \sigma_t^2 = \alpha_0 + \alpha_1 Z_{t-1}^2.$$

Then, the ARCH-M model is given

$$X_t = M_t\beta + g(\sigma_t^2) + Z_t \tag{8.6}$$

where g is a function of the volatility σ_t^2 and where M_t' consists of a vector of regressors, including lagged values of X_t. If $M_t = (1, X_{t-1})$ then we get the AR(1)-ARCH-M model. The most commonly used specification for g is a linear function: $g(\sigma_t^2) = \delta_0 + \delta_1\sigma_t^2$. In the asset pricing literature, higher volatility would require a higher return to compensate the investor for the additional risk. Thus, if X_t denotes the return on some asset, we expect δ_1 to be positive. Note that any time variation in σ_t^2 translates into a serial correlation of $\{X_t\}$ (see Hong 1991, for details). Of course, one could easily generalize the model to allow for more sophisticated mean and variance equations.

8.1.4 The GARCH(1,1) Model

The Generalized Autoregressive Conditional Heteroskedasticity model of order (1,1), GARCH(1,1) model for short, is considered as a benchmark for more general specifications and often serves as a starting point for further empirical investigations. We therefore want to explore its properties in more detail. Many of its properties generalize in a straightforward way to the GARCH(p,q) process. According to Eq. (8.5) the GARCH(1,1) model is defined as:

$$\text{GARCH}(1,1): \quad Z_t = \nu_t\sigma_t \quad \text{with}\ \sigma_t^2 = \alpha_0 + \alpha_1 Z_{t-1}^2 + \beta\sigma_{t-1}^2 \tag{8.7}$$

where α_0, α_1, and $\beta \geq 0$. We assume $\alpha_1 + \beta > 0$ to avoid the degenerate case $\alpha_1 = \beta = 0$ which implies that $\{Z_t\}$ is just a sequence of IID random variables. Moreover, $\nu_t \sim \text{IID}(0,1)$ with ν_t being independent of $Z_{t-j}, j \geq 1$. Note that we do not make further distributional assumption. In particular, ν_t need not required to be normally distributed. For this model, we can formulate a similar theorem as for the ARCH(1) model (see Theorem: 8.1):

Theorem 8.3. *Let $\{Z_t\}$ be a GARCH(1,1) process as defined above. Under the assumption*

$$\mathbb{E}\log(\alpha_1\nu_t^2 + \beta) < 0,$$

the difference equation (8.7) possess strictly stationary solution:

$$Z_t = \nu_t\sqrt{\alpha_0\sum_{j=0}^{\infty}\prod_{i=1}^{j}(\alpha_1\nu_{t-i}^2 + \beta)}$$

where $\prod_i^j = 1$ whenever $i > j$. The solution is also unique given the sequence $\{\nu_t\}$. The solution is unique and (weakly) stationary with variance $\mathbb{E}Z_t^2 = \frac{\alpha_0}{1-\alpha_1-\beta} < \infty$ if $\alpha_1 + \beta < 1$.

Proof. The proof proceeds similarly to Theorem 8.1. For this purpose, we define $Y_t = \sigma_t^2$ and rewrite the GARCH(1,1) model as

$$Y_t = \alpha_0 + \alpha_1 \nu_{t-1}^2 Y_{t-1} + \beta Y_{t-1} = \alpha_0 + (\alpha_1 \nu_{t-1}^2 + \beta) Y_{t-1}.$$

This defines an AR(1) process with time-varying coefficients $\xi_t = \alpha_1 \nu_t^2 + \beta \geq 0$. Iterate this equation backwards k times to obtain:

$$\begin{aligned} Y_t &= \alpha_0 + \alpha_0 \xi_{t-1} + \ldots + \alpha_0 \xi_{t-1} \ldots \xi_{t-k} + \xi_{t-1} \ldots \xi_{t-k} \xi_{t-k-1} Y_{t-k-1} \\ &= \alpha_0 \sum_{j=0}^{k} \prod_{i=1}^{j} \xi_{t-i} + \prod_{i=1}^{k+1} \xi_{t-i} Y_{t-k-1}. \end{aligned}$$

Taking the limit $k \to \infty$, we define the process $\{Y_t'\}$

$$Y_t' = \alpha_0 \sum_{j=0}^{\infty} \prod_{i=1}^{j} \xi_{t-i}. \tag{8.8}$$

The right-hand side of this expression converges almost surely as can be seen from the following argument. Given that $\nu_t \sim$ IID and given the assumption $\mathbb{E} \log(\alpha_1 \nu_t^2 + \beta) < 0$, the strong law of large numbers (Theorem C.5) implies that

$$\limsup_{j \to \infty} \frac{1}{j} \left(\sum_{i=1}^{j} \log(\xi_{t-i}) \right) < 0 \quad \text{a.s.,}$$

or equivalently,

$$\limsup_{j \to \infty} \log \left(\prod_{i=1}^{j} \xi_{t-i} \right)^{1/j} < 0 \quad \text{a.s.}$$

Thus,

$$\limsup_{j \to \infty} \left(\prod_{i=1}^{j} \xi_{t-i} \right)^{1/j} < 1 \quad \text{a.s.}$$

The application of the root test then shows that the infinite series (8.8) converges almost surely. Thus, $\{Y_t'\}$ is well-defined. It is easy to see that $\{Y_t'\}$ is strictly stationary and satisfies the difference equation. Moreover, if $\alpha_1 + \beta < 1$, we get

$$\mathbb{E} Y_t' = \alpha_0 \sum_{j=0}^{\infty} \mathbb{E} \prod_{i=1}^{j} \xi_{t-i} = \alpha_0 \sum_{j=0}^{\infty} (\alpha_1 + \beta)^j = \frac{\alpha_0}{1 - \alpha_1 - \beta}.$$

Thus, $\mathbb{E}Z_t^2 = \frac{\alpha_0}{1-\alpha_1-\beta} < \infty$.

To show uniqueness, we assume that there exists another strictly stationary process $\{Y_t\}$ which also satisfies the difference equation. This implies that

$$|Y_t - Y_t'| = |\xi_{t-1}|\,|Y_{t-1} - Y_{t-1}'| = \left(\prod_{i=1}^{k} \xi_{t-i}\right) |Y_{t-k} - Y_{t-k}'|$$
$$\leq \left(\prod_{i=1}^{k} \xi_{t-i}\right) |Y_{t-k}| + \left(\prod_{i=1}^{k} \xi_{t-i}\right) |Y_{t-k}'|$$

The assumption $\mathbb{E}\log\xi_t = \mathbb{E}\log(\alpha_1\nu_t^2 + \beta) < 0$ together with the strong law of large numbers (Theorem C.5) imply

$$\prod_{i=1}^{k} \xi_{t-i} = \left(\exp\left(\frac{1}{k}\sum_{i=1}^{k}\log\xi_{t-i}\right)\right)^k \longrightarrow 0 \quad \text{a.s.}$$

As both solutions are strictly stationary so that the distribution of $|Y_{t-k}|$ and $|Y_{t-k}'|$ do not depend on t, this implies that both $\left(\prod_{i=1}^{k}\xi_{t-i}\right)|Y_{t-k}|$ and $\left(\prod_{i=1}^{k}\xi_{t-i}\right)|Y_{t-k}'|$ converge in probability to zero. Thus, $Y_t = Y_t'$ a.s. once the sequence ξ_t, respectively ν_t, is given. Because $Z_t = \nu_t\sqrt{Y_t'}$ this completes the proof. □

Remark 8.2. Using Jensen's inequality, we see that

$$\mathbb{E}\log(\alpha_1\nu_t^2 + \beta) \leq \log\mathbb{E}(\alpha_1\nu_t^2 + \beta) = \log(\alpha_1 + \beta).$$

Thus, the condition $\alpha_1 + \beta < 1$ is sufficient, but not necessary, to ensure the existence of a strictly stationary solution. Thus even when $\alpha_1 + \beta = 1$, a strictly stationary solution exists, albeit one with infinite variance. This case is known as the IGARCH model and is discussed below. In the case $\alpha_1 + \beta < 1$, the Borel-Cantelli lemma can be used as Theorem 8.1 to establish the uniqueness of the solution. Further details can be found in the references listed in footnote 7.

Assume that $\alpha_1 + \beta < 1$, then a unique strictly stationary process $\{Z_t\}$ with finite variance which satisfies the above difference equation exists. In particular $Z_t \sim \text{WN}(0, \sigma_Z^2)$ such that

$$\mathbb{V}(Z_t) = \frac{\alpha_0}{1 - \alpha_1 - \beta}.$$

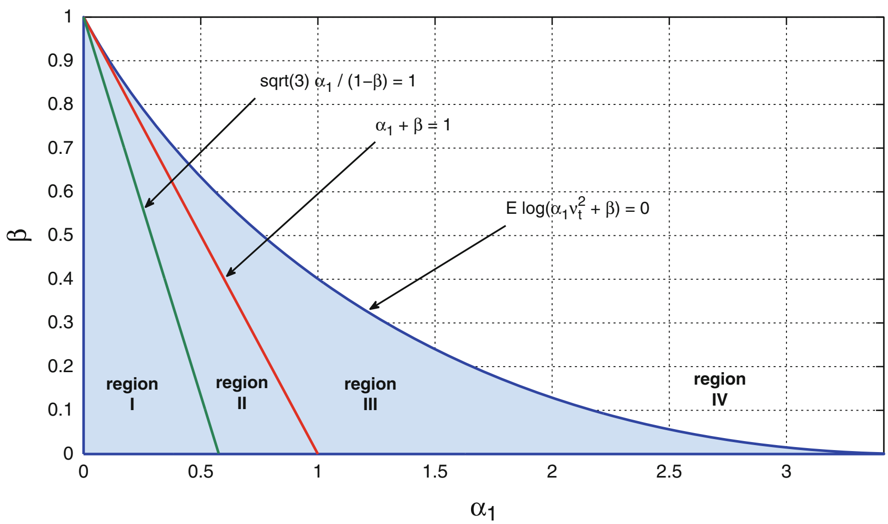

Fig. 8.2 Parameter region for which a strictly stationary solution to the GARCH(1,1) process exists assuming $\nu_t \sim$ IID N(0, 1)

The assumption $1-\alpha_1-\beta > 0$ guarantees that the variance exists. The third moment of Z_t is zero due to the assumption of a symmetric distribution for ν_t. The condition for the existence of the fourth moment is: $\sqrt{3}\frac{\alpha_1}{1-\beta} < 1$.[8] The kurtosis is then

$$\kappa = \frac{\mathbb{E}Z_t^4}{[\mathbb{E}Z_t^2]^2} = 3 \times \frac{1-(\alpha_1+\beta)^2}{1-(\alpha_1+\beta)^2-2\alpha_1^2} > 3$$

if $\mathbb{E}Z_t^4$ exists.[9] Therefore the GARCH(1,1) model also possesses the heavy-tail property because Z_t is more peaked than the normal distribution.

Figure 8.2 shows how the different assumptions and conditions divide up the parameter space. In region I, all conditions are fulfilled. The process has a strictly stationary solution with finite variance and kurtosis. In region II, the kurtosis does no longer exist, but the variance does as $\alpha_1 + \beta < 1$ still holds. In region III, the process has infinite variance, but a strictly stationary solution yet exists. In region IV, no such solution exists.

Viewing the equation for σ_t^2 as a stochastic difference equation, its solution is given by

$$\sigma_t^2 = \frac{\alpha_0}{1-\beta} + \alpha_1 \sum_{j=0}^{\infty} \beta^j Z_{t-1-j}^2. \tag{8.9}$$

[8]A necessary and sufficient condition is $(\alpha_1+\beta)^2 + 2\alpha_1^2 < 1$ (see Zadrozny (2005)).

[9]The condition for the existence of the fourth moment implies $3\alpha_1^2 < (1-\beta)^2$ so that the denominator $1-\beta^2-2\alpha_1\beta-3\alpha_1^2 > 1-\beta^2-2\alpha_1\beta-1-\beta^2+2\beta = 2\beta(1-\alpha_1-\beta) > 0$.

This expression is well-defined because $0 < \beta < 1$ so that the infinite sum converges. The conditional variance given the infinite past is therefore equal to

$$\mathbb{V}(Z_t|Z_{t-1}, Z_{t-2}, \ldots) = \mathbb{E}\left(Z_t^2|Z_{t-1}, Z_{t-2}, \ldots\right) = \frac{\alpha_0}{1-\beta} + \alpha_1 \sum_{j=0}^{\infty} \beta^j Z_{t-1-j}^2.$$

Thus, the conditional variance depends on the entire history of the time series and not just on Z_{t-1} as in the case of the ARCH(1) model. As all coefficients are assumed to be positive, the clustering of volatility is more persistent than for the ARCH(1) model.

Defining a new time series $\{e_t\}$ by $e_t = Z_t^2 - \sigma_t^2 = (\nu_t^2 - 1)(\alpha_0 + \alpha_1 Z_{t-1}^2 + \beta\sigma_{t-1}^2)$, one can verify that Z_t^2 obeys the stochastic difference equation

$$\begin{aligned} Z_t^2 &= \alpha_0 + \alpha_1 Z_{t-1}^2 + \beta\sigma_{t-1}^2 + e_t = \alpha_0 + \alpha_1 Z_{t-1}^2 + \beta(Z_{t-1}^2 - e_{t-1}) + e_t \\ &= \alpha_0 + (\alpha_1 + \beta)Z_{t-1}^2 + e_t - \beta e_{t-1}. \end{aligned} \tag{8.10}$$

This difference equation defines an ARMA(1,1) process if e_t has finite variance which is the case if the fourth moment of Z_t exists. In this case, it is easy to verify that $\{e_t\}$ is white noise. The so-defined ARMA(1,1) process is causal and invertible with respect to $\{e_t\}$ because $0 < \alpha_1 + \beta < 1$ and $0 < \beta < 1$. The autocorrelation function (ACF), $\rho_{Z^2}(h)$, can be computed using the methods laid out Sect. 2.4. This gives

$$\begin{aligned} \rho_{Z^2}(1) &= \frac{(1-\beta^2-\alpha_1\beta)\alpha_1}{1-\beta^2-2\alpha_1\beta} = \frac{(1-\beta\varphi)(\varphi-\beta)}{1+\beta^2-2\varphi\beta}, \\ \rho_{Z^2}(h) &= (\alpha_1+\beta)\rho_{Z^2}(h-1) = \varphi\rho_{Z^2}(h-1), \qquad h = 2, 3, \ldots \end{aligned} \tag{8.11}$$

with $\varphi = \alpha_1 + \beta$ (see also Bollerslev (1988)).

The IGARCH Model

Practice has shown that the sum $\alpha_1 + \beta$ is often close to one. Thus, it seems interesting to examine the limiting case where $\alpha_1 + \beta = 1$. This model was proposed by Engle and Bollerslev (1986) and was termed the *integrated GARCH* (IGARCH) model in analogy to the notion of integrated processes (see Chap. 7). From Eq. (8.10) we get

$$Z_t^2 = \alpha_0 + Z_{t-1}^2 + e_t - \beta e_{t-1}$$

with $e_t = Z_t^2 - \sigma_t^2 = (\nu_t^2 - 1)(\alpha_0 + (1-\beta)Z_{t-1}^2 + \beta\sigma_{t-1}^2)$. As $\{e_t\}$ is white noise, the squared innovations Z_t^2 behave like a random walk with a MA(1) error term. Although the variance of Z_t becomes infinite, the difference equation still allows for a strictly stationary solution provided that $\mathbb{E}\log(\alpha_1\nu_t^2 + \beta) < 0$ (see Theorem 8.3

and the citations in footnote 7 for further details).[10] It has been shown by Lumsdaine (1986) and Lee and Hansen (1994) that standard inferences can still be applied although $\alpha_1 + \beta = 1$. The model may easily generalized to higher lag orders.

Forecasting

On many occasions it is necessary to obtain forecasts of the conditional variance σ_t^2. An example is given in Sect. 8.4 where the value at risk (VaR) of a portfolio several periods ahead must be evaluated. Denote by $\mathbb{P}_t\sigma_{t+h}^2$ the h period ahead forecast based on information available in period t. We assume that predictions are based on the infinite past. Then the one-period ahead forecast based on Z_t and σ_t^2, respectively ν_t and σ_t^2, is:

$$\mathbb{P}_t\sigma_{t+1}^2 = \alpha_0 + \alpha_1 Z_t^2 + \beta\sigma_t^2 = \alpha_0 + (\alpha_1\nu_t^2 + \beta)\sigma_t^2. \tag{8.12}$$

As $\nu_t \sim \text{IID}(0,1)$ and independent of $Z_{t-j}, j \geq 1$,

$$\mathbb{P}_t\sigma_{t+2}^2 = \alpha_0 + (\alpha_1 + \beta)\mathbb{P}_t\sigma_{t+1}^2.$$

Thus, forecast for $h \geq 2$ can be obtained recursively as follows:

$$\begin{aligned}\mathbb{P}_t\sigma_{t+h}^2 &= \alpha_0 + (\alpha_1 + \beta)\mathbb{P}_t\sigma_{t+h-1}^2 \\ &= \alpha_0\sum_{j=0}^{h-2}(\alpha_1 + \beta)^j + (\alpha_1 + \beta)^{h-1}\mathbb{P}_t\sigma_{t+1}^2.\end{aligned} \tag{8.13}$$

Assuming $\alpha_1 + \beta < 1$, the second term in the above expression vanishes as h goes to infinity. Thus, the contribution of the current conditional variance vanishes when we look further and further into the future. The forecast of the conditional variance then approaches the unconditional one: $\lim_{h\to\infty}\mathbb{P}_t\sigma_{t+h}^2 = \frac{\alpha_0}{1-\alpha_1-\beta}$. If $\alpha_1 + \beta = 1$ as in the IGARCH model, the contribution of the current conditional variance is constant, but diminishes to zero relative to the first term. Finally, if $\alpha_1 + \beta > 1$, the two terms are of the same order and we have a particularly persistent situation.

In practice, the parameters of the model are unknown and have therefore be replaced by an estimate. The method can be easily adapted for higher order models. Instead of using the recursive approach outlined above, it is possible to use simulation methods by drawing repeatedly from the actual empirical distribution of the $\hat{\nu}_t = \hat{Z}_t/\hat{\sigma}_t$. This has the advantage to capture deviations from the underlying distributional assumptions (see Sect. 8.4 for a comparison of both methods). Such methods must be applied if nonlinear models for the conditional variance, like the TARCH model, are employed.

[10] As the variance becomes infinite, the IGARCH process is an example of a stochastic process which is strictly stationary, but not stationary.

8.2 Tests for Heteroskedasticity

Before modeling the volatility of a time series it is advisable to test whether heteroskedasticity is actually present in the data. For this purpose the literature proposed several tests of which we are going to examine two. For both tests the null hypothesis is that there is no heretoskedasticity i.e. that there are no ARCH effects. These tests can also be useful in a conventional regression setting.

8.2.1 Autocorrelation of Quadratic Residuals

The first test is based on the autocorrelation function of squared residuals from a preliminary regression. This preliminary regression or mean regression produces a series $\widehat{Z}_t$ which should be approximately white noise if the equation is well specified. Then we can look at the ACF of the squared residuals $\{\widehat{Z}_t^2\}$ and apply the Ljung-Box test (see Eq. (4.4)). Thus the test can be broken down into three steps.

(i) Estimate an ARMA model for $\{X_t\}$ and retrieve the residuals $\widehat{Z}_t$ from this model. Compute $\widehat{Z}_t^2$. These data can be used to estimate σ^2 as

$$\hat{\sigma}^2 = \frac{1}{T}\sum_{t=1}^{T}\widehat{Z}_t^2$$

Note that the ARMA model should be specified such that the residuals are approximately white noise.

[(ii)] Estimate the ACF for the squared residuals in the usual way:

$$\hat{\rho}_{Z^2}(h) = \frac{\sum_{t=h+1}^{T}\left(\widehat{Z}_t^2 - \hat{\sigma}^2\right)\left(\hat{Z}_{t-h}^2 - \hat{\sigma}^2\right)}{\sum_{t=1}^{T}\left(\widehat{Z}_t^2 - \hat{\sigma}^2\right)^2}$$

(iii) It is now possible to use one of the methods laid out in Chap. 4 to test the null hypothesis that $\{Z_t^2\}$ is white noise. It can be shown that under the null hypothesis $\sqrt{T}\hat{\rho}_{Z^2}(h) \xrightarrow{\mathrm{d}} \mathrm{N}(0,1)$. One can therefore construct confidence intervals for the ACF in the usual way. Alternatively, one may use the Ljung-Box test statistic (see Eq. (4.4)) to test the hypothesis that all correlation coefficients up to order N are simultaneously equal to zero.

$$Q' = T(T+2)\sum_{h=1}^{N}\frac{\hat{\rho}_{Z^2}^2(h)}{T-h}.$$

Under the null hypothesis this statistic is distributed as χ_N^2. To carry out the test, N should be chosen rather high, for example equal to $T/4$.

8.2.2 Engle's Lagrange-Multiplier Test

Engle (1982) proposed a Lagrange-Multiplier test. This test rests on an ancillary regression of the squared residuals against a constant and lagged values of $\widehat{Z}_{t-1}^2, \widehat{Z}_{t-2}^2, \ldots, \widehat{Z}_{t-p}^2$ where the $\{\widehat{Z}_t\}$ is again obtained from a preliminary regression. The auxiliary regression thus is

$$\widehat{Z}_t^2 = \alpha_0 + \alpha_1 \widehat{Z}_{t-1}^2 + \alpha_2 \widehat{Z}_{t-2}^2 + \ldots + \alpha_p \widehat{Z}_{t-p}^2 + \varepsilon_t,$$

where ε_t denotes the error term. Then the null hypothesis $\mathrm{H}_0 : \alpha_1 = \alpha_2 = \ldots = \alpha_p = 0$ is tested against the alternative hypothesis $\mathrm{H}_1 : \alpha_j \neq 0$ for at least one j. As a test statistic one can use the coefficient of determination times T, i.e. $T\mathrm{R}^2$. This test statistic is distributed as a χ^2 with p degrees of freedom. Alternatively, one may use the conventional F-test.

8.3 Estimation of GARCH(p,q) Models

8.3.1 Maximum-Likelihood Estimation

The literature has proposed several approaches to estimate models of volatility (see Fan and Yao (2003, 156–162)). The most popular one, however, rest on the method of maximum-likelihood. We will describe this method using the GARCH(p,q) model. Related and more detailed accounts can be found in Weiss (1986), Bollerslev et al. (1994) and Hall and Yao (2003).

In particular we consider the following model:

mean equation: $$X_t = c + \phi_1 X_{t-1} + \ldots + \phi_r X_{t-r} + Z_t,$$

where

$$Z_t = \nu_t \sigma_t \qquad \text{with } \nu_t \sim \mathrm{IIDN}(0,1) \text{ and}$$

variance equation: $$\sigma_t^2 = \alpha_0 + \sum_{j=1}^{p} \alpha_j Z_{t-j}^2 + \sum_{j=1}^{q} \beta_j \sigma_{t-j}^2.$$

The mean equation represents a simple AR(r) process for which we assume that it is causal with respect to $\{Z_t\}$, i.e. that all roots of $\Phi(z)$ are outside the unit circle. The method demonstrated here can be easily generalized to ARMA processes or even ARMA process with additional exogenous variables (so-called ARMAX processes) as noted by Weiss (1986). The method also incorporates the ARCH-in-mean model (see equation (8.6)) which allows for an effect of the conditional variance σ_t on X_t.

In addition, we assume that the coefficients of the variance equation are all positive, that $\sum_{j=1}^{p} \alpha_j + \sum_{j=1}^{q} \beta_j < 1$ and that $\mathbb{E}Z_t^4 < \infty$ exists.[11]

As ν_t is identically and independently standard normally distributed, the distribution of X_t conditional on $\mathcal{X}_{t-1} = \{X_{t-1}, X_{t-2}, \ldots\}$ is normal with mean $c + \phi_1 X_{t-1} + \ldots + \phi_r X_{t-r}$ and variance σ_t^2. The conditional density, $f(X_t|\mathcal{X}_{t-1})$, therefore is:

$$f\left(X_t|\mathcal{X}_{t-1}\right) = \frac{1}{\sqrt{2\pi\sigma_t^2}} \exp\left(-\frac{Z_t^2}{2\sigma_t^2}\right)$$

where Z_t equals $X_t - c - \phi_1 X_{t-1} - \ldots - \phi_r X_{t-r}$ and σ_t^2 is given by the variance equation.[12] The joint density $f(X_1, X_2, \ldots, X_T)$ of a random sample $(X_1, X_2, \ldots, X_T)$ can therefore be factorized as

$$f(X_1, X_2, \ldots, X_T) = f(X_1, X_2, \ldots, X_{s-1}) \prod_{t=s}^{T} f(X_t|\mathcal{X}_{t-1})$$

where s is an integer greater than p. The necessity, not to factorize the first $s-1$ observations, relates to the fact that σ_t^2 can only be evaluated for $s > p$ in the ARCH(p) model. For the ARCH(p) model s can be set to $p+1$. In the case of a GARCH model σ_t^2 is given by weighted infinite sum of the $Z_{t-1}^2, Z_{t-2}^2, \ldots$ (see the expression (8.9) for σ_t^2 in the GARCH(1,1) model). For finite samples, this infinite sum must be approximated by a finite sum of s summands such that the numbers of summands s is increasing with the sample size.(see Hall and Yao (2003)).

We then merge all parameters of the model as follows: $\phi = (c, \phi_1, \ldots, \phi_r)'$, $\alpha = (\alpha_0, \alpha_1, \ldots, \alpha_p)'$ and $\beta = (\beta_1, \ldots, \beta_q)'$. For a given realization $x = (x_1, x_2, \ldots, x_T)$ the *likelihood function* conditional on x, $\mathrm{L}(\phi, \alpha, \beta|x)$, is defined as

$$\mathrm{L}(\phi, \alpha, \beta|x) = f(x_1, x_2, \ldots, x_{s-1}) \prod_{t=s}^{T} f(x_t|\mathcal{X}_{t-1})$$

where in $\mathcal{X}_{t-1}$ the random variables are replaced by their realizations. The likelihood function can be seen as the probability of observing the data at hand given the values for the parameters. The method of maximum likelihood then consist in choosing the parameters (ϕ, α, β) such that the likelihood function is maximized. Thus we chose the parameter so that the probability of observing the data is maximized. In this way

[11] The existence of the fourth moment is necessary for the asymptotic normality of the maximum-likelihood estimator, but not for the consistence. It is possible to relax this assumption somewhat (see Hall and Yao (2003)).

[12] If ν_t is assumed to follow another distribution than the normal, one may use this distribution instead.

we obtain the maximum likelihood estimator. Taking the first s realizations as given deterministic starting values, we then get the *conditional likelihood function*.

In practice we do not maximize the likelihood function but the logarithm of it where we take $f(x_1, \ldots, x_{s-1})$ as a fixed constant which can be neglected in the optimization:

$$\log \mathrm{L}(\phi, \alpha, \beta | x) = \sum_{t=s}^{T} \log f\,(x_t | \mathcal{X}_t)$$

$$= -\frac{T}{2} \log 2\pi - \frac{1}{2} \sum_{t=s}^{T} \log \sigma_t^2 - \frac{1}{2} \sum_{t=s}^{T} \frac{z_t^2}{\sigma_t^2}$$

where $z_t = x_t - c - \phi_1 x_{t-1} - \ldots - \phi_r x_{t-r}$ denotes the realization of Z_t. The maximum likelihood estimator is obtained by maximizing the likelihood function over the admissible parameter space. Usually, the implementation of the stationarity condition and the condition for the existence of the fourth moment turns out to be difficult and cumbersome so that often these conditions are neglected and only checked in retrospect or some ad hoc solutions are envisaged. It can be shown that the (conditional) maximum likelihood estimator leads to asymptotically normally distributed estimates.[13] The maximum likelihood estimator remains meaningful even when $\{\nu_t\}$ is not normally distributed. In this case the quasi maximum likelihood estimator is obtained (see Hall and Yao (2003) and Fan and Yao (2003)).

For numerical reasons it is often convenient to treat the mean equation and the variance equation separately. As the mean equation is a simple AR(r) model, it can be estimated by ordinary least-squares (OLS) in a first step. This leads to consistent parameter estimates. However, due to the heteroskedasticity, this no longer true for the covariance matrix of the coefficients so that the usual t- and F-tests are nor reliable. This problem can be circumvented by the use of the White correction (see White (1980)). In this way it is possible to find an appropriate specification for the mean equation without having to estimate the complete model. In the second step, one can then work with the residuals to find an appropriate ARMA model for the squared residuals. This leads to consistent estimates of the parameters of the variance equation. These estimates are under additional weakly assumptions asymptotically normally distributed (see Weiss (1986)). It should, however, be noted that this way of proceeding is, in contrast to the maximum likelihood estimator, not efficient because it neglect the nonlinear character of the GARCH model. The parameters found in this way can, however, serve as meaningful starting values for the numerical maximization procedure which underlies the maximum likelihood estimation.

[13] Jensen and Rahbek (2004) showed that, at least for the GARCH(1,1) case, the stationarity condition is not necessary.

A final remark concerns the choice of the parameter r, p and q. Similarly to the ordinary ARMA models, one can use information criteria such as the Akaike or the Bayes criterion, to determine the order of the model (see Sect. 5.4).

8.3.2 Method of Moment Estimation

The maximization of the likelihood function requires the use of numerical optimization routines. Depending on the routine actually used and on the starting value, different results may be obtained if the likelihood function is not well-behaved. It is therefore of interest to have alternative estimation methods at hand. The method of moments is such an alternative. It is similar to the Yule-Walker estimator (see Sect. 5.1) applied to the autocorrelation function of $\{Z_t^2\}$. This method not only leads to an analytic solution, but can also be easily implemented. Following Kristensen and Linton (2006), we will illustrate the method for the GARCH(1,1) model.

Equation (8.11) applied to $\rho_{Z^2}(1)$ and $\rho_{Z^2}(2)$ constitutes a nonlinear equation system in the unknown parameters β and α_1. This system can be reparameterized to yield an equation system in $\varphi = \alpha_1 + \beta$ and β which can be reduced to a single quadratic equation in β:

$$\beta^2 - b\beta - 1 = 0 \quad \text{where } b = \frac{\varphi^2 + 1 - 2\rho_{Z^2}(1)\varphi}{\varphi - \rho_{Z^2}(1)}.$$

The parameter b is well-defined because $\varphi = \alpha_1 + \beta \geq \rho_{Z^2}(1)$ with equality only if $\beta = 0$. In the following we will assume that $\beta > 0$. Under this assumption $b > 2$ so that the only solution with the property $0 < \beta < 1$ is given by

$$\beta = \frac{b - \sqrt{b^2 - 4}}{2}.$$

The moment estimator can therefore be constructed as follows:

(i) Estimate the correlations $\rho_{Z^2}(1)$ and $\rho_{Z^2}(2)$ and σ^2 based on the formulas (8.11) in Sect. 8.2.

(ii) An estimate for $\varphi = \alpha_1 + \beta$ is then given by

$$\hat{\varphi} = \widehat{(\alpha_1 + \beta)} = \frac{\hat{\rho}_{Z^2}(2)}{\hat{\rho}_{Z^2}(1)}.$$

(iii) use the estimate $\hat{\varphi}$ to compute an estimate for b:

$$\hat{b} = \frac{\hat{\varphi}^2 + 1 - 2\hat{\rho}_{Z^2}(1)\hat{\varphi}}{\hat{\varphi} - \hat{\rho}_{Z^2}(1)}.$$

The estimate $\hat{\beta}$ for β is then

$$\hat{\beta} = \frac{\hat{b} - \sqrt{\hat{b}^2 - 4}}{2}.$$

(iv) The estimate for α_1 is $\hat{\alpha}_1 = \hat{\varphi} - \hat{\beta}$. Because $\alpha_0 = \sigma^2(1-(\alpha_1+\beta))$, the estimate for α_0 is equal to $\hat{\alpha}_0 = \hat{\sigma}^2(1-\hat{\varphi})$.

Kristensen and Linton (2006) show that, given the existence of the fourth moment of Z_t, this method of moment leads to consistent and asymptotically normal distributed estimates. These estimates may then serve as starting values for the maximization of the likelihood function to improve efficiency.

8.4 Example: Swiss Market Index (SMI)

In this section, we will illustrate the methods discussed previously to analyze the volatility of the Swiss Market Index (SMI). The SMI is the most important stock market index for Swiss blue chip companies. It is constructed solely from stock market prices, dividends are not accounted for. The data are the daily values of the index between the 3rd of January 1989 and the 13th of February 2004. Figure 1.5 shows a plot of the data. Instead of analyzing the level of the SMI, we will investigate the daily return computed as the logged difference. This time series is denoted by X_t and plotted in Fig. 8.3. One can clearly discern phases of high (observations around $t = 2500$ and $t = 3500$) and low ($t = 1000$ and $t = 2000$) volatility. This represents a first sign of heteroskedasticity and positively correlated volatility.

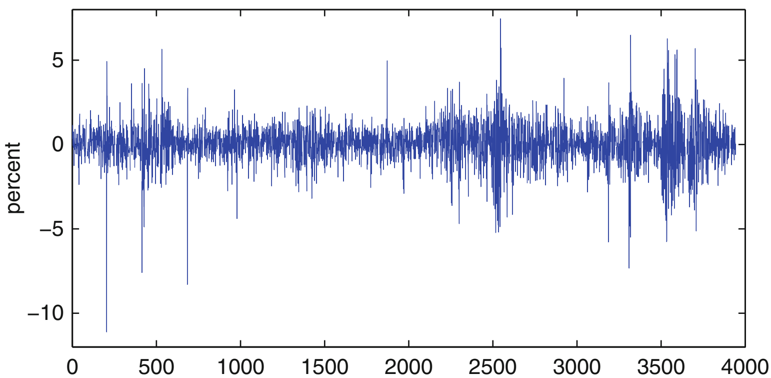

Fig. 8.3 Daily return of the SMI (Swiss Market Index) computed as $\Delta \log(\text{SMI}_t)$ between January 3rd 1989 and February 13th 2004

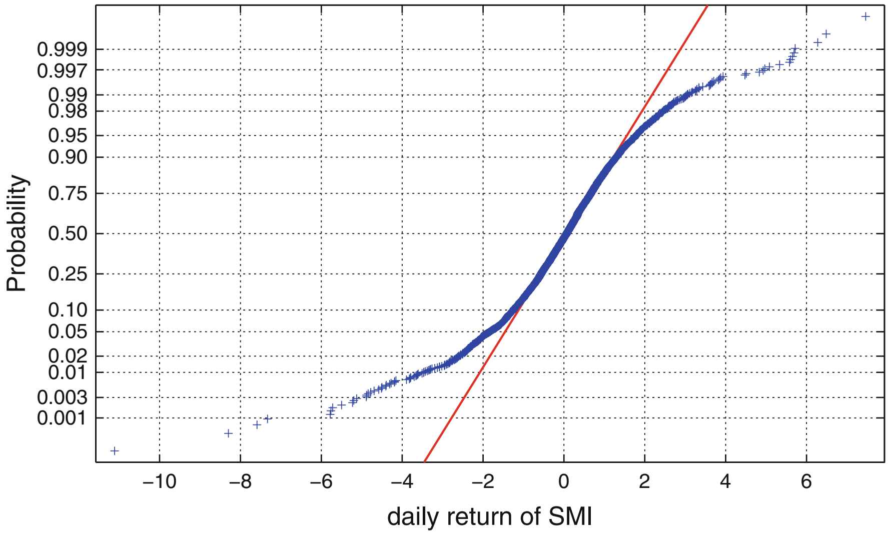

Fig. 8.4 Normal-Quantile Plot of the daily returns of the SMI (Swiss Market Index)

Figure 8.4 shows a normal-quantile plot to compare the empirical distribution of the returns with those from a normal distribution. This plot clearly demonstrates that the probability of observing large returns is bigger than warranted from a normal distribution. Thus the distribution of returns exhibits the heavy-tail property. A similar argument can be made by comparing the histogram of the returns and the density of a normal distribution with the same mean and the same variance. shown in Fig. 8.5. Again one can seen that absolutely large returns are more probable than expected from a normal distribution. Moreover, the histogram shows no obvious sign for an asymmetric distribution, but a higher peakedness.

After the examination of some preliminary graphical devices, we are going to analyze the autocorrelation functions of $\{X_t\}$ and $\{X_t^2\}$. Figure 8.6 shows the estimated ACFs. The estimated ACF of $\{X_t\}$ shows practically no significant autocorrelation so that we can consider $\{X_t\}$ be approximately white noise. The corresponding Ljung-Box statistic with $L = 100$, however, has a value of 129.62 which is just above the 5 % critical value of 124.34. Thus there is some sign of weak autocorrelation. This feature is not in line with efficiency of the Swiss stock market (see Campbell et al. (1997)). The estimated ACF of X_t^2 is clearly outside the 95 % confidence interval for at least up to order 20. Thus we can reject the hypothesis of homoskedasticity in favor of heteroskedasticity. This is confirmed by the Ljung-Box statistic with $L = 100$ with a value of 2000.93 which is much higher than the critical value of 124.34.

After these first investigations, we want to find an appropriate model for the mean equation. We will use OLS with the White-correction. It turns out that a MA(1) model fits the data best although an AR(1) model leads to almost the same results. In the next step we will estimate a GARCH(p,q) model with the method of maximum

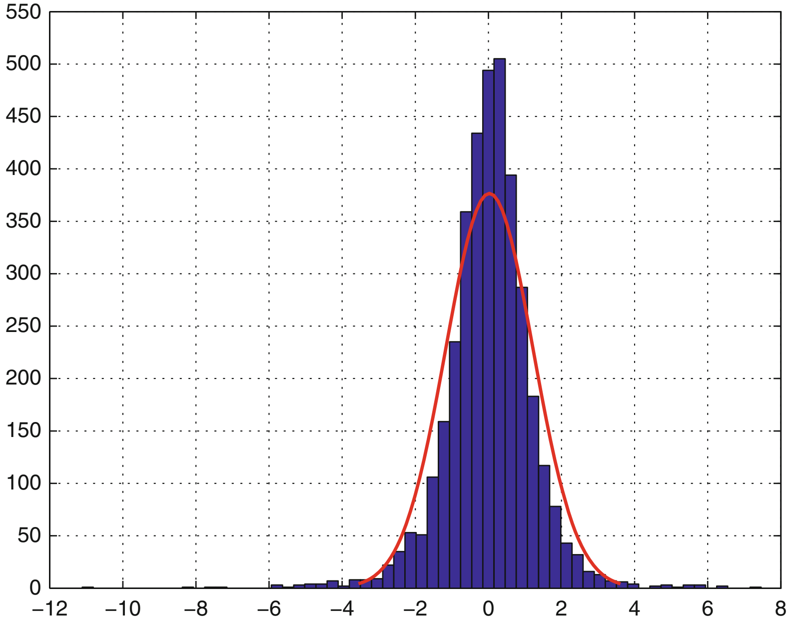

Fig. 8.5 Histogram of the daily returns of the SMI (Swiss Market Index) and the density of a fitted normal distribution (*red line*)

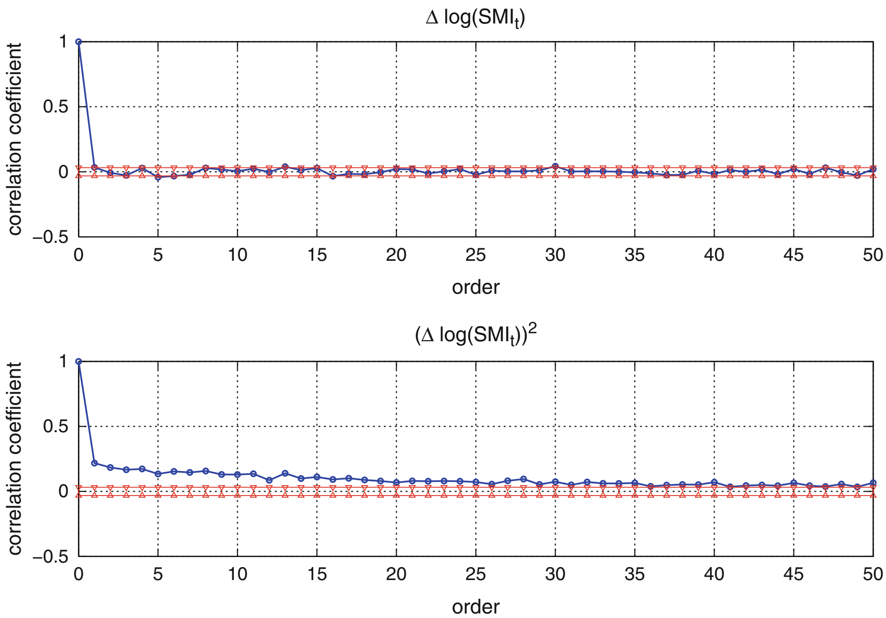

Fig. 8.6 ACF of the daily returns and the squared daily returns of the SMI

Table 8.1 AIC criterion for the variance equation in the GARCH(p,q) model

	q			
p	0	1	2	3
1	3.0886	2.9491	2.9491	2.9482
2	3.0349	2.9496	2.9491	2.9486
3	2.9842	2.9477	2.9472	**2.9460**

Minimum value in bold

Table 8.2 BIC criterion for the variance equation in the GARCH(p,q) model

	q			
p	0	1	2	3
1	3.0952	**2.9573**	2.9590	2.9597
2	3.0431	2.9595	2.9606	2.9617
3	2.9941	2.9592	2.9604	2.9607

Minimum value in bold

likelihood where p is varied between p 1 and 3 and q between 0 and 3. The values of the AIC, respectively BIC criterion corresponding to the variance equation are listed in Tables 8.1 and 8.2.

The results reported in these tables show that the AIC criterion favors a GARCH(3,3) model corresponding to the bold number in Table 8.1 whereas the BIC criterion opts for a GARCH(1,1) model corresponding to the bold number in Table 8.2. It also turns out that high dimensional models, in particular those for which $q > 0$, the maximization algorithm has problems to find an optimum. Furthermore, the roots of the implicit AR and the MA polynomial corresponding to the variance equation of the GARCH(3,3) model are very similar. These two arguments lead us to prefer the GARCH(1,1) over the GARCH(3,3) model. This model was estimated to have the following mean equation:

$$X_t = \underset{(0.0174)}{0.0755} + Z_t + \underset{(0.0184)}{0.0484} Z_{t-1}$$

with the corresponding variance equation

$$\sigma_t^2 = \underset{(0.0046)}{0.0765} + \underset{(0.0095)}{0.1388} Z_{t-1}^2 + \underset{(0.0099)}{0.8081} \sigma_{t-1}^2,$$

where the estimated standard deviations are reported below the corresponding coefficient estimate. The small, but significant value of 0.0484 for the MA(1) coefficient shows that there is a small but systematic correlation of the returns from one day to the next. The coefficients of the GARCH model are all positive and their sum $\alpha_1 + \beta = 0.1388 + 0.8081 = 0.9469$ is statistically below one so that all conditions for a stationary process are fulfilled.[14] Because $\sqrt{3}\frac{\alpha_1}{1-\beta} = \sqrt{3}\frac{0.1388}{1-0.8081} = 1.2528 > 1$, the condition for the existence of the fourth moment of Z_t is violated.

[14]The corresponding Wald test clearly rejects the null hypothesis $\alpha_1 + \beta = 1$ at a significance level of 1 %.

As a comparison we also estimate the GARCH(1,1) model using the methods of moments. First we estimate a MA(1) model for $\Delta \log SMI$. This results in an estimate $\hat{\theta} = 0.034$ (compare this with the ML estimate). The squared residuals have correlation coefficients

$$\hat{\rho}_{Z^2}(1) = 0.228 \quad \text{and} \quad \hat{\rho}_{Z^2}(2) = 0.181.$$

The estimate of b therefore is $\hat{b} = 2.241$ which leads to an estimate of β equal to $\hat{\beta} = 0.615$. This finally results in the estimates of α_1 and α_0 equal to $\hat{\alpha}_1 = 0.179$ and $\hat{\alpha}_0 = 0.287$ with an estimate for σ^2 equal to $\hat{\sigma}^2 = 1.391$. Thus these estimates are quite different from those obtained by the ML method.

Value at Risk

We are now in a position to use our ML estimates to compute the *Value-at-risk* (VaR). The VaR is a very popular measure to estimate the risk of an investment. In our case we consider the market portfolio represented by the stocks in the SMI. The VaR is defined as the maximal loss (in absolute value) of an investment which occurs with probability α over a time horizon h. Thus a 1 % VaR for the return on the SMI for the next day is the threshold value of the return such that one can be confident with 99 % that the loss will not exceed this value. Thus the α VaR at time t for h periods, $\mathrm{VaR}^{\alpha}_{t,t+h}$, is nothing but the α-quantile of the distribution of the forecast of the return in h periods given information X_{t-k}, $k = 0, 1, 2, \ldots$. Formally, we have:

$$\mathrm{VaR}^{\alpha}_{t,t+h} = \inf\left\{x : \mathbf{P}\left[\widetilde{X}_{t+h} \leq x | X_t, X_{t-1}, \ldots\right] \geq \alpha\right\},$$

where $\widetilde{X}_{t+h}$ is the return of the portfolio over an investment horizon of h periods. This return is approximately equal to the sum of the daily returns: $\widetilde{X}_{t+h} = \sum_{j=1}^{h} X_{t+j}$.

The one period forecast error is given by $X_{t+1} - \widetilde{\mathbb{P}}_t X_{t+1}$ which is equal to $Z_{t+1} = \sigma_{t+1}\nu_{t+1}$. Thus the VaR for the next day is

$$\mathrm{VaR}^{\alpha}_{t,t+1} = \inf\left\{x : \mathbf{P}\left[\nu_{t+1} \leq \frac{x - \widetilde{\mathbb{P}}_t X_{t+1}}{\sigma_{t+1}}\right] \geq \alpha\right\}.$$

This entity can be computed by replacing the forecast given the infinite past, $\widetilde{\mathbb{P}}_t X_{t+1}$, by a forecast given the finite sample information X_{t-k}, $k = 0, 1, 2, \ldots, t-1$, $\mathbb{P}_t X_{t+1}$, and by substituting σ_{t+1} by the corresponding forecast from variance equation, $\hat{\sigma}_{t,t+1}$. Thus we get:

$$\widehat{\mathrm{VaR}^{\alpha}_{t,t+1}} = \inf\left\{x : \mathbf{P}\left[\nu_{t+1} \leq \frac{x - \mathbb{P}_t X_{t+1}}{\hat{\sigma}_{t,t+1}}\right] \geq \alpha\right\}.$$

Table 8.3 One percent VaR for the next day of the return to the SMI according to the ARMA(0,1)-GARCH(1,1) model

Date	$\mathbb{P}_t X_{t+1}$	$\hat{\sigma}^2_{t,t+1}$	VaR ($\widehat{\text{VaR}}^{0.01}_{t,t+1}$) Parametric	Non-parametric
31.12.2001	0.28	6.61	5.71	6.30
5.2.2002	−0.109	6.80	6.19	6.79
24.7.2003	0.0754	0.625	1.77	1.95

Table 8.4 One percent VaR for the next 10 days of the return to the SMI according to the ARMA(0,1) -GARCH(1,1) model

Date	$\mathbb{P}_t \widetilde{X}_{t+1}$	VaR ($\widehat{\text{VaR}}^{0.01}_{t,t+10}$) Parametric	Non-parametric
31.12.2001	0.84	18.39	22.28
5.2.2002	0.65	19.41	21.53
24.7.2003	0.78	6.53	7.70

The computation of $\widehat{\text{VaR}}^{\alpha}_{t,t+1}$ requires to determine the α-quantile of the distribution of ν_t. This can be done in two ways. The first one uses the assumption about the distribution of ν_t explicitly. In the simplest case, ν_t is distributed as a standard normal so that the appropriate quantile can be easily retrieved. The 1 % quantile for the standard normal distribution is −2.33. The second approach is a non-parametric one and uses the empirical distribution function of $\hat{\nu}_t = \hat{Z}_t/\hat{\sigma}_t$ to determine the required quantile. This approach has the advantage that deviations from the standard normal distribution are accounted for. In our case, the 1 % quantile is −2.56 and thus considerably lower than the −2.33 obtained from the normal distribution. Thus the VaR is under estimated by using the assumption of the normal distribution.

The corresponding computations for the SMI based on the estimated ARMA(0,1)-GARCH(1,1) model are reported in Table 8.3. A value of 5.71 for 31st of December 2001 means that one can be 99 % sure that the return of an investment in the stocks of the SMI will not be lower than −5.71 %. The values for the non-parametric approach are typically higher. The comparison of the VaR for different dates clearly shows how the risk evolves over time.

Due to the nonlinear character of the model, the VaR for more than one day can only be gathered from simulating the one period returns over the corresponding horizon. Starting from a given date 10'000 realizations of the returns over the next 10 days have been simulated whereby the corresponding values for ν_t are either drawn from a standard normal distribution (parametric case) or from the empirical distribution function of $\hat{\nu}_t$ (non-parametric case). The results from this exercise are reported in Table 8.4. Obviously, the risk is much higher for a 10 day than for a one day investment. Alternatively, one may use the forecasting equation (8.12) and the corresponding recursion formula (8.13).