4 Regime-switching models for volatility

Uncertainty, or risk, is of paramount importance in financial analysis. For example, the Capital Asset-Pricing Model (CAPM) (Sharpe, 1964; Lintner, 1965; Mossin, 1966; Merton, 1973) postulates a direct relationship between the required return on an asset and its risk, where the latter is determined by the covariance of the returns on the particular asset and some benchmark portfolio. Similarly, the most important determinant of the price of an option is the uncertainty associated with the price of the underlying asset, as measured by its volatility.

One of the most prominent stylized facts of returns on financial assets is that their volatility changes over time. In particular, periods of large movements in prices alternate with periods during which prices hardly change (see section 1.2). This characteristic feature commonly is referred to as *volatility clustering*. Even though the time-varying nature of the volatility of financial assets has long been recognized (see Mandelbrot, 1963a, 1963b, 1967; Fama, 1965), explicit modelling of the properties of the volatility process has been taken up only fairly recently.

In this chapter we discuss (extensions of) the class of (Generalized) Autoregressive Conditional Heteroscedasticity ((G)ARCH) models, introduced by Engle (1982) and Bollerslev (1986). Nowadays, models from the GARCH class are the most popular volatility models among practitioners. GARCH models enjoy such popularity because they are capable of describing not only the feature of volatility clustering, but also certain other characteristics of financial time series, such as their pronounced excess kurtosis or fat-tailedness. Still, the standard GARCH model cannot capture other empirically relevant properties of volatility. For example, since Black (1976), negative shocks or news are believed to affect volatility quite differently than positive shocks of equal size (see also section 1.2). In the standard GARCH model, however, the effect of a shock on volatility depends only on its size. The sign of the shock is irrelevant. Another limitation of the standard GARCH model is that it does not imply that expected returns and volatility are related directly, as is the case in the CAPM.

Over the past few years, quite a few nonlinear variants of the basic GARCH model have been proposed, most of them designed to capture such aspects as the asymmetric effect of positive and negative shocks on volatility, and possible correlation between the return and volatility.

The outline of this chapter is as follows. In section 4.1, we discuss representations of the basic GARCH model and several nonlinear extensions. We emphasize which of the stylized facts of returns on financial assets can and cannot be captured by the various models. Testing for GARCH is the subject of section 4.2. We discuss tests for the standard GARCH model and for its nonlinear variants, and we examine the influence of outliers on the various test-statistics. Estimation of ARCH models is discussed in section 4.3. In section 4.4 various diagnostic checks which can be used to evaluate estimated GARCH models are reviewed. In section 4.5 we focus on out-of-sample forecasting: both the consequences for forecasting the conditional mean in the presence of ARCH, as well as forecasting volatility itself are discussed. Measures of persistence of shocks in GARCH models are discussed in section 4.6; we emphasize the role these various elements play in the empirical specification of ARCH models. The final section of this chapter (section 4.7) contains a brief discussion on multivariate GARCH models.

We should remark that the aim of this chapter is not to provide a complete account of the vast literature on GARCH models, but rather to provide an introduction to this area, with a specific focus on asymmetric GARCH models and the impact of outliers. For topics not covered in this chapter, the interested reader should consult one of the many surveys on GARCH models which have appeared in recent years. Bollerslev, Chou and Kroner (1992) provide a comprehensive overview of empirical applications of GARCH models to financial time series. Bollerslev, Engle and Nelson (1994) focus on the theoretical aspects of GARCH models. Gourieroux (1997) discusses in great detail how GARCH models can be incorporated in financial decision problems such as asset-pricing and portfolio management. Additional reviews of GARCH and related models can be found in Bera and Higgins (1993); Diebold and Lopez (1995); Pagan (1996); Palm (1996); and Shephard (1996).

4.1 Representation

As stated in section 2.1, an observed time series y_t can be written as the sum of a predictable and an unpredictable part,

$$y_t = \mathrm{E}[y_t|\Omega_{t-1}] + \varepsilon_t, \tag{4.1}$$

where Ω_{t-1} is the information set consisting of all relevant information up to and including time $t-1$. In previous chapters we have concentrated on different specifications for the predictable part or conditional mean $\mathrm{E}[y_t|\Omega_{t-1}]$, while simply assuming that the unpredictable part or shock ε_t satisfies the white noise properties (2.1)–(2.3). In particular, ε_t was assumed to be both unconditionally and conditionally homoscedastic – that is, $\mathrm{E}[\varepsilon_t^2] = \mathrm{E}[\varepsilon_t^2|\Omega_{t-1}] = \sigma^2$ for all t. Here we relax part of this assumption and allow the *conditional variance* of ε_t to vary over time – that is, $\mathrm{E}[\varepsilon_t^2|\Omega_{t-1}] = h_t$ for some nonnegative function $h_t \equiv h_t(\Omega_{t-1})$. Put differently, ε_t is *conditionally heteroscedastic*. A convenient way to express this in general is

$$\varepsilon_t = z_t\sqrt{h_t}, \tag{4.2}$$

where z_t is independent and identically distributed with zero mean and unit variance. For convenience, we assume that z_t has a standard normal distribution throughout this section. Some remarks on this assumption are made at the end of this section.

From (4.2) and the properties of z_t it follows immediately that the distribution of ε_t conditional upon the history Ω_{t-1} is normal with mean zero and variance h_t. Also note that the *unconditional variance* of ε_t is still assumed to be constant. Using the law of iterated expectations,

$$\sigma^2 \equiv \mathrm{E}[\varepsilon_t^2] = \mathrm{E}[\mathrm{E}[\varepsilon_t^2|\Omega_{t-1}]] = \mathrm{E}[h_t]. \tag{4.3}$$

Hence, we assume that the unconditional expectation of h_t is constant.

To complete the model, we need to specify how the conditional variance of ε_t evolves over time. In this section, we discuss the representation of various linear and nonlinear models for h_t. The properties of the resultant time series ε_t are used to see whether these models can capture (some of) the stylized facts of stock and exchange rate returns.

4.1.1 Linear GARCH models

Engle (1982) introduced the class of AutoRegressive Conditionally heteroscedastic (ARCH) models to capture the volatility clustering of financial time series (even though the first empirical applications did not deal with high-frequency financial data). In the basic ARCH model, the conditional variance of the shock that occurs at time t is a linear function of the squares of past shocks. For example, in the ARCH model of order 1, h_t is specified as

$$h_t = \omega + \alpha_1\varepsilon_{t-1}^2. \tag{4.4}$$

Obviously, the (conditional) variance h_t needs to be nonnegative. In order to guarantee that this is the case for the ARCH(1) model, the parameters in (4.4) have to satisfy the conditions $\omega > 0$ and $\alpha_1 \geq 0$. Where $\alpha_1 = 0$, the conditional variance is constant and, hence, the series ε_t is conditionally homoscedastic.

To understand why the ARCH model can describe volatility clustering, observe that model (4.2) with (4.4) basically states that the conditional variance of ε_t is an increasing function of the square of the shock that occurred in the previous time period. Therefore, if ε_{t-1} is large (in absolute value), ε_t is expected to be large (in absolute value) as well. In other words, large (small) shocks tend to be followed by large (small) shocks, of either sign.

An alternative way to see the same thing is to note that the ARCH(1) model can be rewritten as an AR(1) model for ε_t^2. Adding ε_t^2 to (4.4) and subtracting h_t from both sides gives

$$\varepsilon_t^2 = \omega + \alpha_1 \varepsilon_{t-1}^2 + \nu_t, \tag{4.5}$$

where $\nu_t \equiv \varepsilon_t^2 - h_t = h_t(z_t^2 - 1)$. Notice that $\mathrm{E}[\nu_t | \Omega_{t-1}] = 0$. Using the theory for AR models summarized in chapter 2, it follows that (4.5) is covariance-stationary if $\alpha_1 < 1$. In that case the unconditional mean of ε_t^2, or the unconditional variance of ε_t, can be obtained as

$$\sigma^2 \equiv \mathrm{E}[\varepsilon_t^2] = \frac{\omega}{1 - \alpha_1}. \tag{4.6}$$

Furthermore, (4.5) can be rewritten as

$$\begin{aligned} \varepsilon_t^2 &= (1 - \alpha_1)\frac{\omega}{1 - \alpha_1} + \alpha_1 \varepsilon_{t-1}^2 + \nu_t \\ &= (1 - \alpha_1)\sigma^2 + \alpha_1 \varepsilon_{t-1}^2 + \nu_t \\ &= \sigma^2 + \alpha_1(\varepsilon_{t-1}^2 - \sigma^2) + \nu_t. \end{aligned} \tag{4.7}$$

Assuming that $0 \leq \alpha_1 < 1$, (4.7) shows that if ε_{t-1}^2 is larger (smaller) than its unconditional expected value σ^2, ε_t^2 is expected to be larger (smaller) than σ^2 as well.

The ARCH model cannot only capture the volatility clustering of financial data, but also their excess kurtosis. From (4.2) it can be seen that the kurtosis of ε_t always exceeds the kurtosis of z_t,

$$\mathrm{E}[\varepsilon_t^4] = \mathrm{E}[z_t^4]\mathrm{E}[h_t^2] \geq \mathrm{E}[z_t^4]\mathrm{E}[h_t]^2 = \mathrm{E}[z_t^4]\mathrm{E}[\varepsilon_t^2]^2, \tag{4.8}$$

which follows from Jensen's inequality. As shown by Engle (1982), for the ARCH(1) model with normally distributed z_t the kurtosis of ε_t is equal to

$$\mathrm{K}_\varepsilon = \frac{\mathrm{E}[\varepsilon_t^4]}{\mathrm{E}[\varepsilon_t^2]^2} = \frac{3(1 - \alpha_1^2)}{1 - 3\alpha_1^2}, \tag{4.9}$$

which is finite if $3\alpha_1^2 < 1$. Clearly, K_ε is always larger than the normal value of 3.

Another characteristic of the ARCH(1) model which is worthwhile noting is the implied autocorrelation function for the squared shocks ε_t^2. From the AR(1) representation in (4.5), it follows that the kth order autocorrelation of ε_t^2 is equal to α_1^k. In figures 2.1 and 2.2 it could be seen that the first-order autocorrelation of squared stock and exchange returns generally is quite small, while the subsequent decay is very slow. The small first-order autocorrelation would imply a small value of α_1 in the ARCH(1) model, but this in turn would imply that the autocorrelations would become close to zero quite quickly. Thus it appears that the ARCH(1) model cannot describe the two characteristic features of the empirical autocorrelations of the returns series simultaneously.

To cope with the extended persistence of the empirical autocorrelation function, one may consider generalizations of the ARCH(1) model. One possibility to allow for more persistent autocorrelations is to include additional lagged squared shocks in the conditional variance function. The general ARCH(q) model is given by

$$h_t = \omega + \alpha_1 \varepsilon_{t-1}^2 + \alpha_2 \varepsilon_{t-2}^2 + \cdots + \alpha_q \varepsilon_{t-q}^2. \tag{4.10}$$

To guarantee nonnegativeness of the conditional variance, it is required that $\omega > 0$ and $\alpha_i \geq 0$ for all $i = 1, \ldots, q$. The ARCH(q) model can be rewritten as an AR(q) model for ε_t^2 in exactly the same fashion as writing (4.4) as (4.5), that is,

$$\varepsilon_t^2 = \omega + \alpha_1 \varepsilon_{t-1}^2 + \alpha_2 \varepsilon_{t-2}^2 + \cdots + \alpha_q \varepsilon_{t-q}^2 + \nu_t. \tag{4.11}$$

It follows that the unconditional variance of ε_t is equal to

$$\sigma^2 = \frac{\omega}{1 - \alpha_1 - \cdots - \alpha_q}, \tag{4.12}$$

while the ARCH(q) model is covariance-stationary if all roots of the lag polynomial $1 - \alpha_1 L - \cdots - \alpha_q L^q$ are outside the unit circle. Milhøj (1985) derives conditions for the existence of unconditional moments of ARCH(q) processes.

To capture the dynamic patterns in conditional volatility adequately by means of an ARCH(q) model, q often needs to be taken quite large. It turns out that it can be quite cumbersome to estimate the parameters in such a model, because of the nonnegativity and stationarity conditions that need to be imposed. To reduce the computational problems, it is common to impose some structure on the parameters in the ARCH(q) model, such as $\alpha_i = \alpha(q+1-i)/(q(q+1)/2)$, $i = 1, \ldots, q$, which implies that the parameters of the lagged squared shocks decline linearly and sum to α (see Engle, 1982, 1983). As an alternative solution, Bollerslev (1986) suggested adding lagged conditional variances to the ARCH model instead. For example, adding h_{t-1} to the ARCH(1) model (4.4) results in the Generalized ARCH (GARCH) model of order (1,1)

$$h_t = \omega + \alpha_1 \varepsilon_{t-1}^2 + \beta_1 h_{t-1}. \tag{4.13}$$

The parameters in this model should satisfy $\omega > 0$, $\alpha_1 > 0$ and $\beta_1 \geq 0$ to guarantee that $h_t \geq 0$, while α_1 must be strictly positive for β_1 to be identified (see also (4.16)).

To see why the lagged conditional variance avoids the necessity of adding many lagged squared residual terms to the model, notice that (4.13) can be rewritten as

$$h_t = \omega + \alpha_1 \varepsilon_{t-1}^2 + \beta_1(\omega + \alpha_1 \varepsilon_{t-2}^2 + \beta_1 h_{t-2}), \tag{4.14}$$

or, by continuing the recursive substitution, as

$$h_t = \sum_{i=1}^{\infty} \beta_1^i \omega + \alpha_1 \sum_{i=1}^{\infty} \beta_1^{i-1} \varepsilon_{t-i}^2. \tag{4.15}$$

This shows that the GARCH(1,1) model corresponds to an ARCH(∞) model with a particular structure for the parameters of the lagged ε_t^2 terms.

Alternatively, by adding ε_t^2 to both sides of (4.13) and moving h_t to the right-hand side, the GARCH(1,1) model can be rewritten as an ARMA(1,1) model for ε_t^2 as

$$\varepsilon_t^2 = \omega + (\alpha_1 + \beta_1)\varepsilon_{t-1}^2 + \nu_t - \beta_1 \nu_{t-1}, \tag{4.16}$$

where again $\nu_t = \varepsilon_t^2 - h_t$. Using the theory for ARMA models discussed in section 2.1, it follows that the GARCH(1,1) model is covariance-stationary if and only if $\alpha_1 + \beta_1 < 1$. In that case the unconditional mean of ε_t^2 – or, equivalently, the unconditional variance of ε_t – is equal to

$$\sigma^2 = \frac{\omega}{1 - \alpha_1 - \beta_1}. \tag{4.17}$$

The ARMA(1,1) representation in (4.16) also makes clear why α_1 needs to be strictly positive for identification of β_1. If $\alpha_1 = 0$, the AR and MA polynomials both are equal to $1 - \beta_1 L$. Rewriting the ARMA(1,1) model for ε_t^2 as an MA(∞), these polynomials cancel out,

$$\varepsilon_t^2 = \frac{1 - \beta_1 L}{1 - \beta_1 L} \nu_t = \nu_t, \tag{4.18}$$

which shows that β_1 then is not identified.

As shown by Bollerslev (1986), the unconditional fourth moment of ε_t is finite if $(\alpha_1+\beta_1)^2+2\alpha_1^2<1$. If in addition the z_t are assumed to be normally distributed, the kurtosis of ε_t is given by

$$\mathrm{K}_\varepsilon = \frac{3[1-(\alpha_1+\beta_1)^2]}{1-(\alpha_1+\beta_1)^2-2\alpha_1^2}, \tag{4.19}$$

which again is always larger than the normal value of 3. Notice that if $\beta_1 = 0$, (4.19) reduces to (4.9).

The autocorrelations of ε_t^2 are derived in Bollerslev (1988) and are found to be

$$\rho_1 = \alpha_1 + \frac{\alpha_1^2\beta_1}{1-2\alpha_1\beta_1-\beta_1^2}, \tag{4.20}$$

$$\rho_k = (\alpha_1+\beta_1)^{k-1}\rho_1 \qquad \text{for } k = 2, 3, \ldots \tag{4.21}$$

Even though the autocorrelations still decline exponentially, the decay factor in this case is $\alpha_1+\beta_1$. If this sum is close to one, the autocorrelations will decrease only very gradually. When the fourth moment of ε_t is not finite, the autocorrelations of ε_t^2 are time-varying. Of course, one can still compute the sample autocorrelations in this case. As shown by Ding and Granger (1996), if $\alpha_1+\beta_1<1$ and $(\alpha_1+\beta_1)^2+2\alpha_1^2 \geq 1$, such that the GARCH(1,1) model is covariance-stationary but with infinite fourth moment, the autocorrelations of ε_t^2 behave approximately as

$$\rho_1 \approx \alpha_1 + \beta_1/3, \tag{4.22}$$

$$\rho_k \approx (\alpha_1+\beta_1)^{k-1}\rho_1 \quad \text{for } k = 2, 3, \ldots \tag{4.23}$$

The parameter restriction $(\alpha_1+\beta_1)^2+2\alpha_1^2 = 1$ is equivalent to $1-2\alpha_1\beta_1-\beta_1^2 = 3\alpha_1^2$, from which it follows that (4.22) is identical to (4.20) where this restriction is satisfied. Therefore, the autocorrelations of ε_t^2 can be considered as continuous functions of α_1 and β_1, in the sense that their behaviour does not suddenly change when these parameters take values for which the condition for existence of the fourth moment is no longer satisfied.

The general GARCH(p,q) model is given by

$$\begin{aligned} h_t &= \omega + \sum_{i=1}^{q}\alpha_i\varepsilon_{t-i}^2 + \sum_{i=1}^{p}\beta_i h_{t-i} \\ &= \omega + \alpha(L)\varepsilon_t^2 + \beta(L)h_t, \end{aligned} \tag{4.24}$$

where $\alpha(L) = \alpha_1 L + \cdots + \alpha_q L^q$ and $\beta(L) = \beta_1 L + \cdots + \beta_p L^p$. Assuming that all the roots of $1-\beta(L)$ are outside the unit circle, the model can be rewritten as an infinite-order ARCH model

$$\begin{aligned} h_t &= \frac{\omega}{1-\beta(1)} + \frac{\alpha(L)}{1-\beta(L)}\varepsilon_t^2 \\ &= \frac{\omega}{1-\beta_1-\cdots-\beta_p} + \sum_{i=1}^{\infty}\delta_i\varepsilon_{t-i}^2. \end{aligned} \tag{4.25}$$

For nonnegativeness of the conditional variance it is required that all δ_i in (4.25) are nonnegative. Nelson and Cao (1992) discuss the conditions this implies for the parameters α_i, $i = 1,\dots,q$, and β_i, $i = 1\dots,q$, in the original model (4.24).

Alternatively, the GARCH(p,q) can be interpreted as an ARMA(m,p) model for ε_t^2 given by

$$\varepsilon_t^2 = \omega + \sum_{i=1}^{m}(\alpha_i+\beta_i)\varepsilon_{t-i}^2 - \sum_{i=1}^{p}\beta_i\nu_{t-i} + \nu_t, \tag{4.26}$$

where $m = \max(p,q)$, $\alpha_i \equiv 0$ for $i > q$ and $\beta_i \equiv 0$ for $i > p$. It follows that the GARCH(p,q) model is covariance-stationary if all the roots of $1-\alpha(L)-\beta(L)$ are outside the unit circle.

To determine the appropriate orders p and q in the GARCH(p,q) model, one can use a general-to-specific procedure by starting with a model with p and q set equal to large values, and testing down using likelihood-ratio-type restrictions (see Akgiray, 1989; Cao and Tsay, 1992). Alternatively, one can use modified information criteria, as suggested by Brooks and Burke (1997, 1998).

Even though the general GARCH(p,q) model might be of theoretical interest, the GARCH(1,1) model often appears adequate in practice (see also Bollerslev, Chou and Kroner, 1992). Furthermore, many nonlinear extensions to be discussed below have been considered only for the GARCH(1,1) case.

IGARCH

In applications of the GARCH(1,1) model (4.13) to high-frequency financial time series, it is often found that the estimates of α_1 and β_1 are such that their sum is close or equal to one. Following Engle and Bollerslev (1986), the model that results when $\alpha_1+\beta_1 = 1$ is commonly referred to as Integrated GARCH (IGARCH). The reason for this is that the restriction $\alpha_1+\beta_1 = 1$ implies a unit root in the ARMA(1,1) model for ε_t^2 given in (4.16), which then can be written as

$$(1-L)\varepsilon_t^2 = \omega + \nu_t - \beta_1\nu_{t-1}. \tag{4.27}$$

The analogy with a unit root in an ARMA model for the conditional mean of a time series is however rather subtle. For example, from (4.17) it is seen that the unconditional variance of ε_t is not finite in this case. Therefore, the IGARCH model is not covariance-stationary. However, the IGARCH(1,1) model may still be strictly stationary, as shown by Nelson (1990). This can be illustrated by rewriting (4.13) as

$$\begin{aligned} h_t &= \omega + (\alpha_1 z_{t-1}^2 + \beta_1)h_{t-1} \\ &= \omega + (\alpha_1 z_{t-1}^2 + \beta_1)(\omega + (\alpha_1 z_{t-2}^2 + \beta_1)h_{t-2}) \\ &= \omega(1 + (\alpha_1 z_{t-1}^2 + \beta_1)) + (\alpha_1 z_{t-1}^2 + \beta_1)(\alpha_1 z_{t-2}^2 + \beta_1)h_{t-2}, \end{aligned}$$

and continuing the substitution for h_{t-i}, it follows that

$$h_t = \omega\left(1 + \sum_{i=1}^{t-1}\prod_{j=1}^{i}(\alpha_1 z_{t-j}^2 + \beta_1)\right) + \prod_{i=1}^{t}(\alpha_1 z_{t-i}^2 + \beta_1)h_0. \qquad (4.28)$$

As shown by Nelson (1990), a necessary condition for strict stationarity of the GARCH(1,1) model is $\mathrm{E}[\ln(\alpha_1 z_{t-i}^2 + \beta_1)] < 0$. If this condition is satisfied, the impact of h_0 disappears asymptotically.

As expected, the autocorrelations of ε_t^2 for an IGARCH model are not defined properly. However, Ding and Granger (1996) show that the approximate autocorrelations are given by

$$\rho_k = \frac{1}{3}(1 + 2\alpha)(1 + 2\alpha^2)^{-k/2}. \qquad (4.29)$$

Hence, the autocorrelations still decay exponentially. This is in sharp contrast with the autocorrelations for a random walk model, for which the autocorrelations are approximately equal to 1 (see (2.33)).

FIGARCH

The properties of the conditional variance h_t as implied by the IGARCH model are not very attractive from an empirical point of view. Still, estimates of the parameters of GARCH(1,1) models for high-frequency financial time series invariably yield a sum of α_1 and β_1 close to 1, with α_1 small and β_1 large. This implies that the impact of shocks on the conditional variance diminishes only very slowly. From the ARCH(∞) representation of the GARCH(1,1) model as given in (4.15), the impact of the shock ε_t on h_{t+k} is given by $\alpha_1\beta_1^{k-1}$. With β_1 close to 1, this impact decays very slowly as k increases. Similarly, the autocorrelations for ε_t^2 given in (4.20) and (4.21) die out very slowly if the sum $\alpha_1 + \beta_1$ is close to 1. However, the decay is still at an exponential rate, which might be too fast to mimic the observed autocorrelation patterns of empirical time series, no matter how small the difference $1 - (\alpha_1 + \beta_1)$ is. For example, Ding, Granger and Engle (1993) suggest that the sample autocorrelations of squared – and, especially, absolute – returns decline only at a hyperbolic rate. This type of behaviour of the autocorrelations can be modelled by means of long-memory or fractionally integrated processes, as discussed in section 2.4.

Baillie, Bollerslev and Mikkelsen (1996) propose the class of Fractionally Integrated GARCH (FIGARCH) models. The basic FIGARCH(1,d,0) model is most easily obtained from (4.27), by simply adding an exponent d to the first-difference operator $(1-L)$, that is,

$$(1-L)^d \varepsilon_t^2 = \omega + \nu_t - \beta_1 \nu_{t-1}, \tag{4.30}$$

where $0 < d < 1$. Using the definition of $\nu_t = \varepsilon_t^2 - h_t$, this can be rewritten as an ARCH(∞) process for the conditional variance as

$$\begin{aligned} h_t &= \omega/(1-\beta_1) + (1-(1-L)^d/(1-\beta_1 L))\varepsilon_t^2 \\ &= \omega/(1-\beta_1) + \lambda(L)\varepsilon_t^2, \end{aligned} \tag{4.31}$$

where $\lambda(L) \equiv 1-(1-L)^d/(1-\beta_1 L)$. By using the expansion (2.106) for $(1-L)^d$, it can be shown that for large k

$$\lambda_k \approx [(1-\beta_1)\Gamma(d)^{-1}]k^{d-1}, \tag{4.32}$$

where $\Gamma(\cdot)$ is the gamma function. This expression shows that the effect of ε_t on h_{t+k} decays only at a hyperbolic rate as k increases. FIGARCH models are applied to exchange rates by Baillie, Bollerslev and Mikkelsen (1996), while Bollerslev and Mikkelsen (1996) apply the model to stock returns and option prices.

Ding and Granger (1996) argue that the sample autocorrelation functions of squared returns initially decrease faster than exponentially, and that only at higher lags does the decrease become (much) slower. This pattern suggests that volatility may consist of several components, some of which have a strong effect on volatility in the short run but die out quite rapidly, while others may have a small but persistent effect. To formalize this notion, Ding and Granger (1996) put forward the component GARCH model

$$h_t = \gamma h_{1,t} + (1-\gamma)h_{2,t}, \tag{4.33}$$

$$h_{1,t} = \alpha_1 \varepsilon_{t-1}^2 + (1-\alpha_1)h_{1,t-1}, \tag{4.34}$$

$$h_{2,t} = \omega + \alpha_2 \varepsilon_{t-1}^2 + \beta_2 h_{2,t-1}. \tag{4.35}$$

In this model, the conditional variance is seen to be a weighted sum of two components, one specified as an IGARCH model and the other as a GARCH model. A similar model is applied by Jones, Lamont and Lumsdaine (1998) to investigate whether shocks that occur on specific days, on which announcements of important macroeconomic figures are made, have different effects on volatility than shocks that occur on other days.

GARCH in mean

Many financial theories postulate a direct relationship between the return and risk of financial assets. For example, in the CAPM the excess return on a risky asset is proportional to its nondiversifiable risk, which is measured by the covariance with the market portfolio. The GARCH in mean (GARCH-M) model introduced by Engle, Lilien and Robins (1987) was explicitly designed to capture such direct relationships between return and possibly time-varying risk (as measured by the conditional variance). This is established by including (a function) of the conditional variance h_t in the model for the conditional mean of the variable of interest y_t, for example,

$$y_t = \phi_0 + \phi_1 y_{t-1} + \cdots + \phi_p y_{t-p} + \delta g(h_t) + \varepsilon_t, \tag{4.36}$$

where $g(h_t)$ is some function of the conditional variance of ε_t, h_t, which is assumed to follow a (possibly nonlinear) GARCH process. In most applications, $g(h_t)$ is taken to be the identity function or square root function – that is, $g(h_t) = h_t$ or $g(h_t) = \sqrt{h_t}$. The additional term $\delta g(h_t)$ in (4.36) is often interpreted as some sort of risk premium. As h_t varies over time, so does this risk premium.

To gain some intuition for the properties of y_t as implied by the GARCH-M model, consider (4.36) with $p = 0$ and $g(h_t) = h_t$ and assume that h_t follows an ARCH(1) process

$$y_t = \delta h_t + \varepsilon_t, \tag{4.37}$$

$$h_t = \omega + \alpha_1 \varepsilon_{t-1}^2. \tag{4.38}$$

Substituting (4.38) in (4.37) and using the fact that $\mathrm{E}[\varepsilon_{t-1}^2] = \omega/(1-\alpha_1)$ (see (4.6)), it follows that the unconditional expectation of y_t is equal to

$$\mathrm{E}[y_t] = \delta\omega\left(1 + \frac{\alpha_1}{1-\alpha_1}\right).$$

Similarly, it can be shown that the unconditional variance of y_t is equal to

$$\sigma_y^2 = \frac{\omega}{1-\alpha_1} + \frac{(\delta\alpha_1)^2 2\omega^2}{(1-\alpha_1)^2(1-3\alpha_1^2)},$$

which is larger than the unconditional variance of y_t in the absence of the GARCH-M effect, as in that case $\sigma_y^2 = \frac{\omega}{1-\alpha_1}$. Another consequence of the presence of h_t as regressor in the conditional mean equation (4.37) is that y_t is serially correlated. As shown by Hong (1991),

$$\rho_1 = \frac{2\alpha_1^3\delta^2\omega}{2\alpha_1^2\delta^2\omega + (1-\alpha_1)(1-3\alpha_1^2)} \tag{4.39}$$

$$\rho_k = \alpha_1^{k-1}\rho_1 \qquad k = 2, 3, \ldots \tag{4.40}$$

An overview of applications of GARCH-M models to stock returns, interest rates and exchange rates can be found in Bollerslev, Chou and Kroner (1992).

Stochastic volatility

In the GARCH model, the conditional volatility of the observed time series y_t is driven by the same shocks as its conditional mean. Furthermore, conditional upon the history of the time series as summarized in the information set Ω_{t-1}, current volatility h_t is deterministic. An alternative class of volatility models which has received considerable attention assumes that h_t is subject to an additional contemporaneous shock. The basic stochastic volatility (SV) model, introduced by Taylor (1986), is given by

$$\varepsilon_t = z_t\sqrt{h_t}, \tag{4.41}$$

$$\ln(h_t) = \gamma_0 + \gamma_1 \ln(h_{t-1}) + \gamma_2\eta_t, \tag{4.42}$$

with $z_t \sim$ NID(0, 1), $\eta_t \sim$ NID(0, 1), and η_t and z_t uncorrelated. A heuristic interpretation of the SV model is that the shock η_t represents shocks to the intensity of the flow of new information as measured by h_t, whereas the shock z_t represents the contents (large/small, positive/negative) of the news.

To understand the similarities and differences between the GARCH(1,1) and SV models, it is useful to consider the implied moments and correlation properties of ε_t. First note that, if $|\gamma_1| < 1$ in (4.42), $\ln(h_t)$ follows a stationary AR(1) process and $\ln(h_t) \sim N(\mu_h, \sigma_h^2)$ with

$$\mu_h = \mathrm{E}[\ln(h_t)] = \frac{\gamma_0}{1-\gamma_1}, \tag{4.43}$$

$$\sigma_h^2 = \mathrm{var}[\ln(h_t)] = \frac{\gamma_2^2}{1-\gamma_1^2}. \tag{4.44}$$

Put differently, h_t has a log-normal distribution. For the series ε_t, this implies that

$$\mathrm{E}[\varepsilon_t^r] = 0 \quad \text{for } r \text{ odd}, \tag{4.45}$$

$$\mathrm{E}[\varepsilon_t^2] = \mathrm{E}[z_t^2 h_t] = \mathrm{E}[z_t^2]\mathrm{E}[h_t] = \exp(\mu_h + \sigma_h^2/2), \tag{4.46}$$

$$\mathrm{E}[\varepsilon_t^4] = \mathrm{E}[z_t^4 h_t^2] = \mathrm{E}[z_t^4]\mathrm{E}[h_t^2] = 3\exp(2\mu_h + 2\sigma_h^2). \tag{4.47}$$

In particular, from (4.46) and (4.47) it follows that

$$\mathrm{K}_\varepsilon = \frac{\mathrm{E}[\varepsilon_t^4]}{\mathrm{E}[\varepsilon_t^2]^2} = 3\exp(\sigma_h^2), \tag{4.48}$$

which demonstrates that the SV model implies excess kurtosis in the series ε_t, similar to the GARCH(1,1) model (see (4.19)). A difference, however, is that the GARCH(1,1) model with z_t normally distributed typically cannot capture the excess kurtosis observed in financial time series completely (as will be discussed in more detail below), whereas the SV model can, as $\exp(\sigma_h^2)$ can take any value.

The correlation properties of ε_t^2 can be derived by noting that $\mathrm{E}[\varepsilon_t^2\varepsilon_{t-k}^2] = \mathrm{E}[z_t^2 h_t z_{t-k}^2 h_{t-k}] = \mathrm{E}[h_t^2 h_{t-k}^2]$. It follows that

$$\rho_k = \frac{\exp(\sigma_h^2\gamma_1^k) - 1}{3\exp(\sigma_h^2) - 1} \approx \frac{\exp(\sigma_h^2) - 1}{3\exp(\sigma_h^2) - 1}\gamma_1^k, \tag{4.49}$$

where the approximation is valid for small values of σ_h^2 and/or large γ_1 (see Taylor, 1986, pp. 74–5). Comparing (4.49) with the autocorrelations of ε_t^2 implied by the GARCH(1,1) model as given in (4.20) and (4.21) shows that in both cases the ACF of ε_t^2 is characterized by exponential decay towards zero. For the GARCH(1,1) model, the sum $\alpha_1 + \beta_1$ determines how fast the autocorrelations decline towards zero, whereas in the SV model this is determined by γ_1. When SV models are applied to high-frequency financial time series, the parameter estimates that are typically found imply that the first-order autocorrelation is small, whereas the subsequent decay is very slow. For example, typical parameter estimates are $\hat{\sigma}_\eta = 0.3$ and $\hat{\gamma}_1 = 0.95$, which according to (4.44) and (4.49) imply $\hat{\rho}_1 \approx 0.21$ and $\hat{\rho}_k \approx 0.95^k\hat{\rho}_1$.

The main difference between the GARCH and SV models is found at the estimation stage. In the GARCH model, the parameters can be estimated by straightforward application of maximum likelihood techniques, as will be discussed in section 4.3. This is owing to the fact that even though the conditional volatility h_t appears to be unobserved, it can be reconstructed using the past shocks $\varepsilon_{t-1}, \varepsilon_{t-2}, \ldots$ (assuming these can be obtained from the observed series $y_{t-1}, y_{t-2}, \ldots,$) and h_0. Technically, h_t is measurable with respect to the information set Ω_{t-1}. As a consequence, the distribution of ε_t conditional upon the history Ω_{t-1} can be obtained directly from the distribution of z_t, and the likelihood function can easily be constructed. By contrast, for the SV model the distribution of $\varepsilon_t|\Omega_{t-1}$ cannot be characterized explicitly, owing to the fact that h_t is not only unobserved, but also cannot be reconstructed from the history of the time series. Therefore, standard maximum likelihood techniques cannot be applied to estimate the parameters in SV models. Several alternative procedures have been examined, such as a (simulation-based) generalized method of moments (Melino and Turnbull, 1990; Duffie and Singleton, 1993), quasi-maximum likelihood via the Kalman filter (Harvey, Ruiz and Shephard, 1994), indirect inference (Gourieroux, Monfort and Renault, 1993), simulation-based maximum likelihood (Danielsson and Richard, 1993; Danielsson, 1994) and Bayesian methods (Jacquier, Polson and Rossi, 1994). As of yet, there is no consensus on the appropriate method(s) of estimation and inference in SV models. For that reason, we do not consider these models any further in this chapter and restrict ourselves to the GARCH class. Surveys of the SV literature can be found in Ghysels, Harvey and Renault (1996) and Shephard (1996).

4.1.2 Nonlinear GARCH models

As shown in section 1.2, for stock returns it appears to be the case that volatile periods often are initiated by a large negative shock, which suggests that positive and negative shocks may have an asymmetric impact on the conditional volatility of subsequent observations. This was recognized by Black (1976), who suggested that a possible explanation for this finding might be the way firms are financed. When the value of (the stock of) a firm falls, the debt-to-equity ratio increases, which in turn leads to an increase in the volatility of the returns on equity. As the debt-to-equity ratio is also known as the 'leverage' of the firm, this phenomenon is commonly referred to as the 'leverage effect'.

The GARCH models discussed above cannot capture such asymmetric effects of positive and negative shocks. As the conditional variance depends only on the square of the shock, positive and negative shocks of the same magnitude have the same effect on the conditional volatility – that is, the sign of the shock is not important. Most nonlinear extensions of the GARCH model which have been developed over the years are designed to allow for different effects of positive and negative shocks or other types of asymmetries. In this section we review several of such nonlinear GARCH models. The models that are discussed below are only a small sample from all the different nonlinear GARCH models which have been proposed. For more complete overviews, the interested reader is referred to Hentschel (1995), among others. We generally concentrate on those models that make use of the idea of regime switching, as discussed in chapter 3 for nonlinear models for the conditional mean.

Most nonlinear GARCH models are motivated by the desire to capture the different effects of positive and negative shocks on conditional volatility or other types of asymmetry. A natural question to ask, then, is whether all these models are indeed different from each other, or whether they are more or less similar. A convenient way to compare different GARCH models is by means of the so-called *news impact curve* (NIC), introduced by Pagan and Schwert (1990) and popularized by Engle and Ng (1993). The NIC measures how new information is incorporated into volatility. To be more precise, the NIC shows the relationship between the current shock or news ε_t and conditional volatility 1 period ahead h_{t+1}, holding constant all other past and current information. In the basic GARCH(1,1) model and nonlinear variants thereof, the only relevant information from the past is the current conditional variance h_t. Thus, the NIC for the GARCH(1,1) model (4.13) is defined as

$$NIC(\varepsilon_t | h_t = h) = \omega + \alpha_1 \varepsilon_t^2 + \beta_1 h = A + \alpha_1 \varepsilon_t^2, \tag{4.50}$$

where $A = \omega + \beta_1 h$. Hence, the NIC is a quadratic function centred on $\varepsilon_t = 0$. As the value of the lagged conditional variance h_t affects only the constant A in (4.50), it only shifts the NIC vertically, but does not change its basic shape. In practice, it is customary to take h_t equal to the unconditional variance σ^2.

Exponential GARCH

The earliest variant of the GARCH model which allows for asymmetric effects is the Exponential GARCH (EGARCH) model, introduced by Nelson (1991). The EGARCH(1,1) model is given by

$$\ln(h_t) = \omega + \alpha_1 z_{t-1} + \gamma_1(|z_{t-1}| - E(|z_{t-1}|)) + \beta_1 \ln(h_{t-1}). \qquad (4.51)$$

As the EGARCH model (4.51) describes the relation between past shocks and the *logarithm* of the conditional variance, no restrictions on the parameters α_1, γ_1 and β_1 have to be imposed to ensure that h_t is nonnegative. Using the properties of z_t, it follows that $g(z_t) \equiv \alpha_1 z_t + \gamma_1(|z_t| - E(|z_t|))$ has mean zero and is uncorrelated. The function $g(z_t)$ is piecewise linear in z_t, as it can be rewritten as

$$g(z_t) = (\alpha_1 + \gamma_1) z_t I(z_t > 0) + (\alpha_1 - \gamma_1) z_t I(z_t < 0) - \gamma_1 E(|z_t|).$$

Thus, negative shocks have an impact $\alpha_1 - \gamma_1$ on the log of the conditional variance, while for positive shocks the impact is $\alpha_1 + \gamma_1$. This property of the function $g(z_t)$ leads to an asymmetric NIC. In particular, the NIC for the EGARCH(1,1) model (4.51) is given by

$$NIC(\varepsilon_t | h_t = \sigma^2) = \begin{cases} A \exp\left(\dfrac{\alpha_1 + \gamma_1}{\sigma} \varepsilon_t\right) & \text{for } \varepsilon_t > 0, \\ A \exp\left(\dfrac{\alpha_1 - \gamma_1}{\sigma} \varepsilon_t\right) & \text{for } \varepsilon_t < 0, \end{cases} \qquad (4.52)$$

with $A = \sigma^{2\beta_1} \exp(\omega - \gamma_1 \sqrt{2/\pi})$.

Typical NICs for the GARCH(1,1) and EGARCH(1,1) models are shown in panel (a) of figure 4.1. The parameters in the models have been chosen such that the constants A in (4.52) and (4.50) are the same and, hence, the NICs are equal when $\varepsilon_t = 0$. The shape of the NIC of the EGARCH model is typical for parameterizations with $\alpha_1 < 0$, $0 \leq \gamma_1 < 1$ and $\gamma_1 + \beta_1 < 1$. For such parameter configurations, negative shocks have a larger effect on the conditional variance than positive shocks of the same size. For the range of ε_t for which the NIC is plotted in figure 4.1, it also appears that negative shocks in the EGARCH model have a larger effect on the conditional variance than shocks in the GARCH model, while the reverse holds for positive shocks. However, as ε_t increases, the impact on h_t will eventually become larger in the EGARCH model, as the exponential function in (4.52) dominates the quadratic in (4.50) for large ε_t.

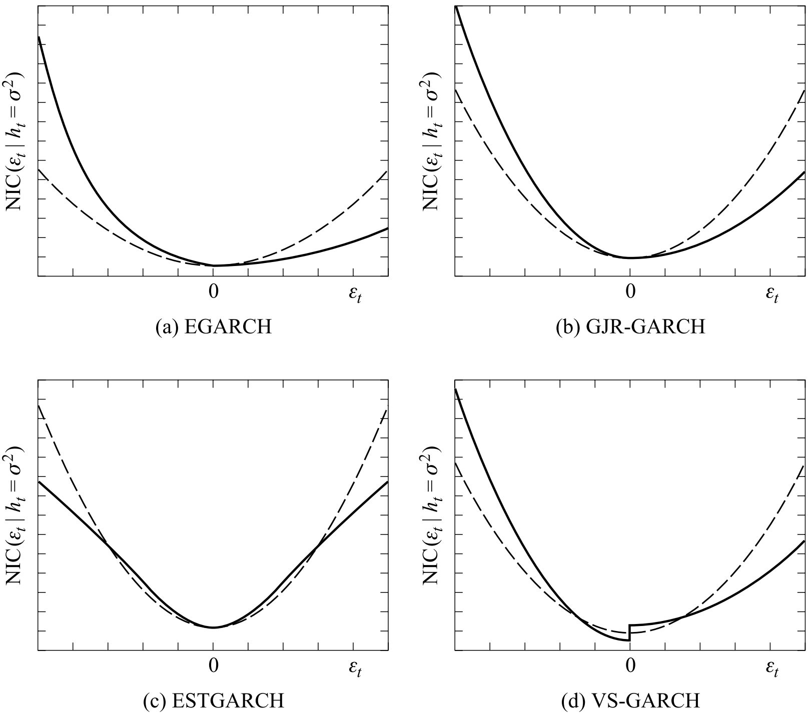

Figure 4.1 Examples of news impact curves for various nonlinear models (solid lines); for comparison, the news impact curve of a GARCH(1,1) model is also shown in each panel (dashed line) (a) The parameters in the EGARCH model (4.51) are such that $\alpha_1 < 0$, $0 \leq \gamma_1 < 1$ and $\gamma_1 + \beta_1 < 1$ (b) The parameters in the GJR-GARCH model (4.53) are such that $\alpha_1 > \gamma_1$, while $(\alpha_1 + \gamma_1)/2$ is equal to α_1 in the GARCH(1,1) model (c) The parameters in the ESTGARCH model (4.55) with (4.58) are such that $\alpha_1 > \gamma_1$ and their average is equal to α_1 in the GARCH(1,1) model (d) The parameters in the VS-GARCH model (4.59) are set such that $\alpha_1 > \gamma_1$ and $\omega + \beta_1 h < \zeta + \delta_1 h$, while the averages of all three pairs of parameters in the two regimes are equal to the corresponding parameter in the GARCH(1,1) model

GJR-GARCH

The model introduced by Glosten, Jagannathan and Runkle (1993) offers an alternative method to allow for asymmetric effects of positive and negative shocks on volatility. The model is obtained from the GARCH(1,1) model (4.13) by assuming that the parameter of ε_{t-1}^2 depends on the sign of the shock, that is,

$$h_t = \omega + \alpha_1 \varepsilon_{t-1}^2 (1 - I[\varepsilon_{t-1} > 0]) + \gamma_1 \varepsilon_{t-1}^2 I[\varepsilon_{t-1} > 0] + \beta_1 h_{t-1}, \tag{4.53}$$

where as usual $I[\cdot]$ is an indicator function. The conditions for nonnegativeness of the conditional variance are $\omega > 0$, $(\alpha_1 + \gamma_1)/2 \geq 0$ and $\beta_1 > 0$. The condition for covariance-stationarity is $(\alpha_1 + \gamma_1)/2 + \beta_1 < 1$. If this condition is satisfied, the unconditional variance of ε_t is $\sigma^2 = \omega/(1 - (\alpha_1 + \gamma_1)/2 - \beta_1)$. The NIC for the GJR-GARCH model follows directly from (4.53) and is equal to

$$NIC(\varepsilon_t | h_t = \sigma^2) = A + \begin{cases} \alpha_1 \varepsilon_t^2 & \text{if } \varepsilon_t < 0, \\ \gamma_1 \varepsilon_t^2 & \text{if } \varepsilon_t > 0, \end{cases} \tag{4.54}$$

where $A = \omega + \beta_1 \sigma^2$. The NIC of the GJR-GARCH model is a quadratic function centred on $\varepsilon_t = 0$, similar to the NIC of the basic GARCH model. However, the slopes of the GJR-GARCH NIC are allowed to be different for positive and negative shocks. Depending on the values of α_1 and γ_1 in (4.53), the NIC (4.54) can be steeper or less steep than the GARCH NIC (4.50). An example of the GJR-GARCH NIC is shown in panel (b) of figure 4.1, where we have set α_1 and γ_1 such that $\alpha_1 > \gamma_1$ while their average is equal to the value of α_1 in the GARCH(1,1) model. In this case, the NIC is steeper than the GARCH NIC for negative news and less steep for positive news. Comparing the NICs of the EGARCH and GJR-GARCH models as shown in panels (a) and (b) of figure 4.1 shows that they are rather similar. Hence, the GJR-GARCH model and the EGARCH model may be considered as alternative models for the same series. It may be difficult to develop criteria that can help to distinguish between the two models.

Smooth Transition GARCH

The GJR-GARCH model (4.53) can be interpreted as a threshold model, as it allows the parameter corresponding to the lagged squared shock to change abruptly from α_1 to γ_1 at $\varepsilon_{t-1} = 0$. Hagerud (1997) and González-Rivera (1998) independently applied the idea of smooth transition, discussed in section 3.1, to allow for a more gradual change of this parameter. The Logistic Smooth Transition GARCH (LSTGARCH) model is given by

$$h_t = \omega + \alpha_1 \varepsilon_{t-1}^2 [1 - F(\varepsilon_{t-1})] + \gamma_1 \varepsilon_{t-1}^2 F(\varepsilon_{t-1}) + \beta_1 h_{t-1}, \tag{4.55}$$

where the function $F(\varepsilon_{t-1})$ is the logistic function

$$F(\varepsilon_{t-1}) = \frac{1}{1 + \exp(-\theta \varepsilon_{t-1})}, \quad \theta > 0. \tag{4.56}$$

As the function $F(\varepsilon_{t-1})$ in (4.56) changes monotonically from 0 to 1 as ε_{t-1} increases, the impact of ε_{t-1}^2 on h_t changes smoothly from α_1 to γ_1. When the parameter θ in (4.56) becomes large, the logistic function approaches a step function which equals 0 for negative ε_{t-1} and 1 for positive ε_{t-1}. In that case, the LSTGARCH model reduces to the GJR-GARCH model (4.53).

The parameter restrictions necessary for h_t to be positive and for the model to be covariance-stationary are the same as for the GJR-GARCH model given above. The NIC for the LSTGARCH model is given by

$$NIC(\varepsilon_t|h_t = \sigma^2) = [A + \alpha_1\varepsilon_t^2][1 - F(\varepsilon_t)] + \gamma_1\varepsilon_t^2]F(\varepsilon_t), \qquad (4.57)$$

where $A = \omega + \beta_1\sigma^2$.

The STGARCH model (4.55) can also be used to describe asymmetric effects of large and small shocks on conditional volatility, by using the exponential function

$$F(\varepsilon_{t-1}) = 1 - \exp(-\theta\varepsilon_{t-1}^2), \quad \theta > 0. \qquad (4.58)$$

The function $F(\varepsilon_{t-1})$ in (4.58) changes from 1 for large negative values of ε_{t-1} to 0 for $\varepsilon_{t-1} = 0$ and increases back again to 1 for large positive values of ε_{t-1}. Thus, the effective parameter of ε_{t-1}^2 in the Exponential STGARCH (ESTGARCH) model given by (4.55) with (4.58) changes from γ_1 to α_1 and back to γ_1 again. Panel (c) of figure 4.1 shows an example of the NIC of the ESTGARCH model.

Volatility-Switching GARCH

The LSTGARCH and GJR-GARCH models assume that the asymmetric behaviour of h_t depends only on the sign of the past shock ε_{t-1}. In applications it is typically found that $\gamma_1 < \alpha_1$, such that a negative shock increases the conditional variance more than a positive shock of the same size. On the other hand, the ESTGARCH model assumes that the asymmetry is caused entirely by the size of the shock. Rabemananjara and Zakoïan (1993) point out that the asymmetric behaviour of h_t may be more complicated and that both the sign and the size of the shock may be important. In particular, they argue that negative shocks increase future conditional volatility more than positive shocks only if the shock is large in absolute value. For small shocks they observe the opposite kind of asymmetry, in that small positive shocks increase the conditional volatility more than small negative shocks.

Fornari and Mele (1996, 1997) discuss a model which allows for such complicated asymmetric behaviour. The model is in fact a generalization of the GJR-GARCH model, and is obtained by allowing all parameters in the conditional variance equation to depend on the sign of the shock ε_{t-1}. The Volatility-Switching GARCH (VS-GARCH) model of order (1,1) is given by

$$\begin{aligned} h_t =&(\omega + \alpha_1\varepsilon_{t-1}^2 + \beta_1 h_{t-1})(1 - I[\varepsilon_{t-1} > 0]) \\ &+ (\zeta + \gamma_1\varepsilon_{t-1}^2 + \delta_1 h_{t-1})I[\varepsilon_{t-1} > 0]. \end{aligned} \qquad (4.59)$$

Fornari and Mele (1997) show that the unconditional variance of ε_t is equal to $\sigma^2 = [(\omega + \zeta)/2]/[1 - (\alpha_1 + \gamma_1)/2 - (\beta_1 + \delta_1)/2]$. The fourth unconditional moment of ε_t, and hence the kurtosis, implied by (4.59) is typically higher than that of a GARCH(1,1) model with parameters equal to the average of the parameters in the two regimes of the VS-GARCH model (see Fornari and Mele, 1997) for the exact expression.

The NIC for the VS-GARCH model is given by

$$NIC(\varepsilon_t | h_t = h) = \begin{cases} \omega + \alpha_1 \varepsilon_t^2 + \beta_1 h & \text{if } \varepsilon_t < 0, \\ \zeta + \gamma_1 \varepsilon_t^2 + \delta_1 h & \text{if } \varepsilon_t > 0. \end{cases} \tag{4.60}$$

This NIC is seen to be an asymmetric quadratic function centred on $\varepsilon_t = 0$, with possibly different slopes for positive and negative shocks. In this respect, the NIC of the VS-GARCH model is identical to the NIC of the GJR-GARCH model. However, as in general $\omega + \beta_1 h \neq \zeta + \delta_1 h$, the NIC (4.60) can be discontinuous at $\varepsilon_t = 0$. The size of the jump at this point depends on the magnitude of the past conditional volatility $h_t = h$. An example of the NIC (4.60) is shown in panel (d) of figure 4.1. The parameters in the VS-GARCH model (4.59) are set such that $\alpha_1 > \gamma_1$ and $\omega + \beta_1 h < \zeta + \delta_1 h$. For such parameter configurations, small positive shocks have a larger impact on the conditional volatility than small negative shocks, while the reverse holds for large shocks. This demonstrates that the VS-GARCH can describe more complicated asymmetric effects of shocks on conditional volatility than just sign or size effects.

Asymmetric Nonlinear Smooth Transition GARCH

Anderson, Nam and Vahid (1999) modify the VS-GARCH model by allowing the transition from one regime to the other to be smooth. The resulting Asymmetric Nonlinear Smooth Transition GARCH (ANST-GARCH) model is given by

$$\begin{aligned} h_t = {} & [\omega + \alpha_1 \varepsilon_{t-1}^2 + \beta_1 h_{t-1}][1 - F(\varepsilon_{t-1})] \\ & + [\zeta + \gamma_1 \varepsilon_{t-1}^2 + \delta_1 h_{t-1}] F(\varepsilon_{t-1}), \end{aligned} \tag{4.61}$$

where $F(\varepsilon_{t-1})$ is the logistic function (4.56). Even though the corresponding NIC,

$$\begin{aligned} NIC(\varepsilon_t | h_t = h) = {} & [\omega + \alpha_1 \varepsilon_t^2 + \beta_1 h][1 - F(\varepsilon_t)] \\ & + [\zeta + \gamma_1 \varepsilon_t^2 + \delta_1 h] F(\varepsilon_t), \end{aligned} \tag{4.62}$$

looks similar to the NIC of the VS-GARCH model at first sight, closer inspection of its properties reveals that it is rather different. In particular, the NIC of the ANST-GARCH model is always continuous at $\varepsilon_t = 0$, but it does not necessarily attain its minimum value at this point (which is the case for the NIC of the VS-GARCH model). It might be that the best news, which is defined as the shock that minimizes next period's conditional volatility, is nonzero. The exact size of the shock ε_t that constitutes the best news depends in a nontrivial way on current conditional volatility. Anderson, Nam and Vahid, (1999) demonstrate that the relationship between the best news and h_t is positive when $\delta_1 < \beta_1$. Some examples of NIC for this model are shown in figure 4.2. Evidently, the NIC changes shape as the current conditional volatility changes. As h_t increases, the asymmetry of the NIC becomes more pronounced while the shock ε_t that minimizes next period's volatility becomes increasingly positive.

Quadratic GARCH

Sentana (1995) introduced the Quadratic GARCH (QGARCH) model as another way to cope with asymmetric effects of shocks on volatility. The QGARCH(1,1) model is specified as

$$h_t = \omega + \gamma_1 \varepsilon_{t-1} + \alpha_1 \varepsilon_{t-1}^2 + \beta_1 h_{t-1}. \tag{4.63}$$

The additional term $\gamma_1 \varepsilon_{t-1}$ makes it possible for positive and negative shocks to have different effects on h_t. To see this, note that the model can be rewritten as

$$h_t = \omega + \left(\frac{\gamma_1}{\varepsilon_{t-1}} + \alpha_1 \right) \varepsilon_{t-1}^2 + \beta_1 h_{t-1}, \tag{4.64}$$

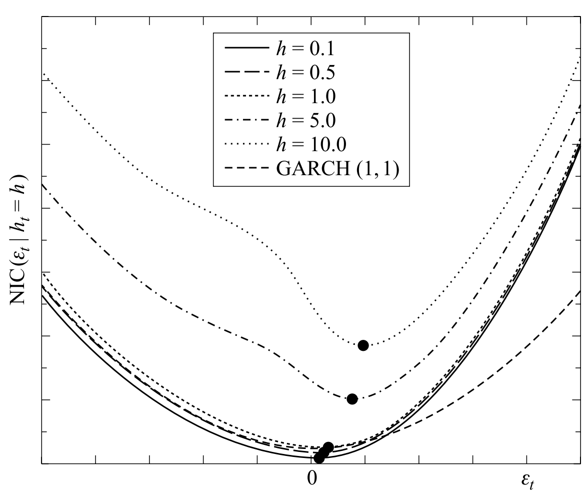

Figure 4.2 Examples of news impact curves for the ANST-GARCH model
The parameters in the model (4.61) are such that $\omega > \zeta$, $\alpha_1 < \gamma_1$ and $\beta_1 > \delta_1$; the location of the best news for each level of past conditional volatility is marked with a solid circle

which shows that the impact of ε_{t-1}^2 on h_t is equal to $\gamma_1/\varepsilon_{t-1} + \alpha_1$. If $\gamma_1 < 0$, the effect of negative shocks on h_t will be larger than the effect of a positive shock of the same size. Notice that in addition the effect depends on the size of the shock.

Alternatively, (4.63) can be expressed as

$$h_t = \omega - \frac{\gamma_1^2}{4\alpha_1} + \alpha_1\left(\varepsilon_{t-1} + \frac{\gamma_1}{2\alpha_1}\right)^2 + \beta_1 h_{t-1} \tag{4.65}$$

(see also Engle and Ng, 1993). This representation shows that in the QGARCH model, the effect of shocks on the conditional variance is symmetric around $\varepsilon_t = -\gamma_1/(2\alpha_1)$.

Apart from the asymmetry, the QGARCH model is very similar to the standard GARCH model. For example, as shown in Sentana (1995), the unconditional variance of ε_t as implied by the QGARCH(1,1) model (4.63) is the same as that implied by the GARCH(1,1) model (4.13) – that is, $\sigma^2 = \omega/(1-\alpha_1-\beta_1)$. Furthermore, the condition for covariance-stationarity of the QGARCH(1,1) model and the condition for existence of the unconditional fourth moment are the same as the corresponding conditions in the GARCH(1,1) model. The kurtosis of ε_t is, however, different from (4.19) and depends on the asymmetry parameter γ_1 as follows,

$$\mathrm{K}_\varepsilon = \frac{3[1-(\alpha_1+\beta_1)^2 + \gamma_1^2(1-\alpha_1-\beta_1)/\omega]}{1-(\alpha_1+\beta_1)^2 - 2\alpha_1^2}. \tag{4.66}$$

In particular, (4.66) shows that the kurtosis of the QGARCH model is an increasing function of the absolute value of γ_1. Therefore, for fixed values of the parameters ω, α_1 and β_1, the kurtosis for the QGARCH model is larger than the kurtosis for the corresponding GARCH model, which results if $\gamma_1 = 0$. Finally, from (4.65) it follows that the NIC of the QGARCH model is the same as the NIC of the basic GARCH model, except that it is centred at $-\gamma_1/(2\alpha_1)$.

Markov-Switching GARCH

In the previous specifications, the parameters in the model change according to the sign and/or the size of the lagged shock ε_{t-1}. Therefore, these models can be interpreted as regime-switching models where the regime is determined by an observable variable, similar in spirit to the SETAR and STAR models for the conditional mean discussed in chapter 3.

An obvious alternative is to assume that the regime is determined by an unobservable Markov-process s_t, as in the Markov-Switching model discussed in subsection 3.1.2. A general Markov-Switching GARCH [MSW-GARCH] model is given by

$$
\begin{aligned}
h_t = {} & [\omega + \alpha_1 \varepsilon_{t-1}^2 + \beta_1 h_{t-1}] I[s_t = 1] \\
& + [\zeta + \gamma_1 \varepsilon_{t-1}^2 + \delta_1 h_{t-1}] I[s_t = 2],
\end{aligned} \tag{4.67}
$$

where s_t is a two-state Markov chain with transition probabilities defined below (3.16). The general form in (4.67) is considered in Klaassen (1999). Various restricted versions are applied in Kim (1993); Cai (1994); Hamilton and Susmel (1994) and Dueker (1997).

Alternative error distributions

So far we have assumed that the innovations z_t in (4.2) are normally distributed, which is equivalent to stating that the conditional distribution of ε_t is normal with mean zero and variance h_t. The *unconditional* distribution of a series ε_t for which the conditional variance follows a GARCH model is nonnormal in this case. In particular, as the kurtosis of ε_t is larger than the normal value of 3, the unconditional distribution has fatter tails than the normal distribution. However, in many applications of the standard GARCH(1,1) model (4.13) to high-frequency financial time series it is found that the unconditional kurtosis of ε_t given in (4.19) is much smaller than the kurtosis of the observed time series. Put differently, the kurtosis of the standardized residuals $\hat{z}_t \equiv \hat{\varepsilon}_t \hat{h}_t^{-1/2}$ is found to be larger than 3. The nonlinear GARCH models discussed above imply a higher kurtosis of ε_t (see, for example, 4.66) for the QGARCH model. Therefore, nonlinear GARCH models might be able to accommodate this deficiency of the standard GARCH model (in addition to the asymmetric response of the conditional variance to positive and negative shocks for which these nonlinear models were originally designed). He and Teräsvirta (1999a, 1999b) derive expressions for the unconditional moments of ε_t (and the autocorrelations of ε_t^2) for various nonlinear GARCH models. In principle, these expressions can be used to select a nonlinear GARCH model which best suits the moment (and correlation) properties of an observed time series, although this has not been thoroughly investigated yet.

An alternative approach which has been followed is to consider alternative distributions for z_t. The unconditional kurtosis of ε_t is an increasing function of the kurtosis of z_t (see Teräsvirta, 1996) and, hence, K_ε can be increased by assuming a leptokurtic distribution for z_t. Following Bollerslev (1987), a popular choice has become the standardized Student-t distribution with η degrees of freedom, that is,

$$
f(z_t) = \frac{\Gamma((\eta+1)/2)}{\sqrt{\pi(\eta-2)}\Gamma(\eta/2)} \left(1 + \frac{z_t^2}{\eta-2}\right)^{-(\eta+1)/2}, \tag{4.68}
$$

where $\Gamma(\cdot)$ is the Gamma function. The Student-t distribution is symmetric around zero (and thus $\mathrm{E}[z_t] = 0$), while it converges to the normal distribution as the number of degrees of freedom η becomes larger. A further characteristic of the Student-t distribution is that only moments up to order η exist. Hence, for $\eta > 4$, the fourth moment of z_t exists and is equal to $3(\eta - 2)/(\eta - 4)$. As this is larger than the normal value of 3, the unconditional kurtosis of ε_t will also be larger than in case z_t followed a normal distribution. The number of degrees of freedom of the Student-t distribution need not be specified in advance. Rather, η can be treated as a parameter and can be estimated along with the other parameters in the model.

4.2 Testing for GARCH

It seems self-evident that a formal test for the presence of conditional heteroscedasticity of ε_t should be part of a specification procedure for GARCH models. In particular, even though it appears obvious from summary statistics and graphs such as those presented in chapter 1 that the conditional volatility of high-frequency financial time series changes over time, one might want to perform such a test prior to actually estimating a GARCH model. In this section we review several tests for linear and nonlinear GARCH models. We also discuss whether these tests are sensitive to various sorts of model misspecification, and elaborate upon the effects of outliers on the test-statistics in some more detail.

4.2.1 Testing for linear GARCH

Engle (1982) developed a test for conditional heteroscedasticity in the context of ARCH models based on the Lagrange Multiplier (LM) principle. The conditional variance h_t in the ARCH(q) model in (4.10) is constant if the parameters corresponding to the lagged squared shocks ε^2_{t-i}, $i = 1, \ldots, q$, are equal to zero. Therefore, the null hypothesis of conditional homoscedasticity can be formulated as $H_0 : \alpha_1 = \ldots = \alpha_q$. The corresponding LM test can be computed as nR^2, where n is the sample size and the R^2 is obtained from a regression of the squared residuals on a constant and q of its lags,

$$\hat{\varepsilon}^2_t = \omega + \alpha_1 \hat{\varepsilon}^2_{t-1} + \cdots + \alpha_q \hat{\varepsilon}^2_{t-q} + u_t, \tag{4.69}$$

where the residuals $\hat{\varepsilon}_t$ are obtained by estimating the model for the conditional mean of the observed time series y_t under the null hypothesis. The LM test-statistic has an asymptotic $\chi^2(q)$ distribution. The test for ARCH can alternatively be interpreted as a test for serial correlation in the squared residuals. In fact, the ARCH test is asymptotically equivalent to the test of McLeod and Li (1983) given in (2.45) (see Granger and Teräsvirta, 1993, pp. 93–4). Lee (1991) shows that the LM test against this GARCH(p, q) alternative is the same as the LM test against the alternative of ARCH(q) errors.

Example 4.1: Testing for ARCH in stock and exchange rate returns We apply the LM test for ARCH(q) to the weekly returns on stock indexes and exchange rates. We calculate the test-statistics for two sample periods of 5 years, from January 1986–December 1990 and January 1991–December 1995. For simplicity, we assume that the conditional mean of the time series can be adequately described by an AR(k) model, where the autoregressive-order k is determined by the AIC. The second–fourth columns of tables 4.1 and 4.2 contain p-values for the LM test for ARCH(q) with $q = 1, 5$ and 10 for the stock and exchange rate returns, respectively.

The results in tables 4.1 and 4.2 show that there is substantial evidence for the presence of ARCH, especially if we allow $q > 1$, which should capture GARCH-type properties. Notice that for the stock returns the p-values for the

Table 4.1 *Testing for ARCH in weekly stock index returns*

	Standard test			Robust test			No. obs. with
Stock market q	1	5	10	1	5	10	$w_r(r_t) < 0.05$
Sample 1986–90							
Amsterdam	0.000	0.000	0.000	0.125	0.506	0.752	6
Frankfurt	0.020	0.024	0.125	0.359	0.125	0.034	8
Hong Kong	0.006	0.104	0.422	0.004	0.006	0.029	13
London	0.000	0.000	0.000	0.487	0.490	0.563	4
New York	0.000	0.000	0.000	0.832	0.825	0.415	6
Paris	0.277	0.125	0.038	0.070	0.811	0.340	6
Singapore	0.001	0.060	0.330	0.159	0.012	0.036	17
Tokyo	0.000	0.000	0.000	0.003	0.036	0.024	22
Sample 1991–95							
Amsterdam	0.565	0.016	0.199	1.000	0.023	0.311	2
Frankfurt	0.881	0.448	0.606	0.870	0.747	0.854	2
Hong Kong	0.560	0.010	0.004	0.503	0.054	0.176	7
London	0.642	0.922	0.971	0.963	0.093	0.224	2
New York	0.041	0.334	0.052	0.915	0.835	0.284	9
Paris	0.593	0.426	0.456	0.593	0.426	0.456	0
Singapore	0.271	0.026	0.044	0.128	0.015	0.122	2
Tokyo	0.162	0.000	0.000	0.295	0.310	0.083	5

Notes: p-values of the standard and outlier-robust variants of the LM test for ARCH(q) for weekly stock index returns.
The tests are applied to residuals from an AR(k) model, with k determined by the AIC. The last column reports the number of observations (out of 260) which receive a weight less than 0.05 in the robust estimation procedure for the AR(k) model.

Table 4.2 *Testing for ARCH in weekly exchange rate returns*

		Standard test			Robust test			No. obs. with
Currency	q	1	5	10	1	5	10	$w_r(r_t) < 0.05$
Sample 1986–90								
Australian dollar		0.384	0.558	0.664	0.984	0.972	0.438	9
British pound		0.902	0.393	0.516	0.711	0.862	0.941	2
Canadian dollar		0.714	0.000	0.000	0.165	0.606	0.803	5
Dutch guilder		0.790	0.323	0.645	0.326	0.398	0.693	1
French franc		0.494	0.191	0.458	0.213	0.198	0.398	1
German Dmark		0.820	0.325	0.617	0.394	0.373	0.617	1
Japanese yen		0.105	0.257	0.584	0.610	0.948	0.941	3
Swiss franc		0.588	0.190	0.581	0.755	0.342	0.818	0
Sample 1991–95								
Australian dollar		0.184	0.414	0.131	0.198	0.462	0.171	0
British pound		0.000	0.000	0.000	0.425	0.207	0.468	8
Canadian dollar		0.406	0.297	0.401	0.756	0.068	0.283	0
Dutch guilder		0.937	0.018	0.097	0.908	0.571	0.723	6
French franc		0.925	0.001	0.019	0.680	0.841	0.531	5
German Dmark		0.951	0.013	0.084	0.973	0.713	0.644	4
Japanese yen		0.658	0.131	0.061	0.833	0.541	0.976	11
Swiss franc		0.556	0.058	0.184	0.700	0.561	0.517	5

Notes: p-values of the standard and outlier-robust variants of the LM test for ARCH(q) for weekly exchange rate returns.
The tests are applied to residuals from an AR(k) model, with k determined by the AIC.
The last column reports the number of observations (out of 260) which receive a weight less than 0.05 in the robust estimation procedure for the AR(k) model.

tests are, in general, much smaller for the first subsample 1986–90. For the exchange rate returns, the evidence for ARCH seems largely confined to the second subsample 1991–95.

4.2.2 *Testing for nonlinear GARCH*

With respect to the specification of nonlinear GARCH models discussed in the previous subsection, there are two possible routes one might follow. First, one can start with specifying and estimating a linear GARCH model and subsequently test the need for asymmetric of other nonlinear components in the model. The test-statistics that are involved in this approach are discussed in detail in section 4.4. Second, one can test the null hypothesis of conditional homoscedasticity directly against the alternative of asymmetric ARCH. In this section we present test-statistics which might be used for this purpose.

Engle and Ng (1993) discuss tests to check whether positive and negative shocks have a different impact on the conditional variance. Let S^-_{t-1} denote a dummy variable which takes the value 1 when $\hat{\varepsilon}_{t-1}$ is negative and 0 otherwise, where $\hat{\varepsilon}_t$ are the residuals from estimating a model for the conditional mean of y_t under the assumption of conditional homoscedasticity. The tests examine whether the squared residual $\hat{\varepsilon}^2_t$ can be predicted by S^-_{t-1}, $S^-_{t-1}\hat{\varepsilon}_{t-1}$, and/or $S^+_{t-1}\hat{\varepsilon}_{t-1}$, where $S^+_{t-1} \equiv 1 - S^-_{t-1}$. The test-statistics are computed as the t-ratio of the parameter ϕ_1 in the regression

$$\hat{\varepsilon}^2_t = \phi_0 + \phi_1 \hat{w}_{t-1} + \xi_t, \tag{4.70}$$

where $\hat{w}_{t-1}$ is one of the three measures of asymmetry defined above and ξ_t the residual.

Where $\hat{w}_t = S^-_{t-1}$ in (4.70), the test is called the Sign Bias (SB) test, as it simply tests whether the magnitude of the square of the current shock ε_t (and, hence, the conditional variance h_t) depends on the sign of the lagged shock ε_{t-1}. In case $\hat{w}_t = S^-_{t-1}\hat{\varepsilon}_{t-1}$ or $\hat{w}_t = S^+_{t-1}\hat{\varepsilon}_{t-1}$, the tests are called the Negative Size Bias (NSB) and Positive Size Bias (PSB) tests, respectively. These tests examine whether the effect of negative or positive shocks on the conditional variance also depends on their size. As the SB-, NSB- and PSB-statistics are t-ratios, they follow a standard normal distribution asymptotically.

The tests can also be conducted jointly, by estimating the regression

$$\hat{\varepsilon}^2_t = \phi_0 + \phi_1 S^-_{t-1} + \phi_2 S^-_{t-1}\hat{\varepsilon}_{t-1} + \phi_3 S^+_{t-1}\hat{\varepsilon}_{t-1} + \xi_t. \tag{4.71}$$

The null hypothesis $H_0 : \phi_1 = \phi_2 = \phi_3 = 0$ can be evaluated by computing n times the R^2 from this regression. The resultant test-statistic has an asymptotic χ^2 distribution with 3 degrees of freedom.

Example 4.2: Testing for Sign and Size Bias in stock and exchange rate returns We apply the SB, NSB, PSB tests and the general test for asymmetry based on (4.71) to weekly stock and exchange rate returns over the 10-year sample period from January 1986 until December 1995. The tests are computed for the residuals from an AR(k) model, where the order k is determined by minimizing the AIC. The values of the test-statistics along with the corresponding p-values are given in table 4.3. Clearly, there is substantial evidence of asymmetric ARCH effects. Comparing the p-values of the SB test with those of the NSB and PSB tests, for the majority of these series size effects appear to be more important than sign effects.

Table 4.3 *Testing for asymmetric ARCH effects in weekly stock index and exchange rate returns*

	SB test		NSB test		PSB test		General test	
Stock market	Test	p-value	Test	p-value	Test	p-value	Test	p-value
Amsterdam	1.93	0.027	−22.73	0.000	6.43	0.000	381.78	0.000
Frankfurt	1.65	0.049	−21.15	0.000	6.89	0.000	381.28	0.000
Hong Kong	1.95	0.025	−25.86	0.000	3.15	0.001	375.69	0.000
London	1.18	0.118	−21.34	0.000	7.20	0.000	373.42	0.000
New York	1.94	0.026	−24.44	0.000	4.96	0.000	377.85	0.000
Singapore	1.06	0.145	−25.81	0.000	4.95	0.000	380.05	0.000
Tokyo	1.73	0.042	−16.99	0.000	17.75	0.000	433.18	0.000
Exchange rate								
Australian dollar	−2.87	0.002	−4.12	0.000	49.69	0.000	431.72	0.000
British pound	−2.31	0.010	−4.39	0.000	51.49	0.000	425.98	0.000
Canadian dollar	−2.22	0.013	−10.70	0.000	22.42	0.000	435.07	0.000
Dutch guilder	−0.67	0.250	−9.43	0.000	28.03	0.000	423.55	0.000
French franc	−0.68	0.247	−9.71	0.000	26.42	0.000	418.71	0.000
German Dmark	−0.594	0.278	−9.38	0.000	28.17	0.000	422.17	0.000
Japanese yen	0.80	0.212	−13.90	0.000	20.91	0.000	450.40	0.000
Swiss franc	−0.10	0.462	−12.11	0.000	20.73	0.000	442.13	0.000

Notes: Sign Bias (SB), Negative Size Bias (NSB), Positive Size Bias (PSB) tests and general test for asymmetric volatility effects for weekly stock and exchange rate returns. The sample runs from January 1986 until December 1995.
The tests are applied to residuals from an AR(k) model, with k determined by the AIC.

An alternative to the tests of Engle and Ng (1993) are LM tests against various forms of asymmetric ARCH. Sentana (1995) discusses a test of homoscedasticity against the alternative of quadratic ARCH (QARCH). Consider the QARCH(q) model, which can be obtained from (4.63) by setting $\beta_1 = 0$ and adding lagged shocks $\varepsilon_{t-2}, \ldots, \varepsilon_{t-q}$ and their squares, that is,

$$\begin{aligned} h_t &= \omega + \gamma_1\varepsilon_{t-1} + \gamma_2\varepsilon_{t-2} + \cdots + \gamma_q\varepsilon_{t-q} \\ &\quad + \alpha_1\varepsilon_{t-1}^2 + \alpha_2\varepsilon_{t-2}^2 + \cdots + \alpha_q\varepsilon_{t-q}^2, \end{aligned} \tag{4.72}$$

where $\alpha_1 = \cdots = \alpha_q = \gamma_1 = \cdots = \gamma_q = 0$, h_t is constant. A LM-statistic to test these parameter restrictions can be computed as n times the R^2 from a regression of the squared residuals $\hat{\varepsilon}_t^2$ on $\hat{\varepsilon}_{t-1}, \ldots, \hat{\varepsilon}_{t-q}$ and $\hat{\varepsilon}_{t-1}^2, \ldots, \hat{\varepsilon}_{t-q}^2$. Asymptotically, the statistic is χ^2 distributed with $2q$ degrees of freedom.

Hagerud (1997) suggests two test-statistics to test constant conditional variance against Smooth Transition ARCH (STARCH). The STARCH(q) model is given by

$$\begin{aligned} h_t = \omega + \alpha_1 \varepsilon_{t-1}^2 [1 - F(\varepsilon_{t-1})] + \gamma_1 \varepsilon_{t-1}^2 F(\varepsilon_{t-1}) \\ + \cdots + \alpha_q \varepsilon_{t-q}^2 [1 - F(\varepsilon_{t-q})] + \gamma_q \varepsilon_{t-q}^2 F(\varepsilon_{t-q}), \end{aligned} \quad (4.73)$$

where $F(\cdot)$ is either the logistic function (4.56) or the exponential function (4.58). The null hypothesis of conditional homoscedasticity can again be specified as $H_0 : \alpha_1 = \cdots = \alpha_q = \gamma_1 = \cdots = \gamma_q = 0$. The testing problem is complicated in this case as the parameter θ in the transition function $F(\cdot)$ is not identified under the null hypothesis. This identification problem is similar to the one discussed in subsection 3.3.2 in case of testing linearity of the conditional mean against STAR-type alternatives. The solution here is also the same – that is, the transition function can be approximated by a low-order Taylor approximation. In case of the Logistic STARCH (LSTARCH) model (4.73) with (4.56), this results in the auxiliary model

$$h_t = \omega + \alpha_1^* \varepsilon_{t-1}^2 + \cdots + \alpha_q^* \varepsilon_{t-q}^2 + \gamma_1^* \varepsilon_{t-1}^3 + \cdots + \gamma_q^* \varepsilon_{t-q}^3. \quad (4.74)$$

An LM-statistic to test the equivalent null hypothesis $H_0^* : \alpha_1^* = \cdots = \alpha_q^* = \gamma_1^* = \cdots = \gamma_q^* = 0$ can be obtained as n times the R^2 from the regression of $\hat{\varepsilon}_t^2$ on $\hat{\varepsilon}_{t-1}^2, \ldots, \hat{\varepsilon}_{t-q}^2$ and $\hat{\varepsilon}_{t-1}^3, \ldots, \hat{\varepsilon}_{t-q}^3$. Asymptotically, the statistic is χ^2 distributed with $2q$ degrees of freedom. In case of the Exponential STARCH (ESTARCH) model (4.73) with (4.58), the auxiliary model is similar to (4.74) except that ε_{t-i}^4, $i = 1, \ldots, q$, are included instead of ε_{t-i}^3, $i = 1, \ldots, q$. The null hypothesis of constant conditional variance can be tested by n times the R^2 of the auxiliary regression of $\hat{\varepsilon}_t^2$ on $\hat{\varepsilon}_{t-1}^2, \ldots, \hat{\varepsilon}_{t-q}^2$ and $\hat{\varepsilon}_{t-1}^4, \ldots, \hat{\varepsilon}_{t-q}^4$. This statistic also has an asymptotic χ^2 distribution with $2q$ degrees of freedom.

Example 4.3: Testing for nonlinear ARCH in stock and exchange rate returns We apply the LM tests for QARCH(q), LSTARCH(q) and ESTARCH(q) to the weekly returns on stock indices and exchange rates, using the same strategy as in example 4.1. The left-hand side panels of tables 4.4 and 4.5 contain p-values for the above LM tests with $q = 5$ for the stock and exchange rate returns, respectively. The results in these tables suggest that there is ample evidence of nonlinear ARCH.

Table 4.4 *Testing for nonlinear ARCH in weekly stock index returns*

	Standard tests			Robust tests		
Stock market	LM_Q	LM_L	LM_E	LM_Q	LM_L	LM_E
Sample 1986–90						
Amsterdam	0.000	0.000	0.000	0.509	0.276	0.181
Frankfurt	0.033	0.042	0.000	0.183	0.020	0.291
Hong Kong	0.000	0.000	0.000	0.068	0.063	0.039
London	0.001	0.000	0.000	0.603	0.742	0.908
New York	0.000	0.000	0.000	0.427	0.580	0.778
Paris	0.101	0.136	0.187	0.957	0.858	0.936
Singapore	0.015	0.022	0.000	0.081	0.067	0.026
Tokyo	0.000	0.000	0.002	0.032	0.068	0.107
Sample 1991–5						
Amsterdam	0.064	0.077	0.042	0.088	0.112	0.108
Frankfurt	0.112	0.075	0.492	0.181	0.067	0.446
Hong Kong	0.049	0.005	0.002	0.154	0.144	0.008
London	0.097	0.111	0.720	0.022	0.020	0.127
New York	0.710	0.802	0.336	0.903	0.891	0.960
Paris	0.409	0.312	0.323	0.409	0.312	0.323
Singapore	0.005	0.004	0.008	0.046	0.013	0.101
Tokyo	0.000	0.000	0.000	0.214	0.150	0.123

Notes: p-values of the standard and outlier-robust variants of the LM test for QARCH(q) [LM_Q], LSTARCH(q) [LM_L] and ESTARCH(q) [LM_E] for $q = 5$ applied to weekly stock index returns.
The tests are applied to residuals from an AR(k) model, with k determined by the AIC.

4.2.3 Testing for ARCH in the presence of misspecification

The small sample properties of the LM test for linear (G)ARCH have been investigated quite extensively. In particular, it has been found that rejection of the null hypothesis of homoscedasticity might be due to other sorts of model misspecification, such as neglected serial correlation, nonlinearity and omitted variables in the model for the conditional mean. For example, Engle, Hendry and Trumble (1985), Bera, Higgins and Lee (1992) and Sullivan and Giles (1995) show that in the presence of neglected serial correlation, the LM test tends to overreject the null hypothesis. Bera and Higgins (1997) discuss the similarity between ARCH and bilinear processes such as (2.37) and suggest that the two may easily be mistaken. Giles, Giles and Wong (1993) provide simulation evidence on the effects of omitted variables, which demonstrates that this may

Table 4.5 *Testing for nonlinear ARCH in weekly exchange rate returns*

	Standard tests			Robust tests		
Currency	LM_Q	LM_L	LM_E	LM_Q	LM_L	LM_E
Sample 1986–90						
Australian dollar	0.154	0.581	0.514	0.944	0.947	0.997
British pound	0.458	0.202	0.315	0.567	0.752	0.698
Canadian dollar	0.006	0.000	0.001	0.769	0.569	0.303
Dutch guilder	0.563	0.198	0.524	0.545	0.270	0.533
French franc	0.280	0.088	0.374	0.234	0.069	0.504
German Dmark	0.601	0.214	0.549	0.513	0.246	0.475
Japanese yen	0.553	0.535	0.514	0.977	0.978	0.954
Swiss franc	0.454	0.540	0.390	0.715	0.543	0.394
Sample 1991–5						
Australian dollar	0.747	0.841	0.690	0.805	0.741	0.676
British pound	0.000	0.000	0.000	0.223	0.068	0.390
Canadian dollar	0.263	0.035	0.137	0.108	0.070	0.178
Dutch guilder	0.012	0.002	0.001	0.772	0.432	0.494
French franc	0.003	0.000	0.000	0.955	0.881	0.978
German Dmark	0.011	0.002	0.001	0.875	0.705	0.664
Japanese yen	0.198	0.385	0.041	0.697	0.584	0.646
Swiss franc	0.145	0.103	0.116	0.729	0.665	0.284

Notes: p-values of the standard and outlier-robust variants of the LM tests for QARCH(q) [LM_Q], LSTARCH(q) [LM_L] and ESTARCH(q) [LM_E] for $q = 5$ applied to weekly exchange rate returns.
The tests are applied to residuals from an AR(k) model, with k determined by the AIC.

also lead to significant ARCH-statistics. Lumsdaine and Ng (1999) investigate the properties of the LM test for ARCH in the presence of misspecification in the conditional mean model at a general level. They conclude that model misspecification causes the regression residuals $\hat{\varepsilon}_t$ to be serially correlated even if the true errors ε_t are not. Consequently, the squared regression residuals also exhibit spurious correlation and, hence, model misspecification necessarily leads to positive (and never negative) size distortion for the LM test. In the next subsection we elaborate upon the properties of the LM test in the presence of outliers in more detail.

Example 4.4: Properties of the ARCH tests in case of neglected nonlinearity
To illustrate the properties of the LM tests for ARCH, QARCH, LSTARCH and ESTARCH in the presence of misspecification, the following simulation experiment is performed. We generate 5,000 series of length $n = 250$ from

Table 4.6 *Testing for ARCH and QARCH in simulated SETAR series*

Intercepts					
$\phi_{0,1}$	$\phi_{0,2}$	LM_A	LM_Q	LM_L	LM_E
0	0	24.04	23.72	28.82	25.68
−0.3	0.1	67.46	68.24	78.22	65.10
−0.3	−0.1	78.46	90.18	85.92	73.06
0.3	−0.1	9.06	9.96	6.98	22.34

Notes: Rejection frequencies of the null hypothesis of conditional homoscedasticity against ARCH(1) [LM_A], QARCH(1) [LM_Q], LSTARCH(1) [LM_L] and ESTARCH(1) [LM_E], for series of length $n = 250$ generated from the SETAR model (3.1), with $\phi_{1,1} = -0.5$, $\phi_{1,2} = 0.5$, $c = 0$ and $\varepsilon_t \sim \text{NID}(0, 0.25^2)$, based on 5,000 replications.

the SETAR model (3.1), with $\phi_{1,1} = -0.5$, $\phi_{1,2} = 0.5$, $c = 0$ and $\varepsilon_t \sim \text{NID}(0, 0.25^2)$. For the intercepts $\phi_{0,1}$ and $\phi_{0,2}$ we take the values that were used to generate the example series discussed in subsection 3.1.1. For each series, we erroneously specify and estimate a linear AR(1) model for the conditional mean, and test the residuals for conditional heteroscedasticity by means of the various LM tests against ARCH with $q = 1$. Table 4.6 contains the rejection frequencies of the null hypothesis using the 5 per cent asymptotic critical value. These frequencies vary substantially, depending on the values of $\phi_{0,1}$ and $\phi_{0,2}$, but in all cases they are above the nominal significance level.

4.2.4 *Testing for ARCH in the presence of outliers*

The adverse effects of outliers on estimates of models for the conditional mean and specification tests for such models have been discussed in chapters 2 and 3, respectively. In the light of these results, it should come as no surprise that outliers, additive outliers in particular, affect the tests for ARCH as well. van Dijk, Franses and Lucas (1999b) show that the behaviour of the LM test for ARCH based on the regression (4.69) in the presence of AOs is very similar to the behaviour of the test for STAR nonlinearity under such circumstances, as discussed in subsection 3.3.4. If the AOs are neglected, the LM test rejects the null hypothesis of conditional homoscedasticity too often when it is in fact true, while the test has difficulty detecting genuine GARCH effects, in the sense that the power of the test is reduced considerably.

An alternative test-statistic which is robust to the presence of AOs can be obtained by employing the robust estimation techniques discussed in section 2.5. For example, where an AR(1) model is entertained for the conditional mean of the series y_t, one can use the standardized residuals r_t defined just below (2.115) and the weights $w_r(r_t)$ to construct the weighted residuals $w_r(r_t)r_t \equiv \psi(r_t)$. A robust equivalent to the LM test for ARCH(q) is obtained as n times the R^2 of a regression of $\psi(r_t)^2$ on a constant and $\psi(r_{t-1})^2, \ldots, \psi(r_{t-q})^2$. Under conventional assumptions, the outlier-robust LM test has a $\chi^2(q)$ distribution asymptotically. A similar procedure can be followed to obtain outlier-robust tests against the alternative of nonlinear ARCH. For example, a robust test against LSTARCH(q) can be computed as n times the R^2 of the auxiliary regression of $\psi(r_t)^2$ on $\psi(r_{t-1})^2, \ldots, \psi(r_{t-q})^2$ and $\psi(r_{t-1})^3, \ldots, \psi(r_{t-q})^3$.

The analysis of van Dijk, Franses and Lucas (1999b) concerns AOs that occur in isolation. Additionally, Franses, van Dijk and Lucas (1998) demonstrate that the effects on the standard LM test can be even more dramatic if outliers appear as consecutive observations. Reassuringly, the outlier-robust test is affected to a much lesser extent by such patches of outliers. Of course, patches of outliers are difficult to distinguish from genuine GARCH effects, as they look very similar upon casual inspection of a graph of a time series. However, combining the outcomes of the standard and robust tests for ARCH can sometimes provide a way to tell the two apart, as illustrated in the example below.

Example 4.5: Properties of the ARCH tests in the presence of neglected outliers To examine the effects of outliers on the LM tests for ARCH, we conduct the following simulation experiment. We generate artificial time series from an AR(1) model,

$$y_t = \phi_1 y_{t-1} + \varepsilon_t, \quad t = 1, \ldots, n, \tag{4.75}$$

with $\phi_1 = 0.5$. The shocks ε_t are either drawn from a standard normal distribution or defined as $\varepsilon_t = z_t\sqrt{h_t}$, with h_t generated according to a GARCH(1,1) model as in (4.13) and $z_t \sim$ NID(0, 1). In the GARCH(1,1) model for h_t, we set $\alpha_1 = 0.25$, $\beta_1 = 0.65$ and $\omega = 1 - \alpha_1 - \beta_1$, such that ε_t has unconditional variance equal to 1. The sample size n is set equal to 100, 250 or 500 observations. Next, we add a single outlier of size $\zeta = 0, 3, 5$ or 7 to the observation at the middle of the sample, $y_{n/2}$. For the resultant series, we estimate an AR(1) model and obtain the residuals $\hat{\varepsilon}_t$. Given the AR(1) model for the conditional mean, the AO at $t = n/2$ potentially yields two large consecutive residuals. We compute the standard and outlier-robust test-statistics against qth order ARCH and nonlinear ARCH for $q = 1$ to see whether the tests are affected by these large residuals. Table 4.7 shows the rejection frequencies of the null hypothesis at a nominal significance level of 5 per cent, based on 5,000 replications.

Table 4.7 *Rejection frequencies of standard and robust tests for (nonlinear) ARCH in the presence of outliers*

		Standard tests				Robust tests			
		LM_A	LM_Q	LM_L	LM_E	LM_A	LM_Q	LM_L	LM_E
$\varepsilon_t \sim NID(0, 1)$									
100	0	3.62	3.78	3.62	3.94	4.02	4.14	3.64	4.30
	3	7.84	8.38	10.82	9.82	4.86	4.96	5.42	5.40
	5	20.22	18.90	27.10	22.92	3.96	4.02	4.22	5.02
	7	17.62	16.06	15.98	16.82	3.54	3.96	3.98	4.54
250	0	4.36	4.60	4.86	4.64	4.48	4.68	4.64	4.54
	3	9.46	9.42	13.02	13.70	5.10	5.40	5.30	5.30
	5	35.04	32.06	40.08	43.98	4.86	4.78	4.68	4.88
	7	51.98	45.76	50.48	52.10	4.54	4.86	4.78	4.74
500	0	3.88	4.30	4.38	4.26	4.46	4.32	4.08	3.76
	3	7.42	7.38	10.56	11.94	4.62	4.58	4.44	4.20
	5	37.46	33.62	45.46	57.52	4.54	4.18	3.74	3.92
	7	68.46	62.92	70.04	71.60	4.42	4.28	4.08	3.94
$\varepsilon_t \sim GARCH(1, 1)$									
100	0	43.46	42.64	42.86	46.06	27.96	25.36	26.50	26.68
	3	38.56	36.72	37.20	40.48	27.24	24.76	26.08	26.48
	5	31.66	28.46	29.10	31.18	28.92	26.40	27.24	27.94
	7	19.28	18.02	18.02	19.00	29.64	27.22	28.10	29.06
250	0	83.74	81.24	81.84	85.86	61.96	55.92	57.18	59.14
	3	81.80	78.66	80.22	83.44	61.78	55.46	56.46	58.48
	5	78.82	74.24	74.90	77.06	62.50	56.04	57.54	60.42
	7	70.98	64.74	64.98	66.28	63.28	56.84	58.16	60.88
500	0	98.90	98.18	98.26	99.06	89.20	84.66	84.78	87.20
	3	98.24	97.66	97.86	98.70	88.62	83.80	84.22	86.98
	5	97.20	96.06	96.96	97.46	89.28	85.00	85.44	87.38
	7	95.64	93.90	94.80	95.22	89.50	85.30	85.84	87.82

Notes: Rejection frequencies of the standard and outlier-robust LM statistics for ARCH(1) [LM_A], QARCH(1) [LM_Q], LSTARCH(1) [LM_L] and ESTARCH(1) [LM_E], using the 5 per cent asymptotic critical value.
Time series are generated according to the AR(1) model (4.75) and an AO of magnitude ζ is added to the observation in the middle of the sample.
The empirical power in the lower panel is not adjusted for empirical size.

The results in the upper panel of table 4.7 show that the rejection frequencies under the null hypothesis for the standard tests are much larger than the nominal significance level if the magnitude of the outlier is substantial. The robust tests do not suffer from size distortion at all. The lower panel shows that the power of the robust test where no outliers occur is smaller than the power of the standard test. This illustrates that protection against aberrant observations comes at a cost in terms of a decrease in power where no outliers occur. The rejection frequencies of the non-robust test decrease in the presence of outliers, especially for sample sizes $n = 100$ and 250.

The different effects of isolated outliers and patches of outliers on the LM test for linear ARCH are illustrated by means of a similar simulation experiment. Artificial time series of length $n = 500$ are obtained from (4.75) with $\phi_1 = 0$ and the properties of the shocks ε_t are as specified above. Next, we add either $m = 5$ or 10 isolated outliers, or $m = 1$ or 2 patches of $k = 2, 3$ or 5 outliers at random places in the series. The absolute magnitude of the outliers is set equal to $\zeta = 3, 5$ or 7, while the sign of each outlier is positive or negative with equal probability. The ARCH tests are applied to series from which the mean has been removed, either by estimating it with OLS or the outlier-robust GM estimator. For each replication we record whether we find ARCH with both the standard and robust tests, denoted as (Y,Y), ARCH with the standard test but not with the robust test [(Y,N)], or one of the other combinations [(N,Y) or (N,N)]. The results for this experiment are reported in table 4.8.

From table 4.8, several conclusions emerge. First, the size of both the standard and robust tests, which can be obtained by adding up the entries in the columns headed (Y,Y) and (Y,N) or (Y,Y) and (N,Y), respectively, is hardly affected by the occurrence of isolated outliers. Note that for the standard test this differs from the results in table 4.7. This is due to the fact that in this case the true value of the autoregressive parameter ϕ_1 is assumed to be known (and equal to zero). In contrast to the limited impact of neglecting isolated AOs in white noise series on the standard ARCH test, it is seen that in case of clustering of AOs the standard LM test is affected to a much larger extent. For almost all combinations of m, k and ζ considered here, the test-statistic is severely oversized. In fact, the empirical rejection frequency equals 100 per cent already in case of a single patch of 3 outliers or 2 patches of 2 outliers (of absolute magnitude 5 or 7) out of the 500 observations. In sharp contrast with these findings for the standard test, the empirical size of the robust test is usually close to the nominal 5 per cent significance level.

If isolated outliers occur, the power of the standard test decreases quite dramatically, while the power of the robust test remains high. In the presence of patchy outliers, the power of both the standard and robust tests is very high.

Table 4.8 *Properties of standard and robust tests for ARCH in the presence of patchy outliers*

			$\varepsilon_t \sim$ NID(0, 1)				$\varepsilon_t \sim$ GARCH(1, 1)			
m	k	ζ	(Y,Y)	(Y,N)	(N,Y)	(N,N)	(Y,Y)	(Y,N)	(N,Y)	(N,N)
1	2	3	3.8	25.8	2.6	67.8	90.1	9.7	0.2	0.0
		5	5.6	90.4	0.2	3.8	90.7	9.2	0.1	0.0
		7	5.9	93.8	0.0	0.3	90.7	9.2	0.1	0.0
	3	3	4.9	49.6	2.2	43.3	90.9	9.0	0.0	0.1
		5	6.2	93.1	0.0	0.7	92.1	7.8	0.1	0.0
		7	6.3	93.6	0.0	0.1	92.2	7.7	0.1	0.0
	5	3	5.5	75.4	1.4	17.7	89.3	10.4	0.1	0.2
		5	4.6	95.4	0.0	0.0	91.7	8.1	0.0	0.2
		7	4.9	95.1	0.0	0.0	91.8	8.0	0.0	0.2
2	2	3	3.9	49.3	1.9	44.9	89.6	10.3	0.1	0.0
		5	4.8	95.2	0.0	0.0	91.4	8.6	0.1	0.0
		7	4.8	95.2	0.0	0.0	91.6	8.4	0.0	0.0
	3	3	6.9	76.1	0.7	16.3	88.3	11.7	0.0	0.0
		5	6.1	93.9	0.0	0.0	90.3	9.7	0.0	0.0
		7	6.2	93.8	0.0	0.0	90.2	9.8	0.0	0.0
	5	3	10.3	87.3	0.0	2.4	89.9	10.1	0.0	0.0
		5	5.2	94.8	0.0	0.0	93.0	7.0	0.0	0.0
		7	5.3	94.7	0.0	0.0	93.0	7.0	0.0	0.0
5	1	3	1.4	2.8	4.7	91.1	79.8	9.9	7.9	2.4
		5	0.3	3.3	5.4	91.0	42.3	3.7	47.0	7.0
		7	0.4	4.1	5.3	90.2	17.0	0.9	72.0	10.1
10	1	3	0.9	3.4	4.0	91.7	71.5	7.8	16.9	3.8
		5	0.2	4.9	5.6	89.3	30.2	2.4	60.8	6.6
		7	0.4	6.1	5.2	88.3	13.6	1.2	76.8	8.4

Notes: Rejection frequencies of the standard and robust LM tests against ARCH(1), based on 1,000 replications.
The cells report the number of times that a certain outcome occurs when the test statistics are evaluated at the 5 per cent nominal significance level. For example, (Y,N) means that the standard LM test detects ARCH (Y) while the robust test does not (N).
The series are generated according to an AR(1) process.
m patches of k outliers of absolute magnitude ζ are added at random places in the series.

The above simulation results suggest how the outcomes of the standard and robust ARCH tests can sometimes be helpful to distinguish genuine GARCH effects from outliers. If the robust test finds no ARCH, there probably is no ARCH, and when it finds ARCH, there probably is. Furthermore, the result

(Y,N), meaning finding ARCH with the standard test but not with the robust test, can most likely be seen as evidence against ARCH in favour of a short sequence of extraordinary observations. The opposite result (N,Y), meaning finding ARCH with the robust test but not with the standard test, can be interpreted as evidence of ARCH, possibly contaminated with a few isolated outliers. In both cases it is recommended to have a closer look at the weights from the robust regression and the corresponding observations in the original time series, before carrying on with any subsequent analyses.

Example 4.1/4.3: Testing for (nonlinear) ARCH in stock and exchange rate returns The sample periods of 5 years for which the LM tests for (nonlinear) ARCH(q) were computed for the weekly returns on stock indices and exchange rates are such that one of the samples is more or less regular, whereas the other clearly contains an unusual event, which might be regarded as an outlier. For the stock index returns, the unusual event is the crash on 19 October 1987, which is part of the first sample from January 1986–December 1990. For the exchange rates, the unusual event is the speculative attack on a number of European currencies in September 1992, which is contained in the second sample running from January 1991 until December 1995. Columns 5–7 of tables 4.1 and 4.2 contain p-values for the outlier-robust variants of the LM test for ARCH(q) with $q = 1, 5$ and 10 for the stock and exchange rate returns, respectively. The same columns of tables 4.4 and 4.5 contain results for the robust test for QARCH(q). The overwhelming evidence for, possibly nonlinear, ARCH found by the standard tests becomes somewhat weaker when we consider the robust tests. In the rightmost columns of tables 4.1 and 4.2, we present the fraction of estimated weights in the robust estimation method that is smaller than 0.05. Apparently, only a few observations may cause the nonrobust tests to reject the null hypothesis of conditional homoscedasticity.

The differences that may occur across robust and nonrobust tests should be interpreted with great care. In particular, the evidence for time variation in the conditional volatility of high-frequency financial time series is so overwhelming that it may seem odd to attribute this entirely to aberrant observations. It is therefore not recommended to discard the GARCH model altogether in case the robust test fails to reject the null hypothesis. It may be better to conclude that AOs are possibly relevant and should be taken into account when estimating the parameters in a GARCH model. Some methods which are potentially useful for this purpose are discussed in subsection 4.3.3.

4.3 Estimation

In this section we discuss estimation of the parameters in GARCH models. General principles are discussed first, followed by some remarks on simplifications which are available in case of linear GARCH models and on estimation in the presence of outliers.

4.3.1 General principles

Consider the general nonlinear autoregressive model of order p,

$$y_t = G(x_t; \xi) + \varepsilon_t, \tag{4.76}$$

where $x_t = (1, y_{t-1}, \ldots, y_{t-p})'$ and the skeleton $G(x_t; \xi)$ is a general nonlinear function of the parameters ξ that is at least twice continuously differentiable. The conditional variance h_t of ε_t is assumed to follow a possibly nonlinear GARCH model with parameters ψ. For example, where a QGARCH(1,1) model (4.63) is specified for h_t, $\psi = (\omega, \alpha_1, \gamma_1, \beta_1)'$. The parameters in the models for the conditional mean and conditional variance are gathered in the vector $\theta \equiv (\xi', \psi')'$. The true parameter values are denoted $\theta_0 = (\xi_0', \psi_0')'$. The parameters in θ can conveniently be estimated by maximum likelihood (ML). The conditional log likelihood for the tth observation is equal to

$$l_t(\theta) = \ln f(\varepsilon_t/\sqrt{h_t}) - \ln \sqrt{h_t}, \tag{4.77}$$

where $f(\cdot)$ denotes the density of the i.i.d. shocks z_t. For example, if z_t is assumed to be normally distributed,

$$l_t(\theta) = -\frac{1}{2}\ln 2\pi - \frac{1}{2}\ln h_t - \frac{\varepsilon_t^2}{2h_t}. \tag{4.78}$$

The maximum likelihood estimate (MLE) for θ, which we denote as $\hat{\theta}_{\mathrm{ML}}$, is found by maximizing the log likelihood function for the full sample, which is simply the sum of the conditional log likelihoods as given in (4.77). The MLE solves the first-order condition

$$\sum_{t=1}^{n} \frac{\partial l_t(\theta)}{\partial \theta} = 0. \tag{4.79}$$

The vector of derivatives of the log likelihood with respect to the parameters is usually referred to as the score $s_t(\theta) \equiv \partial l_t(\theta)/\partial\theta$. For the model (4.76) with $\varepsilon_t = z_t\sqrt{h_t}$, the score can be decomposed as $s_t(\theta) = (\partial l_t(\theta)/\partial\xi', \partial l_t(\theta)/\partial\psi')$, where

$$\frac{\partial l_t(\theta)}{\partial \xi} = \frac{\varepsilon_t}{h_t}\,\frac{\partial G(x_t;\xi)}{\partial \xi} + \frac{1}{2h_t}\left(\frac{\varepsilon_t^2}{h_t} - 1\right)\frac{\partial h_t}{\partial \xi}, \tag{4.80}$$

$$\frac{\partial l_t(\theta)}{\partial \psi} = \frac{1}{2h_t}\left(\frac{\varepsilon_t^2}{h_t} - 1\right)\frac{\partial h_t}{\partial \psi}. \tag{4.81}$$

The second term on the right-hand-side of (4.80) arises because the conditional variance h_t in general depends on ε_{t-1}, and thus on the parameters in the conditional mean for y_t, as $\varepsilon_{t-1} = y_{t-1} - G(x_{t-1}; \xi)$.

As the first-order conditions in (4.79) are nonlinear in the parameters, an iterative optimization procedure has to be used to obtain the MLE $\hat{\theta}_{\mathrm{ML}}$. If the conditional distribution $f(\cdot)$ is correctly specified, the resulting estimates are consistent and asymptotically normal. The asymptotic covariance matrix of $\sqrt{n}(\hat{\theta}_{\mathrm{ML}} - \theta_0)$ is then equal to A_0^{-1}, the inverse of the information matrix evaluated at the true parameter vector θ_0,

$$A_0 = -\frac{1}{n}\sum_{t=1}^{n} \mathrm{E}\left(\frac{\partial^2 l_t(\theta_0)}{\partial\theta\partial\theta'}\right) = \frac{1}{n}\sum_{t=1}^{n} \mathrm{E}(H_t(\theta_0)). \tag{4.82}$$

The negative of the matrix of second-order partial derivatives of the log likelihood with respect to the parameters, $H_t(\theta) \equiv -\partial^2 l_t(\theta)/\partial\theta\partial\theta'$, is called the Hessian. The matrix A_0 can be consistently estimated by its sample analogue

$$A_n(\hat{\theta}_{\mathrm{ML}}) = -\frac{1}{n}\sum_{t=1}^{n}\left(\frac{\partial^2 l_t(\hat{\theta}_{\mathrm{ML}})}{\partial\theta\partial\theta'}\right). \tag{4.83}$$

As argued in section 4.1.2, conditional normality of ε_t is often not a very realistic assumption for high-frequency financial time series, as the resulting model fails to capture the kurtosis in the data. Instead, one sometimes assumes that z_t is drawn from a (standardized) Student-t distribution given in (4.68) or any other distribution. The parameters in the GARCH models can then be estimated by maximizing the log likelihood corresponding with this particular distribution. As one can never be sure that the specified distribution of z_t is the correct one, an alternative approach is to ignore the problem and base the likelihood on the normal distribution as in (4.78). This method usually is referred to as quasi-maximum likelihood estimation (QMLE). In general, the resulting estimates still are consistent and asymptotically normal, provided that the models for the conditional mean and conditional variance are correctly specified. Weiss (1984, 1986) has demonstrated this for ARCH(q) models as in (4.10), while Bollerslev and Wooldridge (1992), Lee and Hansen (1994) and Lumsdaine (1996) have obtained the same result where h_t follows a GARCH(1,1) model as in (4.13), under varying assumptions on the properties of z_t.

Interestingly, consistency and asymptotic normality of the QMLE estimates do not require that the parameters in the GARCH(1,1) model satisfy the covariance-stationarity condition $\alpha_1 + \beta_1 < 1$, but they continue to hold for the IGARCH(1,1) model. This is another difference with unit root models for the conditional mean. Recall from chapter 2 that the properties of the estimates of, for example, autoregressive parameters change dramatically where the model contains a unit root.

As the true distribution of z_t is not assumed to be the same as the normal distribution which is used to construct the likelihood function, the standard errors of the parameters have to be adjusted accordingly. In particular, the asymptotic covariance matrix of $\sqrt{n}(\hat{\theta} - \theta_0)$ is equal to $A_0^{-1} B_0 A_0^{-1}$, where A_0 is the information matrix (4.82) and B_0 is the expected value of the outer product of the gradient matrix,

$$B_0 = \frac{1}{n} \sum_{t=1}^{n} \mathrm{E}\left(\frac{\partial l_t(\theta_0)}{\partial \theta} \frac{\partial l_t(\theta_0)}{\partial \theta'} \right) = \frac{1}{n} \sum_{t=1}^{n} \mathrm{E}(s_t(\theta_0) s_t(\theta_0)'). \tag{4.84}$$

The asymptotic covariance matrix can be estimated consistently by using the sample analogues for both A_0, as given in (4.83), and B_0, given by

$$B_n(\hat{\theta}_{\mathrm{ML}}) = \frac{1}{n} \sum_{t=1}^{n} \left(\frac{\partial l_t(\hat{\theta}_{\mathrm{ML}})}{\partial \theta} \frac{\partial l_t(\hat{\theta}_{\mathrm{ML}})}{\partial \theta'} \right) = \frac{1}{n} \sum_{t=1}^{n} s_t(\hat{\theta}_{\mathrm{ML}}) s_t(\hat{\theta}_{\mathrm{ML}})'. \tag{4.85}$$

The finite sample properties of the quasi-maximum likelihood estimates for GARCH(1,1) models are considered in Engle and González-Rivera (1991) and Bollerslev and Wooldridge (1992). It appears that as long as the distribution of z_t is symmetric, QMLE is reasonably accurate and close to the estimates obtained from exact MLE methods, while for skewed distributions this is no longer the case. Lumsdaine (1995) investigates the finite sample properties of the MLE method where the series follow an IGARCH model and she concludes that this method is quite accurate.

The iterative optimization procedures that can be used to estimate the parameters typically require the first- and second-order derivatives of the log likelihood with respect to θ – that is, the score $s_t(\theta)$ and Hessian matrix $H_t(\theta)$ defined above. For example, the iterations in the well known Newton–Raphson method take the form

$$\hat{\theta}^{(m)} = \hat{\theta}^{(m-1)} - \lambda \left(\sum_{t=1}^{n} H_t(\hat{\theta}^{(m-1)}) \right)^{-1} \sum_{t=1}^{n} s_t(\hat{\theta}^{(m-1)}), \tag{4.86}$$

where $\hat{\theta}^{(m)}$ is the estimate of the parameter vector obtained in the mth iteration and the scalar λ denotes a step size. In the algorithm of Berndt *et al.* (1974) (BHHH), which is by far the most popular method to estimate GARCH models, the Hessian $H_t(\hat{\theta}^{(m-1)})$ in (4.86) is replaced by the outer product of the gradient matrix $B_n(\hat{\theta}^{(m-1)})$ obtained from (4.85). It is common to use numerical approximations to these quantities, as the analytical derivatives are fairly complex and contain recursions which are thought to be too cumbersome to compute. However, Fiorentini, Calzolari and Panatoni (1996) show that this is not the case and suggest that it might be advantageous to use analytic derivatives. In general, convergence of the optimization algorithm requires much less iteration, whereas the standard errors of the parameter estimates are far more accurate.

At the outset of this section, it was assumed that the conditional mean function $G(x_t; \xi)$ is at least twice continuously differentiable with respect to the parameters ξ. The STAR and Markov-Switching models discussed in chapter 3 obviously satisfy this requirement. Specification of STAR models for the conditional mean combined with GARCH models for the conditional variance is discussed in detail in Lundbergh and Teräsvirta (1998a). The parameters of such models can be estimated using the (Q)MLE method described above. To estimate the parameters of a model with Markov-Switching in either the conditional mean or variance (or both), the algorithm discussed in section 3.2.3 can be used (see Hamilton and Susmel, 1994, and Dueker, 1997). The SETAR model is not continuous and, hence, the parameters of a SETAR-GARCH models cannot be estimated by the (quasi-)maximum likelihood method outlined above. Li and Li (1996) suggest an alternative estimation procedure for such models, see also Liu, Li and Li (1997).

Example 4.6: Nonlinear GARCH models for Tokyo stock index returns
We estimate several nonlinear variants of the GARCH(1,1) model for weekly returns on the Tokyo stock index (as the relevant tests computed earlier indicate their potential usefulness), for the 10-year sample from January 1986 until December 1995. For convenience, we assume that the conditional mean of the returns is constant and need not be described by a linear or nonlinear model. Parameter estimates for GARCH, GJR-GARCH, QGARCH and VS-GARCH models are given in table 4.9. In the GJR-GARCH model, the estimate of α_1 is larger than the estimate of γ_1, which implies that negative shocks have a larger effect on conditional volatility than positive shocks of the same magnitude. The negative estimate of γ_1 in the QGARCH model implies that the news impact curve (NIC) is shifted to the right, relative to the standard GARCH(1,1) model, implying that positive shocks have smaller impact on the conditional variance than negative shocks. The VS-GARCH model also implies rather different behaviour of h_t following negative and positive shocks. In particular, if the shock in the previous period was negative, the conditional volatility process is explosive (as $\hat{\alpha}_1 + \hat{\beta}_1 > 1$).

The conditional standard deviations as implied by the GARCH(1,1) model and the three nonlinear variants discussed here are shown in figure 4.3. These plots confirm that the main difference between the linear GARCH model and the GJR-GARCH and QGARCH model is the response of conditional volatility to positive shocks. For example, in the second week of August 1992 and the first week of July 1995, the Nikkei index experienced large positive returns of 12.1 per cent and 11.3 per cent, respectively. According to the GARCH model,

Table 4.9 *Estimates of nonlinear GARCH(1,1) models for weekly returns on the Tokyo stock index*

	ω	α_1	β_1	ζ	γ_1	δ_1	$\bar{l}_t(\hat{\theta}_{\mathrm{ML}})$	$\mathrm{SK}_{\hat{z}}$	$\mathrm{K}_{\hat{z}}$
GARCH	0.313	0.192	0.789				−2.404	−0.64	4.20
	(0.117)	(0.040)	(0.039)						
	[0.134]	[0.040]	[0.038]						
	{0.155}	{0.042}	{0.039}						
GJR-GARCH	0.328	0.231	0.815		0.067		−2.393	−0.50	4.18
	(0.105)	(0.048)	(0.035)		(0.035)				
	[0.133]	[0.052]	[0.038]		[0.035]				
	{0.172}	{0.058}	{0.044}		{0.036}				
QGARCH	0.394	0.158	0.804		−0.296		−2.394	−0.47	4.07
	(0.113)	(0.038)	(0.036)		(0.078)				
	[0.143]	[0.038]	[0.039]		[0.092]				
	{0.184}	{0.038}	{0.044}		{0.110}				
VS-GARCH	0.000	0.214	0.920	0.552	0.062	0.748	−2.390	−0.49	4.52
	(0.328)	(0.045)	(0.076)	(0.290)	(0.035)	(0.061)			
	[0.002]	[0.049]	[0.050]	[0.198]	[0.035]	[0.052]			
	{0.000}	{0.059}	{0.056}	{0.235}	{0.039}	{0.056}			

Notes: The estimation sample runs from January 1986 until December 1995.
Figures in round, straight and curly brackets are standard errors based on the outer product of the gradient matrix $B_n(\hat{\theta}_{\mathrm{ML}})$ as given in (4.85), the Hessian matrix $A_n(\hat{\theta}_{\mathrm{ML}})$ as given in (4.83) and the robust quasi-maximum likelihood covariance estimator $A_n^{-1}B_nA_n^{-1}$, respectively.
$\bar{l}_t(\hat{\theta}_{\mathrm{ML}})$ denotes the mean log likelihood evaluated at the maximum likelihood estimates, $\mathrm{SK}_{\hat{z}}$ and $\mathrm{K}_{\hat{z}}$ denote the skewness and kurtosis of the standardized residuals $\hat{z}_t = \hat{\varepsilon}_t\hat{h}_t^{-1/2}$, respectively.

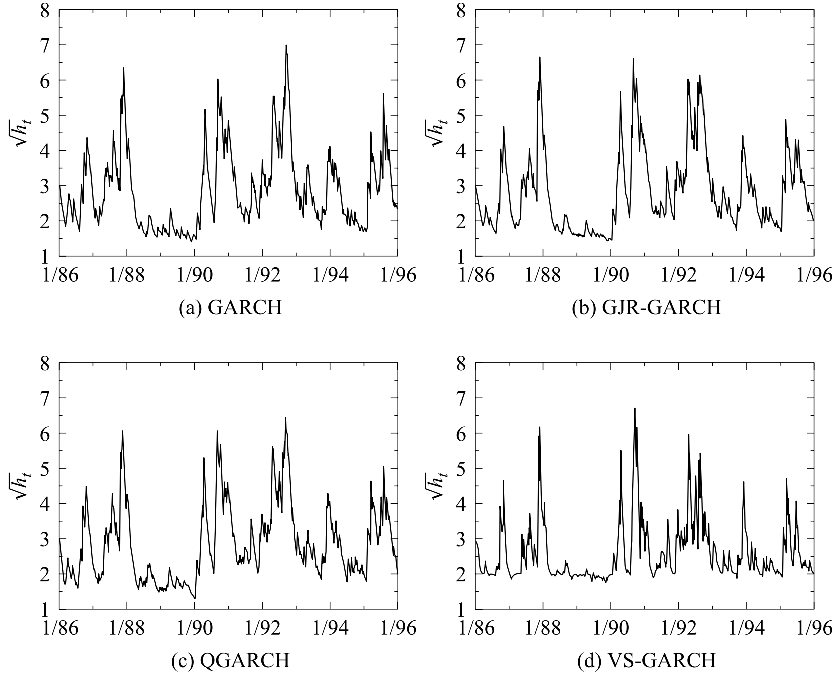

Figure 4.3 Conditional standard deviation for weekly returns on the Tokyo stock index as implied by estimated nonlinear GARCH(1,1) models

the volatility of subsequent returns increases considerably, whereas according to the QGARCH and especially the GJR-GARCH model this was not the case.

4.3.2 Estimation of linear GARCH models

As the parameter vector θ consists of the parameters in the models for the conditional mean and variance, the Hessian matrix $H_t(\theta)$ can be partitioned as

$$H_t(\theta) = \begin{pmatrix} \dfrac{\partial^2 l_t(\theta)}{\partial\xi\partial\xi'} & \dfrac{\partial^2 l_t(\theta)}{\partial\xi\partial\psi'} \\ \dfrac{\partial^2 l_t(\theta)}{\partial\psi\partial\xi'} & \dfrac{\partial^2 l_t(\theta)}{\partial\psi\partial\psi'} \end{pmatrix} = \begin{pmatrix} H_t^{\xi\xi}(\theta) & H_t^{\xi\psi}(\theta) \\ H_t^{\xi\psi}(\theta)' & H_t^{\psi\psi}(\theta) \end{pmatrix}. \tag{4.87}$$

Where the conditional variance h_t is a symmetric function of ε_t, it can be shown that the expected values of the elements in the block $H_t^{\xi\psi}(\theta)$ are equal to zero (see Engle, 1982). It then follows that the Hessian and, based on (4.82), the information matrix are block-diagonal, which implies that consistent and asymptotically efficient estimates of the parameters ξ and ψ can be obtained separately. In this case, the parameters in the model for the conditional mean can be estimated in a first step by (nonlinear) least squares. In a second step, the parameters in the GARCH model are estimated with maximum likelihood, using the residuals $\hat{\varepsilon}_t$ obtained in the first step. Of the GARCH models discussed in section 4.1, only the basic ARCH and GARCH models and the ESTGARCH model describe the conditional variance as a symmetric function of ε_t. For the other nonlinear GARCH models, the information matrix is not block-diagonal and the parameters in the model for the conditional mean and variance have to be estimated jointly.

Example 4.7: GARCH models for stock index and exchange rate returns
To illustrate the methods discussed in this section, we estimate GARCH(1,1) models for some selected weekly stock index and exchange rate returns. Again we assume that the conditional mean of the series is constant. The parameter estimates that are reported in table 4.10 illustrate the typical findings in empirical applications of the GARCH(1,1) model. For all time series, the estimate of

Table 4.10 *Estimates of GARCH(1,1) models for weekly stock index and exchange rate returns*

Stock index	ω	α_1	β_1	Exchange rate	ω	α_1	β_1
Frankfurt	1.560	0.144	0.635	British pound	0.171	0.071	0.856
	(0.490)	(0.019)	(0.085)		(0.061)	(0.025)	(0.042)
	[0.790]	[0.045]	[0.142]		[0.105]	[0.028]	[0.062]
	{1.742}	{0.118}	{0.330}		{0.191}	{0.037}	{0.105}
New York	0.082	0.099	0.888	French franc	0.367	0.097	0.748
	(0.033)	(0.019)	(0.025)		(0.195)	(0.035)	(0.108)
	[0.054]	[0.038]	[0.041]		[0.174]	[0.037]	[0.093]
	{0.097}	{0.092}	{0.086}		{0.160}	{0.043}	{0.083}
Paris	0.468	0.074	0.852	German Dmark	0.405	0.088	0.754
	(0.315)	(0.039)	(0.080)		(0.239)	(0.039)	(0.122)
	[0.256]	[0.029]	[0.057]		[0.184]	[0.036]	[0.089]
	{0.255}	{0.023}	{0.043}		{0.153}	{0.038}	{0.068}

Estimates of GARCH(1,1) models, $h_t = \omega + \alpha_1 \varepsilon_{t-1}^2 + \beta_1 h_{t-1}$, for weekly stock index and exchange rate returns.
The sample runs from January 1986 until December 1995.
Figures in round, straight and curly brackets are standard errors based on the outer product of the gradient matrix $B_n(\hat{\theta}_{\mathrm{ML}})$ as given in (4.85), the Hessian matrix $A_n(\hat{\theta}_{\mathrm{ML}})$ as given in (4.83) and the robust quasi-maximum likelihood covariance estimator $A_n^{-1} B_n A_n^{-1}$, respectively.

α_1 is fairly small, the estimate of β_1 is large and the sum $\alpha_1 + \beta_1$ is close to unity. The standard errors of the parameter estimates which are given in brackets demonstrate that large differences may exist between the different methods to compute them.

Figure 4.4 shows the conditional standard deviation as implied by the estimated GARCH(1,1) models together with the absolute values of the returns series. One feature which stands out from all graphs is that the models tend to overestimate the conditional volatility during relatively quiet periods. The absolute returns suggest that periods of large changes in the stock indexes and exchange rates are relatively short-lived, and the return to more quiet spells occurs quickly. By contrast, the estimates of the parameters in the GARCH(1,1) model imply that conditional volatility is persistent, in the sense that shocks to the conditional variance die out very slowly. Notice again, though, that the GARCH model describes *unobserved* volatility. It may well be that this latent variable displays strong persistence, but that this does not feed through the returns because of small values of z_t.

4.3.3 Robust estimation of GARCH models

In subsection 4.2.4, it was argued that neglected outliers might easily be mistaken for conditional heteroscedasticity. The evidence for time variation in the conditional volatility of high-frequency financial time series is so overwhelming, though, that it may seem odd to maintain that all of this is caused entirely by one-time exogenous events. However, even if conditional heteroscedasticity is a characteristic of the time series under study, it might still happen that outliers occur (see also Friedman and Laibson, 1989, for a theoretical motivation). Hence, it is of interest to consider statistical methods for inference in GARCH models that are applicable where such aberrant observations are present.

Several approaches to handle outliers in GARCH models have been investigated. Sakata and White (1998) consider outlier-robust estimation for GARCH models, using techniques similar to the ones discussed in subsection 3.2.4. Hotta and Tsay (1998) derive test-statistics to detect outliers in a GARCH model, distinguishing between outliers which do and which do not affect the conditional volatility. Franses and Ghijsels (1999) apply the outlier detection method of Chen and Liu (1993) to GARCH models. For illustrative purposes, the latter method is discussed in more detail below.

Consider the GARCH(1,1) model

$$\varepsilon_t = z_t\sqrt{h_t}, \tag{4.88}$$

$$h_t = \omega + \alpha_1\varepsilon_{t-1}^2 + \beta_1 h_{t-1}, \tag{4.89}$$

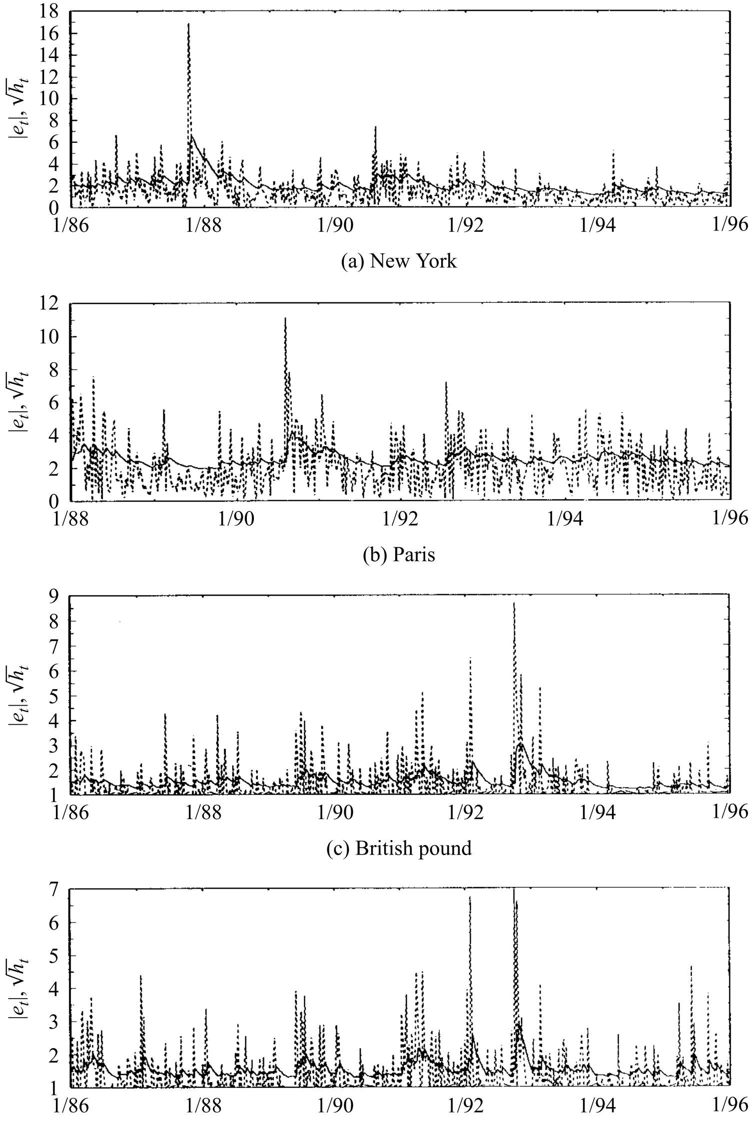

Figure 4.4 Conditional standard deviation for weekly stock index returns as implied by estimated GARCH(1,1) models

where $\omega > 0$, $\alpha_1 > 0$, $\beta_1 > 0$ and $\alpha_1 + \beta_1 < 1$, such that the model is covariance-stationary. For simplicity, we ignore the conditional mean of the observed time series y_t and simply assume that it is equal to 0 – that is, $\varepsilon_t = y_t$. As shown in subsection 4.1.1, the GARCH(1,1) model (4.89) can be rewritten as an ARMA(1,1) model for ε_t^2,

$$\varepsilon_t^2 = \omega + (\alpha_1 + \beta_1)\varepsilon_{t-1}^2 + \nu_t - \beta_1 \nu_{t-1}, \tag{4.90}$$

where $\nu_t = \varepsilon_t^2 - h_t$. Franses and Ghijsels (1999) exploit this analogy of the GARCH model with an ARMA model to adapt the method of Chen and Liu (1993) to detect and correct (additive) outliers in GARCH models. Specifically, suppose that instead of the true series ε_t one observes the series e_t which is defined by

$$e_t^2 = \varepsilon_t^2 + \zeta I[t = \tau], \tag{4.91}$$

where $I[t = \tau]$ is the indicator function defined as $I[t = \tau] = 1$ if $t = \tau$ and zero otherwise, and where ζ is a nonzero constant. Define the lag polynomial $\pi(L)$ as

$$\begin{aligned} \pi(L) &= (1 - \beta_1 L)^{-1}(1 - (\alpha_1 + \beta_1)L) \\ &= \left(1 + \beta_1 L + \beta_1^2 L^2 + \beta_1^3 L^3 + \cdots\right)(1 - (\alpha_1 + \beta_1)L) \\ &= 1 - \alpha_1 L - \alpha_1\beta_1 L^2 - \alpha_1\beta_1^2 L^3 - \cdots . \end{aligned} \tag{4.92}$$

The polynomial $\pi(L)$ allows (4.90) to be written as $\nu_t = -\omega/(1-\beta_1) + \pi(L)\varepsilon_t^2$. Similarly, where the GARCH(1,1) model is applied to the observed series e_t^2, it is straightforward to show that the corresponding residuals v_t are given by

$$\begin{aligned} v_t &= \frac{-\omega}{1-\beta_1} + \pi(L)e_t^2 \\ &= \frac{-\omega}{1-\beta_1} + \pi(L)(\varepsilon_t^2 + \zeta I[t = \tau]) \\ &= \nu_t + \pi(L)\zeta I[t = \tau]. \end{aligned} \tag{4.93}$$

The last line of (4.93) can be interpreted as a regression model for v_t, that is,

$$v_t = \zeta x_t + \nu_t, \tag{4.94}$$

with

$$\begin{aligned} x_t &= 0 \qquad \text{for } t < \tau, \\ x_\tau &= 1, \\ x_{\tau+k} &= -\pi_k \quad \text{for } k = 1, 2, \ldots . \end{aligned}$$

The magnitude ζ of the outlier at time $t = \tau$ then can be estimated as

$$\hat{\zeta}(\tau) = \left(\sum_{t=\tau}^{n} x_t^2\right)^{-1} \left(\sum_{t=\tau}^{n} x_t v_t\right). \tag{4.95}$$

For fixed τ, the t-statistic of $\hat{\zeta}(\tau)$, denoted as $t_{\hat{\zeta}(\tau)}$, has an asymptotic standard normal distribution. Hence, one can test for an outlier at time $t = \tau$ by comparing $t_{\hat{\zeta}(\tau)}$ with the normal critical value. In practice, the timing of possible outliers is of course unknown. In that case, an intuitively plausible test-statistic is the maximum of the absolute values of the t-statistic over the entire sample, that is,

$$t_{\max}(\hat{\zeta}) \equiv \max_{1\leq\tau\leq n} |t_{\hat{\zeta}(\tau)}|. \tag{4.96}$$

The distribution of $t_{\max}(\hat{\zeta})$ is nonstandard. Usually it is compared with a pre-specified critical value C to determine whether an outlier has occurred.

The outlier detection method for GARCH(1,1) models then consists of the following steps.

(1) Estimate a GARCH(1,1) model for the observed series e_t and obtain estimates of the conditional variance $\hat{h}_t$ and $\hat{v}_t \equiv e_t^2 - \hat{h}_t$.

(2) Obtain estimates $\hat{\zeta}(\tau)$ for all possible $\tau = 1, \ldots, n$, using (4.95) and compute the test-statistic $t_{\max}(\hat{\zeta})$ from (4.96). If the value of the test-statistic exceeds the pre-specified critical value C an outlier is detected at the observation for which the t-statistic of $\hat{\zeta}$ is maximized (in absolute value), say $\hat{\tau}$.

(3) Replace $e_{\hat{\tau}}^2$ with $e_{\hat{\tau}}^{*2} \equiv e_{\hat{\tau}}^2 - \hat{\zeta}(\hat{\tau})$ and define the outlier corrected series e_t^* as $e_t^* = e_t$ for $t \neq \hat{\tau}$ and

$$e_{\hat{\tau}}^* = \operatorname{sgn}(e_{\hat{\tau}})\sqrt{e_{\hat{\tau}}^{*2}}.$$

(4) Return to step (1) to estimate a GARCH(1,1) model for the series e_t^*.

The iterations terminate if the $t_{\max}(\hat{\zeta})$ statistic no longer exceeds the critical value C.

Example 4.8: Outlier detection in GARCH models for stock index returns
We apply the outlier detection method for GARCH(1,1) models to weekly stock index returns using a critical value $C = 10$. This choice for C is based on the outcome of the following simulation experiment. We generate 1,000 series of $n = 250$ and 500 observations from the GARCH(1,1) model (4.88) with (4.89) for various values of α_1 and β_1 and set $\omega = 1 - \alpha_1 - \beta_1$. For each series, we estimate a GARCH(1,1) model and compute the outlier detection statistic $t_{\max}(\hat{\zeta})$ as given in (4.96). In this way we obtain an estimate of the distribution

Table 4.11 *Percentiles of the distribution of the outlier detection statistic in GARCH(1,1) models*

		$n = 250$				$n = 500$			
α_1	β_1	0.80	0.90	0.95	0.99	0.80	0.90	0.95	0.99
0.10	0.50	7.68	8.80	9.67	12.50	8.61	9.96	10.94	14.86
0.10	0.60	7.79	8.84	9.95	12.61	8.70	9.90	11.11	14.99
0.10	0.70	7.94	8.99	10.42	13.27	8.90	10.23	11.35	15.66
0.10	0.80	8.21	9.59	10.59	15.09	9.36	10.73	12.27	16.93
0.20	0.50	8.86	10.54	11.68	16.46	10.35	12.01	15.23	20.65
0.20	0.60	9.11	10.92	12.82	17.91	10.72	12.97	15.45	23.23
0.20	0.70	9.87	11.83	13.96	18.98	11.82	14.29	16.93	25.01

Notes: Percentiles of the distribution of the $t_{\max}(\hat{\zeta})$-statistic (4.96) for detection of AOs in the GARCH(1,1) model (4.88) and (4.89), based on 1,000 replications.

of the $t_{\max}(\hat{\zeta})$-statistic under the null hypothesis that no outliers are present. Table 4.11 shows some percentiles of this distribution for several values of α_1 and β_1. It is seen that the value of $C = 10$ is reasonably close to the 90th percentile of this distribution for most parameter combinations. Notice that the value $C = 4$, as recommended in Franses and Ghijsels (1999), would imply a much larger nominal size.

Table 4.12 shows estimates of GARCH(1,1) models for weekly returns on the Amsterdam (AEX) and New York (S&P 500) stock indexes over the sample period January 1986–December 1990, before and after applying the outlier-correction method. For both series, two AOs are detected in the weeks ending 21 and 28 October. The outlier-statistic $t_{\max}(\hat{\zeta})$ takes the values 14.59 and 22.14 for the AEX returns, with corresponding magnitudes of the AOs $\hat{\zeta} = -12.11$ and -18.74. For the returns on the S&P 500, $t_{\max}(\hat{\zeta}) = 29.92$ and 11.52 with outliers of size $\hat{\zeta} = -15.73$ and -9.62, respectively. The parameter estimates in table 4.12 demonstrate that these can be heavily influenced by only very few aberrant observations. Also notice that removing these outliers causes the skewness and kurtosis of the standardized residuals $z_t = \varepsilon_t/\sqrt{h_t}$ to be closer to the normal values of 0 and 3, respectively.

4.4 Diagnostic checking

Just as it is good practice to check the adequacy of a time series model for the conditional mean by computing a number of misspecification tests, such

Table 4.12 *Estimates of GARCH(1,1) models for weekly returns on the Amsterdam and New York stock indexes, before and after outlier correction*

	μ	ω	α_1	β_1	$SK_{\hat{z}}$	$K_{\hat{z}}$
Amsterdam						
Before	0.203	2.220	0.311	0.419	−0.75	5.01
	(0.162)	(0.677)	(0.064)	(0.106)		
After	0.093	0.893	0.110	0.750	−0.56	4.03
	(0.164)	(0.414)	(0.058)	(0.109)		
New York						
Before	0.402	2.248	0.381	0.302	−1.09	5.51
	(0.136)	(0.629)	(0.054)	(0.107)		
After	0.254	0.236	0.052	0.902	−0.62	3.75
	(0.145)	(0.190)	(0.032)	(0.060)		

Notes: Estimates of GARCH(1,1) models for weekly returns on the Amsterdam and New York stock indexes, $y_t = \mu + \varepsilon_t$, with $\varepsilon_t = z_t\sqrt{h_t}$ and $h_t = \omega + \alpha_1\varepsilon_{t-1}^2 + \beta_1 h_{t-1}$.
Models are estimated for the sample January 1986–December 1990.
Standard errors based on the outer product of the gradient are given in parentheses.
The final two columns contain the skewness and kurtosis of the standardized residuals $\hat{z}_t = \hat{\varepsilon}_t\hat{h}_t^{-1/2}$, respectively.

diagnostic checking should also be part of a specification strategy for models for the conditional variance. In this section we discuss tests which might be used for this purpose.

Testing properties of standardized residuals

One of the assumptions which is made in GARCH models is that the innovations $z_t = \varepsilon_t h_t^{-1/2}$ are independent and identically distributed. Hence, if the model is correctly specified, the standardized residuals $\hat{z}_t = \hat{\varepsilon}_t\hat{h}_t^{-1/2}$ should possess the classical properties of well behaved regression errors, such as constant variance, lack of serial correlation, and so on. Standard test-statistics as discussed in section 2.2 can be used to determine whether this is the case or not.

Of particular interest is to test whether the standardized residuals still contain signs of conditional heteroscedasticity. Li and Mak (1994) and Lundbergh and Teräsvirta (1998b) develop statistics which can be used to test for remaining ARCH in the standardized residuals, which are variants of the statistics of McLeod and Li (1983) and Engle (1982), respectively. For example, the LM test for remaining ARCH(m) in $\hat{z}_t$ proposed by Lundbergh and Teräsvirta (1998b) can be computed as nR^2, where R^2 is obtained from the auxiliary regression

$$\hat{z}_t^2 = \phi_0 + \phi_1 \hat{z}_{t-1}^2 + \cdots + \phi_m \hat{z}_{t-m}^2 + \lambda' \hat{x}_t + u_t, \tag{4.97}$$

where the vector $\hat{x}_t$ consists of the partial derivatives of the conditional variance h_t with respect to the parameters in the original GARCH model, evaluated under the null hypothesis – that is, $\hat{x}_t \equiv \hat{h}_t^{-1} \partial \hat{h}_t / \partial \theta$. For example, in the case of a GARCH(1,1) model

$$h_t = \omega + \alpha_1 \varepsilon_{t-1}^2 + \beta_1 h_{t-1}, \tag{4.98}$$

it follows that

$$\frac{\partial h_t}{\partial \theta'} = \left(1, \varepsilon_{t-1}^2, h_{t-1}\right) + \beta_1 \frac{\partial h_{t-1}}{\partial \theta'}. \tag{4.99}$$

As the pre-sample conditional variance h_0 is usually computed as the sample average of the squared residuals, $h_0 = 1/n \sum_{t=1}^{n} \varepsilon_t^2$, h_0 does not depend on θ, and $\partial h_0 / \partial \theta = 0$. This allows (4.99) to be computed recursively. Alternatively, the partial derivatives can be obtained by recursive substitution as

$$\hat{x}_t' = \left(\frac{\sum_{i=1}^{t-1} \hat{\beta}^{i-1}}{\hat{h}_t}, \frac{\sum_{i=1}^{t-1} \hat{\beta}^{i-1} \hat{\varepsilon}_{t-i}^2}{\hat{h}_t}, \frac{\sum_{i=1}^{t-1} \hat{\beta}^{i-1} \hat{h}_{t-i}}{\hat{h}_t} \right). \tag{4.100}$$

The test-statistic based on (4.97), which tests the null hypothesis $H_0 : \phi_1 = \cdots = \phi_m = 0$ is asymptotically χ^2 distributed with m degrees of freedom.

Testing for higher-order GARCH

The statistic discussed above tests for correlation in the squared standardized residuals. This is closely related to the LM-statistics discussed by Bollerslev (1986), which can be used to test a GARCH(p, q) specification against either a GARCH$(p + r, q)$ or GARCH$(p, q + s)$ alternative. The test-statistics are given by n times the R^2 from the auxiliary regression (4.97), with the lagged squared standardized residuals $\hat{z}_{t-i}^2$, $i = 1, \ldots, m$ replaced by $\hat{\varepsilon}_{t-q-1}^2, \ldots, \varepsilon_{t-q-r}^2$ or $\hat{h}_{t-p-1}, \ldots, \hat{h}_{t-p-s}$, respectively.

Misspecification tests for linear GARCH models

As discussed in subsection 3.1.2, in specifying a suitable model for the conditional mean of a time series, it is common practice to start with a linear model and consider nonlinear models only if diagnostic checks indicate that the linear model is inadequate. A similar strategy can be pursued when specifying a model for the conditional variance. That is, one may start with specifying and estimating a linear GARCH model and move on to nonlinear variants only if certain misspecification tests suggest that symmetry of the conditional variance function is an untenable assumption.

One possible method to test a linear GARCH specification against nonlinear alternatives is by means of the Sign Bias, Negative Size Bias and Positive Size Bias tests of Engle and Ng (1993), discussed in subsection 4.2.2. In this case, the squared standardized residuals $\hat{z}_t^2$ should be taken as the dependent variable in the regressions that are involved, while the partial derivatives $\hat{x}_t \equiv \hat{h}_t^{-1}\partial\hat{h}_t/\partial\theta$ should be added as regressors. The analogue of (4.70) is given by

$$\hat{z}_t^2 = \phi_0 + \phi_1\hat{w}_{t-1} + \lambda'\hat{x}_t + \xi_t, \tag{4.101}$$

where $\hat{w}_{t-1}$ is taken equal to one of the three measures of asymmetry, S_{t-1}^-, $S_{t-1}^-\hat{\varepsilon}_{t-1}$ or $S_{t-1}^+\hat{\varepsilon}_{t-1}$. Similarly, the analogue of (4.71) is

$$\hat{z}_t^2 = \phi_0 + \phi_1 S_{t-1}^- + \phi_2 S_{t-1}^-\hat{\varepsilon}_{t-1} + \phi_3 S_{t-1}^+\hat{\varepsilon}_{t-1} + \lambda'\hat{x}_t + \xi_t. \tag{4.102}$$

Hagerud (1997) examines the ability of the Sign Bias, Negative Size Bias and Positive Size Bias tests to detect various of the asymmetric GARCH effects discussed in subsection 4.1.2 by means of Monte Carlo simulation. In general, the power of the statistics is not very high. Moreover, rejection of the null hypothesis by one or several of the tests does not give much information concerning which nonlinear GARCH model might be the appropriate alternative. It turns out to be difficult to obtain such information on the basis of statistical tests, even if one uses statistics which are designed against a particular alternative model.

Hagerud (1997) develops statistics to test the linear GARCH(1,1) model against the QGARCH(1,1) model given in (4.63) and the LSTGARCH(1,1) model given in (4.55). The test against the QGARCH(1,1) alternative can be computed as nR^2 from the regression of the squared standardized residuals $\hat{z}_t^2$ on the elements of $\hat{x}_t$ given in (4.100) and $\left(\sum_{i=1}^{t-1}\hat{\beta}^{i-1}\hat{\varepsilon}_{t-i}\right)/\hat{h}_t$. This latter quantity is the partial derivative of the conditional variance h_t with respect to the parameter γ_1 in (4.63), evaluated under the null hypothesis $\gamma_1 = 0$. The test-statistic is asymptotically χ^2-distributed with 1 degree of freedom.

If one wants to test the null hypothesis of GARCH(1,1) against the LSTGARCH(1,1) alternative, the same identification problem occurs as encountered when testing homoscedasticity against the LSTARCH alternative, discussed in subsection 4.2.2. The solution is again to replace the logistic function $F(\varepsilon_{t-1})$ in (4.55) with a first-order Taylor approximation, yielding the auxiliary model

$$h_t = \omega^* + \alpha_1^*\varepsilon_{t-1}^2 + \gamma_1^*\varepsilon_{t-1}^3 + \beta_1 h_{t-1}, \tag{4.103}$$

where ω^*, α_1^* and γ_1^* are functions of the parameters in the original model. The null hypothesis $H_0 : \gamma_1^* = 0$ can now be tested by computing nR^2 from the regression of $\hat{z}_t^2$ on $\hat{x}_t$ as given in (4.100) and $\left(\sum_{i=1}^{t-1} \hat{\beta}^{i-1}\hat{\varepsilon}_{t-i}^3\right)/\hat{h}_t$, which is the partial derivative of h_t with respect to γ_1^* evaluated under $\gamma_1^* = 0$. The resultant test statistic has an asymptotic χ^2 distribution with 1 degree of freedom under the null hypothesis. Even though these test-statistics are derived against an explicit alternative model, it turns out that they also reject the null hypothesis quite often in case of time series generated from other nonlinear GARCH models (see Hagerud, 1997). Hence, these tests cannot be used to distinguish between different forms of nonlinear GARCH, but should instead be interpreted as tests against general asymmetry.

Testing parameter constancy

In empirical applications, GARCH models are frequently estimated for time series which cover a very long period of time, sometimes up to 75 years (see, for example Ding, Granger and Engle, 1993). It is hard to imagine that the properties of a time series over such a long time period can be captured by a model with constant parameters. In fact, it has been argued that the typical estimates of the parameters in the GARCH model, implying very strong persistence of shocks, might be caused by occasional shifts in the parameters (see Lamoureux and Lastrapes, 1990; Franses, 1995, among others). Chu (1995) develops a test for parameter constancy in GARCH models against a single structural break, while Lundbergh and Teräsvirta (1998b) consider testing parameter constancy against the alternative of smoothly changing parameters. In the latter case, constancy of the parameters in a GARCH(1,1) model is tested against the alternative

$$h_t = [\omega + \alpha_1\varepsilon_{t-1}^2 + \beta_1 h_{t-1}][1 - F(t)] + [\zeta + \gamma_1\varepsilon_{t-1}^2 + \delta_1 h_{t-1}]F(t), \tag{4.104}$$

where $F(.)$ is the logistic function given in (4.56). Notice that the model in (4.104) is similar to the ANST-GARCH model given in (4.61). Testing the null hypothesis of parameter constancy, or $H_0 : \omega = \zeta$, $\alpha_1 = \gamma_1$ and $\beta_1 = \delta_1$ is complicated by the fact that the parameter θ in $F(t)$ is not identified under the null hypothesis. Replacing the function $F(t)$ by a first-order Taylor approximation yields the auxiliary model,

$$h_t = \omega^* + \alpha_1^*\varepsilon_{t-1}^2 + \beta_1^* h_{t-1} + \zeta^* t + \gamma_1^*\varepsilon_{t-1}^2 t + \delta_1^* h_{t-1} t. \tag{4.105}$$

The null hypothesis $H_0 : \zeta^* = \gamma_1^* = \delta_1^* = 0$ can now be tested by an LM-statistic in a straightforward manner. It also is possible to test constancy of individual parameters in a GARCH(1,1) model, assuming that the remaining ones are constant.

Example 4.9: Evaluating estimated GARCH models for stock index and exchange rate returns Table 4.13 contains p-values of diagnostic tests for the GARCH(1,1) models that are estimated for weekly stock index and exchange rate returns. Especially for the stock index returns, the small p-values indicate that the models suffer from various kinds of misspecification. As most tests also have power against alternatives other than the one for which they are designed, the statistics are not helpful in deciding exactly in which direction one should proceed.

4.5 Forecasting

The presence of time-varying volatility has some pronounced consequences for out-of-sample forecasting. Most of these effects can be understood intuitively.

Table 4.13 *Diagnostic tests for estimated GARCH models for weekly stock index and exchange rate returns*

	Stock index			Exchange rate		
Test	FFT	NY	PRS	BP	FFR	DM
No remaining ARCH ($m = 1$)	0.004	0.145	0.572	0.332	0.055	0.078
Higher-order ARCH ($r = 1$)	0.488	0.068	0.423	0.954	0.032	0.035
Higher-order GARCH ($s = 1$)	0.490	0.068	0.418	0.954	0.032	0.035
Sign Bias	0.183	0.025	0.823	0.562	0.703	0.842
Positive Size Bias	0.067	0.085	0.407	0.158	0.845	0.802
Negative Size Bias	0.042	0.005	0.731	0.050	0.037	0.106
Sign and Size Bias	0.222	0.037	0.800	0.118	0.031	0.067
QGARCH	0.006	0.147	0.002	0.308	0.641	0.171
LSTGARCH	0.001	0.963	0.062	0.633	0.861	0.370
Parameter constancy						
Intercept	0.060	0.022	0.811	0.367	0.809	0.916
ARCH parameter	0.048	0.020	0.552	0.895	0.908	0.932
All parameters	0.065	0.088	0.799	0.338	0.928	0.979
Standardized residuals						
Skewness	−0.680	−1.156	−0.471	0.818	0.338	0.264
Kurtosis	4.881	7.932	3.888	6.312	4.245	4.209
Normality test	0.000	0.000	0.000	0.000	0.000	0.000

Notes: p-values of diagnostic tests for estimated GARCH(1,1) models, $h_t = \omega + \alpha_1 \varepsilon_{t-1}^2 + \beta_1 h_{t-1}$, for weekly stock index and exchange rate returns.
FFT = Frankfurt, NY = New York, PRS = Paris, BP = British pound, FFR = French franc, DM = German Dmark.

First, the optimal h-step-ahead predictor of y_{t+h}, $\hat{y}_{t+h|t}$, is given by the conditional mean, regardless of whether the shocks ε_t are conditionally heteroscedastic or not. The methods discussed in sections 2.3 and 3.5 for forecasting with linear and nonlinear models, respectively, under the assumption of homoscedasticity can still be used in case of heteroscedastic shocks. The analytical expressions for $y_{t+h|t}$ in case of linear models discussed in section 2.3 do not depend on the conditional distribution of y_{t+h}, which implies that the numerical value of $\hat{y}_{t+h|t}$ is the same. For nonlinear models this need not be the case. Second, the conditional variance of the associated h-step-ahead forecast error $e_{t+h|t} = y_{t+h} - \hat{y}_{t+h|t}$ becomes time-varying, which in fact was one of the main motivations for proposing the ARCH model (see Engle, 1982). This makes sense as $e_{t+h|t}$ is a linear combination of the shocks that occur between the forecast origin and the forecast horizon, $\varepsilon_{t+1}, \ldots, \varepsilon_{t+h}$. As the conditional variance of these shocks is time-varying, the conditional variance of any function of these shocks is time-varying as well. Below we discuss these results in detail for the case where the observed time series y_t follows an AR(1) model,

$$y_t = \phi_1 y_{t-1} + \varepsilon_t, \tag{4.106}$$

and the conditional variance of the shocks ε_t is described by a GARCH(1,1) model

$$h_t = \omega + \alpha_1 \varepsilon_{t-1}^2 + \beta_1 h_{t-1}. \tag{4.107}$$

The general case of ARMA(k,l)-GARCH(p,q) models is discussed in Baillie and Bollerslev (1992).

Forecasting the conditional mean in the presence of conditional heteroscedasticity

Let $\hat{y}_{t+h|t}$ denote the h-step-ahead forecast of y_t which minimizes the squared prediction error (SPE)

$$\text{SPE}(h) \equiv \text{E}[e_{t+h|t}^2] = \text{E}[(y_{t+h} - \hat{y}_{t+h|t})^2], \tag{4.108}$$

where $e_{t+h|t}$ is the h-step-ahead forecast error $e_{t+h|t} = y_{t+h} - \hat{y}_{t+h|t}$. Baillie and Bollerslev (1992) show that the forecast that minimizes (4.108) is the same irrespective of whether the shocks ε_t in (4.106) are conditionally homoscedastic or conditionally heteroscedastic. Thus, the optimal h-step-ahead forecast of y_{t+h} is its conditional expectation at time t, that is,

$$\hat{y}_{t+h|t} = \text{E}[y_{t+h}|\Omega_t]. \tag{4.109}$$

For the AR(1) model (4.106) this means that the optimal 1-step-ahead forecast is given by $\hat{y}_{t+1|t} = \phi_1 y_t$. The forecasts for $h > 1$ steps ahead can be obtained from the recursive relationship $\hat{y}_{t+h|t} = \phi_1 \hat{y}_{t+h-1|t}$, or can be computed directly as

$$\hat{y}_{t+h|t} = \phi_1^h y_t. \tag{4.110}$$

For the h-step-ahead prediction error, it follows that

$$\begin{aligned} e_{t+h|t} = y_{t+h} - \hat{y}_{t+h|t} &= \phi_1 y_{t+h-1} + \varepsilon_{t+h} - \phi_1^h y_t \\ &= \phi_1^2 y_{t+h-2} + \phi_1 \varepsilon_{t+h-1} + \varepsilon_{t+h} - \phi_1^h y_t \\ &= \ldots \\ &= \phi_1^h y_t + \sum_{i=1}^{h} \phi_1^{h-i} \varepsilon_{t+i} - \phi_1^h y_t \\ &= \sum_{i=1}^{h} \phi_1^{h-i} \varepsilon_{t+i}. \end{aligned} \tag{4.111}$$

The *conditional* SPE of $e_{t+h|t}$ is given by

$$\begin{aligned} \mathrm{E}[e_{t+h|t}^2 | \Omega_t] &= \mathrm{E}\left[\left(\sum_{i=1}^{h} \phi_1^{h-i} \varepsilon_{t+i}\right)^2 | \Omega_t\right] \\ &= \sum_{i=1}^{h} \phi_1^{2(h-i)} \mathrm{E}[\varepsilon_{t+i}^2 | \Omega_t] \\ &= \sum_{i=1}^{h} \phi_1^{2(h-i)} \mathrm{E}[h_{t+i} | \Omega_t]. \end{aligned} \tag{4.112}$$

In the case of homoscedastic errors, the conditional SPE for the optimal h-step-ahead forecast is constant, as $\mathrm{E}[h_{t+i}|\Omega_t]$ is constant and equal to the unconditional variance of ε_t, σ^2. Expression (4.112) shows that in the case of heteroscedastic errors, the conditional SPE is varying over time. To see the relation between the two, rewrite (4.112) as

$$\mathrm{E}[e_{t+h|t}^2 | \Omega_t] = \sum_{i=1}^{h} \phi_1^{2(h-i)} \sigma^2 + \sum_{i=1}^{h} \phi_1^{2(h-i)} (\mathrm{E}[h_{t+i} | \Omega_t] - \sigma^2). \tag{4.113}$$

The first term on the right-hand-side of (4.113) is the conventional SPE for homoscedastic errors. Notice that the second term on the right-hand-side can be both positive and negative, depending on the conditional expectation of future volatility. Hence, the conditional SPE in the case of heteroscedastic errors can be both larger and smaller than in the case of homoscedastic errors.

Recall that in the homoscedastic case, the SPE converges to the unconditional variance of the model as the forecast horizon increases, that is,

$$\lim_{h\to\infty} \mathrm{E}[e^2_{t+h|t}|\Omega_t] = \lim_{h\to\infty} \sum_{i=1}^{h} \phi_1^{2(h-i)}\sigma^2 = \frac{\sigma^2}{1-\phi^2} \equiv \sigma_y^2. \tag{4.114}$$

Moreover, the convergence is monotonic, in the sense that the h-step-ahead SPE is always smaller than the unconditional variance σ_y^2, while the h-step-ahead SPE is larger than the $(h-1)$-step SPE for all finite horizons h. The convergence of the SPE to the unconditional variance of the time series also holds in the present case of heteroscedastic errors. This follows from the fact that the forecasts of the conditional variance $\mathrm{E}[h_{t+i}|\Omega_t]$ converge to the unconditional variance σ^2 (as will be shown explicitly below). However, the convergence need no longer be monotonic, in the sense that the h-step conditional SPE may be smaller than the $(h-1)$-step conditional SPE. In fact, the conditional SPE may be larger than the unconditional variance of the time series for certain forecast horizons. Intuitively, in periods of large uncertainty, characterized by large values of the conditional variance h_t, it is extremely difficult to forecast the conditional mean of the series y_t accurately. In such cases, the forecast uncertainty may be larger at shorter forecast horizons compared to longer horizons.

The conditional SPE as given in (4.112) might be used to construct prediction intervals. The conditional distribution of the h-step-ahead prediction error $e_{t+h|t}$ is, however, nonnormal and, consequently, the conventional forecasting interval discussed in section 2.3, is not a reliable measure of the true forecast uncertainty. Granger, White and Kamstra (1989) discuss an alternative approach based on quantile estimators (see also Taylor, 1999).

An additional complication in using (4.112) is that conditional expectations of the future conditional variances h_{t+i} at time t are required. How to obtain such forecasts of future volatility is discussed next.

Forecasting the conditional variance

In the case of the GARCH(1,1) model, the conditional expectation of h_{t+s} – or, put differently, the optimal s-step-ahead forecast of the conditional variance – can be computed recursively from

$$\hat{h}_{t+s|t} = \omega + \alpha_1 \hat{\varepsilon}^2_{t+s-1|t} + \beta_1 \hat{h}_{t+s-1|t}, \tag{4.115}$$

where $\hat{\varepsilon}^2_{t+i|t} = \hat{h}_{t+i|t}$ for $i > 0$ by definition, while $\hat{\varepsilon}^2_{t+i|t} = \varepsilon^2_{t+i}$ and $\hat{h}_{t+i|t} = h_{t+i}$ for $i \leq 0$. Alternatively, by recursive substitution in (4.115) we obtain

$$\hat{h}_{t+s|t} = \omega \sum_{i=0}^{s-1} (\alpha_1+\beta_1)^i + (\alpha_1+\beta_1)^{s-1} h_{t+1}, \tag{4.116}$$

which allows the s-step-ahead forecast to be computed directly from h_{t+1}. Notice that h_{t+1} is contained in the information set Ω_t, as it can be computed from observations $y_t, y_{t-1}, \ldots$ (given knowledge of the parameters in the model). If the GARCH(1,1) model is covariance-stationary with $\alpha_1 + \beta_1 < 1$, (4.116) can be rewritten as

$$\hat{h}_{t+s|t} = \sigma^2 + (\alpha_1 + \beta_1)^{s-1}(h_{t+1} - \sigma^2), \tag{4.117}$$

where $\sigma^2 = \omega/(1 - \alpha_1 - \beta_1)$ is the unconditional variance of ε_t, which shows that the forecasts for the conditional variance are similar to forecasts from an AR(1) model with mean σ^2 and AR parameter $\alpha_1 + \beta_1$. The right-hand-side of (4.117) follows from the fact that $\sum_{i=1}^{s-1} r^i = (1 - r^{s-1})/(1 - r)$ for all r with $|r| < 1$. Also note that for the IGARCH model with $\alpha_1 + \beta_1 = 1$, (4.116) simplifies to

$$\hat{h}_{t+s|t} = \omega(s - 1) + h_{t+1}, \tag{4.118}$$

which shows that the forecasts for the conditional variance increase linearly as the forecast horizon s increases, provided $\omega > 0$.

To express the uncertainty in the s-step-ahead forecast of the conditional variance, one might consider the associated forecast error $v_{t+s|t} \equiv h_{t+s} - \hat{h}_{t+s|t}$. Subtracting the expression of the s-step-ahead forecast in (4.115) from the definition of the GARCH(1,1) model for h_{t+s}, $h_{t+s} = \omega + \alpha_1\varepsilon^2_{t+s-1} + \beta_1 h_{t+s-1}$, we obtain

$$\begin{aligned}
v_{t+s|t} &\equiv h_{t+s} - \hat{h}_{t+s|t} \\
&= \alpha_1(\varepsilon^2_{t+s-1} - \hat{\varepsilon}^2_{t+s-1|t}) + \beta_1(h_{t+s-1} - \hat{h}_{t+s-1|t}) \\
&= \alpha_1(\varepsilon^2_{t+s-1} - h_{t+s-1} + h_{t+s-1} - \hat{h}^2_{t+s-1|t}) \\
&\quad + \beta_1(h_{t+s-1} - \hat{h}_{t+s-1|t}) \\
&= \alpha_1 v_{t+s-1} + (\alpha_1 + \beta_1)v_{t+s-1|t},
\end{aligned} \tag{4.119}$$

where we have used the fact that $\hat{\varepsilon}^2_{t+i|t} = \hat{h}_{t+i|t}$ for $i > 0$ and the definition $\nu_t \equiv \varepsilon^2_t - h_t$. By continued recursive substitution, we finally arrive at

$$\begin{aligned}
v_{t+s|t} &= \alpha_1 \nu_{t+s-1} + (\alpha_1 + \beta_1)\alpha_1 \nu_{t+s-2} + \cdots \\
&\quad + (\alpha_1 + \beta_1)^{s-2}\alpha_1 \nu_{t+1} \\
&= \alpha_1 \sum_{i=1}^{s-1} (\alpha_1 + \beta_1)^{i-1} \nu_{t+s-i}.
\end{aligned} \tag{4.120}$$

As the ν_ts are serially uncorrelated and can be written as $\nu_t = h_t(z_t^2 - 1)$, it follows that the conditional SPE of the s-step-ahead forecast $\hat{h}_{t+s|t}$ is given by

$$\mathrm{E}[v_{t+s|t}^2|\Omega_t] = (\kappa - 1)\alpha_1^2 \sum_{i=1}^{s-1} (\alpha_1 + \beta_1)^{2(i-1)} \mathrm{E}[h_{t+s-i}^2|\Omega_t], \quad (4.121)$$

where κ is the kurtosis of z_t. Baillie and Bollerslev (1992) give expressions for the expectation of h_t^2, the conditional fourth moment of ε_t, which is required to evaluate (4.121). However, even if the value of the conditional SPE is available, it is quite problematic to use it to construct a confidence interval for the s-step-ahead forecast $\hat{h}_{t+s|t}$, because the distribution of h_{t+s} conditional upon Ω_t is highly nonnormal.

Forecasting conditional volatility for nonlinear GARCH models

In the discussion of out-of-sample forecasting with regime-switching models for the conditional mean in section 3.5, it was argued that one should resort to numerical techniques to obtain multiple-step-ahead forecasts because analytic expressions are impossible to obtain. By contrast, for most nonlinear GARCH models discussed in this chapter, out-of-sample forecasts for the conditional variance can be computed analytically in a straightforward manner. As an example, consider the GJR-GARCH(1,1) model

$$h_t = \omega + \alpha_1 \varepsilon_{t-1}^2 (1 - I[\varepsilon_{t-1} > 0]) + \gamma_1 \varepsilon_{t-1}^2 I[\varepsilon_{t-1} > 0] + \beta_1 h_{t-1}. \quad (4.122)$$

Assuming that the distribution of z_t is symmetric around 0, the 2-step-ahead forecast of h_{t+2} is given by

$$\begin{aligned}\hat{h}_{t+2|t} &= \mathrm{E}[\omega + \alpha_1 \varepsilon_{t+1}^2 (1 - I[\varepsilon_{t+1} > 0]) \\ &\quad + \gamma_1 \varepsilon_{t+1}^2 I[\varepsilon_{t+1} > 0] + \beta_1 h_{t+1}|\Omega_t] \\ &= \omega + ((\alpha_1 + \gamma_1)/2 + \beta_1) h_{t+1}, \end{aligned} \quad (4.123)$$

which follows from observing that ε_{t+1}^2 and the indicator function $I[\varepsilon_{t+1} > 0]$ are uncorrelated and $\mathrm{E}[I[\varepsilon_{t+1} > 0]] = P(\varepsilon_{t+1} > 0) = 0.5$, and again using $\mathrm{E}[\varepsilon_{t+1}^2|\Omega_t] = h_{t+1}$. In general, s-step-ahead forecasts can be computed either recursively as

$$\hat{h}_{t+s|t} = \omega + ((\alpha_1 + \gamma_1)/2 + \beta_1)\hat{h}_{t+s-1|t}, \quad (4.124)$$

or directly from

$$\hat{h}_{t+s|t} = \omega \sum_{i=0}^{s-1} ((\alpha_1 + \gamma_1)/2 + \beta_1)^i + ((\alpha_1 + \gamma_1)/2 + \beta_1)^{s-1} h_{t+1}, \quad (4.125)$$

compare the analogous expressions for the GARCH(1,1) model, as given in (4.115) and (4.116).

For the LSTGARCH model given in (4.55) the same formulae apply. This follows because ε_{t+i}^2 and the logistic function $F(\varepsilon_{t+i}) = [1+\exp(-\theta\varepsilon_{t+i})]^{-1}$ are uncorrelated, combined with the fact that $F(\varepsilon_{t+i})$ is anti-symmetric around the expected value of ε_{t+i} $(= 0)$ and, thus, $\mathrm{E}[F(\varepsilon_{t+i})] = F(\mathrm{E}[\varepsilon_{t+i}]) = 0.5$. In general, a function $G(x)$ is said to be anti-symmetric around a if $G(x+a) - G(a) = -(G(-x+a) - G(a))$ for all x. If furthermore x is symmetrically distributed with mean a it holds that $\mathrm{E}[G(x)] = G(\mathrm{E}[x]) = G(a)$.

Using the properties of the indicator function $I[\varepsilon_t > 0]$ and the logistic function $F(\varepsilon_t)$ noted above, it is straightforward to show that s-step-ahead forecasts for the conditional variance from the VS-GARCH model in (4.59) and the ANST-GARCH model in (4.61) can be computed either recursively from

$$\hat{h}_{t+s|t} = \omega + ((\alpha_1 + \gamma_1)/2 + (\beta_1 + \delta_1)/2)\hat{h}_{t+s-1|t}, \tag{4.126}$$

or directly from

$$\begin{aligned}\hat{h}_{t+s|t} = \omega \sum_{i=0}^{s-1} ((\alpha_1 + \gamma_1)/2 + (\beta_1 + \delta_1)/2)^i \\ + ((\alpha_1 + \gamma_1)/2 + (\beta_1 + \delta_1)/2)^{s-1} h_{t+1}.\end{aligned} \tag{4.127}$$

For the QGARCH(1,1) model given in (4.63), the asymmetry term $\gamma_1\varepsilon_{t-1}$ does not affect the forecasts for the conditional variance, as the conditional expectation of ε_{t+i} with $i > 0$ is 0 by assumption. Hence, point forecasts for the conditional variance can be obtained using the expressions given in (4.115) and (4.116) for the GARCH(1,1) model. Note, however, that $\gamma_1\varepsilon_{t-1}$ does alter the forecast error $v_{t+s|t}$, which now becomes

$$\begin{aligned} v_{t+s|t} &\equiv h_{t+s} - \hat{h}_{t+s|t} \\ &= \gamma_1\varepsilon_{t+s-1} + \alpha_1(\varepsilon_{t+s-1}^2 - \hat{\varepsilon}_{t+s-1|t}^2) \\ &\quad + \beta_1(h_{t+s-1} - \hat{h}_{t+s-1|t}) \\ &= \gamma_1\varepsilon_{t+s-1} + \alpha_1\nu_{t+s-1} + (\alpha_1 + \beta_1)v_{t+s-1|t} \\ &= \alpha_1 \sum_{i=1}^{s-1} (\alpha_1 + \beta_1)^{i-1}\nu_{t+s-i} + \gamma_1 \sum_{i=1}^{s-1} (\alpha_1 + \beta_1)^{i-1}\varepsilon_{t+s-i}. \end{aligned} \tag{4.128}$$

As $\mathrm{E}[\nu_{t+s-i}|\Omega_t] = \mathrm{E}[\varepsilon_{t+s-i}|\Omega_t] = 0$ for all $i = 1, \ldots, s-1$, the forecasts are unbiased, in the sense that $\mathrm{E}[v_{t+s|t}|\Omega_t] = 0$. However, the conditional variance of $v_{t+s|t}$ and, hence, the uncertainty in the forecast of h_{t+s} will be larger than in the GARCH(1,1) case.

Of the nonlinear GARCH models discussed in subsection 4.1.2, the only model for which analytic expressions for multiple-step-ahead forecasts of the conditional variance cannot be obtained is the ESTGARCH model given in (4.55) with (4.58). The exponential function $F(\varepsilon_{t+1}) = 1 - \exp(-\theta\varepsilon_{t+1}^2)$ is correlated with ε_{t+1}^2, and it also is not the case that $\mathrm{E}[F(\varepsilon_{t+1})] = F(\mathrm{E}[\varepsilon_{t+1}])$. Therefore, it is not possible to derive a recursive or direct formula for the s-step-ahead forecast $\hat{h}_{t+s|t}$ in this case. Instead, forecasts for future conditional variances have to be obtained by means of simulation.

Finally, analytic expressions for multiple-step-ahead forecasts of the conditional variance for the Markov-Switching GARCH model given in (4.67) can be obtained by exploiting the properties of the Markov-process (see Hamilton and Lin, 1996; Dueker, 1997; Klaassen, 1999).

Evaluating forecasts of conditional volatility

As just discussed, it is quite difficult to select a suitable nonlinear GARCH model on the basis of specification tests only. The out-of-sample forecasting ability of various GARCH models is an alternative approach to judge the adequacy of different models. Obviously, if a volatility model is to be of any use to practitioners in financial markets, it should be capable of generating accurate predictions of future volatility.

Whereas forecasting the future conditional volatility from (nonlinear) GARCH models is fairly straightforward, evaluating the forecasts is a more challenging task. In the following we assume that a GARCH model has been estimated using a sample of n observations, whereas observations at $t = n+1, \ldots n+m+s-1$ are held back for evaluation of s-step-ahead forecasts for the conditional variance.

Most studies use statistical criteria such as the mean squared prediction error (MSPE), which for a set of m s-step-ahead forecasts is computed as

$$\mathrm{MSPE} = \frac{1}{m}\sum_{j=0}^{m-1}(\hat{h}_{n+s+j|n+j} - h_{n+s+j})^2 \tag{4.129}$$

(see Akgiray, 1989; West and Cho, 1995; Brailsford and Faff, 1996; Franses and van Dijk, 1996, among many others), or the regression

$$h_{n+s+j} = a + b\hat{h}_{n+s+j|n+j} + e_{n+s+j}, \quad j = 0, \ldots, m-1 \tag{4.130}$$

(see Pagan and Schwert, 1990; Day and Lewis, 1992; Cumby, Figlewski and Hasbrouck, 1993; Lamoureux and Lastrapes, 1993; Jorion, 1995). In this case $\hat{h}_{n+s+j|n+j}$ is an unbiased forecast of h_{n+s+j}, $a = 0$, $b = 1$ and

$E(e_{n+s+j}) = 0$ in (4.130). The MSPE as defined in (4.129) cannot be computed as the true volatility h_{n+s+j} is unobserved, whereas for the same reason the parameters in (4.130) cannot be estimated. To make these forecast evaluation criteria operational, h_t is usually replaced by the squared shock $\varepsilon^2_{n+s+j} = z^2_{n+s+j} h_{n+s+j}$. As $\mathrm{E}[z^2_{n+s+j}] = 1$, ε^2_{n+s+j} is an unbiased estimate of h_{n+s+j}.

A common finding from forecast competitions is that all GARCH models provide seemingly poor volatility forecasts and explain only very little of the variability of asset returns, in the sense that the MSPE (or any other measure of forecast accuracy) is very large while the R^2 from the regression (4.130) is very small, typically below 0.10. In addition, the forecasts from GARCH appear to be biased, as it commonly found that $\hat{a} \neq 0$ in (4.130). Andersen and Bollerslev (1998) and Christodoulakis and Satchell (1998) demonstrated that this poor forecasting performance is caused by the fact that the unobserved true volatility h_{n+s+j} is approximated with the squared shock ε^2_{n+s+j}. As shown by Andersen and Bollerslev (1998), for a GARCH(1,1) model with a finite unconditional fourth moment the population R^2 from the regression (4.130) for $s = 1$ and h_{n+s+j} replaced by ε^2_{n+s+j} is equal to

$$R^2 = \frac{\alpha_1^2}{1 - \beta_1^2 - 2\alpha_1\beta_1}. \tag{4.131}$$

As the condition for a finite unconditional fourth moment in the GARCH(1,1) model is given by $\kappa\alpha_1^2 + \beta_1^2 + 2\alpha_1\beta_1 < 1$, it follows that the population R^2 is bounded from above by $1/\kappa$. Where z_t is normally distributed, the R^2 cannot be larger than 1/3, while the upper bound is even smaller if, for example, z_t is assumed to be Student-t-distributed.

Christodoulakis and Satchell (1998) explain the occurrence of apparent bias in GARCH volatility forecasts by noting that

$$\ln(\varepsilon^2_{n+s+j}) = \ln(h_{n+s+j}) + \ln(z^2_{n+s+j}), \tag{4.132}$$

or

$$\ln(\varepsilon^2_{n+s+j}) - \ln(\hat{h}_{n+s+j|n+j}) = \big(\ln(h_{n+s+j}) - \ln(\hat{h}_{n+s+j|n+j})\big) + \ln(z^2_{n+s+j}). \tag{4.133}$$

As $\ln(x) \approx -(1-x)$ for small x, the left-hand-side of (4.133) is approximately equal to the observed bias $\varepsilon^2_{n+s+j} - \hat{h}_{n+s+j|n+j}$. If the GARCH forecasts are unbiased, the first term on the right-hand-side of (4.133) is equal to zero. Hence, the expected observed bias is equal to $\mathrm{E}[\ln(z^2_{n+s+j})]$, which in the case of normally distributed z_t is equal to –1.27.

Andersen and Bollerslev (1998) suggest that a (partial) solution to the above-mentioned problems might be to estimate the unobserved volatility with data which is sampled more frequently than the time series of interest. For example, if y_t is a time series of weekly returns, the corresponding daily returns – if available – might be used to obtain a more accurate measure of the weekly volatility.

Other criteria have also been considered to evaluate the forecasts from GARCH models. Examples are the profitability of trading or investment strategies which make use of GARCH models to forecast conditional variance (see Engle *et al.*, 1993), or utility-based measures (see West, Edison and Cho, 1993).

Example 4.10: Forecasting the volatility of the Tokyo stock index As an alternative way to evaluate the nonlinear GARCH models estimated previously for weekly returns on the Tokyo stock index we compare their out-of-sample forecasting performance. To obtain forecasts of the conditional volatility we follow the methodology used in Franses and van Dijk (1996) (see also Donaldson and Kamstra, 1997). GARCH, GJR-GARCH, QGARCH and VS-GARCH models are estimated using a moving window of 5 years of data (260 observations). We start with the sample ranging from the first week of January 1986 until the last week of December 1990. The fitted models then are used to obtain 1- to 5-steps-ahead forecasts of h_t – that is, the conditional variance during the first 5 weeks of 1991. Next, the window is moved 1 week into the future, by deleting the observation from the first week of January 1986 and adding the observation for the first week of January 1991. The various GARCH models are re-estimated on this sample, and are used to obtain forecasts for h_t during the second until the sixth week of 1991. This procedure is repeated until the final estimation sample consists of observations from the first week of 1991 until the last week of 1995. In this way, we obtain 260 1- to 5-steps-ahead forecasts of the conditional variance. To evaluate and compare the forecasts from the different models, several forecast evaluation criteria are computed, with true volatility measured by the squared realized return. Table 4.14 reports the ratio of the forecast error criteria of the nonlinear GARCH models to those of the GARCH(1,1) model. For example, the figure 0.94 in the row $h = 1$ and column MSPE for the GJR-GARCH model means that the MSPE for 1-step-ahead forecasts from this model is 6 per cent smaller than the corresponding criterion for forecasts from the linear GARCH model. It is seen that the GJR-GARCH and QGARCH model perform better on all four criteria across all forecast horizons considered. Also reported are p-values for the predictive accuracy test of Diebold and Mariano (1995) as given in (2.75), based on both absolute and squared prediction errors. These p-values suggest that, even though the difference in the forecast error criteria is substantial, the forecasts from the linear and nonlinear GARCH models need not be significantly different from each other.

Table 4.14 *Forecast evaluation of nonlinear GARCH models for weekly returns on the Tokyo stock index, as compared to the GARCH (1,1) model*

Model	h	MSPE	MedSPE	DM(S)	MAPE	MedAPE	DM(A)
GJR-GARCH	1	0.94	0.74	0.05	0.95	0.86	0.05
	2	0.93	0.77	0.15	0.93	0.88	0.06
	3	0.92	0.75	0.27	0.91	0.87	0.07
	4	0.89	0.80	0.26	0.92	0.90	0.12
	5	0.89	0.75	0.35	0.90	0.87	0.12
QGARCH	1	0.95	0.83	0.19	0.93	0.91	0.01
	2	0.94	0.87	0.32	0.91	0.93	0.01
	3	0.92	0.92	0.25	0.90	0.96	0.02
	4	0.94	0.88	0.48	0.91	0.94	0.05
	5	0.89	0.86	0.37	0.89	0.93	0.06
VS-GARCH	1	1.01	0.54	0.87	0.86	0.74	0.00
	2	1.03	0.36	0.77	0.82	0.60	0.00
	3	1.04	0.37	0.84	0.80	0.61	0.00
	4	1.02	0.32	0.96	0.79	0.56	0.00
	5	1.05	0.27	0.83	0.77	0.52	0.01

Notes: Forecast evaluation criteria for nonlinear GARCH models for weekly returns on the Tokyo stock index.
Out-of-sample forecasts are constructed for the period January 1991–December 1995, with models estimated on a rolling window of 5 years.
The columns labelled DM(S) and DM(A) contain p-values for the prediction accuracy test given in (2.75), based on squared and absolute prediction errors, respectively.

Finally, table 4.15 contains parameter estimates, with standard errors in parentheses, and R^2 measures from the regression of (demeaned) returns on the forecasts of conditional variance. Only for the linear GARCH model are the estimate of the intercept $\hat{a}$ and the slope $\hat{b}$ significantly different from 0 and 1, respectively. Hence it appears that forecasts from the GARCH model are not unbiased, whereas for the nonlinear GARCH models they generally are. The low values of the R^2 suggest that the models explain only a small fraction of the variability in the conditional variance of the returns, which confirms previous findings as discussed above.

4.6 Impulse response functions

Of particular interest in the context of GARCH models is the influence of shocks on future conditional volatility, or the persistence of shocks. A natural measure of this influence is the expectation of the conditional volatility s-periods

Table 4.15 *Forecast evaluation of nonlinear GARCH models for weekly returns on the Tokyo stock index*

Model	h	1	2	3	4	5
GARCH	$\hat{a}$	3.58	2.93	3.68	4.80	4.13
		(1.71)	(1.72)	(1.76)	(1.80)	(1.80)
	$\hat{b}$	0.56	0.61	0.53	0.42	0.47
		(0.12)	(0.12)	(0.12)	(0.12)	(0.12)
	R^2	0.07	0.09	0.07	0.04	0.05
GJR-GARCH	$\hat{a}$	2.21	1.04	1.70	2.07	1.42
		(1.77)	(1.78)	(1.84)	(1.90)	(1.93)
	$\hat{b}$	0.75	0.88	0.81	0.78	0.86
		(0.14)	(0.15)	(0.15)	(0.16)	(0.17)
	R^2	0.10	0.12	0.10	0.08	0.09
QGARCH	$\hat{a}$	1.99	0.39	0.51	3.45	0.34
		(1.93)	(1.96)	(2.05)	(2.19)	(2.18)
	$\hat{b}$	0.82	0.98	0.98	0.66	0.99
		(0.17)	(0.18)	(0.19)	(0.20)	(0.20)
	R^2	0.08	0.11	0.05	0.04	0.08
VS-GARCH	$\hat{a}$	2.56	−1.00	0.55	1.68	2.10
		(2.08)	(2.19)	(2.39)	(2.54)	(2.62)
	$\hat{b}$	1.13	1.93	1.87	1.82	1.89
		(0.28)	(0.35)	(0.44)	(0.52)	(0.60)
	R^2	0.06	0.11	0.07	0.04	0.04

Notes: Summary statistics for regressions of observed (demeaned) squared return on forecast of conditional variance from nonlinear GARCH models for weekly returns on the Tokyo stock index.
Out-of-sample forecasts are constructed for the period January 1991–December 1995, with models estimated on a rolling window of 5 years.

ahead, conditional on a particular current shock and current conditional volatility, that is

$$\mathrm{E}[h_{t+s}|\varepsilon_t = \delta, h_t = h]. \tag{4.134}$$

Notice that for $s = 1$, (4.134) is in fact the definition of the news impact curve (NIC) discussed in subsection 4.1.2. The NIC measures the direct impact of a shock on the conditional variance. By examining how the conditional expectation (4.134) changes as s increases, one can obtain an impression of the dynamic effect of a particular shock or, more generally, about the propagation of shocks. It turns out that for most (nonlinear) GARCH models a simple recursive relationship between the conditional expectations at different horizons can be derived. For example, for the GARCH(1,1) model it is straightforward to show that

$$\mathrm{E}[h_{t+s}|\varepsilon_t = \delta, h_t = h] = (\alpha_1 + \beta_1)\mathrm{E}[h_{t+s-1}|\varepsilon_t = \delta, h_t = h], \tag{4.135}$$

for all $s \geq 2$. Hence, the effect of the shock ε_t on h_{t+s} decays exponentially with rate $\alpha_1 + \beta_1$. Similarly, for the GJR-GARCH model given in (4.53),

$$\begin{aligned}&\mathrm{E}[h_{t+s}|\varepsilon_t = \delta, h_t = h]\\ &\quad = ((\alpha_1 + \gamma_1)/2 + \beta_1)\mathrm{E}[h_{t+s-1}|\varepsilon_t = \delta, h_t = h].\end{aligned} \tag{4.136}$$

The so-called *conditional volatility profile* (4.134) is discussed in detail in Gallant, Rossi and Tauchen (1993) for univariate volatility models. Hafner and Herwartz (1998a, 1998b) provide a generalization to multivariate GARCH models.

The conditional expectation (4.134) is closely related to impulse response functions. For example, Lin (1997) defines the impulse response function for (multivariate) linear GARCH models as

$$TIRF_h(s, \delta) = \mathrm{E}[h_{t+s}|\varepsilon_t = \delta, h_t = h] - \mathrm{E}[h_{t+s}|\varepsilon_t = 0, h_t = h]. \tag{4.137}$$

Notice that this definition is similar to the traditional impulse response function for the conditional mean of a time series as given in (3.95), in the sense that the conditional expectation (4.134) is compared with the conditional expectation given that the current shock is equal to zero. Given that the conditional variance essentially measures the expected value of ε_t^2, this might not be the most appropriate benchmark profile for comparison. It appears more natural to set ε_t to its expected value h_t in the second term on the right-hand-side of (4.137).

Alternatively, one can compare $\mathrm{E}[h_{t+s}|\varepsilon_t = \delta, h_t = h]$ with the expectation of h_{t+s} conditional only on the current conditional variance as in the generalized impulse response functions (GIRF) discussed in section 3.6. That is, a GIRF for the conditional variance may be defined as

$$GIRF_h(s, \delta, h) = \mathrm{E}[h_{t+s}|\varepsilon_t = \delta, h_t = h] - \mathrm{E}[h_{t+s}|h_t = h]. \tag{4.138}$$

The second conditional expectation can also be computed analytically for most nonlinear GARCH models, although the algebra can become quite tedious. An alternative is to resort to simulation techniques, as discussed in section 3.6.

4.7 On multivariate GARCH models

The GARCH models discussed so far all are univariate. Given the interpretation of shocks as news and the fact that at least certain news items affect various assets simultaneously, it might be suggested that the volatility of different assets moves together over time. Consequently, it is of interest to consider multivariate models to describe the volatility of several time series jointly, to exploit possible linkages which exist. An alternative motivation for multivariate models is that an important subject of financial economics is the construction of portfolios from various assets. The covariances among the assets play a crucial role in this decision problem, as in the CAPM, for example. Multivariate GARCH models can be used to model the time-varying behaviour of these conditional covariances.

A general multivariate GARCH model for the k-dimensional process $\boldsymbol{\varepsilon}_t = (\varepsilon_{1t}, \ldots, \varepsilon_{kt})'$ is given by

$$\boldsymbol{\varepsilon}_t = \mathbf{z}_t \mathbf{H}_t^{1/2}, \tag{4.139}$$

where z_t is a k-dimensional i.i.d. process with mean zero and covariance matrix equal to the identity matrix I_k. From these properties of z_t and (4.139), it follows that $\mathrm{E}[\boldsymbol{\varepsilon}_t|\Omega_{t-1}] = \mathbf{0}$ and $\mathrm{E}[\boldsymbol{\varepsilon}_t\boldsymbol{\varepsilon}_t'|\Omega_{t-1}] = \mathbf{H}_t$. To complete the model, a parameterization for the conditional covariance matrix $\mathbf{H}_t$ needs to be specified. This turns out to be a nontrivial task. As in the univariate GARCH models, one may want to allow $\mathbf{H}_t$ to depend on lagged shocks $\boldsymbol{\varepsilon}_{t-i}$, $i = 1, \ldots, q$, and on lagged conditional covariance matrices $\mathbf{H}_{t-i}$, $i = 1, \ldots, p$. Several parameterizations have been proposed, some of which are discussed below. To simplify the exposition, we discuss only the case $p = q = 1$.

The vec model

Let vech$(\cdot)$ denote the operator which stacks the lower portion of a matrix in a vector. As the conditional covariance matrix is symmetric, vech$(\mathbf{H}_t)$ contains all unique elements of $\mathbf{H}_t$. A general representation for the multivariate analogue of the GARCH(1,1) model in (4.13) is given by

$$\mathrm{vech}(\mathbf{H}_t) = \boldsymbol{W}^* + \boldsymbol{A}_1^*\mathrm{vech}(\boldsymbol{\varepsilon}_{t-1}\boldsymbol{\varepsilon}_{t-1}') + \boldsymbol{B}_1^*\mathrm{vech}(\mathbf{H}_{t-1}), \tag{4.140}$$

where $\boldsymbol{W}^*$ is a $k(k+1)/2 \times 1$ vector and $\boldsymbol{A}_1^*$ and $\boldsymbol{B}_1^*$ are $(k(k+1)/2 \times k(k+1)/2)$ matrices. This general model, which is called the vec representation by Engle and Kroner (1995), is very flexible as it allows all elements of $\mathbf{H}_t$ to depend on all elements of the cross-products of $\boldsymbol{\varepsilon}_{t-1}$ and all elements of the lagged conditional covariance matrix $\mathbf{H}_t$.

The vec model has two important drawbacks. First, the number of parameters in (4.140) equals $(k(k+1)/2)(1+2(k(k+1)/2))$, which becomes excessively large as k increases. For example, in the bivariate case, (4.140) takes the form,

$$\begin{pmatrix} h_{11,t} \\ h_{12,t} \\ h_{22,t} \end{pmatrix} = \begin{pmatrix} \omega_{11}^* \\ \omega_{12}^* \\ \omega_{22}^* \end{pmatrix} + \begin{pmatrix} \alpha_{11}^* & \alpha_{12}^* & \alpha_{13}^* \\ \alpha_{21}^* & \alpha_{22}^* & \alpha_{23}^* \\ \alpha_{31}^* & \alpha_{32}^* & \alpha_{33}^* \end{pmatrix} \begin{pmatrix} \varepsilon_{1,t-1}^2 \\ \varepsilon_{1,t-1}\varepsilon_{2,t-1} \\ \varepsilon_{2,t-1}^2 \end{pmatrix} + \begin{pmatrix} \beta_{11}^* & \beta_{12}^* & \beta_{13}^* \\ \beta_{21}^* & \beta_{22}^* & \beta_{23}^* \\ \beta_{31}^* & \beta_{32}^* & \beta_{33}^* \end{pmatrix} \begin{pmatrix} h_{11,t-1} \\ h_{12,t-1} \\ h_{22,t-1} \end{pmatrix}. \tag{4.141}$$

For this simplest possible case, the model already contains 21 parameters which have to be estimated. Estimation of this general model may therefore be quite problematic. The second shortcoming of the vec model is that conditions on the matrices $\boldsymbol{A}_1^*$ and $\boldsymbol{B}_1^*$ which guarantee positive semi-definiteness of the conditional covariance matrix $\mathbf{H}_t$ are not easy to impose.

The diagonal model

Bollerslev, Engle and Wooldridge (1988) suggest reducing the number of parameters in the multivariate GARCH model by constraining the matrices $\boldsymbol{A}_1^*$ and $\boldsymbol{B}_1^*$ in (4.140) to be diagonal. In this case, the conditional covariance between $\varepsilon_{i,t}$ and $\varepsilon_{j,t}$, $h_{ij,t}$, depends only on lagged cross-products of the two shocks involved and lagged values of the covariance itself,

$$h_{ij,t} = \omega_{ij} + \alpha_{ij}\varepsilon_{i,t-1}\varepsilon_{j,t-1} + \beta_{ij}h_{ij,t-1}, \tag{4.142}$$

where α_{ij} and β_{ij} are the (i, j)th element of the $(k \times k)$ matrices $\boldsymbol{A}_1$ and $\boldsymbol{B}_1$, respectively, which are implicitly defined by $\boldsymbol{A}_1^* = \text{diag}(\text{vech}(\boldsymbol{A}_1))$ and $\boldsymbol{B}_1^* = \text{diag}(\text{vech}(\boldsymbol{B}_1))$. These definitions allow the complete model to be written as

$$\mathbf{H}_t = \boldsymbol{W} + \boldsymbol{A}_1 \odot (\boldsymbol{\varepsilon}_{t-1}\boldsymbol{\varepsilon}_{t-1}') + \boldsymbol{B}_1 \odot \mathbf{H}_{t-1}, \tag{4.143}$$

where $\odot$ again denotes the Hadamard or element-by-element product.

The number of parameters in the diagonal GARCH(1,1) model equals $3(k(k+1)/2)$. For the bivariate case, setting all off-diagonal elements α_{ij}^* and β_{ij}^*, $i \neq j$, in (4.141) equal to zero, it is seen that 9 parameters remain to be estimated. An additional advantage of the diagonal model is that conditions which ensure that the conditional covariance matrix is positive definite are quite easy to check. In particular, $\mathbf{H}_t$ is positive definite if $\boldsymbol{W}$ is positive definite and $\boldsymbol{A}_1$ and $\boldsymbol{B}_1$ are positive semi-definite, see Attanasio (1991).

On the other hand, the diagonal model may be considered too restrictive, as it does not allow the conditional variance of one series to depend on the history of other variables in the system.

The BEKK model

An alternative representation of the multivariate GARCH(1,1) model is given by

$$\mathbf{H}_t = \boldsymbol{W} + \boldsymbol{A}_1'\boldsymbol{\varepsilon}_{t-1}\boldsymbol{\varepsilon}_{t-1}'\boldsymbol{A}_1 + \boldsymbol{B}_1'\mathbf{H}_{t-1}\boldsymbol{B}_1, \tag{4.144}$$

where $\boldsymbol{W}$, $\boldsymbol{A}_1$ and $\boldsymbol{B}_1$ are $(k \times k)$ matrices, with $\boldsymbol{W}$ symmetric and positive definite. Engle and Kroner (1995) discuss this formulation, which they dub the BEKK representation after Baba *et al.* (1991). As the second and third terms on the right-hand-side of (4.144) are expressed as quadratic forms, $\mathbf{H}_t$ is guaranteed to be positive definite without the need for imposing constraints on the parameter matrices $\boldsymbol{A}_i$ and $\boldsymbol{B}_i$. This is the main advantage of the BEKK representation. On the other hand, the number of parameters in (4.144) equals $2k^2 + k/2$, which still becomes very large as k increases. In the bivariate case, the BEKK model is given by

$$\begin{aligned}\begin{pmatrix} h_{11,t} & h_{12,t} \\ h_{12,t} & h_{22,t} \end{pmatrix} &= \begin{pmatrix} w_{11} & w_{12} \\ w_{12} & w_{22} \end{pmatrix} \\ &\quad + \begin{pmatrix} \alpha_{11} & \alpha_{12} \\ \alpha_{21} & \alpha_{22} \end{pmatrix}' \begin{pmatrix} \varepsilon_{1,t-1}^2 & \varepsilon_{1,t-1}\varepsilon_{2,t-1} \\ \varepsilon_{1,t-1}\varepsilon_{2,t-1} & \varepsilon_{2,t-1}^2 \end{pmatrix} \begin{pmatrix} \alpha_{11} & \alpha_{12} \\ \alpha_{21} & \alpha_{22} \end{pmatrix} \\ &\quad + \begin{pmatrix} \beta_{11} & \beta_{12} \\ \beta_{21} & \beta_{22} \end{pmatrix}' \begin{pmatrix} h_{11,t-1} & h_{12,t-1} \\ h_{12,t-1} & h_{22,t-1} \end{pmatrix} \begin{pmatrix} \beta_{11} & \beta_{12} \\ \beta_{21} & \beta_{22} \end{pmatrix},\end{aligned} \tag{4.145}$$

which is seen to contain 12 parameters, compared with 21 for the vec model in (4.141). By applying the vech operator to (4.144), the model can be expressed in the vec representation (4.140). As shown by Engle and Kroner (1995), this vec representation is unique. Conversely, every vec model which can be rewritten as a BEKK representation renders positive definite covariance matrices $\mathbf{H}_t$.

The constant correlation model

Bollerslev (1990) put forward an alternative way to simplify the general model (4.140), by assuming that the conditional correlations between the elements of $\boldsymbol{\varepsilon}_t$ are time-invariant. This implies that the conditional covariance $h_{ij,t}$ between ε_{it} and ε_{jt} is proportional to the product of their conditional standard deviations. The individual conditional variances are assumed to follow univariate GARCH(1,1) models. The diagonal model is given by

$$h_{ii,t} = \omega_{ii} + \alpha_{ii}\varepsilon_{i,t-1}^2 + \beta_{ii}h_{ii,t-1} \quad \text{for } i = 1, \ldots, k, \tag{4.146}$$

$$h_{ij,t} = \rho_{ij}\sqrt{h_{ii,t}}\sqrt{h_{jj,t}} \quad \text{for all } i \neq j, \tag{4.147}$$

or alternatively,

$$\mathbf{H}_t = \boldsymbol{D}_t^{1/2} \boldsymbol{R} \mathbf{D}_t^{1/2}, \tag{4.148}$$

where $\mathbf{D}_t$ is a $(k \times k)$ matrix with the conditional variances $h_{ii,t}$ on the diagonal, and $\boldsymbol{R}$ is a $(k \times k)$ matrix containing the correlations ρ_{ij}. For example, in the bivariate case

$$\mathbf{H}_t = \begin{pmatrix} \sqrt{h_{11,t}} & 0 \\ 0 & \sqrt{h_{22,t}} \end{pmatrix} \begin{pmatrix} 1 & \rho_{12} \\ \rho_{12} & 1 \end{pmatrix} \begin{pmatrix} \sqrt{h_{11,t}} & 0 \\ 0 & \sqrt{h_{22,t}} \end{pmatrix}. \tag{4.149}$$

It is seen from (4.149) that the dynamic properties of the covariance matrix $\mathbf{H}_t$ are determined entirely by the conditional variances in $\mathbf{D}_t$. All that is required for the conditional covariance matrix implied by the constant correlation model to be positive definite is that the univariate GARCH models for $h_{ii,t}$ render positive conditional variances and that the correlation matrix $\boldsymbol{R}$ is positive definite.

The factor model

One of the main motivations for considering multivariate GARCH models is that the volatility of different assets is driven or affected by the same sources of news. This can be made more explicit in the model by assuming the presence of so-called common factors, as suggested in Diebold and Nerlove (1989). An r-factor multivariate GARCH model can be represented as

$$\boldsymbol{\varepsilon}_t = \boldsymbol{B} \boldsymbol{f}_t + \boldsymbol{v}_t, \tag{4.150}$$

where $\boldsymbol{B}$ is a $(k \times r)$ matrix of full-column rank of so-called factor loadings, $\boldsymbol{f}_t$ is a $r \times 1$ vector containing the common factors and $\boldsymbol{v}_t$ is a $k \times 1$ vector of idiosyncratic noise. It is assumed that $\boldsymbol{f}_t$ and $\boldsymbol{v}_t$ have conditional mean zero and conditional variance matrices given by $\boldsymbol{\Lambda}_t$ and $\boldsymbol{\Sigma}_t$, respectively. Both $\boldsymbol{\Lambda}_t$ and $\boldsymbol{\Sigma}_t$ are diagonal, while it is also common to assume that $\boldsymbol{\Sigma}_t$ is constant over time. Finally, $\boldsymbol{f}_t$ and $\boldsymbol{v}_t$ are assumed to be uncorrelated. The preceding assumptions imply that the conditional covariance matrix of $\boldsymbol{\varepsilon}_t$ is given by

$$\mathbf{H}_t = \mathbf{B} \boldsymbol{\Lambda}_t \mathbf{B}' + \boldsymbol{\Sigma}. \tag{4.151}$$

For example, in the bivariate case, where we can have only one common factor, this amounts to

$$\begin{pmatrix} h_{11,t} & h_{12,t} \\ h_{12,t} & h_{22,t} \end{pmatrix} = \begin{pmatrix} \beta_1^2 & \beta_1\beta_2 \\ \beta_1\beta_2 & \beta_2^2 \end{pmatrix} \lambda_t + \begin{pmatrix} \sigma_1^2 & 0 \\ 0 & \sigma_2^2 \end{pmatrix}. \tag{4.152}$$

The conditional variance of ε_{it}, $i = 1, 2$, is composed of the variance of the news which is specific to the ith asset, σ_i^2 and the conditional variance of the common news factor f_t. A nonzero conditional covariance between ε_{1t} and ε_{2t} is caused solely by this common factor. For the common factor(s) a standard GARCH model can be postulated. Estimation of factor GARCH models is considered in Lin (1992). Applications to financial time series can be found in Engle, Ng and Rothschild (1990) and Ng, Engle and Rothschild (1992).

Common heteroscedasticity

An important consequence of the presence of common factors is that there are linear combinations of the elements of $\boldsymbol{\varepsilon}_t$ which have constant conditional variance. For example, in the bivariate factor GARCH model the conditional variance of the linear combination $\beta_2\varepsilon_{1,t} - \beta_1\varepsilon_{2,t}$ is given by $(\beta_2, -\beta_1)\mathbf{H}_t(\beta_2, -\beta_1)'$. Substituting $\mathbf{H}_t$ from (4.153), it follows that this conditional variance is constant and equal to $\beta_2^2\sigma_1^2 + \beta_1^2\sigma_2^2$. Engle and Susmel (1993) suggest a test for such common heteroscedasticity which avoids actually estimating a factor GARCH model. The test is based upon the intuitive idea that, for fixed τ, a test for ARCH in the linear combination $\varepsilon_{1,t} - \tau\varepsilon_{2,t}$ can be obtained as nR^2 where R^2 is the coefficient of determination of the regression of $\varepsilon_{1,t} - \tau\varepsilon_{2,t}$ on lagged squares and cross-products of $\varepsilon_{1,t}$ and $\varepsilon_{2,t}$ – that is, $\varepsilon_{j,t-i}^2$, $j = 1, 2$, $i = 1, \ldots, p$ and $\varepsilon_{1,t-i}\varepsilon_{2,t-i}$, $i = 1, \ldots, p$, similar to the univariate LM test for ARCH discussed in section 4.2. As τ is unknown, the test-statistic is obtained as the minimum of the point-wise statistic, where a grid search over τ is performed. As shown by Engle and Kozicki (1993), the resulting test-statistic follows a chi-squared distribution with $3p - 1$ degrees of freedom asymptotically.

Example 4.11: Testing for common ARCH in stock index and exchange rate returns The test for common ARCH is computed for the indexes of the four 'big' stock markets – Frankfurt, London, New York and Tokyo – for the sample January 1986–December 1990, and for four European exchange rates British pound, Dutch guilder, French franc and German Dmark for the sample January 1991–December 1995. Table 4.16 shows p-values of the LM test for $q = 1, 5$ and 10, as well as the estimates of the parameter τ for which the $TR^2(\tau)$ function is minimized.

For the stock indexes, the null hypothesis of common ARCH can be rejected for most combinations of stock markets. An exception are the stock markets in Frankfurt and London. For the European exchange rates there appears to be more evidence for common ARCH. In particular, the conditional heteroscedasticity in the Dutch guilder, French franc and German Dmark appears to have a common source, whereas the conditional volatility of the British pound seems to behave independently from these currencies from the European continent.

Table 4.16 *Testing for common ARCH effects in weekly stock index and exchange rate returns*

	p-value			$\hat{\tau}$		
q	1	5	10	1	5	10
Stock indexes						
Frankfurt/London	0.911	0.908	0.982	1.346	2.860	3.197
Frankfurt/New York	0.028	0.434	0.771	2.300	2.610	2.596
Frankfurt/Tokyo	0.018	0.000	0.010	3.077	2.016	1.877
London/New York	0.128	0.008	0.039	0.889	0.579	0.586
London/Tokyo	0.088	0.016	0.201	0.913	0.208	−0.353
New York/Tokyo	0.045	0.039	0.286	1.005	1.408	1.536
Exchange rates						
British pound/Dutch guilder	0.002	0.003	0.159	−0.521	−2.586	−2.307
British pound/French franc	0.000	0.000	0.032	−0.310	−0.829	−0.832
British pound/German Dmark	0.002	0.002	0.131	−0.508	−2.377	−2.220
Dutch guilder/French franc	0.993	0.841	0.938	−0.549	0.609	0.435
Dutch guilder/German Dmark	0.680	0.175	0.453	1.228	1.024	1.019
French franc/German Dmark	0.990	0.889	0.983	1.761	1.567	1.813

Notes: p-values of the LM test for common ARCH of order q for weekly stock index and exchange rate returns.
The sample runs from January 1986 until December 1990 for the stock indexes, and from January 1991 until December 1995 for the exchange rates.
The tests are applied to residuals from an AR(k) model, with k determined by the AIC.
The final three columns report the value of the parameter τ for which the TR^2 function is minimized.

Finally, multivariate nonlinear GARCH models have only recently been considered, see Hafner and Herwartz (1998a, 1998b) and Kroner and Ng (1998). Kroner and Ng (1998) also discuss multivariate analogues of the Sign and Sign-Bias tests discussed in section 4.4, which can be used to test for the presence of asymmetric effects of positive and negative or large and small shocks. It seems that much further research is needed on issues as representation, specification, estimation, inference and forecasting for these models.