16

Multivariate ARCH and GARCH Models

16.1 Background

In the previous chapters, we have discussed modelling the conditional mean of the data generation process of a multiple time series, conditional on the past at each particular time point. In that context, the variance or covariance matrix of the conditional distribution was assumed to be time invariant. In fact, in much of the discussion, the residuals or forecast errors were assumed to be independent white noise. Such a simplification is useful and justified in many applications.

There are also situations, however, when such an assumption is problematic, for instance, when financial time series are being analyzed. To see this, consider the monthly returns of the DAX (German stock index) for the period 1965–1995 depicted in Figure 16.1. The autocorrelations are all within the $\pm 2/\sqrt{T}$ band and, hence, in accordance with the results discussed in Chapter 4, Section 4.4, one may conclude that the returns are not autocorrelated. If they were not only uncorrelated but also independent, then their squares were independent too. That this is not the case is clearly seen in the third panel of Figure 16.1, where also the autocorrelations of the squared returns are given. Consequently, in this case, assuming independent observations or, equivalently, independent residuals in the AR(0) model $y_t = \nu + u_t$ is clearly problematic. Because we have used the independence assumption in deriving the $\pm 2/\sqrt{T}$ confidence bounds in Chapter 4, Section 4.4, the conclusion of uncorrelated returns may also be questioned in this case.

The correlations in the squares of the DAX returns shown in Figure 16.1 indicate that there is conditional heteroskedasticity. With a little imagination, it can also be seen in the figure that the volatility in the DAX returns changes over time. It is lower in the first half of the sample period than in the second half. Similar characteristics in many time series, in particular in financial market series, have motivated the development of specific models for conditionally heteroskedastic data.

H. Lütkepohl, *New Introduction to Multiple Time Series Analysis*

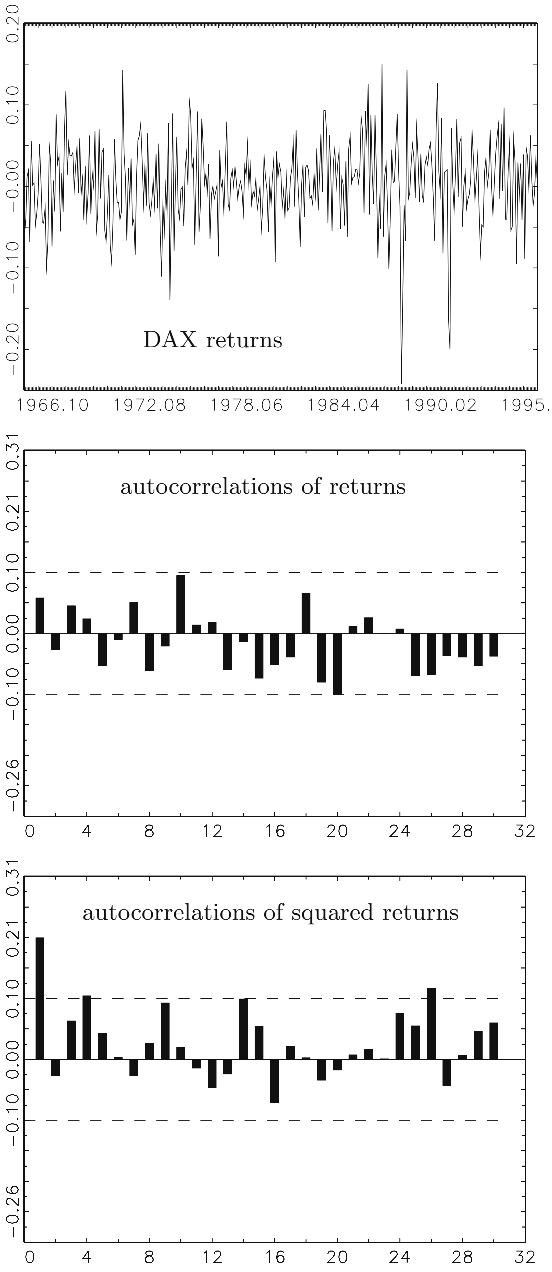

Fig. 16.1. Monthly DAX returns for the years 1965–1995 with autocorrelations and autocorrelations of squared returns.

It may be tempting to argue that the conditional mean is the optimal forecast and, hence, changes in volatility are of less importance from a forecasting point of view. This position ignores, however, that the forecast error variances, that is, the variances of the conditional distributions are needed for setting up forecast intervals. Taking into account conditional heteroskedasticity is therefore important also when forecasts of the variables under investigation are desired. Moreover, for example in financial analysis, forecasts of the future volatility of a series under consideration are often of interest to assess the risk associated with certain assets. In that case, variance forecasts are of direct interest, of course. Furthermore, the volatility in a market and, hence, the risk associated with investments in a particular market may have a direct effect on the expectations of the market participants. Hence, there may be a feedback from the second to the first moments. Therefore, the emphasis on a more detailed modelling of the volatility of time series was a natural development which was boosted by Engle's (1982) invention of ARCH (*autoregressive conditional heteroskedasticity*) models. By now the acronym ARCH stands for a wide range of models for changing conditional volatility. Moreover, there is also some literature on multivariate extensions which are the central topic of this chapter.

Because many series have a close relationship, it is obvious to conjecture that an increase, say, in the volatility of one series may have an impact on the volatility of another series as well. For example, this may occur in exchange rates of different currencies, in interest rates for bonds of different times to maturity, or in returns on stocks in a specific segment of the market. Therefore, multivariate models for conditional heteroskedasticity are of interest.

In the following, a brief review of some facts on univariate ARCH and generalized ARCH (GARCH) models is given and then multivariate extensions will be discussed. Part of this chapter reports results from an article by Engle & Kroner (1995). There are also a number of review articles which cover multivariate ARCH and GARCH models among other things. Examples are Bollerslev, Engle & Nelson (1994), Bera & Higgins (1993), Bauwens, Laurent & Rombouts (2004), Bollerslev, Chou & Kroner (1992), and Pagan (1996). The latter two articles also survey some of the applied literature.

16.2 Univariate GARCH Models

16.2.1 Definitions

Consider the univariate serially uncorrelated, zero mean process u_t. For instance, u_t may represent the residuals of an autoregressive process. The u_t are said to follow an *autoregressive conditionally heteroskedastic process of order* q (ARCH(q)) if the conditional distribution of u_t, given its past $\Omega_{t-1} := \{u_{t-1}, u_{t-2}, \dots\}$, has zero mean and the conditional variance is

$$\sigma^2_{t|t-1} := \text{Var}(u_t|\Omega_{t-1}) = E(u_t^2|\Omega_{t-1}) = \gamma_0 + \gamma_1 u_{t-1}^2 + \cdots + \gamma_q u_{t-q}^2, \quad (16.2.1)$$

that is, $u_t|\Omega_{t-1} \sim (0, \sigma^2_{t|t-1})$. Another, sometimes quite useful way to define an ARCH process is to specify

$$u_t = \sigma_{t|t-1}\varepsilon_t, \quad \varepsilon_t \sim \text{i.i.d.}(0,1). \tag{16.2.2}$$

Here the i.i.d. assumption for ε_t is slightly more restrictive than the previous definition which makes statements about the first two moments of the conditional distribution only. In the following, the definition (16.2.2) will be used. The u_t's, generated in this way, will be serially uncorrelated with mean zero.

Originally, Engle (1982), in his seminal paper on ARCH models, assumed the conditional distribution to be normal so that

$$\varepsilon_t \sim \text{i.i.d.}\,\mathcal{N}(0,1) \quad \text{and} \quad u_t|\Omega_{t-1} \sim \mathcal{N}(0, \sigma^2_{t|t-1}). \tag{16.2.3}$$

Although different distributions were considered later as well, even with this special distributional assumption the model is capable of generating series with characteristics similar to those of many observed time series. In particular, it is capable to generate series with volatility clustering and outliers similar to the DAX series in Figure 16.1. Even if the conditional distribution underlying an ARCH(q) model is normal, the unconditional distribution will generally be nonnormal. In particular, it is *leptokurtic*, that is, it has more mass around zero and in the tails than the normal distribution and, hence, it can produce occasional outliers.

It turns out, however, that, for many series, ARCH processes with fairly large orders are necessary to capture the dynamics in the conditional variances. Therefore, Bollerslev (1986) and Taylor (1986) proposed to gain greater parsimony by extending the model in a similar manner as the AR model when moving to mixed ARMA models. They suggested the *generalized ARCH* (GARCH) model with conditional variances given by

$$\sigma^2_{t|t-1} = \gamma_0 + \gamma_1 u^2_{t-1} + \cdots + \gamma_q u^2_{t-q} + \beta_1 \sigma^2_{t-1|t-2} + \cdots + \beta_m \sigma^2_{t-m|t-m-1}. \tag{16.2.4}$$

These models are briefly denoted by GARCH(q, m). They generate processes with existing *un*conditional variance if and only if the coefficient sum

$$\gamma_1 + \cdots + \gamma_q + \beta_1 + \cdots + \beta_m < 1. \tag{16.2.5}$$

If this condition is satisfied, u_t has a constant *un*conditional variance given by

$$\sigma^2_u = \frac{\gamma_0}{1 - \gamma_1 - \cdots - \gamma_q - \beta_1 - \cdots - \beta_m}. \tag{16.2.6}$$

The similarity of GARCH models and ARMA models for the conditional mean can be seen by defining $v_t := u^2_t - \sigma^2_{t|t-1}$, substituting $u^2_t - v_t$ for $\sigma^2_{t|t-1}$ in (16.2.4) and rearranging terms. Thereby we get

$$u_t^2 = \gamma_0 + (\beta_1 + \gamma_1)u_{t-1}^2 + \cdots + (\beta_q + \gamma_q)u_{t-q}^2 + v_t - \beta_1 v_{t-1} - \cdots - \beta_m v_{t-m} \quad (16.2.7)$$

which is formally an ARMA(q, m) model for u_t^2. Here it is assumed without loss of generality that $q \geq m$ and $\beta_j := 0$ for $j > m$.

16.2.2 Forecasting

Although the conditional expectation of the process u_t given Ω_{t-h} is zero and, hence, the optimal h-step forecasts are all zero for $h = 1, 2, \ldots$, there is an important difference to the situation where the u_t are an independent white noise process. If the u_t are Gaussian $\mathcal{N}(0, \sigma_u^2)$, a 1-step ahead $(1 - \alpha)100\%$ forecast interval has the form

$$u_t(1) \pm c_{1-\alpha/2}\sigma_u,$$

where $u_t(1)$ denotes the forecast, as usual, and $c_{1-\alpha/2}$ is the relevant $1 - \alpha/2$ percentage point of the normal distribution (see Section 2.2.3). Thus, the forecast intervals are of constant width, regardless of the forecast origin t. In contrast, if u_t is a GARCH(q, m) process, the correct 1-step ahead $(1-\alpha)100\%$ forecast interval is

$$u_t(1) \pm c_{1-\alpha/2}\sigma_{t+1|t}, \quad (16.2.8)$$

where the length depends on the history of the process because the conditional standard deviation, $\sigma_{t+1|t}$, varies over time.

To illustrate this phenomenon, suppose the mean-adjusted DAX returns were generated by a GARCH$(1, 1)$ model with conditionally normal components and conditional variances

$$\sigma_{t|t-1}^2 = 0.0003 + 0.120u_{t-1}^2 + 0.771\sigma_{t-1|t-2}^2.$$

This model was actually fitted to the monthly DAX returns by Lütkepohl (1997) for the period 1960–1991. The 1-step ahead 95% forecast intervals are shown in Figure 16.2. The unconditional variance is in this case

$$\sigma_u^2 = \frac{\gamma_0}{1 - \gamma_1 - \beta_1} = \frac{0.0003}{1 - 0.120 - 0.771} = 2.75 \times 10^{-3}.$$

Assuming mistakenly that the data is i.i.d. normal and using the foregoing white noise variance, results in the constant forecast intervals also shown in Figure 16.2. It is important to note the implications of these results. The constant intervals completely ignore the variations in volatility, whereas the GARCH intervals clearly reflect the greater forecast uncertainty in times of high volatility and are narrower in times where the stock market is less volatile.

As mentioned earlier, if the residuals follow a normal GARCH process, the unconditional distribution of the observations will generally be nonnormal.

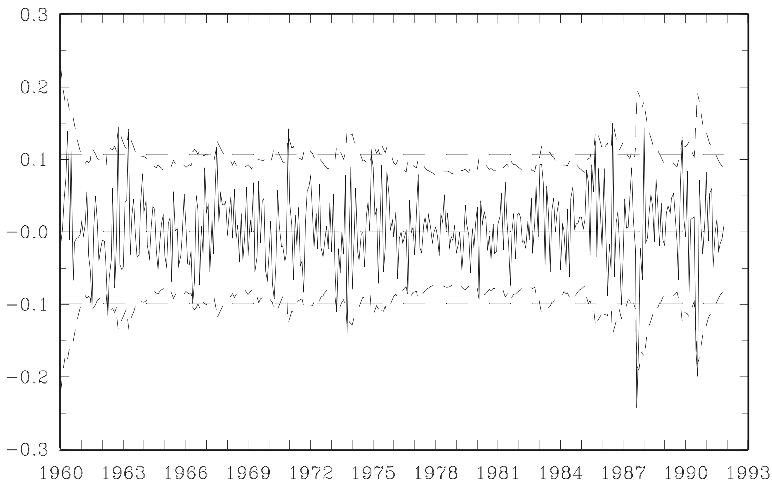

Fig. 16.2. 95% 1-step ahead forecast intervals for the DAX returns obtained under GARCH (- - -) and constant (– –) variance assumptions.

Hence, the constant forecast intervals which have been computed under normality assumptions may not have the desired 95% probability content because of the false distributional assumption. The nonnormal unconditional distribution of GARCH processes also complicates multi-step interval forecasting. Formulas and properties of multi-step forecasts were discussed by Baillie & Bollerslev (1992). Without going into details, it may be worth noting that for a stationary process, when the forecast horizon increases, the optimal forecast will always approach the process mean with the unconditional variance being the forecast error variance and the forecast error distribution approaching the unconditional process distribution, which will generally be nonnormal if the conditional distribution is normal.

We will now discuss how to extend these concepts to the case of vector processes. In that context, we will also address the issue of estimating the parameters of a GARCH model.

16.3 Multivariate GARCH Models

Multivariate extensions of ARCH and GARCH models may be defined in principle similarly to VAR and VARMA models. Early articles on multivariate ARCH and GARCH models are Engle, Granger & Kraft (1986), Diebold & Nerlove (1989), Bollerslev, Engle & Wooldridge (1988). There are a number of complications in analyzing and estimating such models which will be discussed now. The simpler multivariate ARCH models will be considered first.

16.3.1 Multivariate ARCH

Suppose that $u_t = (u_{1t}, \dots, u_{Kt})'$ is a K-dimensional zero mean, serially uncorrelated process which may be the residual process of some dynamic model and which can be represented as

$$u_t = \Sigma_{t|t-1}^{1/2} \varepsilon_t, \tag{16.3.1}$$

where ε_t is K-dimensional i.i.d. white noise, $\varepsilon_t \sim$ i.i.d. $(0, I_K)$, and $\Sigma_{t|t-1}$ is the conditional covariance matrix of u_t, given $u_{t-1}, u_{t-2}, \dots$. As usual, $\Sigma_{t|t-1}^{1/2}$ is the symmetric positive definite square root of $\Sigma_{t|t-1}$ (see Appendix A.9.2 for details on the square root of a positive definite matrix). Obviously, the u_t's have a conditional distribution, given $\Omega_{t-1} := \{u_{t-1}, u_{t-2}, \dots\}$, of the form

$$u_t|\Omega_{t-1} \sim (0, \Sigma_{t|t-1}). \tag{16.3.2}$$

They represent a multivariate ARCH(q) process if

$$\text{vech}(\Sigma_{t|t-1}) = \gamma_0 + \Gamma_1 \text{vech}(u_{t-1}u_{t-1}') + \cdots + \Gamma_q \text{vech}(u_{t-q}u_{t-q}'), \tag{16.3.3}$$

where vech again denotes the half-vectorization operator which stacks the columns of a square matrix from the diagonal downwards in a vector, γ_0 is a $\frac{1}{2}K(K+1)$-dimensional vector of constants and the Γ_j's are $(\frac{1}{2}K(K+1) \times \frac{1}{2}K(K+1))$ coefficient matrices. Different conditional distributions have been assumed and analyzed. For example, a multivariate normal conditional distribution may be considered, i.e., $\varepsilon_t \sim \mathcal{N}(0, I_K)$, so that $u_t|\Omega_{t-1} \sim \mathcal{N}(0, \Sigma_{t|t-1})$. Although this distribution is perhaps not the most suitable one for many financial time series, it will play a role when parameter estimation is discussed in Section 16.4. Conditional distributions of processes representing financial time series are often better represented by more heavy-tailed distributions such as t-distributions with a small degrees of freedom parameter.

As an example, consider a bivariate ($K = 2$) ARCH(1) process,

$$\text{vech}\begin{bmatrix} \sigma_{11,t|t-1} & \sigma_{12,t|t-1} \\ \sigma_{12,t|t-1} & \sigma_{22,t|t-1} \end{bmatrix} = \begin{bmatrix} \sigma_{11,t|t-1} \\ \sigma_{12,t|t-1} \\ \sigma_{22,t|t-1} \end{bmatrix}$$
$$= \begin{bmatrix} \gamma_{10} \\ \gamma_{20} \\ \gamma_{30} \end{bmatrix} + \begin{bmatrix} \gamma_{11} & \gamma_{12} & \gamma_{13} \\ \gamma_{21} & \gamma_{22} & \gamma_{23} \\ \gamma_{31} & \gamma_{32} & \gamma_{33} \end{bmatrix} \begin{bmatrix} u_{1,t-1}^2 \\ u_{1,t-1}u_{2,t-1} \\ u_{2,t-1}^2 \end{bmatrix}.$$

Obviously, even this simple model for a bivariate series has a fair number of parameters which makes it difficult to handle. In particular, the implications of a general model of this type for the relationships between the variables and their higher order moment properties are not obvious. Therefore, more restricted models have been proposed. For instance, Bollerslev et al. (1988) considered diagonal ARCH processes where the Γ_j matrices are all diagonal. In the first order case, the model has the form

$$\begin{bmatrix} \sigma_{11,t|t-1} \\ \sigma_{12,t|t-1} \\ \sigma_{22,t|t-1} \end{bmatrix} = \begin{bmatrix} \gamma_{10} \\ \gamma_{20} \\ \gamma_{30} \end{bmatrix} + \begin{bmatrix} \gamma_{11} & 0 & 0 \\ 0 & \gamma_{22} & 0 \\ 0 & 0 & \gamma_{33} \end{bmatrix} \begin{bmatrix} u_{1,t-1}^2 \\ u_{1,t-1}u_{2,t-1} \\ u_{2,t-1}^2 \end{bmatrix}.$$

Even simple processes of this type can generate rich volatility dynamics. Still, despite their simpler structure, processes of this type involve nontrivial technical problems. One of them is that the parameters have to be such that the conditional covariance matrices $\Sigma_{t|t-1}$ are all positive definite. To guarantee this property, Baba, Engle, Kraft & Kroner (1990) and Engle & Kroner (1995) investigated the following variant of a multivariate ARCH model,

$$\Sigma_{t|t-1} = \Gamma_0^* + \Gamma_1^{*\prime} u_{t-1}u_{t-1}' \Gamma_1^* + \cdots + \Gamma_q^{*\prime} u_{t-q}u_{t-q}' \Gamma_q^*, \tag{16.3.4}$$

where the Γ_j^*'s are each $(K \times K)$ matrices. This particular multivariate model has been christened *BEKK model*. Here the $\Sigma_{t|t-1}$ are positive definite if Γ_0^* has this property which may be enforced by writing it in a product form, $\Gamma_0^* = C_0^{*\prime} C_0^*$ with triangular C_0^* matrix. Another advantage of this model is that it is relatively parsimonious. For instance, for a bivariate process with $K = 2$ and $q = 1$, there are only 7 parameters, whereas the full model has 12 coefficients. Moreover, in contrast to the diagonal model, it can produce quite rich interactions between the conditional second order moments.

16.3.2 MGARCH

In principle, multivariate ARCH models may be generalized in the same way as in the univariate case. In the multivariate GARCH (MGARCH) model for u_t, the conditional covariance matrices have the form

$$\text{vech}(\Sigma_{t|t-1}) = \gamma_0 + \sum_{j=1}^{q} \Gamma_j \text{vech}(u_{t-j}u_{t-j}') + \sum_{j=1}^{m} G_j \text{vech}(\Sigma_{t-j|t-j-1}), \tag{16.3.5}$$

where the G_j's are also fixed $(\frac{1}{2}K(K+1) \times \frac{1}{2}K(K+1))$ coefficient matrices. For example, for a bivariate GARCH$(1,1)$ model,

$$\text{vech} \begin{bmatrix} \sigma_{11,t|t-1} & \sigma_{12,t|t-1} \\ \sigma_{12,t|t-1} & \sigma_{22,t|t-1} \end{bmatrix} = \begin{bmatrix} \sigma_{11,t|t-1} \\ \sigma_{12,t|t-1} \\ \sigma_{22,t|t-1} \end{bmatrix}$$

$$= \begin{bmatrix} \gamma_{10} \\ \gamma_{20} \\ \gamma_{30} \end{bmatrix} + \begin{bmatrix} \gamma_{11} & \gamma_{12} & \gamma_{13} \\ \gamma_{21} & \gamma_{22} & \gamma_{23} \\ \gamma_{31} & \gamma_{32} & \gamma_{33} \end{bmatrix} \begin{bmatrix} u_{1,t-1}^2 \\ u_{1,t-1}u_{2,t-1} \\ u_{2,t-1}^2 \end{bmatrix}$$
$$+ \begin{bmatrix} g_{11} & g_{12} & g_{13} \\ g_{21} & g_{22} & g_{23} \\ g_{31} & g_{32} & g_{33} \end{bmatrix} \begin{bmatrix} \sigma_{11,t-1|t-2} \\ \sigma_{12,t-1|t-2} \\ \sigma_{22,t-1|t-2} \end{bmatrix}.$$

A VARMA representation of an MGARCH process may be obtained analogously to the univariate case (see (16.2.7)) by defining $\mathbf{x}_t :=$ vech$(u_t u_t')$ and $v_t := \mathbf{x}_t -$ vech$(\Sigma_{t|t-1})$. Using these specifications and substituting $\mathbf{x}_t - v_t$ for vech$(\Sigma_{t|t-1})$, (16.3.5) can be rewritten as

$$\mathbf{x}_t = \gamma_0 + \sum_{j=1}^{\max(q,m)} (\Gamma_j + G_j)\mathbf{x}_{t-j} + v_t - \sum_{j=1}^{m} G_j v_{t-j},$$

where $\Gamma_j = 0$ for $j > q$ and $G_j = 0$ for $j > m$. This representation is occasionally useful in deriving properties of MGARCH processes (e.g., Section 16.6.1).

Engle & Kroner (1995) showed that the MGARCH process u_t with conditional covariances as given in (16.3.5) is stationary if and only if all eigenvalues of the matrix

$$\sum_{j=1}^{q} \Gamma_j + \sum_{j=1}^{m} G_j \tag{16.3.6}$$

have modulus less than one.

The parameter space of an MGARCH model has a large dimension in general and needs to be restricted to guarantee uniqueness of the representation and to obtain suitable properties of the conditional covariances. To reduce the parameter space, Bollerslev et al. (1988) discussed diagonal MGARCH models, where the Γ_j's and G_i's in (16.3.5) are diagonal matrices. Alternatively, a BEKK GARCH model of the following form may be useful:

$$\Sigma_{t|t-1} = C_0^{*\prime} C_0^* + \sum_{n=1}^{N}\sum_{j=1}^{q} \Gamma_{jn}^{*\prime} u_{t-j}u_{t-j}' \Gamma_{jn}^* + \sum_{n=1}^{N}\sum_{j=1}^{m} G_{jn}^{*\prime} \Sigma_{t-j|t-j-1} G_{jn}^*, \tag{16.3.7}$$

where again C_0^* is a triangular $(K \times K)$ matrix and the coefficient matrices Γ_{jn}^*, G_{jn}^* are also $(K \times K)$. Given the similarity of MGARCH and VARMA models, it is clear from Chapter 12, Section 12.1, that restrictions have to be imposed on the coefficient matrices to ensure uniqueness of the parameterization. Engle & Kroner (1995) gave the following properties of BEKK GARCH models which also address the uniqueness problem:

(1) From the stationarity condition (16.3.6), the BEKK model is seen to be stationary if all eigenvalues of the matrix

$$\sum_{n=1}^{N}\sum_{j=1}^{q}\Gamma_{jn}^{*\prime}\otimes\Gamma_{jn}^{*\prime}+\sum_{n=1}^{N}\sum_{j=1}^{m}G_{jn}^{*\prime}\otimes G_{jn}^{*\prime} \tag{16.3.8}$$

have modulus less than one.

(2) The BEKK model nests all positive definite diagonal GARCH models, that is, every diagonal GARCH model with positive definite conditional covariance matrices has a BEKK representation.

(3) The BEKK model (16.3.7) generates positive definite covariance matrices $\Sigma_{t|t-1}$ if $\Sigma_{0|-1},\Sigma_{-1|-2},\dots,\Sigma_{-m+1|-m}$ are positive definite and if at least one of the matrices C_0^*, G_{jn}^*, $j=1,\dots,m$, $n=1,\dots,N$, is nonsingular (see Engle & Kroner (1995, Proposition 2.5)).

(4) In the class of BEKK GARCH(1,1) models with $N=1$, the representation

$$\Sigma_{t|t-1}=C_0^{*\prime}C_0^*+\Gamma_{11}^{*\prime}u_{t-1}u_{t-1}^{\prime}\Gamma_{11}^*+G_{11}^{*\prime}\Sigma_{t-1|t-2}G_{11}^*$$

is unique if all diagonal elements of C_0^* are positive and $\gamma_{11,1}^*$, $g_{11,1}^*>0$. Here $\gamma_{11,1}^*$ and $g_{11,1}^*$ represent the upper left-hand elements of Γ_{11}^* and G_{11}^*, respectively.

(5) For a more general BEKK GARCH(1,1) model with

$$\Sigma_{t|t-1}=C_0^{*\prime}C_0^*+\sum_{n=1}^{N}\Gamma_{1n}^{*\prime}u_{t-1}u_{t-1}^{\prime}\Gamma_{1n}^*+\sum_{n=1}^{N}G_{1n}^{*\prime}\Sigma_{t-1|t-2}G_{1n}^*,$$

uniqueness is achieved by the following restrictions:
  (a) All diagonal elements of C_0^* are positive.
  (b) $\Gamma_{1n}^*=G_{1n}^*=0$ for $n>K^2$.
  (c) In the matrices $\Gamma_{1n_j}^*$ with $n_j=K(j-1)+1,\dots,Kj$, and $j=1,\dots,K$, the first $j-1$ columns and the first $n_j-K(j-1)-1$ rows are zero. Moreover, the lower right hand element of $\Gamma_{1n_j}^*$, $\gamma_{KK,n_j}^*>0$.
  (d) Restrictions analogous to those for the Γ_{1n}^* also hold for the G_{1n}^*.

  (Engle & Kroner (1995, Proposition 2.3)).

For illustrative purposes, suppose $K=3$ so that $N=K^2=9$ and $n_1=1,2,3$; $n_2=4,5,6$; $n_3=7,8,9$. Hence, a unique representation is obtained if the zero restrictions shown in the following matrices are imposed:

$$\Gamma_{11}^*=\begin{bmatrix}\gamma_{11,1}^* & \gamma_{12,1}^* & \gamma_{13,1}^*\\ \gamma_{21,1}^* & \gamma_{22,1}^* & \gamma_{23,1}^*\\ \gamma_{31,1}^* & \gamma_{32,1}^* & \gamma_{33,1}^*\end{bmatrix},\qquad \Gamma_{12}^*=\begin{bmatrix}0 & 0 & 0\\ \gamma_{21,2}^* & \gamma_{22,2}^* & \gamma_{23,2}^*\\ \gamma_{31,2}^* & \gamma_{32,2}^* & \gamma_{33,2}^*\end{bmatrix},$$

$$\Gamma_{13}^*=\begin{bmatrix}0 & 0 & 0\\ 0 & 0 & 0\\ \gamma_{31,3}^* & \gamma_{32,3}^* & \gamma_{33,3}^*\end{bmatrix},\qquad \Gamma_{14}^*=\begin{bmatrix}0 & \gamma_{12,4}^* & \gamma_{13,4}^*\\ 0 & \gamma_{22,4}^* & \gamma_{23,4}^*\\ 0 & \gamma_{32,4}^* & \gamma_{33,4}^*\end{bmatrix},$$

$$\varGamma_{15}^* = \begin{bmatrix} 0 & 0 & 0 \\ 0 & \gamma_{22,5}^* & \gamma_{23,5}^* \\ 0 & \gamma_{32,5}^* & \gamma_{33,5}^* \end{bmatrix}, \qquad \varGamma_{16}^* = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & \gamma_{32,6}^* & \gamma_{33,6}^* \end{bmatrix},$$

$$\varGamma_{17}^* = \begin{bmatrix} 0 & 0 & \gamma_{13,7}^* \\ 0 & 0 & \gamma_{23,7}^* \\ 0 & 0 & \gamma_{33,7}^* \end{bmatrix}, \qquad \varGamma_{18}^* = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & \gamma_{23,8}^* \\ 0 & 0 & \gamma_{33,8}^* \end{bmatrix},$$

$$\varGamma_{19}^* = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & \gamma_{33,9}^* \end{bmatrix}.$$

The same zero restrictions are also imposed on the G_{1n}^*. Of course, in a specific case, there may be further zero restrictions on the coefficient matrices. In particular, N may be less than K^2. An example of this type is given in Section 16.4.2.

Given the correspondence between GARCH and VARMA models, it should be clear from the discussion of uniqueness of VARMA representations in Chapter 12 that a unique parameterization of a multivariate GARCH representation is not a trivial matter. Whether the constraints given here are the most operational ones in practice remains to be seen. If a unique representation is set up, estimation becomes possible. This issue will be discussed in Section 16.4.

16.3.3 Other Multivariate ARCH and GARCH Models

Although the BEKK model with low orders may be a relatively parsimonious representation of the conditional covariance structure of a process, the number of parameters still grows quickly with the dimension of the underlying system. Therefore, in practice, it is only feasible if systems with just a few variables are under consideration and further simplifications were proposed to alleviate modelling of higher dimensional processes. Some of them can be viewed as special BEKK models. For example, Lin (1992) specified a *factor GARCH model*, where the $\varGamma_{1n}^*$'s and G_{1n}^*'s in a BEKK GARCH$(1,1)$ model are of the form

$$\varGamma_{1n}^* = \gamma_n \eta_n \xi_n' \quad \text{and} \quad G_{1n}^* = g_n \eta_n \xi_n', \quad n = 1, \dots, N. \tag{16.3.9}$$

Here γ_n and g_n are scalars and η_n and ξ_n are $(K \times 1)$ vectors satisfying $\xi_n' \xi_n = 1$, $\eta_n' \xi_n = 1$ for $n = 1, \dots, N$ and $\eta_n' \xi_k = 0$ for $n \neq k$. Thus, the $\varGamma_{1n}^*$'s and G_{1n}^*'s have all rank 1.

In some proposals, the conditional covariance matrix has the form

$$\varSigma_{t|t-1} = Q H_{t|t-1} Q', \tag{16.3.10}$$

where Q is $(K \times K)$ and does not depend on t, whereas $H_{t|t-1}$ is a positive definite $(K \times K)$ matrix which may depend on t. For example, Vrontos, Dellaportas & Politis (2003) proposed to use a triangular matrix Q and specified

$$H_{t|t-1} = \text{diag}(\sigma^2_{1t|t-1}, \ldots, \sigma^2_{Kt|t-1}) \tag{16.3.11}$$

to be a diagonal matrix with univariate GARCH conditional variances $\sigma^2_{kt|t-1}$ on the diagonal. A closely related model, the so-called *generalized orthogonal GARCH model*, was proposed by van der Weide (2002).

Clearly, restricting the second moment dynamics to a transformation of univariate GARCH models as in (16.3.11) is restrictive and, in particular, it limits the covariance dynamics in a potentially undesired way. Therefore, the alternative specification

$$\Sigma_{t|t-1} = D_t R_t D_t \tag{16.3.12}$$

was proposed, where restrictions of different forms are specified for the $(K \times K)$ matrices D_t and R_t. For example, if $R_t = R$ is a time invariant correlation matrix and $D_t = \text{diag}(\sigma_{1t|t-1}, \ldots, \sigma_{Kt|t-1})$ is a diagonal matrix with time varying conditional standard deviations on the diagonal, Bollerslev's (1990) *constant conditional correlation (CCC) MGARCH model* is obtained. Clearly, in this model, the time invariant R is the correlation matrix corresponding to the covariance matrix $\Sigma_{t|t-1}$ for all t. Engle (2002) extended the model by allowing for richer dynamics and proposed the so-called *dynamic conditional correlation (DCC) model*. A related model was also proposed by Tse & Tsui (2002).

In financial markets, it has been observed frequently that positive and negative shocks or news have quite different effects (Black (1976)). This so-called *leverage effect* can be introduced in different ways in MGARCH models. For example, Hafner & Herwartz (1998b) and Herwartz & Lütkepohl (2000) generalized a univariate proposal by Glosten, Jagannathan & Runkle (1993) and replaced

$$\Gamma^{*\prime}_{11} u_{t-1} u'_{t-1} \Gamma^*_{11} \quad \text{by} \quad \Gamma^{*\prime}_{11} u_{t-1} u'_{t-1} \Gamma^*_{11} + \Gamma^{*\prime}_{-} u_{t-1} u'_{t-1} \Gamma^*_{-} \mathbb{I}\left(\sum_{k=1}^{K} u_{kt} < 0\right) \tag{16.3.13}$$

in a BEKK model with $N = 1$. Here $\mathbb{I}(\cdot)$ denotes an indicator function which takes the value 1 if the argument is valid and 0 otherwise and Γ^*_{-} is an additional $(K \times K)$ coefficient matrix. Another approach to allow for asymmetry is to use the so-called *exponential GARCH* (EGARCH) model proposed by Nelson (1991). A multivariate version was considered by Braun, Nelson & Sunier (1995).

A range of other models was also proposed and the literature on MGARCH models has grown rapidly over the last years. A recent survey was provided by Bauwens et al. (2004), where more information on the aforementioned models, further proposals and references can be found.

16.4 Estimation

16.4.1 Theory

Using Bayes' theorem, the joint density function of $u_1, \ldots, u_T$ is $f(u_1, \ldots, u_T) = f(u_1)f(u_2|u_1)\cdots f(u_T|u_{T-1}, \ldots, u_1)$. Thus, if in (16.3.1) $\varepsilon_t \sim$ i.i.d. $\mathcal{N}(0, I_K)$ so that the conditional distribution of u_t given Ω_{t-1} is Gaussian and if the u_t are observed quantities, the log-likelihood function of the general GARCH model described by (16.3.5), for a sample $u_1, \ldots, u_T$, is given by

$$\ln l(\delta) = \sum_{t=1}^{T} \ln l_t(\delta), \tag{16.4.1}$$

where $\delta = \text{vec}(\gamma_0, \Gamma_1, \ldots, \Gamma_q, G_1, \ldots, G_m)$ is the vector of unknown parameters and

$$\ln l_t(\delta) = -\frac{K}{2}\ln 2\pi - \frac{1}{2}\ln|\Sigma_{t|t-1}| - \frac{1}{2}u_t'\Sigma_{t|t-1}^{-1}u_t, \quad t = 1, \ldots, T, \tag{16.4.2}$$

where the required initial values for specifying $\Sigma_{t|t-1}$ are assumed to be available. Similarly, the log-likelihood may be set up for special cases such as diagonal or BEKK models.

The likelihood function may be maximized with respect to the parameters δ by using numerical methods. A closed form solution does not exist because of the nonlinearity of the function. For uniqueness of the maximum and, hence, the existence of a unique ML estimator, it is important that an identified, unique parameterization is used, e.g., the BEKK form of the model with the restrictions discussed in Section 16.3.2. Of course, if the log-likelihood function (16.4.1)/(16.4.2) is used although the true distribution of the ε_t is nonnormal, the resulting estimators will just be quasi ML estimators. Comte & Lieberman (2003) showed that quasi ML estimators have the following properties.

Proposition 16.1 (*Properties of Quasi ML Estimators of GARCH Models*)
Let u_t be a BEKK GARCH process satisfying the following conditions:

(a) The parameter space is compact and identification restrictions are imposed.
(b) The eigenvalues of the matrix (16.3.6) have modulus less than one.
(c) $\varepsilon_t = (\varepsilon_{1t}, \ldots, \varepsilon_{Kt})' \sim$ i.i.d. $(0, I_K)$ with $\varepsilon_{it}, \varepsilon_{jt}$ independent for $i \neq j$ $(i, j = 1, \ldots, K)$ and such that u_t admits moments of at least order 8. Moreover, the ε_t are continuous random variables with a density which is positive in a neighborhood of the origin.
(d) The initial values $u_t, t \leq 0$, are such that the process u_t is strictly stationary.

Then the quasi ML estimator $\widetilde{\delta}$ of δ obtained by maximizing the Gaussian likelihood function exists and is strongly consistent,

$\widetilde{\delta} \stackrel{a.s.}{\to} \delta.$

Moreover, $\widetilde{\delta}$ has an asymptotic normal distribution,

$$\sqrt{T}(\widetilde{\delta} - \delta) \stackrel{d}{\to} \mathcal{N}(0, C_1^{-1} C_0 C_1^{-1}), \tag{16.4.3}$$

where

$$C_1 = -E\left(\frac{\partial^2 \ln l_t(\delta)}{\partial\delta\partial\delta'}\right) \quad \text{and} \quad C_0 = E\left(\frac{\partial \ln l_t(\delta)}{\partial\delta}\frac{\partial \ln l_t(\delta)}{\partial\delta'}\right). \tag{16.4.4}$$

■

A number of comments are worth making regarding this proposition.

Remark 1 It can be shown that $C_0 = C_1$ if ε_t is normally distributed. Hence, in this case, the asymptotic distribution in (16.4.3) becomes $\mathcal{N}(0, C_1^{-1})$, that is, the covariance matrix is the inverse asymptotic information matrix. ■

Remark 2 The condition of a compact parameter space is typical for nonlinear estimation problems. Although not totally satisfactory, it is not regarded as very problematic because the compact subset of the Euclidean space to which it refers may be so large that the condition is not really restrictive. The assumption regarding the initial values is also not restrictive if the stationarity of the process is accepted. It can be replaced by the assumption that the initial values are fixed, nonstochastic values. ■

Remark 3 In contrast, the assumptions regarding the ε_t are not fully satisfactory. In particular, the requirement that moments of order 8 have to exist for u_t is undesirable for financial time series where the existence of higher order moments is regarded as problematic. On the other hand, the theorem improves on previously available results which shows how difficult it is to derive asymptotic properties of the estimators of MGARCH processes. A number of other authors have derived more specialized results, notably for univariate processes (see the review articles mentioned at the end of Section 16.1). For multivariate GARCH processes, consistency of the quasi ML estimators was shown by Jeantheau (1998) under the main assumption of a strictly stationary and ergodic process. Ling & McAleer (2003) derived asymptotic normality of quasi ML estimators under less restrictive moment assumptions for a VARMA process with CCC GARCH residuals. ■

Typically, the u_t are residuals of some dynamic model. Suppose they are the errors of a VAR(p) process, possibly with integrated or cointegrated variables. Thus, we have a model of the form

$$y_t = \nu + A_1 y_{t-1} + \cdots + A_p y_{t-p} + u_t.$$

In this case, the VAR parameters have to be estimated in addition to the coefficients associated with the u_t process. Setting up the corresponding Gaussian likelihood or quasi likelihood function is not difficult. However, the optimization may be a formidable task. Assuming that the numerical problems can be solved, there is some hope that the asymptotics can also be resolved because, under quite general conditions, the asymptotic information matrix of the VAR parameters and the GARCH parameters is block diagonal so that the estimators of the VAR coefficients are asymptotically independent of the GARCH parameter estimators. This result also suggests a two step estimation procedure in which the VAR coefficients $\nu, A_1, \dots, A_p$ are estimated by LS or, if restrictions are imposed on the parameters, by EGLS and then a GARCH model is fitted to the residuals of the first stage estimation.

Given that normality of the conditional distribution of the u_t is often difficult to justify, in particular, in financial applications, it may also be worth pointing out that ML estimation with other distributions has been studied. The survey by Bauwens et al. (2004) provides further information and references on these issues as well as computational aspects of ML and quasi ML estimation.

16.4.2 An Example

Two series of daily stock returns (first differences of ln prices) will be used to illustrate the previous theoretical considerations. In particular, returns of VW (Volkswagen) common stock (y_{1t}) and preference stock (y_{2t}) for the period January 1987–December 1992 (1579 observations) are used.[1] The two series are plotted in Figure 16.3. The corresponding logarithms of the price series are both strongly related to the performance of the VW company and, hence, they are likely to be related to each other. Therefore, it makes sense to analyze the stocks as a bivariate series. The ln price series were previously analyzed by Herwartz & Lütkepohl (2000). In contrast to these authors, we consider the bivariate series $y_t = (y_{1t}, y_{2t})'$ of returns. Although the ln prices may be cointegrated, a preliminary analysis has shown that there is weak evidence of cointegration at best. Therefore, it seems justified to focus on the returns in the following.

The two series of stock returns display some changes in their volatility and there are also some unusually large (in absolute value) observations. Such values are often classified as outliers. Thus, based on the graphs in Figure 16.3, one may not expect the series to be generated by a Gaussian process and ARCH or GARCH models may be used to capture the volatility dynamics.

Fitting VAR(p) models of increasing order to the bivariate series y_t, it turns out that AIC and HQ recommend an order of $p = 3$ while SC suggests $p = 0$. Therefore, the residuals of the following estimated VAR(3) model (with t-values in parentheses) will be used in the following bivariate GARCH analysis:

[1] The price series are from Deutsche Finanzdatenbank Karlsruhe.

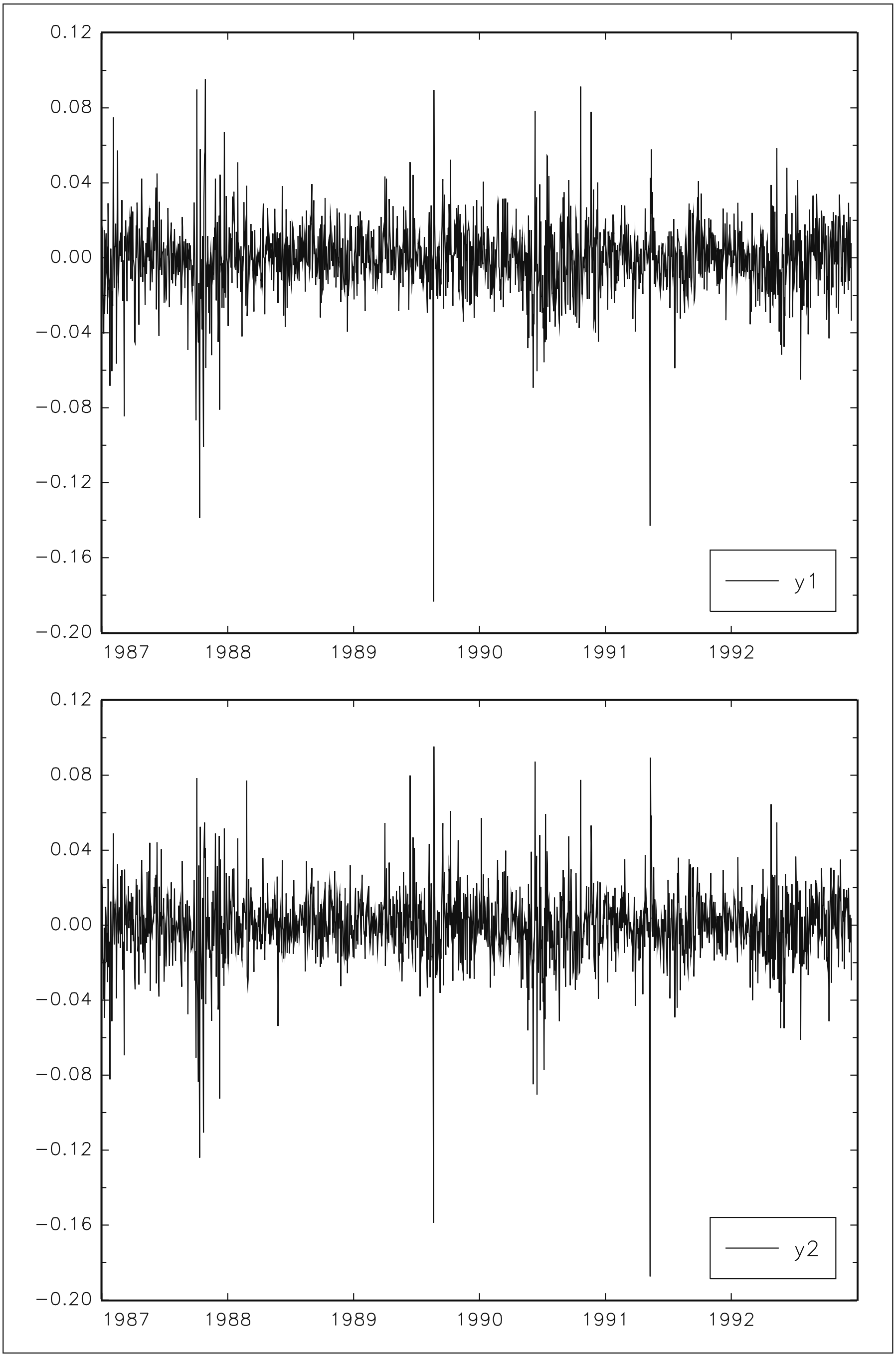

Fig. 16.3. Daily returns of VW common (y_1) and preference stock (y_2).

$$y_t = \begin{bmatrix} \underset{(-0.5)}{-0.24 \times 10^{-3}} \\ \underset{(-0.8)}{-0.42 \times 10^{-3}} \end{bmatrix} + \begin{bmatrix} \underset{(-0.0)}{-0.00} & \underset{(0.3)}{0.02} \\ \underset{(2.0)}{0.12} & \underset{(-2.3)}{-0.13} \end{bmatrix} y_{t-1}$$

$$+ \begin{bmatrix} \underset{(-3.2)}{-0.18} & \underset{(2.8)}{0.16} \\ \underset{(-1.3)}{-0.08} & \underset{(0.6)}{0.03} \end{bmatrix} y_{t-2} + \begin{bmatrix} \underset{(-1.3)}{-0.08} & \underset{(1.9)}{0.11} \\ \underset{(-0.2)}{-0.01} & \underset{(0.1)}{0.01} \end{bmatrix} y_{t-3} + \widehat{u}_t. \quad (16.4.5)$$

The residual series are plotted in Figure 16.4. They still show volatility clusters and outliers. Hence, there may be conditional heteroskedasticity in the residuals of model (16.4.5). In that case, it may not be a good strategy to choose the VAR order first by one of our standard model selection criteria, as we have done it here. Alternatively, it may be preferable to derive criteria that allow a simultaneous determination of the joint model for the conditional first and second moments (see Brooks & Burke (2003)). We will nevertheless use the residuals from the model (16.4.5) in the subsequent analysis for illustrative purposes.

Based on the residuals of the model (16.4.5), the following BEKK GARCH-(1, 1) model was estimated (with t-values in parentheses):

$$\Sigma_{t|t-1} = \begin{bmatrix} \underset{(2.6)}{0.004} & \underset{(3.2)}{0.005} \\ 0 & \underset{(5.5)}{0.003} \end{bmatrix} \begin{bmatrix} \underset{(2.6)}{0.004} & 0 \\ \underset{(3.2)}{0.005} & \underset{(5.5)}{0.003} \end{bmatrix}$$

$$+ \begin{bmatrix} \underset{(1.9)}{0.254} & \underset{(-0.0)}{-0.004} \\ \underset{(0.1)}{0.040} & \underset{(1.1)}{0.332} \end{bmatrix} \widehat{u}_{t-1}\widehat{u}'_{t-1} \begin{bmatrix} \underset{(1.9)}{0.254} & \underset{(0.1)}{0.040} \\ \underset{(-0.0)}{-0.004} & \underset{(1.1)}{0.332} \end{bmatrix}$$

$$+ \begin{bmatrix} \underset{(7.8)}{0.941} & \underset{(0.2)}{0.023} \\ \underset{(-0.2)}{-0.019} & \underset{(17.6)}{0.864} \end{bmatrix} \Sigma_{t-1|t-2} \begin{bmatrix} \underset{(7.8)}{0.941} & \underset{(-0.2)}{-0.019} \\ \underset{(0.2)}{0.023} & \underset{(17.6)}{0.864} \end{bmatrix}. \quad (16.4.6)$$

Note that the uniqueness conditions mentioned in Section 16.3.2 are satisfied here. It is not clear, however, that in the present situation the t-ratios have standard normal limiting distributions because the assumptions of Proposition 16.1 are violated. In particular, we are working with residuals from a previously fitted model rather than with original observations. Still the sizes of the t-ratios underneath the coefficient estimates indicate some interaction in the conditional second moments.

It would be helpful to have tools for checking the model quality and for analyzing the relationships summarized in the model. For model checking, the estimated $\varepsilon_t = \Sigma_{t|t-1}^{-1/2} u_t$ can be used. Standardized estimated ε_t (divided

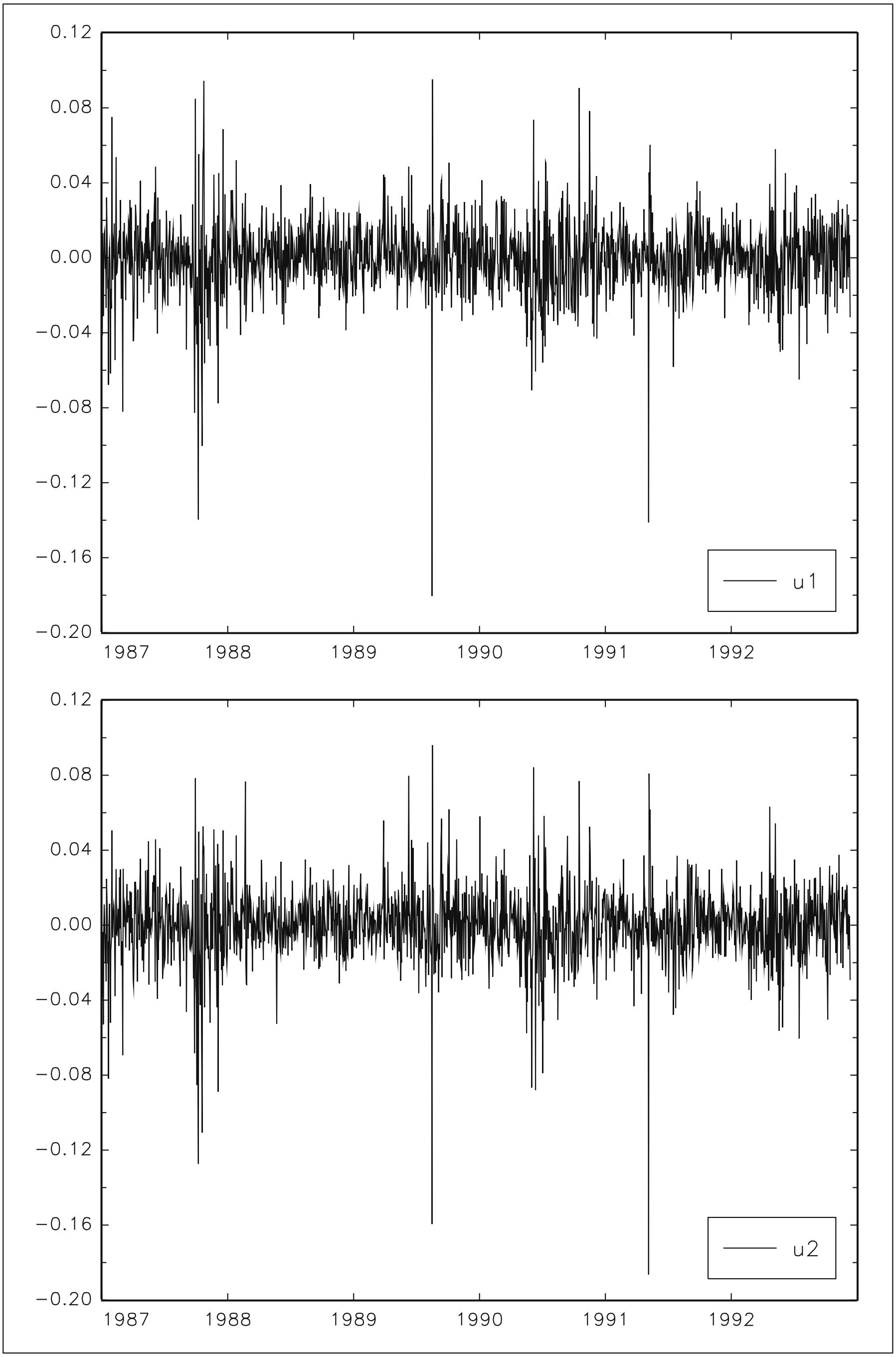

Fig. 16.4. Residual series of model (16.4.5).

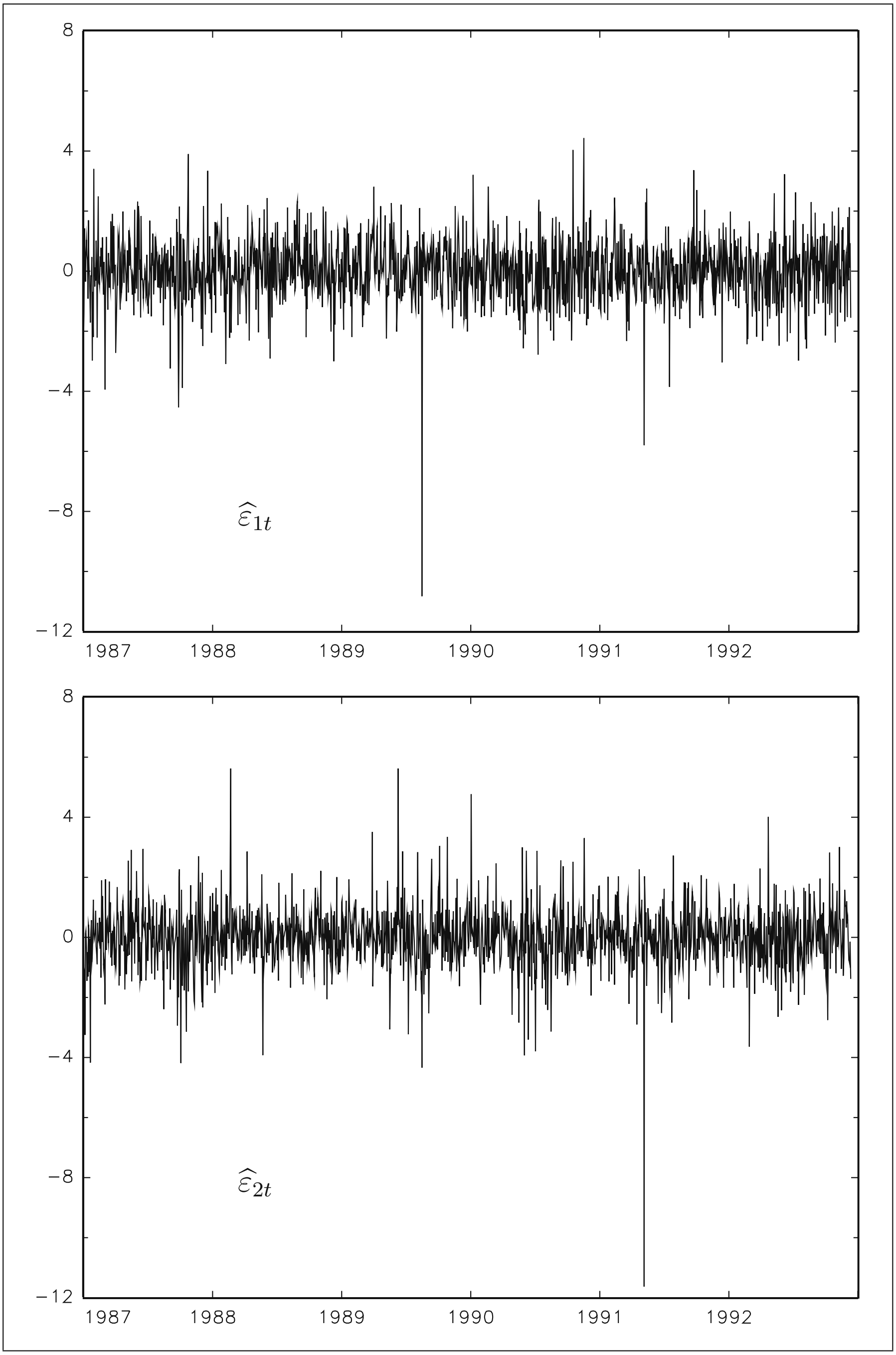

Fig. 16.5. Standardized residuals of model (16.4.6) ($\widehat{\varepsilon}_{1t}$ upper panel, $\widehat{\varepsilon}_{2t}$ lower panel).

by their estimated standard deviations) are plotted in Figure 16.5. Volatility clusters are not quite so obvious anymore as in Figure 16.4. On the other hand, outliers are still present which sheds doubt on the normality of the ε_t. Some tests for model adequacy will be discussed in the next section.

16.5 Checking MGARCH Models

16.5.1 ARCH-LM and ARCH-Portmanteau Tests

Before an MGARCH model is fitted to the residuals of a VAR or VECM, one may want to check if ARCH effects are present in the residuals. An LM test is a standard tool for this purpose (e.g., Doornik & Hendry (1997)). The idea is to consider the auxiliary model

$$\text{vech}(u_t u_t') = \beta_0 + B_1 \text{vech}(u_{t-1} u_{t-1}') + \cdots + B_q \text{vech}(u_{t-q} u_{t-q}') + \text{error}_t, \tag{16.5.1}$$

where β_0 is $\frac{1}{2}K(K+1)$-dimensional and the B_j's are $(\frac{1}{2}K(K+1) \times \frac{1}{2}K(K+1))$ coefficient matrices ($j = 1, \dots, q$). If all the B_j matrices are zero, there is no ARCH in the residuals. Therefore, the pair of hypotheses

$$H_0 : B_1 = \cdots = B_q = 0 \quad \text{versus} \quad H_1 : B_1 \neq 0 \text{ or } \cdots \text{ or } B_q \neq 0, \tag{16.5.2}$$

is checked. It turns out that the corresponding LM statistic can be determined by replacing all unknown u_t's in (16.5.1) by estimated residuals from a VAR or VECM, say, and estimating the parameters in the resulting auxiliary model by LS. Denoting the resulting residual covariance matrix estimator based on (16.5.1) by $\widehat{\Sigma}_{\text{vech}}$ and the corresponding matrix obtained for $q = 0$ by $\widehat{\Sigma}_0$, the relevant LM statistic can be shown to be of the form

$$LM_{MARCH}(q) = \frac{1}{2}TK(K+1) - T\text{tr}(\widehat{\Sigma}_{\text{vech}} \widehat{\Sigma}_0^{-1}). \tag{16.5.3}$$

Under the null hypothesis, the statistic has an asymptotic $\chi^2(qK^2(K+1)^2/4)$-distribution, if u_t satisfies standard conditions (see Doornik & Hendry (1997, Sec. 10.9.2.4)).

In (16.5.1), each of the B_j matrices is of dimension $(\frac{1}{2}K(K+1) \times \frac{1}{2}K(K+1))$ and, hence, the auxiliary model involves a large number of parameters even if the order q and the dimension of the process K are only moderate. Therefore the test is not suitable to check for large q, unless the sample size is very large too. It is possible, however, to apply the test to each of the K residual series individually.

From the VARMA representation of an MGARCH process in Section 16.3.2, it can be seen that there is no ARCH in the process u_t, if the process $\mathbf{x}_t := \text{vech}(u_t u_t')$ has no serial correlation. This observation suggests that we may apply a portmanteau test to $\mathbf{x}_t$ to check for ARCH in u_t. Thus, one may use

$$Q_h^{ARCH} := T \sum_{i=1}^{h} \operatorname{tr}(C_i' C_0^{-1} C_i C_0^{-1}) \tag{16.5.4}$$

or the associated modified version

$$\bar{Q}_h^{ARCH} := T^2 \sum_{i=1}^{h} (T-i)^{-1} \operatorname{tr}(C_i' C_0^{-1} C_i C_0^{-1}) \tag{16.5.5}$$

where now $C_i = T^{-1} \sum_{t=i+1}^{T} (\mathbf{x}_t - \overline{\mathbf{x}})(\mathbf{x}_{t-i} - \overline{\mathbf{x}})'$ $(i = 0, 1, \ldots, h)$.

The asymptotic χ^2-distributions of these tests follow from the results in Section 4.4.3, if $\mathbf{x}_t$ is indeed white noise. In practice, it will usually be replaced by a quantity based on estimation residuals $\widehat{u}_t$. A rigorous treatment of the properties of the statistics in that case seems to be still missing. In principle, the ARCH-portmanteau test can also be applied to the individual residual series.

16.5.2 LM and Portmanteau Tests for Remaining ARCH

In practice, it is also useful to check for remaining ARCH in the residuals of a fitted ARCH or GARCH model. Such tests are of particular importance in the present context because low order multivariate models are typically fitted as a first attempt to account for conditionally heteroskedastic residuals. Higher order models often have an excessive number of parameters and the estimates are difficult to compute numerically. Therefore, it makes sense to start with low order models and increase the order only if the low order model cannot capture the second order moment dynamics in the data properly. Hence, tests for remaining ARCH in the residuals of an MGARCH model are needed.

Both the ARCH-LM and the ARCH-portmanteau tests have been used for this purpose. In that case, the u_t's in the $\mathbf{x}_t$ vectors are replaced by the estimated ε_t from (16.3.1). In other words, $\widetilde{\varepsilon}_t := \widetilde{\Sigma}_{t|t-1}^{-1/2} \widetilde{u}_t$ is used instead of u_t. Here ML estimators are signified by a tilde. Whereas the LM tests maintain their validity under general conditions in the present situation (Engle & Kroner (1995)), the same is not true for the portmanteau tests. They have still found widespread use in applied work (see Tse & Tsui (1999) for references and further discussion).

Again, it may be useful to apply the tests not only to the multivariate residual vectors but also to the univariate components separately. There are also other tests for remaining ARCH which have a sounder theoretical basis than the portmanteau tests (see Bauwens et al. (2004) for a review and Lundbergh & Teräsvirta (2002) for a discussion of the univariate case).

16.5.3 Other Diagnostic Tests

Other diagnostic tools for checking the validity of fitted MGARCH models are also available. In fact, some of the residual diagnostics for VAR models are also applicable here. Instead of the u_t's, the ε_t's should now be used as the basic residuals. For example, they can be used to perform tests for nonnormality. Although the necessary extensions are in many cases possible, some care is needed in the present context. It can by no means be taken for granted that all the procedures work properly. The case of normality tests and related caveats when they are applied to GARCH residuals was discussed by Fiorentini, Sentana & Calzolari (2004).

16.5.4 An Example

As an example, we consider again the VW stock returns. In Section 16.4.2, we have fitted an MGARCH$(1,1)$ model because the residuals of the model (16.4.5) appeared to have volatility clusters. Now we can use ARCH-LM tests and formally test for conditional heteroskedasticity of the u_t's. Some results are presented in Table 16.1. Both bivariate and univariate tests applied to the individual residual series clearly reject the no-ARCH null hypothesis. Thus, there is strong evidence in favor of conditionally heteroskedastic residuals. We do not present results of the ARCH-portmanteau test because its validity is not clear.

Table 16.1. ARCH-LM tests for $\widehat{u}_t$ residuals from (16.4.5)

	bivariate		$\widehat{u}_{1t}$		$\widehat{u}_{2t}$	
test	$LM(1)$	$LM(4)$	$LM(1)$	$LM(4)$	$LM(1)$	$LM(4)$
test value	147.4	245.5	54.9	62.1	56.1	70.8
asymptotic distribution	$\chi^2(9)$	$\chi^2(36)$	$\chi^2(1)$	$\chi^2(4)$	$\chi^2(1)$	$\chi^2(4)$
p-value	0.00	0.00	0.00	0.00	0.00	0.00

Of course, the fact that there may be ARCH in the residuals does not necessarily mean that an MGARCH$(1,1)$ process is a suitable model. Therefore, we also applied tests for remaining ARCH to the residuals $\widetilde{\varepsilon}_t = \widetilde{\Sigma}_{t|t-1}^{-1/2}\widetilde{u}_t$ based on (16.4.6). The test results are given in Table 16.2. Now none of the ARCH-LM tests rejects the null hypothesis at conventional significance levels. On the other hand, applying nonnormality tests confirms what could have been conjectured by looking at the residuals in Figure 16.5, namely that, due to the outliers, normality of the conditional distribution is not likely to be a reasonable assumption. Thus, it may be worth trying some other distribution for the ε_t or some other model than the standard BEKK GARCH$(1,1)$ we have presented in Section 16.4.2.

Table 16.2. ARCH-LM and nonnormality ($\widehat{\lambda}_{sk}$) tests for $\widehat{\varepsilon}_t$ residuals associated with (16.4.6)

	bivariate			$\widehat{\varepsilon}_{1t}$			$\widehat{\varepsilon}_{2t}$		
test	$LM(1)$	$LM(4)$	$\widehat{\lambda}_{sk}$	$LM(1)$	$LM(4)$	$\widehat{\lambda}_{sk}$	$LM(1)$	$LM(4)$	$\widehat{\lambda}_{sk}$
test value	10.43	33.48	12770	0.067	0.364	4440	0.099	0.385	10314
asymp. distr.	$\chi^2(9)$	$\chi^2(36)$	$\chi^2(4)$	$\chi^2(1)$	$\chi^2(4)$	$\chi^2(2)$	$\chi^2(1)$	$\chi^2(4)$	$\chi^2(2)$
p-value	0.32	0.59	0.00	0.80	0.99	0.00	0.75	0.98	0.00

16.6 Interpreting GARCH Models

16.6.1 Causality in Variance

As we have seen in Chapter 2, Section 2.3.1, Granger's definition of causality is based on forecasts. We have also seen in Chapter 2 that, under suitable conditions, optimal forecasts are obtained as conditional expectations. Therefore, Granger-causality may be defined in terms of conditional expectations. In other words, we may define a time series variable x_t to be causal for z_t, if

$$E(z_{t+1}|z_t, z_{t-1}, \dots) \neq E(z_{t+1}|z_t, z_{t-1}, \dots, x_t, x_{t-1}, \dots). \tag{16.6.1}$$

This definition suggests a direct extension to higher order conditional moments. We define x_t to be *causal* for z_t *in r-th moment* if

$$E(z_{t+1}^r|z_t, z_{t-1}, \dots) \neq E(z_{t+1}^r|z_t, z_{t-1}, \dots, x_t, x_{t-1}, \dots). \tag{16.6.2}$$

Thus, (16.6.1) defines causality in mean and considering the central second moments in (16.6.2) gives a definition of causality in variance which is analogous to the previous definition of Granger-causality. The interpretation is also analogous to that of Granger-causality in mean. In other words, if x_t is causal-in-variance for z_t, the conditional volatility of z_t can be predicted more precisely by taking into account present and past information in x_t than without taking this information into account.

If the conditional covariance structure can be described by multivariate ARCH or MGARCH models, the restrictions implied by these definitions are also similar to those for Granger-causality in VAR and VARMA models (see Comte & Lieberman (2000)). In other words, they can be described in terms of zero restrictions on the ARCH or MGARCH parameters. Depending on the specific parameterization of the MGARCH model, the restrictions can be nonlinear in the present situation, however. Tests for causality in variance were proposed and investigated by Cheung & Ng (1996), Hong (2001), and Pantelidis & Pittis (2004).

It is also possible to generalize the causality definition and specify, for example, conditions for both the conditional first and second order moments (e.g., Granger, Robins & Engle (1986)). More generally, one may consider the full conditional distributions rather than just specific moments. In other words, one may define x_t to be causal for z_t if

$$F_{z_{t+1}|z_t,z_{t-1},\dots}(\cdot) \neq F_{z_{t+1}|z_t,z_{t-1},\dots,x_t,x_{t-1},\dots}(\cdot), \tag{16.6.3}$$

where $F_{z|x}(\cdot)$ denotes the conditional distribution function of z given x. Generalizing these concepts to the case where x_t and z_t are vectors is theoretically straightforward, as in the case of Granger-causality in mean.

16.6.2 Conditional Moment Profiles and Generalized Impulse Responses

Impulse responses were used among other tools for analyzing the relations between the variables of linear models such as VARs and VECMs. In linear models, they have the advantage of being time invariant and their shape is invariant to the size and direction of the impulses. These features enable the analyst to represent the reactions of the variables to impulses hitting the system in a small set of graphs. GARCH models are nonlinear models, however. In such models, the situation is quite different. In general, in a nonlinear model, the marginal effect of an impulse will depend on the state of the system when the impulse arrives. Thus, it depends on the history of the variables and it may be different in each time point during the sample. Moreover, the shape of the impulse responses will generally depend on the size and direction of the impulse. For example, quite different reactions may be obtained from positive and negative impulses. In a linear model, a negative impulse of one unit induces the same responses of the variables with opposite sign as a positive impulse of one unit. In contrast, in a nonlinear model, a positive impulse may, e.g., induce almost no reaction of the variables whereas a corresponding negative impulse hitting the system at the same state may lead to a strong reaction. These features are quite plausible in some systems. For example, if the impulses represent news arriving in a financial market, positive news may have a quite different effect than negative news. Hence, nonlinear models clearly have their attractive features for describing economic systems or phenomena.

Still, the greater flexibility of nonlinear models makes them more difficult to interpret properly. In fact, it is not obvious how to define impulse responses of nonlinear models in a meaningful manner. Gallant, Rossi & Tauchen (1993) proposed so-called *conditional moment profiles* which may give useful information on important features and implications of nonlinear multiple time series models. In the spirit of their definition, we consider quantities of the general form

$$E[g(y_{t+h})|y_t + \xi, \Omega_{t-1}] - E[g(y_{t+h})|y_t, \Omega_{t-1}], \quad h = 1, 2, \dots, \tag{16.6.4}$$

where $g(\cdot)$ denotes some function of interest, ξ represents the impulses hitting the system at time t, and $\Omega_{t-1} := (y_{t-1}, y_{t-2}, \dots)$ denotes the history of the variables at time t. In other words, the conditional expectation of some quantity of interest, given the history of y_t in period t, is compared to the conditional expectation that is obtained if a shock ξ occurs at time t. For example, defining

$$g(y_{t+h}) = [y_{t+h} - E(y_{t+h}|\Omega_{t+h-1})][y_{t+h} - E(y_{t+h}|\Omega_{t+h-1})]' \qquad (16.6.5)$$

results in conditional volatility profiles, which may be compared to a baseline profile obtained for a specific history of the process and a zero impulse. Clearly, in general the conditional moment profiles depend on the history Ω_{t-1} as well as the impulse ξ. Similar quantities were also considered by Koop, Pesaran & Potter (1996) who called them *generalized impulse responses*.

If models with ARCH or MGARCH errors are used to describe the volatility dynamics of a financial market, the volatility resulting from the arrival of news may be of interest (see Engle & Ng (1993)). In this case, using the function (16.6.5) and comparing conditional covariance matrices

$$\Sigma_{t+h|t} = E\{[y_{t+h} - E(y_{t+h}|\Omega_{t+h-1})][y_{t+h} - E(y_{t+h}|\Omega_{t+h-1})]'|y_t, \Omega_{t-1}\},$$

based on the actual history at time t, to

$$\Sigma^{\xi}_{t+h|t} = E\{[y_{t+h} - E(y_{t+h}|\Omega_{t+h-1})][y_{t+h} - E(y_{t+h}|\Omega_{t+h-1})]'|y_t + \xi, \Omega_{t-1}\},$$

for $h = 1, 2, \dots$, may give an impression of the reactions of the market under consideration. For instance, for the BEKK GARCH$(1,1)$ model, we get

$$\Sigma_{t+1|t} = C_0^{*\prime} C_0^* + \Gamma_{11}^{*\prime} E(u_t u_t'|y_t, \Omega_{t-1})\Gamma_{11}^* + G_{11}^{*\prime} \Sigma_{t|t-1} G_{11}^*. \qquad (16.6.6)$$

The quantities in (16.6.6) are usually computed using the estimates of the conditional mean equation and the relevant GARCH volatility model. The matrix $E(u_t u_t'|y_t, \Omega_{t-1}) = u_t u_t'$ is replaced by $\widehat{u}_t \widehat{u}_t'$, where the $\widehat{u}_t$ are typically residuals from estimating the conditional mean model. If the corresponding quantities related to an impulse ξ are considered, the impulse is simply added to the $\widehat{u}_t$. Because $\Sigma_{t+h|t}, h = 2, 3, \dots$, is a convenient estimator for $E(u_{t+h} u_{t+h}'|y_t, \Omega_{t-1})$, recursive forecasts of future volatility, conditional on information which is available at time t, are computed as:

$$\Sigma_{t+h|t} = C_0^{*\prime} C_0^* + \Gamma_{11}^{*\prime} \Sigma_{t+h-1|t} \Gamma_{11}^* + G_{11}^{*\prime} \Sigma_{t+h-1|t} G_{11}^*, \quad h = 2, 3, \dots. \qquad (16.6.7)$$

From the conditional covariance matrices, conditional moment profiles are obtained as differences

$$\begin{bmatrix} \phi_{11t,h}(\xi) & \dots & \phi_{1Kt,h}(\xi) \\ \vdots & \ddots & \vdots \\ \phi_{1Kt,h}(\xi) & \dots & \phi_{KKt,h}(\xi) \end{bmatrix} = \Sigma^{\xi}_{t+h|t} - \Sigma_{t+h|t}. \qquad (16.6.8)$$

Although these quantities may be interesting to look at, they depend on t, h, and ξ. Hence, there is a separate impulse response function for each given t and ξ. In empirical work, it will therefore be necessary to summarize the wealth of information in the conditional moment profiles in a meaningful way. In a study of two stock price series, Herwartz & Lütkepohl (2000), for example, considered the following summary statistics:

- Averages over all histories for different impulse vectors ξ, $\bar{\phi}_{ij.,h}(\xi) = T^{-1}\sum_{t=1}^{T}\phi_{ijt,h}(\xi)$.
- Averages over a large range of different impulse vectors ξ_r, $\bar{\phi}_{ijt,h}(\cdot) = R^{-1}\sum_{r=1}^{R}\phi_{ijt,h}(\xi_r)$, for given values of t and h. Here R is the number of impulses considered. The impulse ξ_r may, for instance, be obtained from the estimated model residuals.

Although these summary statistics condense the information in the conditional moment profiles considerably, they are still a rich source of information which can be presented in graphs or further condensed by fitting nonparametric density functions or using other summary statistics (see Herwartz & Lütkepohl (2000)).

Of course, in practice, an additional obstacle is that the actual data generation process is unknown and estimated models are available at best. In that case, the conditional moment profiles or generalized impulse responses will be computed from estimated quantities only. They are therefore also estimates and it would be useful to have measures for their sampling variability. It is not clear how this additional information is computed and presented in the best way in practice. In any case, if only the estimated quantities are available and presented, it is useful to keep in mind these further limitations when the results are interpreted.

It is naturally of interest to better understand what the various models for conditional volatility can tell us about the relations between variables and, hence, about what is actually going on in a particular market or segment of the economy. Therefore it is not surprising that the interpretation of MGARCH models is a field of active research. Some important recent contributions in addition to those noted earlier are Engle & Ng (1993), Lin (1997), and Hafner & Herwartz (1998a).

16.7 Problems and Extensions

There are a number of problems associated with ARCH and GARCH modelling. Some of them have been mentioned in earlier sections of this chapter but may be worth emphasizing again. In addition, there are some problems which we have not addressed so far.

First, due to the highly nonlinear form of the log-likelihood function and the potentially large number of parameters in a multivariate GARCH model which have to satisfy a number of restrictions, computing ML estimates is a difficult task. Therefore it is highly desirable to develop fast and robust optimization algorithms which work well under these particular conditions. A review and comparison of some of the available software was provided by Brooks, Burke & Persand (2003).

Secondly, a sound analysis of conditional heteroskedasticity in a multivariate time series context requires that at least the asymptotic properties of the estimators are known. As we have seen in Section 16.4.1, some asymptotic theory is available for quasi ML estimators of specific MGARCH models. Unfortunately, the required conditions are not satisfactory in all situations. Hence, developing asymptotic theory under more general conditions is desirable.

Third, a toolkit for model specification and model checking is available, as we have seen in Section 16.5. There are some open questions regarding the statistical properties of these tools, however. Moreover, given the wealth of possible model specifications, some more refined tools are desirable that help the analyst to find the best specification for a particular data set and analysis objective and for discriminating between alternative models.

Fourth, although a range of proposals have been made on how to interpret multivariate GARCH models, the available tools leave room for improvements. The nonlinearity of these models makes it more difficult to extract the essential features than in linear models for the conditional mean.

Finally, there are many features in financial and other economic data which are not described well by the GARCH models considered in this chapter. Therefore, a range of other models have been proposed that can capture specific aspects of the distributional properties of financial series in a more satisfactory way. For example, exogenous variables may be included in a multivariate GARCH model (see Engle & Kroner (1995)). Also, as mentioned in Section 16.1, the volatility in a series may have an impact on the conditional mean. To account for this possibility, it may be useful to allow conditional variances to enter the conditional mean function (Engle, Lilien & Robins (1987)). These so-called *ARCH-in-mean (ARCH-M) models* may also be generalized to the multivariate case.

Stochastic volatility models represent another approach to modelling time-varying volatility. In this approach, the conditional covariance matrix depends on an unobserved latent process and not on past observations as in the ARCH model. For instance, in the univariate case, letting $\varepsilon_t \sim \text{i.i.d.}\,\mathcal{N}(0,1)$ and specifying $u_t := \sigma_t \varepsilon_t$, the logarithm of the conditional standard deviation is assumed to be generated as

$$\ln \sigma_t = \varphi \ln \sigma_{t-1} + \eta \kappa_t,$$

where $\kappa_t \sim \text{i.i.d.}\,\mathcal{N}(0,1)$ and φ and η are constant parameters. A survey of multivariate stochastic volatility models was given by Ghysels, Harvey & Renault (1996). It may also be worth noting that, in some sense, random coefficient autoregressive models may be regarded as extensions of multivariate ARCH models (see Wong & Li (1997)).

16.8 Exercises

Problem 16.1
Write down BEKK GARCH models explicitly for the following combinations of N and q in (16.3.7):

$$(N, q) = (1, 1), (2, 1), (1, 2), (2, 2).$$

Problem 16.2
Write down the factor MGARCH model (16.3.9) explicitly for $N = 2$.

Problem 16.3
Write down in detail all elements of $\Sigma_{t|t-1}$ of a factor MGARCH model as proposed by Vrontos et al. (2003) (see Section 16.3.3) for the case of a bivariate series ($K = 2$).

Problem 16.4
Consider the DEM/USD and GBP/USD exchange rate series from
www.jmulti.de → datasets
(File exrate.dat) and perform the following analysis steps:

(a) Eliminate all rows with missing values from the exchange rate data set.
(b) Determine the VAR order by model selection criteria.
(c) Plot the autocorrelations series and the mean-adjusted squared series. Interpret the plots.
(d) Use ARCH-LM and ARCH-portmanteau tests for the mean-adjusted series and interpret the results. Apply the tests to the bivariate and the two univariate series separately and compare the results.
(e) Fit a bivariate BEKK GARCH(1, 1) model to the bivariate series.
(f) Perform model specification tests based on the residuals of the estimated MGARCH model and interpret the results.

(Hint: A similar data set was analyzed by Herwartz (2004).)