14

GARCH Models

14.1 Introduction

As seen in earlier chapters, financial market data often exhibits volatility clustering, where time series show periods of high volatility and periods of low volatility; see, for example, Fig. 14.1. In fact, with economic and financial data, time-varying volatility is more common than constant volatility, and accurate modeling of time-varying volatility is of great importance in financial engineering.

As we saw in Chap. 12, ARMA models are used to model the conditional expectation of a process given the past, but in an ARMA model the conditional variance given the past is constant. What does this mean for, say, modeling stock returns? Suppose we have noticed that recent daily returns have been unusually volatile. We might expect that tomorrow's return is also more variable than usual. However, an ARMA model cannot capture this type of behavior because its conditional variance is constant. So we need better time series models if we want to model the nonconstant volatility. In this chapter we look at GARCH time series models that are becoming widely used in econometrics and finance because they have randomly varying volatility.

ARCH is an acronym meaning Auto-Regressive Conditional Heteroskedasticity. In ARCH models the conditional variance has a structure very similar to the structure of the conditional expectation in an AR model. We first study the first order ARCH(1) model, which is the simplest GARCH model, and analogous to an AR(1) model. Then we look at ARCH(p) models, which are analogous to AR(p) models, and GARCH (Generalized ARCH) models, which model conditional variances much as the conditional expectation is modeled by an ARMA model. Finally, we consider several multivariate GARCH processes.

D. Ruppert, D.S. Matteson, *Statistics and Data Analysis for Financial Engineering*, Springer Texts in Statistics,
DOI 10.1007/978-1-4939-2614-5_14

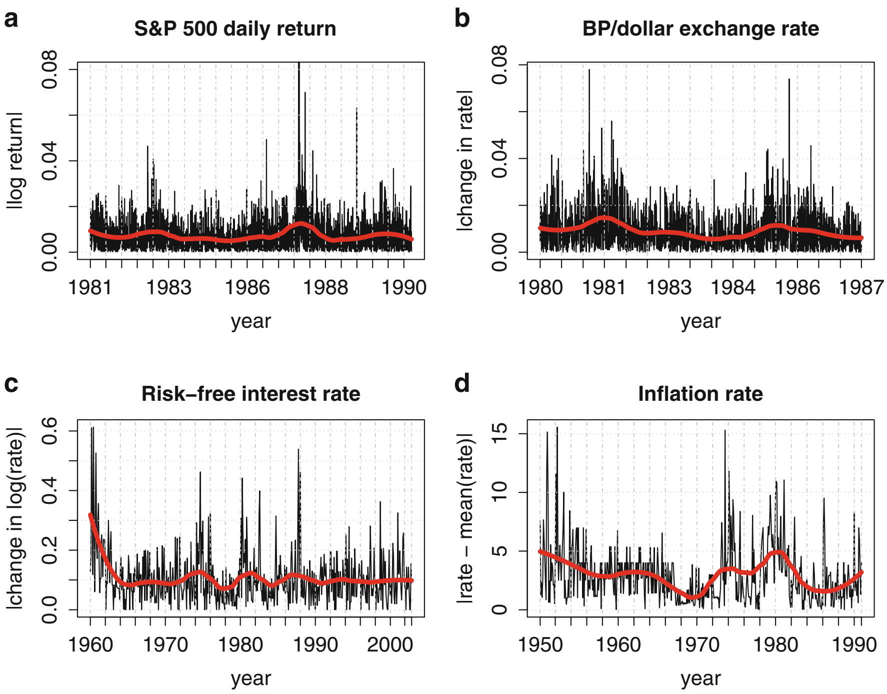

Fig. 14.1. *Examples of financial markets and economic data with time-varying volatility: (a) absolute values of S&P* 500 *log returns; (b) absolute values of changes in the BP/dollar exchange rate; (c) absolute values of changes in the log of the risk-free interest rate; (d) absolute deviations of the inflation rate from its mean. Loess (see Section 21.2) smooths have been added in red.*

14.2 Estimating Conditional Means and Variances

Before looking at GARCH models, we study some general principles about modeling nonconstant conditional variance. Consider regression modeling with a *constant* conditional variance, $\mathrm{Var}(Y_t|\ X_{1,t},\ldots,\ X_{p,t}) = \sigma^2$. Then the general form for the regression of Y_t on $X_{1,t},\ldots,\ X_{p,t}$ is

$$Y_t = f(X_{1,t},\ldots,X_{p,t}) + \epsilon_t, \tag{14.1}$$

where ϵ_t is independent of $X_{1,t},\ldots,X_{p,t}$ and has expectation equal to 0 and a constant conditional variance σ^2_ϵ. The function $f(\cdot)$ is the conditional expectation of Y_t given $X_{1,t},\ldots,X_{p,t}$. Moreover, the conditional variance of Y_t is σ^2_ϵ.

Equation (14.1) can be modified to allow conditional heteroskedasticity. Let $\sigma^2(X_{1,t},\ldots,X_{p,t})$ be the conditional variance of Y_t given $X_{1,t},\ldots,\ X_{p,t}$. Then the model

$$Y_t = f(X_{1,t},\ldots,X_{p,t}) + \epsilon_t\,\sigma(X_{1,t},\ldots,X_{p,t}), \tag{14.2}$$

where ϵ_t has conditional (given $X_{1,t}, \ldots, X_{p,t}$) mean equal to 0 and conditional variance equal to 1, gives the correct conditional mean and variance of Y_t.

The function $\sigma(X_{1,t}, \ldots, X_{p,t})$ should be nonnegative since it is a standard deviation. If the function $\sigma(\cdot)$ is linear, then its coefficients must be constrained to ensure nonnegativity. Such constraints are cumbersome to implement, so nonlinear nonnegative functions are usually used instead. Models for conditional variances are often called *variance function models*. The GARCH models of this chapter are an important class of variance function models.

14.3 ARCH(1) Processes

Suppose for now that $\epsilon_1, \epsilon_2, \ldots$ is Gaussian white noise with unit variance. Later we will allow the noise to be i.i.d. white noise with a possibly nonnormal distribution, such as, a standardized t-distribution. Then

$$E(\epsilon_t | \epsilon_{t-1}, \ldots) = 0,$$

and

$$\mathrm{Var}(\epsilon_t | \epsilon_{t-1}, \ldots) = 1. \tag{14.3}$$

Property (14.3) is called *conditional homoskedasticity*.

The process a_t is an ARCH(1) process under the model

$$a_t = \epsilon_t \sqrt{\omega + \alpha a_{t-1}^2}, \tag{14.4}$$

which is a special case of (14.2) with f equal to 0 and σ equal to $\sqrt{\omega + \alpha a_{t-1}^2}$. We require that $\omega > 0$ and $\alpha \geq 0$ so that $\omega + \alpha a_{t-1}^2 > 0$ for all t. It is also required that $\alpha < 1$ in order for $\{a_t\}$ to be stationary with a finite variance. Equation (14.4) can be written as

$$a_t^2 = \epsilon_t^2 \, (\omega + \alpha a_{t-1}^2),$$

which is similar to an AR(1), but in a_t^2, not a_t, and with multiplicative noise with a mean of 1 rather than additive noise with a mean of 0. In fact, the ARCH(1) model induces an ACF for a_t^2 that is the same as an AR(1)'s ACF, as we will see from the calculations below.

Define

$$\sigma_t^2 = \mathrm{Var}(a_t | a_{t-1}, \ldots)$$

to be the conditional variance of a_t given past values. Since ϵ_t is independent of a_{t-1} and $E(\epsilon_t^2) = \mathrm{Var}(\epsilon_t) = 1$, we have

$$E(a_t | a_{t-1}, \ldots) = 0, \tag{14.5}$$

and

$$\begin{aligned}\sigma_t^2 &= E\{(\omega + \alpha a_{t-1}^2)\,\epsilon_t^2 | a_{t-1}, a_{t-2}, \ldots\} \\ &= (\omega + \alpha a_{t-1}^2) E\{\epsilon_t^2 | a_{t-1}, a_{t-2}, \ldots\} \\ &= \omega + \alpha a_{t-1}^2. \end{aligned} \tag{14.6}$$

Equation (14.6) is crucial to understanding how GARCH processes work. If a_{t-1} has an unusually large absolute value, then σ_t is larger than usual and so a_t is also expected to have an unusually large magnitude. This volatility propagates since when a_t has a large magnitude that makes σ_{t+1}^2 large, then a_{t+1} tends to be large in magnitude, and so on. Similarly, if a_{t-1}^2 is unusually small, then σ_t^2 is small, and a_t^2 is also expected to be small, and so forth. Because of this behavior, unusual volatility in a_t tends to persist, though not forever. The conditional variance tends to revert to the unconditional variance provided that $\alpha < 1$, so that the process is stationary with a finite variance.

The unconditional, that is, marginal, variance of a_t denoted by $\gamma_a(0)$ is obtained by taking expectations in (14.6), which gives us

$$\gamma_a(0) = \omega + \alpha\gamma_a(0)$$

for a stationary model. This equation has a positive solution if $\alpha < 1$:

$$\gamma_a(0) = \omega/(1-\alpha).$$

If $\alpha = 1$, then $\gamma_a(0)$ is infinite, but a_t is stationary nonetheless and is called an integrated GARCH (I-GARCH) model.

Straightforward calculations using (14.5) show that the ACF of a_t is

$$\rho_a(h) = 0 \text{ if } \quad h \neq 0.$$

In fact, any process in which the conditional expectation of the present observation given the past is constant is an uncorrelated process.

In introductory statistics courses, it is often mentioned that independence implies zero correlation but not vice versa. A process, such as a GARCH process, in which the conditional mean is constant but the conditional variance is nonconstant is an example of an uncorrelated but dependent process. The dependence of the conditional variance on the past causes the process to be dependent. The independence of the conditional mean on the past is the reason that the process is uncorrelated.

Although a_t is an uncorrelated process, the process a_t^2 has a more interesting ACF. If $\alpha < 1$, then

$$\rho_{a^2}(h) = \alpha^{|h|}, \quad \forall\, h.$$

If $\alpha \geq 1$, then a_t^2 either is nonstationary or has an infinite variance, so it does not have an ACF. This geometric decay in the ACF of a_t^2 for an ARCH(1) process is analogous to the geometric decay in the ACF of an AR(1) process. To complete the analogy, define $\eta_t = a_t^2 - \sigma_t^2$, and note that $\{\eta_t\}$ is a mean zero weak white noise process, but not an i.i.d. white noise process. Adding η_t to both sides of (14.6) and simplifying we have

$$\sigma_t^2 + \eta_t = a_t^2 = \omega + \alpha a_{t-1}^2 + \eta_t, \tag{14.7}$$

which is a direct representation of $\{a_t^2\}$ as an AR(1) process.

14.4 The AR(1)+ARCH(1) Model

As we have seen, an AR(1) process has a nonconstant conditional mean but a constant conditional variance, while an ARCH(1) process is just the opposite. If both the conditional mean and variance of the data depend on the past, then we can combine the two models. In fact, we can combine any ARMA model with any of the GARCH models in Sect. 14.6. In this section we combine an AR(1) model with an ARCH(1) model.

Let a_t be an ARCH(1) process so that $a_t = \epsilon_t \sqrt{\omega + \alpha a_{t-1}^2}$, where ϵ_t is i.i.d. $N(0, 1)$, and suppose that

$$y_t - \mu = \phi(y_{t-1} - \mu) + a_t.$$

The process y_t is an AR(1) process, except that the noise term (a_t) is not i.i.d. white noise, but rather an ARCH(1) process which is only weak white noise.

Because a_t is an uncorrelated process, it has the same ACF as independent white noise, and therefore, y_t has the same ACF as an AR(1) process with independent white noise

$$\rho_y(h) = \phi^{|h|} \quad \forall\, h,$$

in the stationary case. Moreover, a_t^2 has the ARCH(1) ACF:

$$\rho_{a^2}(h) = \alpha^{|h|} \quad \forall\, h.$$

The ACF of y_t^2 also decays with $|h|$ at a geometric rate in the stationary case, provided some additional assumptions hold, however, the exact expressions are more complicated (see Palma and Zevallos, 2004). We need to assume that both $|\phi| < 1$ and $\alpha < 1$ in order for y_t to be stationary with a finite variance. Of course, $\omega > 0$ and $\alpha \geq 0$ are also assumed for positiveness of the conditional variance process σ_t^2. The process y_t is such that its conditional mean and variance, given the past, are both nonconstant, so a wide variety of time series can be modeled.

Example 14.1. A simulated ARCH(1) process and AR(1)+ARCH(1) process

A simulated ARCH(1) process is shown in Fig. 14.2. Panel (a) shows the i.i.d. white noise process ϵ_t, (b) shows $\sigma_t = \sqrt{1 + 0.55a_{t-1}^2}$, the conditional standard deviation process, and (c) shows $a_t = \sigma_t \epsilon_t$, the ARCH(1) process. As discussed in the previous section, an ARCH(1) process can be used as the noise term of an AR(1) process. This process is shown in panel (d). The AR(1) parameters are $\mu = 0.1$ and $\phi = 0.8$. The unconditional variance of a_t is $\gamma_a(0) = 1/(1 - 0.55) = 2.22$, so the unconditional standard deviation is $\sqrt{2.22} = 1.49$. Panels (e)–(h) are sample ACF plots of the ARCH and AR+ARCH processes and squared processes. Notice that for the ARCH series, the process is uncorrelated but the squared series has autocorrelation. Also notice that for the AR(1)+ARCH(1) series the ACFs of the process and the squared process, panels (g) and (h), both show autocorrelation. While the true ACFs have an exact geometric decay, this is only approximately true for the sample ACFs in panels (f)–(h); similarly, negative values are not present in the true ACFs, but the sample ACF has sampling error and may result in negative values. The processes were all started at 0 and simulated for 10,200 observations. The first 10,000 observations were treated as a burn-in period and discarded. □

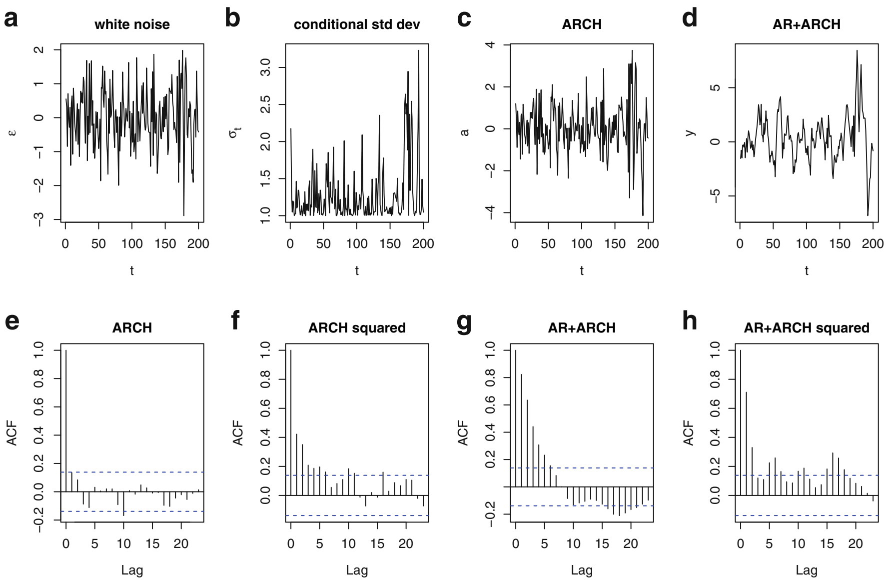

Fig. 14.2. *Simulation of 200 observations from an ARCH(1) process and an AR(1)+ARCH(1) process. The parameters are $\omega = 1$, $\alpha = 0.55$, $\mu = 0.1$, and $\phi = 0.8$. Sample ACF plots of the ARCH and AR+ARCH processes and squared processes are shown in the bottom row.*

14.5 ARCH(p) Models

As before, let ϵ_t be Gaussian white noise with unit variance. Then a_t is an ARCH(p) process if

$$a_t = \sigma_t \epsilon_t,$$

where

$$\sigma_t = \sqrt{\omega + \sum_{i=1}^{p} \alpha_i a_{t-i}^2}$$

is the conditional standard deviation of a_t given the past values $a_{t-1}, a_{t-2}, \ldots$ of this process. Like an ARCH(1) process, an ARCH(p) process is uncorrelated and has a constant mean (both conditional and unconditional) and a constant unconditional variance, but its conditional variance is nonconstant. In fact, the ACF of a_t^2 has the same structure as the ACF of an AR(p) process; see Sect. 14.9.

14.6 ARIMA(p_M, d, q_M)+GARCH(p_V, q_V) Models

A deficiency of ARCH(p) models is that the conditional standard deviation process has high-frequency oscillations with high volatility coming in short bursts. This behavior can be seen in Fig. 14.2b. GARCH models permit a wider range of behavior, in particular, more persistent volatility. The GARCH(p, q) model is

$$a_t = \sigma_t \epsilon_t,$$

in which

$$\sigma_t = \sqrt{\omega + \sum_{i=1}^{p} \alpha_i a_{t-i}^2 + \sum_{j=1}^{q} \beta_j \sigma_{t-j}^2}\,. \tag{14.8}$$

Because past values of the σ_t process are fed back into the present value (with nonnegative coefficients β_j), the conditional standard deviation can exhibit more persistent periods of high or low volatility than seen in an ARCH process. In the stationary case, the process a_t is uncorrelated with a constant unconditional mean and variance and a_t^2 has an ACF like an ARMA process (see Sect. 14.9). GARCH models include ARCH models as a special case, and we use the term "GARCH" to refer to both ARCH and GARCH models.

A very general time series model lets a_t be GARCH(p_V, q_V) and uses a_t as the noise term in an ARIMA(p_M, d, q_M) model. The subscripts on p and q distinguish between the conditional variance (V) or GARCH parameters and the conditional mean (M) or ARIMA parameters. We will call such a process an ARIMA(p_M, d, q_M)+GARCH(p_V, q_V) model.

Figure 14.3 is a simulation of 500 observations from a GARCH(1,1) process and from a AR(1)+GARCH(1,1) process. The GARCH parameters are $\omega = 1$, $\alpha = 0.08$, and $\beta = 0.9$. The large value of β causes σ_t to be highly correlated with σ_{t-1} and gives the conditional standard deviation process a relatively long-term persistence, at least compared to its behavior under an ARCH model. In particular, notice that the conditional standard deviation is less "bursty" than for the ARCH(1) process in Fig. 14.2.

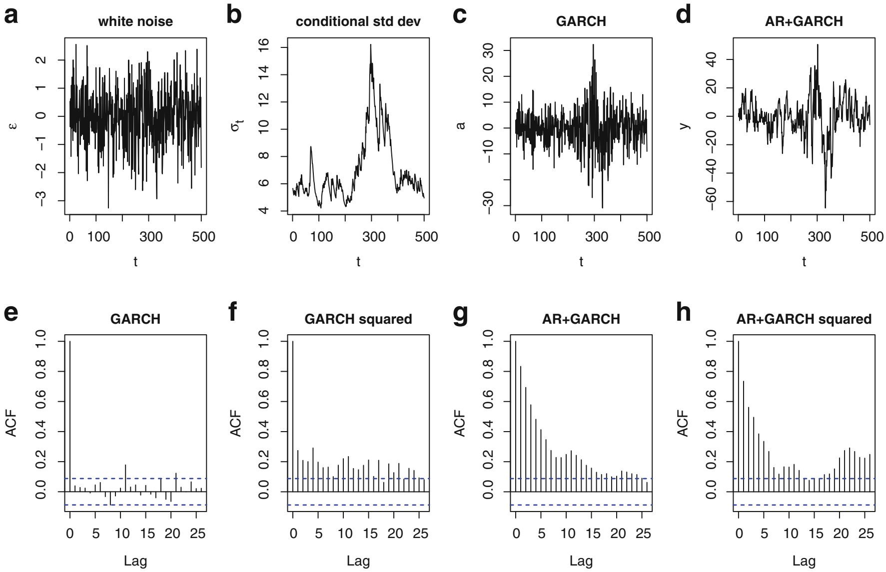

Fig. 14.3. *Simulation of GARCH(1,1) and AR(1)+GARCH(1,1) processes. The parameters are* $\omega = 1$*,* $\alpha = 0.08$*,* $\beta = 0.9$*, and* $\phi = 0.8$*.*

14.6.1 Residuals for ARIMA(p_M, d, q_M)+GARCH(p_V, q_V) Models

When one fits an ARIMA(p_M, d, q_M)+GARCH(p_V, q_V) model to a time series Y_t, there are two types of residuals. The ordinary residual, denoted $\widehat{a}_t$, is the difference between Y_t and its conditional expectation. As the notation implies, $\widehat{a}_t$ estimates a_t. A standardized residual, denoted $\widehat{\epsilon}_t$, is an ordinary residual $\widehat{a}_t$ divided by its estimated conditional standard deviation $\widehat{\sigma}_t$. A standardized residual estimates ϵ_t. The standardized residuals should be used for model checking. If the model fits well, then neither $\widehat{\epsilon}_t$ nor $\widehat{\epsilon}_t^2$ should exhibit serial correlation. Moreover, if ϵ_t has been assumed to have a normal distribution, then this assumption can be checked by a normal plot of the standardized residuals $\widehat{\epsilon}_t$. The $\widehat{a}_t$ are the residuals of the ARIMA process and are used when forecasting via the methods in Sect. 12.12.

14.7 GARCH Processes Have Heavy Tails

Researchers have long noticed that stock returns have "heavy-tailed" or "outlier-prone" probability distributions, and we have seen this ourselves in earlier chapters. One reason for outliers may be that the conditional variance is not constant, and the outliers occur when the variance is large, as in the normal mixture example of Sect. 5.5. In fact, GARCH processes exhibit heavy tails even if $\{\epsilon_t\}$ is Gaussian. Therefore, when we use GARCH models, we can model both the conditional heteroskedasticity and the heavy-tailed distributions of financial market data. Nonetheless, many financial time series have tails that are heavier than implied by a GARCH process with Gaussian $\{\epsilon_t\}$. To handle such data, one can assume that, instead of being Gaussian white noise, $\{\epsilon_t\}$ is an i.i.d. white noise process with a heavy-tailed distribution.

14.8 Fitting ARMA+GARCH Models

Example 14.2. AR(1)+GARCH(1,1) model fit to daily BMW stock log returns

This example uses the daily BMW stock log returns. The `ugarchfit()` function from `R`'s `rugarch` package is used to fit an AR(1)+GARCH(1,1) model to this series. Although `ugarchfit()` allows the white noise to have a nonGaussian distribution, we begin this example using Gaussian white noise (the default). First the model is specified using the `ugarchspec()` function; for an AR(1)+GARCH(1,1) model we specify `armaOrder=c(1,0)` and `garchOrder=c(1,1)`. The commands and abbreviated output are below.

```
library(rugarch)
data(bmw, package="evir")
arma.garch.norm = ugarchspec(mean.model=list(armaOrder=c(1,0)),
                              variance.model=list(garchOrder=c(1,1)))
bmw.garch.norm = ugarchfit(data=bmw, spec=arma.garch.norm)
show(bmw.garch.norm)
```

```
GARCH Model : sGARCH(1,1)
Mean Model : ARFIMA(1,0,0)
Distribution : norm

Optimal Parameters
------------------------------------
        Estimate  Std. Error  t value Pr(>|t|)
mu      0.000453    0.000175   2.5938 0.009493
ar1     0.098135    0.014261   6.8813 0.000000
omega   0.000009    0.000000  23.0613 0.000000
alpha1  0.099399    0.005593  17.7730 0.000000
beta1   0.863672    0.006283 137.4591 0.000000
```

```
LogLikelihood : 17752

Information Criteria
------------------------------------
Akaike       -5.7751
Bayes        -5.7696
Shibata      -5.7751
Hannan-Quinn -5.7732
```

In the output, $\widehat{\phi}_1$ is denoted by `ar1`, the estimated mean $\widehat{\mu}$ is `mean`, and $\widehat{\omega}$ is called `omega`. Note that $\widehat{\phi}_1 = 0.0981$ and is statistically significant, implying that there is a small amount of positive autocorrelation. Both α_1 and β_1 are highly significant and $\widehat{\beta}_1 = 0.8636$, which implies rather persistent volatility clustering. There are two additional information criteria reported, Shibata's information criterion and Hannan–Quinn information criterion (HQIC). These are less widely used than AIC and BIC and will not be discussed here.

In the output from `ugarchfit()`, the AIC and BIC values have been normalized by dividing by n, so these values should be multiplied by $n = 6146$ to have their usual values. In particular, AIC and BIC will not be so close to each other after multiplication by 6146. The daily BMW stock log return series Y_t, with two estimated conditional standard deviations superimposed, and the estimated conditional standard deviation series $\widehat{\sigma}_t$ (vs. the absolute value of the log return series $|Y_t|$) are shown in the top row of Fig. 14.4.

The output also includes the following tests applied to the standardized and squared standardized residuals.

```
Weighted Ljung-Box Test on Standardized Residuals
------------------------------------
                        statistic p-value
Lag[1]                     0.7786  0.3776
Lag[2*(p+q)+(p+q)-1][2]    0.9158  0.7892
Lag[4*(p+q)+(p+q)-1][5]    3.3270  0.3536
d.o.f=1
H0 : No serial correlation
Weighted Ljung-Box Test on Standardized Squared Residuals
------------------------------------
                        statistic p-value
Lag[1]                      0.277  0.5987
Lag[2*(p+q)+(p+q)-1][5]     1.026  0.8537
Lag[4*(p+q)+(p+q)-1][9]     1.721  0.9356
d.o.f=2
Weighted ARCH LM Tests
------------------------------------
            Statistic Shape Scale P-Value
ARCH Lag[3]    0.1922 0.500 2.000  0.6611
ARCH Lag[5]    1.1094 1.440 1.667  0.7008
ARCH Lag[7]    1.2290 2.315 1.543  0.8737
Adjusted Pearson Goodness-of-Fit Test:
```

```
------------------------------------
  group statistic p-value(g-1)
1    20     493.1    1.563e-92
2    30     513.4    5.068e-90
3    40     559.3    2.545e-93
4    50     585.6    5.446e-93
```

Weighted versions of the Ljung-Box (and ARCH-LM) test statistics[1] and their approximate p-values all indicate that the estimated model for the conditional mean and variance are adequate for removing serial correlation from the series and squared series, respectively. The sample ACF of the standardized residuals $\widehat{\epsilon}_t$, and the squared standardized residuals $\widehat{\epsilon}_t^2$ are shown in the middle row of Fig. 14.4. The Goodness-of-Fit tests[2] compare the empirical distribution of the standardized residuals with the theoretical ones from the specified density, which is Gaussian by default. The small p-values strongly reject the null hypothesis that the white noise standardized innovation process $\{\epsilon_t\}$ is Gaussian. Empirical density estimates and a normal quantile plot of the standardized residuals $\widehat{\epsilon}_t$ are shown in the bottom row of Fig. 14.4.

Figure 14.5 shows a t-plot with 4 df for the standardized residuals $\widehat{\epsilon}_t$. Unlike the normal quantile plot in the last panel of Fig. 14.4, this plot is nearly a straight line except for four outliers in the left tail. The sample size is 6146, so the outliers are a very small fraction of the data. Thus, it seems like a t-distribution would be suitable for the innovation process ϵ_t. A t-distribution was fit to the standardized residuals by maximum likelihood using the `fitdistr()` function from the Rpackage `MASS`.

```
library(MASS)
e = residuals(bmw.garch.norm, standardize=TRUE)
fitdistr(e,"t")
```

```
        m          s          df
   -0.0243     0.7269     4.1096
  ( 0.0109)  ( 0.0121)  ( 0.2359)
```

The MLE of the degrees-of-freedom parameter was 4.1. This confirms the good fit by this distribution seen in Fig. 14.5. The AR(1)+GARCH(1,1) model was refit assuming t-distributed errors, so `distribution.model = "std"` in `ugarchspec()`. The commands and abbreviated results are below.

```
arma.garch.t = ugarchspec(mean.model=list(armaOrder=c(1,0)),
                          variance.model=list(garchOrder=c(1,1)),
                          distribution.model = "std")
bmw.garch.t = ugarchfit(data=bmw,spec=arma.garch.t)
show(bmw.garch.t)
```

[1] Weighted Ljung-Box and ARCH-LM statistics of Fisher and Gallagher (2012) are provided by the `ugarchfit()` function to better account for the distribution of the statistics when applied to residuals from a fitted model; their use and interpretation remains unchanged.

[2] These Chi-squared tests are based on the tests of Palm (1996); `group` indicates the number of bins used in the implementation.

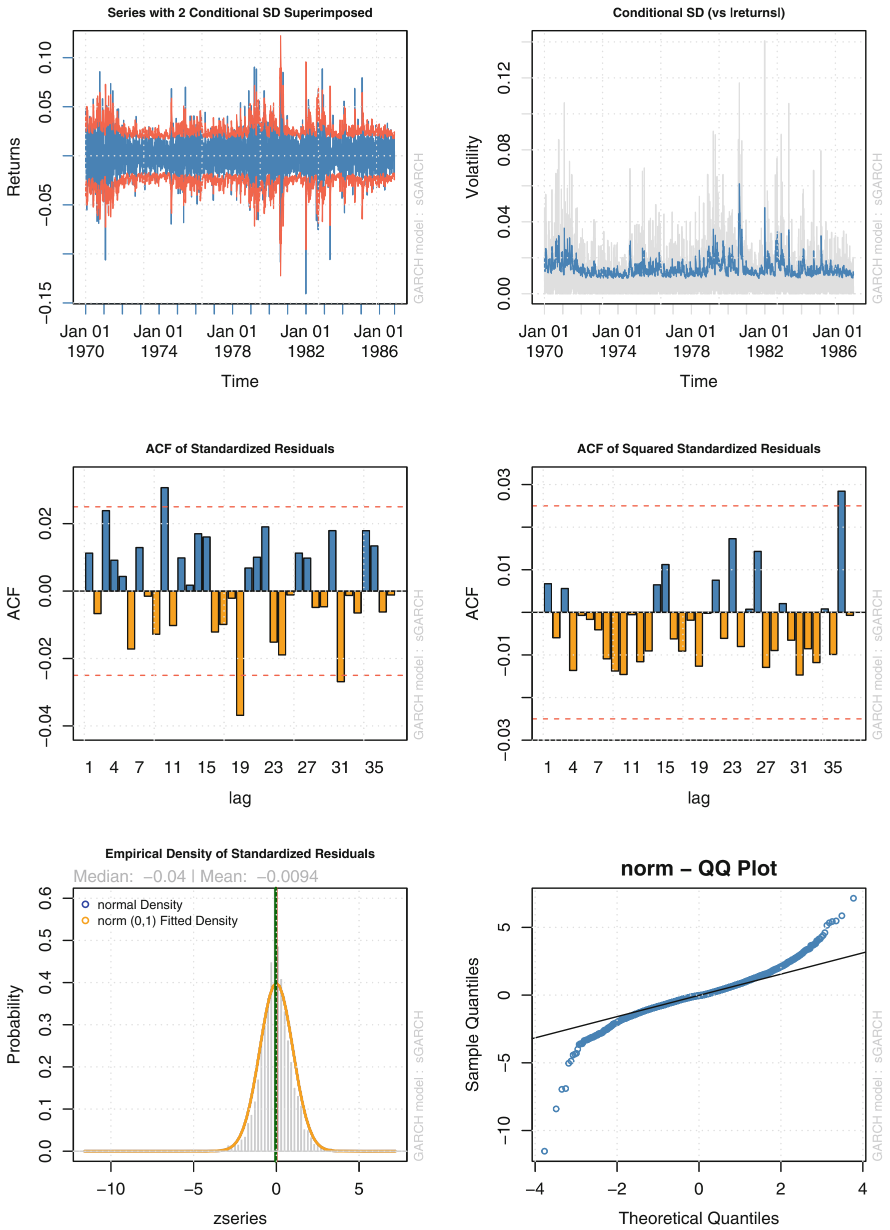

Fig. 14.4. *The daily BMW stock log return series Y_t, with two estimated conditional standard deviations superimposed; the estimated conditional standard deviation $\widehat{\sigma}_t$ series (vs. the absolute value of the log return series $|Y_t|$); the sample ACF of the standardized residuals $\widehat{\epsilon}_t$ and the squared standardized residuals $\widehat{\epsilon}_t^2$; empirical density estimates of the standardized residuals $\widehat{\epsilon}_t$; and a normal quantile plot of the standardized residuals $\widehat{\epsilon}_t$.*

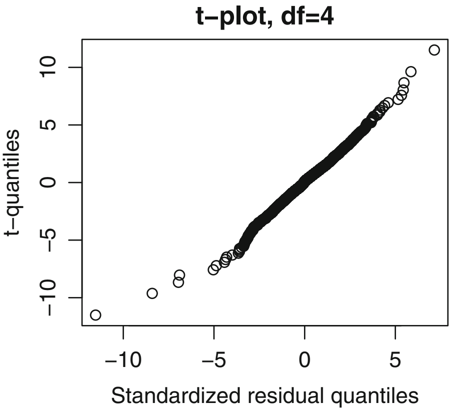

Fig. 14.5. *A t-plot with 4 df for the standardized residuals* $\widehat{\epsilon}_t$ *from an AR(1)+GARCH(1,1) model fit to daily BMW stock log return; the reference lines go through the first and third quartiles.*

```
GARCH Model : sGARCH(1,1)
Mean Model : ARFIMA(1,0,0)
Distribution : std

Optimal Parameters
------------------------------------
        Estimate  Std. Error  t value Pr(>|t|)
mu      0.000135    0.000144  0.93978 0.347333
ar1     0.063911    0.012521  5.10436 0.000000
omega   0.000006    0.000003  1.69915 0.089291
alpha1  0.090592    0.012479  7.25936 0.000000
beta1   0.889887    0.014636 60.80228 0.000000
shape   4.070078    0.301306 13.50813 0.000000

LogLikelihood : 18152

Information Criteria
------------------------------------
Akaike       -5.9048
Bayes        -5.8983
Shibata      -5.9048
Hannan-Quinn -5.9026

Weighted Ljung-Box Test on Standardized Residuals
------------------------------------
                        statistic   p-value
Lag[1]                      9.640 1.904e-03
Lag[2*(p+q)+(p+q)-1][2]     9.653 3.367e-09
Lag[4*(p+q)+(p+q)-1][5]    11.983 1.455e-04
```

```
d.o.f=1
H0 : No serial correlation

Weighted Ljung-Box Test on Standardized Squared Residuals
------------------------------------
                          statistic p-value
Lag[1]                       0.5641  0.4526
Lag[2*(p+q)+(p+q)-1][5]      1.2964  0.7898
Lag[4*(p+q)+(p+q)-1][9]      2.0148  0.9032
d.o.f=2

Adjusted Pearson Goodness-of-Fit Test:
------------------------------------
  group statistic p-value(g-1)
1    20     229.0    5.460e-38
2    30     279.6    8.428e-43
3    40     313.8    1.230e-44
4    50     374.6    1.037e-51
```

The weighted Ljung–Box tests for the residuals have small p-values. These are due to small autocorrelations that should not be of practical importance. The sample size here is 6146 so, not surprisingly, small autocorrelations are statistically significant. The goodness-of-fit test statistics are much smaller but still significant; the large sample size again makes rejection likely even when the discrepancies are negligible from a practical standpoint. However, both AIC and BIC decreased substantially, and the refit model with a t conditional distribution offers an improvement over the original fit with a Gaussian conditional distribution. □

14.9 GARCH Models as ARMA Models

The similarities seen in this chapter between GARCH and ARMA models are not a coincidence. If a_t is a GARCH process, then a_t^2 is an ARMA process, but with weak white noise, not i.i.d. white noise. To show this, we will start with the GARCH(1,1) model, where $a_t = \sigma_t \epsilon_t$. Here ϵ_t is i.i.d. white noise and

$$E(a_t^2|\mathcal{F}_{t-1}) = \sigma_t^2 = \omega + \alpha a_{t-1}^2 + \beta \sigma_{t-1}^2, \tag{14.9}$$

where $\mathcal{F}_{t-1}$ is the information set at time $t-1$. Define $\eta_t = a_t^2 - \sigma_t^2$. Since $E(\eta_t|\mathcal{F}_{t-1}) = E(a_t^2|\mathcal{F}_{t-1}) - \sigma_t^2 = 0$ by (A.33), η_t is an uncorrelated process, that is, a weak white noise process. The conditional heteroskedasticity of a_t is inherited by η_t, so η_t is not i.i.d. white noise.

Simple algebra shows that

$$\sigma_t^2 = \omega + (\alpha + \beta) a_{t-1}^2 - \beta \eta_{t-1} \tag{14.10}$$

and therefore

$$a_t^2 = \sigma_t^2 + \eta_t = \omega + (\alpha + \beta)a_{t-1}^2 - \beta\eta_{t-1} + \eta_t. \tag{14.11}$$

Assume that $\alpha + \beta < 1$. If $\upsilon = \omega/\{1 - (\alpha + \beta)\}$, then

$$a_t^2 - \upsilon = (\alpha + \beta)(a_{t-1}^2 - \upsilon) + \beta\eta_{t-1} + \eta_t. \tag{14.12}$$

From (14.12) one sees that a_t^2 is an ARMA(1,1). Using the notation of (12.25), the mean is $\mu = \upsilon$, the AR(1) coefficient is $\phi = \alpha+\beta$ and the MA(1) coefficient is $\theta = -\beta$.

For the general case, assume that σ_t follows (14.8) such that

$$\sigma_t^2 = \omega + \sum_{i=1}^{p} \alpha_i a_{t-i}^2 + \sum_{j=1}^{q} \beta_j \sigma_{t-j}^2 \,. \tag{14.13}$$

To simplify notation, if $q > p$, then define $\alpha_i = 0$ for $i = p + 1, \ldots, q$. Similarly, if $p > q$, then define $\beta_j = 0$ for $j = q + 1, \ldots, p$. Define $\upsilon = \omega/\{1 - \sum_{i=1}^{\max(p,q)}(\alpha_i + \beta_i)\}$. Straightforward algebra similar to the GARCH(1,1) case shows that

$$a_t^2 - \upsilon = \sum_{i=1}^{\max(p,q)} (\alpha_i + \beta_i)(a_{t-i}^2 - \upsilon) - \sum_{j=1}^{q} \beta_j \eta_{t-j} + \eta_t, \tag{14.14}$$

so that a_t^2 is an ARMA(max$(p, q), q)$ process with mean $\mu = \upsilon$, AR coefficients $\phi_i = \alpha_i + \beta_i$ and MA coefficients $\theta_j = -\beta_j$. As a byproduct of these calculations, we obtain a necessary condition for a_t to be stationary:

$$\sum_{i=1}^{\max(p,q)} (\alpha_i + \beta_i) < 1. \tag{14.15}$$

14.10 GARCH(1,1) Processes

The GARCH(1,1) is the most widely used GARCH process, so it is worthwhile to study it in some detail. If a_t is GARCH(1,1), then as we have just seen, a_t^2 is ARMA(1,1). Therefore, the ACF of a_t^2 can be obtained from formulas (12.31) and (12.32). After some algebra, one finds that

$$\rho_{a^2}(1) = \frac{\alpha(1 - \alpha\beta - \beta^2)}{1 - 2\alpha\beta - \beta^2} \tag{14.16}$$

and

$$\rho_{a^2}(h) = (\alpha + \beta)^{h-1} \rho_{a^2}(1), \quad h \geq 2. \tag{14.17}$$

These formulas also hold in an AR(1)+GARCH(1,1) model, and the ACF of y_t^2 also decays with $h \geq 2$ at a geometric rate in the stationary case, provided some additional assumptions hold, however, the exact expressions are more complicated (see Palma and Zevallos, 2004).

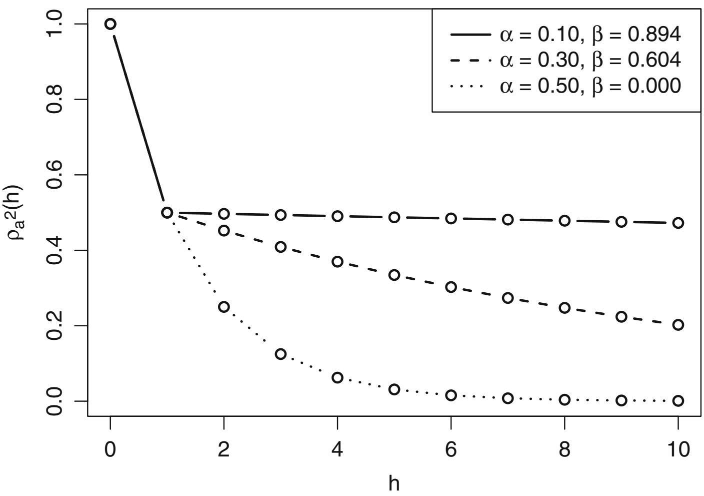

Fig. 14.6. *ACFs of three GARCH(1,1) processes with* $\rho_{a^2}(1) = 0.5$.

By (14.16), there are infinitely many values of (α, β) with the same value of $\rho_{a^2}(1)$. By (14.17), a higher value of $\alpha + \beta$ means a slower decay of $\rho_{a^2}(\cdot)$ after the first lag. This behavior is illustrated in Fig. 14.6, which contains the ACF of a_t^2 for three GARCH(1,1) processes with a lag-1 autocorrelation of 0.5. The solid curve has the highest value of $\alpha + \beta$ and the ACF decays very slowly. The dotted curve is a pure ARCH(1) process and has the most rapid decay.

In Example 14.2, an AR(1)+GARCH(1,1) model was fit to the BMW daily log returns. The GARCH parameters were estimated to be $\widehat{\alpha} = 0.10$ and $\widehat{\beta} = 0.86$. By (14.16) the $\widehat{\rho}_{a^2}(1) = 0.197$ for this process and the high value of $\widehat{\beta}$ suggests slow decay. The sample ACF of the squared residuals [from an AR(1) model] is plotted in Fig. 14.7. In that figure, we see the lag-1 autocorrelation is slightly below 0.2 and after one lag the ACF decays slowly, exactly as expected.

The capability of the GARCH(1,1) model to fit the lag-1 autocorrelation and the subsequent rate of decay separately is important in practice. It appears to be the main reason that the GARCH(1,1) model fits so many financial time series.

14.11 APARCH Models

In some financial time series, large negative returns appear to increase volatility more than do positive returns of the same magnitude. This is called the

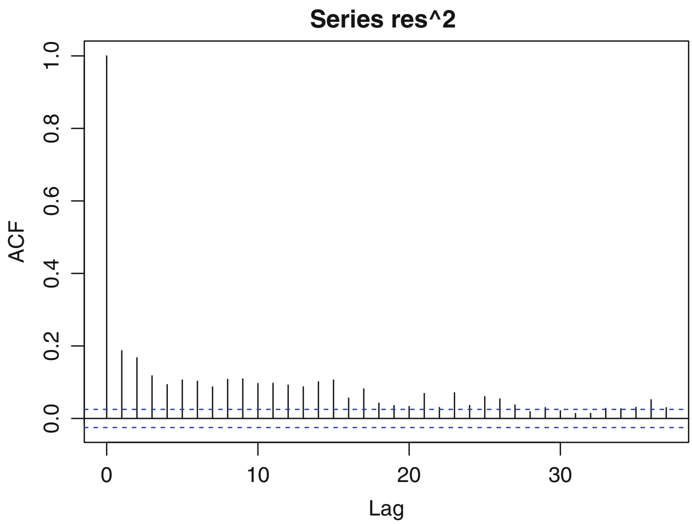

Fig. 14.7. *ACF of the squared residuals from an AR(1) fit to the BMW log returns.*

leverage effect. Standard GARCH models, that is, the models given by (14.8), cannot model the leverage effect because they model σ_t as a function of past values of a_t^2—whether the past values of a_t are positive or negative is not taken into account. The problem here is that the square function x^2 is symmetric in x. The solution is to replace the square function with a flexible class of nonnegative functions that include asymmetric functions. The APARCH (asymmetric power ARCH) models do this. They also offer more flexibility than GARCH models by modeling σ_t^δ, where $\delta > 0$ is another parameter.

The APARCH(p, q) model for the conditional standard deviation is

$$\sigma_t^\delta = \omega + \sum_{i=1}^{p} \alpha_i(|a_{t-i}| - \gamma_i a_{t-i})^\delta + \sum_{j=1}^{q} \beta_j \sigma_{t-j}^\delta, \tag{14.18}$$

where $\delta > 0$ and $-1 < \gamma_i < 1$, $i = 1, \ldots, p$. Note that $\delta = 2$ and $\gamma = \cdots = \gamma_p = 0$ give a standard GARCH model.

The effect of a_{t-i} upon σ_t is through the function g_{γ_i}, where $g_\gamma(x) = |x| - \gamma x$. Figure 14.8 shows $g_\gamma(x)$ for several values of γ. When $\gamma > 0$, $g_\gamma(-x) > g_\gamma(x)$ for any $x > 0$, so there is a leverage effect. If $\gamma < 0$, then there is a leverage effect in the opposite direction to what is expected—positive past values of a_t increase volatility more than negative past values of the same magnitude.

Example 14.3. AR(1)+APARCH(1,1) fit to daily BMW stock log returns

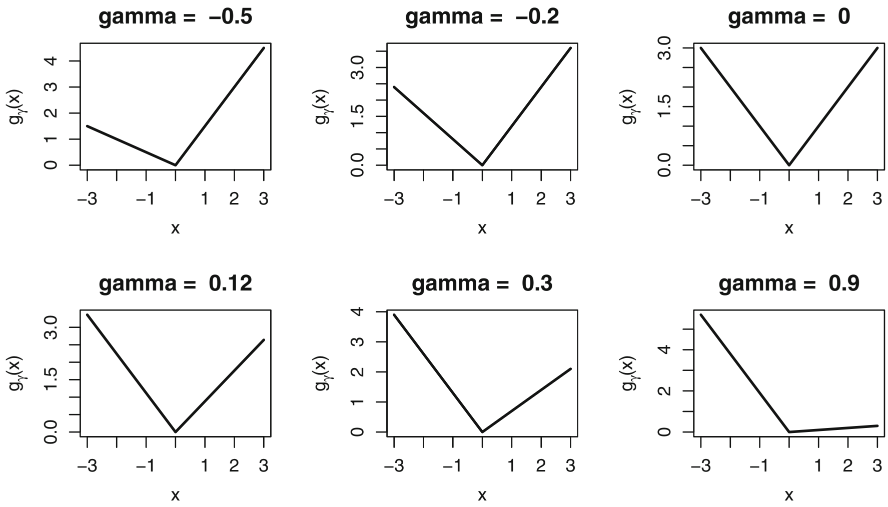

Fig. 14.8. *Plots of $g_\gamma(x)$ for various values of γ.*

In this example, an AR(1)+APARCH(1,1) model with t-distributed errors is fit to the BMW log returns. The commands and abbreviated output from `ugarchfit()` is below. The estimate of δ is 1.48 with a standard error of 0.14, so there is strong evidence that δ is not 2, the value under a standard GARCH model. Also, $\widehat{\gamma}_1$ is 0.12 with a standard error of 0.045, so there is a statistically significant leverage effect, since we reject the null hypothesis that $\gamma_1 = 0$. However, the leverage effect is small, as can be seen in the plot in Fig. 14.8 with $\gamma = 0.12$. The leverage might not be of practical importance.

```
arma.aparch.t = ugarchspec(mean.model=list(armaOrder=c(1,0)),
                variance.model=list(model="apARCH",
                                    garchOrder=c(1,1)),
                           distribution.model = "std")
bmw.aparch.t = ugarchfit(data=bmw, spec=arma.aparch.t)
show(bmw.aparch.t)
```

```
GARCH Model : apARCH(1,1)
Mean Model : ARFIMA(1,0,0)
Distribution : std
Optimal Parameters
------------------------------------
        Estimate  Std. Error  t value Pr(>|t|)
mu      0.000048    0.000147   0.3255 0.744801
ar1     0.063666    0.012352   5.1543 0.000000
```

```
omega   0.000050    0.000032   1.5541 0.120158
alpha1  0.098839    0.012741   7.7574 0.000000
beta1   0.899506    0.013565  66.3105 0.000000
gamma1  0.121947    0.044664   2.7303 0.006327
delta   1.476643    0.142442  10.3666 0.000000
shape   4.073809    0.234417  17.3784 0.000000

LogLikelihood : 18161

Information Criteria
------------------------------------
Akaike       -5.9073
Bayes        -5.8985
Shibata      -5.9073
Hannan-Quinn -5.9042

Weighted Ljung-Box Test on Standardized Residuals
------------------------------------
                        statistic   p-value
Lag[1]                      9.824 1.723e-03
Lag[2*(p+q)+(p+q)-1][2]     9.849 2.003e-09
Lag[4*(p+q)+(p+q)-1][5]    12.253 1.100e-04
d.o.f=1
H0 : No serial correlation

Weighted Ljung-Box Test on Standardized Squared Residuals
------------------------------------
                        statistic p-value
Lag[1]                      1.456  0.2276
Lag[2*(p+q)+(p+q)-1][5]     2.363  0.5354
Lag[4*(p+q)+(p+q)-1][9]     3.258  0.7157
d.o.f=2
```

As mentioned earlier, in the output from `ugarchfit()`, the `Information Criteria` values have been normalized by dividing by n, though this is not noted in the output.

The normalized BIC for this model (-5.8985) is very nearly the same as the normalized BIC for the GARCH model with t-distributed errors (-5.8983), but after multiplying by $n = 6146$, the difference in the BIC values is 1.23. The difference between the two normalized AIC values, -5.9073 and -5.9048, is even larger, 15.4, after multiplication by n. Therefore, AIC and BIC support using the APARCH model instead of the GARCH model.

ACF plots (not shown) for the standardized residuals and their squares showed little correlation, so the AR(1) model for the conditional mean and the APARCH(1,1) model for the conditional variance fit well. Finally, `shape` is the estimated degrees of freedom of the t-distribution and is 4.07 with a small standard error, so there is very strong evidence that the conditional distribution is heavy-tailed. □

14.12 Linear Regression with ARMA+GARCH Errors

When using time series regression, one often observes autocorrelated residuals. For this reason, linear regression with ARMA disturbances was introduced in Sect. 13.3.3. The model considered was

$$Y_t = \beta_0 + \beta_1 X_{t,1} + \cdots + \beta_p X_{t,p} + e_t, \tag{14.19}$$

where

$$(1 - \phi_1\, B - \cdots - \phi_p\, B^p)(e_t - \mu) = (1 + \theta_1\, B + \ldots + \theta_q\, B^q)a_t, \tag{14.20}$$

and $\{a_t\}$ is i.i.d. white noise. This model is sufficient for serially correlated errors, but it does not accommodate volatility clustering, which is often found in the residuals.

One solution is to model the noise as an ARMA+GARCH process. Therefore, we will now assume that, instead of being i.i.d. white noise, $\{a_t\}$ is a GARCH process so that

$$a_t = \sigma_t \epsilon_t, \tag{14.21}$$

where

$$\sigma_t = \sqrt{\omega + \sum_{i=1}^{p} \alpha_i a_{t-i}^2 + \sum_{j=1}^{q} \beta_j \sigma_{t-j}^2}, \tag{14.22}$$

and $\{\epsilon_t\}$ is i.i.d. white noise. The model given by (14.19)–(14.22) is a *linear regression model with ARMA+GARCH disturbances.*

Some software, including the `ugarchfit()` function from `R`'s `rugarch` package, can fit the linear regression model with ARMA+GARCH disturbances in one step. Another solution is to adjust or correct the estimated covariance matrix of the regression coefficients, via the HAC estimator from Sect. 13.3.2, by using the `NeweyWest()` function from the `R` package `sandwich`. However, if such software is not available, then a three-step estimation method is the following:

1. estimate the parameters in (14.19) by ordinary least-squares;
2. fit model (14.20)–(14.22) to the ordinary least-squares residuals;
3. reestimate the parameters in (14.19) by weighted least-squares with weights equal to the reciprocals of the conditional variances from step 2.

Example 14.4. Regression analysis with ARMA+GARCH errors of the Nelson–Plosser data

In Example 9.9, we saw that a parsimonious model for the yearly log returns on the stock index `diff(log(sp))` used `diff(log(ip))` and `diff(bnd)` as predictors. Figure 14.9 contains ACF plots of the residuals [panel (a)] and squared residuals [panel (b)]. Externally studentized residuals were used, but the plots for the raw residuals are similar. There is some autocorrelation in both the residuals and squared residuals.

```
nelsonplosser = read.csv("nelsonplosser.csv", header = TRUE)
new_np = na.omit(nelsonplosser)
attach(new_np)
fit.lm1 = lm(diff(log(sp)) ~ diff(log(ip)) + diff(bnd))
summary(fit.lm1)
```

```
Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)    0.01657    0.02100   0.789 0.433316
diff(log(ip))  0.69748    0.16834   4.143 0.000113 ***
diff(bnd)     -0.13224    0.06225  -2.124 0.037920 *
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1   1

Residual standard error: 0.1509 on 58 degrees of freedom
Multiple R-squared:  0.3087,Adjusted R-squared:  0.2848
F-statistic: 12.95 on 2 and 58 DF,  p-value: 2.244e-05
```

The `auto.arima()` function from `R`'s `forecast` package selected an MA(1) model [i.e., ARIMA(0,0,1)] for the residuals. Next an MA(1)+ARCH(1) model was fit to the regression model's raw residuals. Sample ACF plots of the standardized residuals from the MA(1)+ARCH(1) model are in Fig. 14.9c and d. One sees essentially no short-term autocorrelation in the ARMA+GARCH standardized or squared standardized residuals, which indicates that the ARMA+GARCH model accounts for the observed dependence in the regression residuals satisfactorily. A normal plot showed that the standardized residuals are close to normally distributed, which is not unexpected for yearly log returns.

Finally, the linear model was refit with the reciprocals of the conditional variances as weights. The estimated regression coefficients are given below along with their standard errors and p-values.

```
fit.lm3 = lm(diff(log(sp)) ~ diff(log(ip)) + diff(bnd),
             weights = 1/sigma.arch^2)
summary(fit.lm3)
```

```
Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)    0.03216    0.02052   1.567  0.12263
diff(log(ip))  0.55464    0.16942   3.274  0.00181 **
diff(bnd)     -0.12215    0.05827  -2.096  0.04051 *
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1   1

Residual standard error: 1.071 on 57 degrees of freedom
Multiple R-squared:  0.2416,Adjusted R-squared:  0.2149
F-statistic: 9.077 on 2 and 57 DF,  p-value: 0.0003783
```

There are no striking differences between these results and the unweighted fit in Example 9.9. In this situation, the main reason for using the GARCH

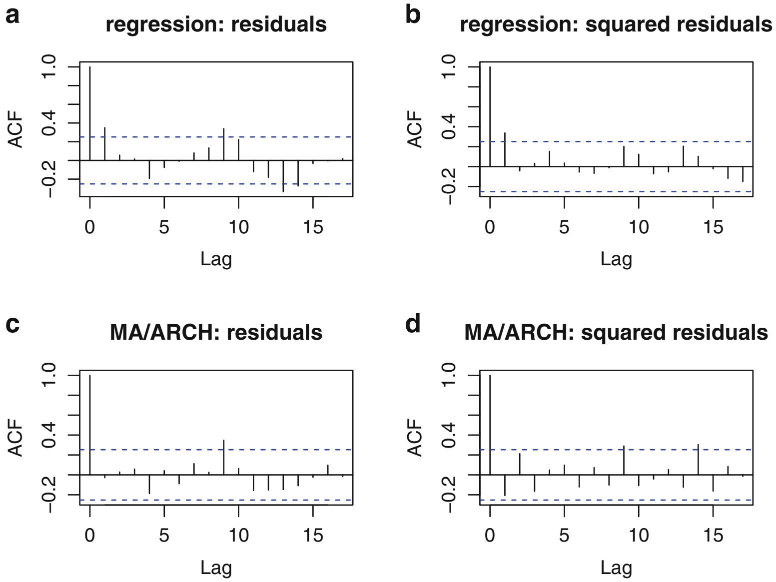

Fig. 14.9. *(a) Sample ACF of the externally studentized residuals and (b) their squared values, from a linear model; (c) Sample ACF of the standardized residuals and (d) their squared values, from an MA(1)+ARCH(1) fit to the regression residuals.*

model for the residuals would be in providing more accurate prediction intervals if the model were to be used for forecasting; see Sect. 14.13. □

14.13 Forecasting ARMA+GARCH Processes

Forecasting ARMA+GARCH processes is in one way similar to forecasting ARMA processes—point estimates, e.g., forecasts of the conditional mean, are the same because a GARCH process is weak white noise. What differs between forecasting ARMA+GARCH and ARMA processes is the behavior of the prediction intervals. In times of high volatility, prediction intervals using an ARMA+GARCH model will widen to take into account the higher amount of uncertainty. Similarly, the prediction intervals will narrow in times of lower volatility. Prediction intervals using an ARMA model without conditional heteroskedasticity cannot adapt in this way.

To illustrate, we will compare the prediction of a Gaussian white noise process and the prediction of a GARCH(1,1) process with Gaussian innovations.

Both have an ARMA(0,0) model for the conditional mean so their forecasts are equal to the marginal mean, which will be called μ. For Gaussian white noise, the prediction limits are $\mu \pm z_{\alpha/2}\sigma$, where σ is the marginal standard deviation. For a GARCH(1,1) process $\{Y_t\}$, the prediction limits at time origin n for h-steps ahead forecasting are $\mu \pm z_{\alpha/2}\sigma_{n+h|n}$ where $\sigma_{n+h|n}$ is the conditional standard deviation of Y_{n+h} given the information available at time n. As h increases, $\sigma_{n+h|n}$ converges to σ, so for long lead times the prediction intervals for the two models are similar. For shorter lead times, however, the prediction limits can be quite different.

Example 14.5. Forecasting BMW log returns

In this example, we will return to the daily BMW stock log returns used in several earlier examples. We have seen in Example 14.2 that an AR(1)+GARCH(1,1) model fits the returns well. Also, the estimated AR(1) coefficient is small, less than 0.1. Therefore, it is reasonable to use a GARCH (1,1) model for forecasting.

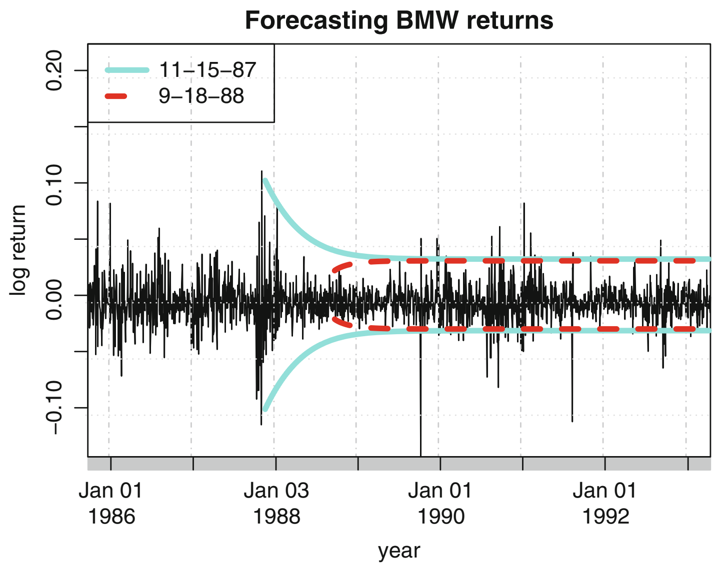

Fig. 14.10. *Prediction limits for forecasting daily BMW stock log returns from two different time origins.*

Figure 14.10 plots the returns from 1986 until 1992. Forecast limits are also shown for two time origins, November 15, 1987 and September 18, 1988. At the first time origin, which is soon after Black Monday, the markets were very volatile. The forecast limits are wide initially but narrow as the conditional standard deviation converges downward to the marginal standard deviation. At the second time origin, the markets were less volatile than usual and the prediction intervals are narrow initially but then widen. In theory, both sets of prediction limits should converge to the same values, $\mu \pm z_{\alpha/2}\sigma$ where σ is the marginal standard deviation for a stationary process. In this example, they do not quite converge to each other because the estimates of σ differ between the two time origins. □

14.14 Multivariate GARCH Processes

Financial asset returns tend to move together over time, as do their respective volatilities, across both assets and markets. Modeling a time-varying conditional covariance matrix, or volatility matrix, is important in many financial applications, including asset pricing, hedging, portfolio selection, and risk management.

Multivariate volatility modeling has major challenges to overcome. First, the curse of dimensionality; there are $d(d+1)/2$ variances and covariances for a d-dimensional process, e.g., 45 for $d = 9$, all of which may vary over time. Further, unlike returns, all of these variances and covariances are unobserved, or latent. Many parameterizations for the evolution of the volatility matrix use such a large number of parameters that estimation becomes infeasible for $d > 10$. In addition to empirical adequacy (i.e., goodness of fit of the model to the data), ease and feasibility of estimation are important considerations.

Analogous to positivity constraints in univariate GARCH models, a well-defined multivariate volatility matrix process must be positive-definite at each time point, and model-based forecasts should as well. From a practical perspective, a well-defined inverse of a volatility matrix is frequently needed in applications. Additionally, a positive conditional variance estimate for a portfolio's return, which are a linear combination of asset returns, is essential; fortunately, this is guaranteed by positive definiteness.

14.14.1 Multivariate Conditional Heteroscedasticity

Figures 14.11a and b are time series plots of daily returns (in percentage) for IBM stock and the Center for Research in Security Prices (CRSP) value-weighted index, including dividends, from January 3, 1989 to December 31, 1998, respectively. The data are from the `Ecdat` package in `R`. Each series clearly exhibits volatility clustering. Let $\boldsymbol{Y}_t$ denote the vector time series of these returns.

```
data(CRSPday, package="Ecdat")
CRSPday = ts(CRSPday, start = c(1989, 1), frequency = 253)
ibm  = CRSPday[,5] * 100
crsp = CRSPday[,7] * 100
Y = cbind(ibm, crsp)
par(mfrow = c(2,1))
plot(Y[,1], type='l', xlab="year", ylab="return (%)", main="(a)")
plot(Y[,2], type='l', xlab="year", ylab="return (%)", main="(b)")
```

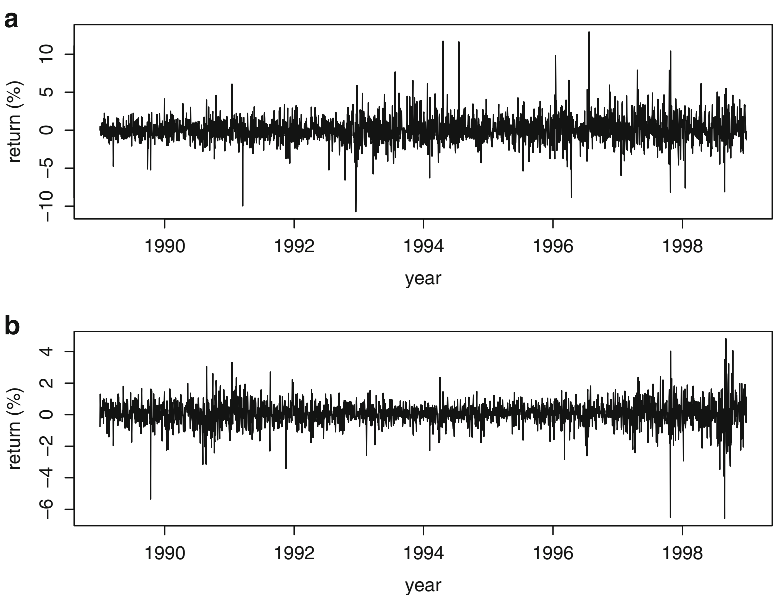

Fig. 14.11. Daily returns (in percentage) for (a) IBM stock and (b) the CRSP value-weighted index, including dividends.

Figures 14.12a and b are the sample ACF plots for the IBM stock and CRSP index returns, respectively. There is some evidence of minor serial correlation at low lags. Next, we consider the lead-lag linear relationship between pairs of returns. Figure 14.12c is the sample cross-correlation function (CCF) between IBM and CRSP. The lag zero estimate for contemporaneous correlation is approximately 0.49. There is also some evidence of minor cross-correlation at low lags.

```
layout(rbind(c(1,2), c(3,3)),widths=c(1,1,2),heights=c(1,1))
acf(as.numeric(Y[,1]), ylim=c(-0.1,0.1), main="(a)")
acf(as.numeric(Y[,2]), ylim=c(-0.1,0.1), main="(b)")
ccf(as.numeric(Y[,1]),as.numeric(Y[,2]),
    type=c("correlation"), main="(c)", ylab="CCF", lag=20)
cor(ibm, crsp)
```

```
[1] 0.4863639
```

The multivariate Ljung-Box test (see Sect. 13.4.3) is applied to simultaneously test that the first K auto-correlations, as well as the lagged cross-correlations, are all zero. The multivariate Ljung-Box test statistic at lag five is 50.15. The associate p-value is very close to zero, which provides strong evidence to reject the null hypothesis and indicates there is significant serial correlation in the vector process.

```
source("SDAFE2.R")
mLjungBox(Y, 5)
```

```
  K  Q(K) d.f. p-value
1 5 50.15   20       0
```

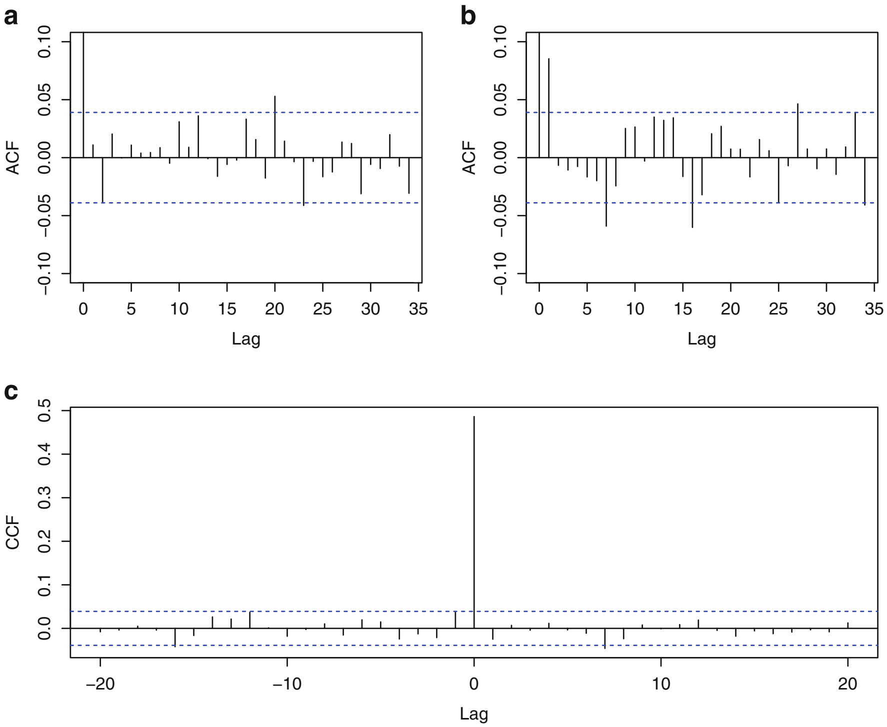

Fig. 14.12. ACFs for (a) the IBM stock and (b) CRSP index returns; (c) CCF between IBM and CRSP returns.

For simplicity, we use ordinary least squares to fit a VAR(1) model (see Sect. 13.4.4) to remove the minor serial correlation and focus on the conditional variance and covariance. Let $\widehat{\boldsymbol{a}}_t$ denote the estimated residuals from the regression; $\widehat{\boldsymbol{a}}_t$ is an estimate of the innovation process $\boldsymbol{a}_t$, which is described more fully below. The multivariate Ljung-Box test statistic at lag five is now 16.21, which has an approximate p-value of 0.704, indicating there is no significant serial correlation in the vector residual process.

```
fit.AR1 = ar(Y, aic = FALSE, order.max=1)
A = fit.AR1$resid[-1,]
mLjungBox(A, 5)
```

```
  K  Q(K) d.f. p-value
1 5 16.21   20   0.704
```

Although the residual series $\boldsymbol{a}_t$ is serially uncorrelated, Fig. 14.13 shows it is not an independent process. Figures 14.13a and b are sample ACF plots for the squared residual series $\widehat{a}_{it}^2$. They both show substantial positive autocorrelation because of the volatility clustering. Figure 14.13c is the sample CCF for the squared series; this figure shows there is a dynamic relationship between the squared series at low lags. Figure 14.13d is the sample ACF for the product series $\widehat{a}_{1t}\widehat{a}_{2t}$ and shows that there is also evidence of positive autocorrelation in the conditional covariance series. The multivariate volatility models described below attempt to account for these forms of dependence exhibited in the vector residual series.

14.14.2 Basic Setting

Let $\boldsymbol{Y}_t = (Y_{1,t}, \ldots, Y_{d,t})'$ denote a d-dimensional vector process and let $\mathcal{F}_t$ denote the information set at time index t, generated by $\boldsymbol{Y}_t, \boldsymbol{Y}_{t-1}, \ldots$. We may partition the process as

$$\boldsymbol{Y}_t = \boldsymbol{\mu}_t + \boldsymbol{a}_t, \tag{14.23}$$

in which $\boldsymbol{\mu}_t = \mathrm{E}(\boldsymbol{Y}_t|\mathcal{F}_{t-1})$ is the conditional mean vector at time index t, and $\{\boldsymbol{a}_t\}$ is the mean zero weak white noise innovation vector process with unconditional covariance matrix $\boldsymbol{\Sigma}_{\boldsymbol{a}} = \mathrm{Cov}(\boldsymbol{a}_t)$. Let

$$\boldsymbol{\Sigma}_t = \mathrm{Cov}(\boldsymbol{a}_t|\mathcal{F}_{t-1}) = \mathrm{Cov}(\boldsymbol{Y}_t|\mathcal{F}_{t-1}) \tag{14.24}$$

denote the conditional covariance or volatility matrix at time index t. Multivariate time series modeling is primarily concerned with the time evolutions of $\boldsymbol{\mu}_t$ and $\boldsymbol{\Sigma}_t$, the conditional mean and conditional covariance matrix. For a stationary process, the unconditional mean and unconditional covariance matrix are constant, even though the conditional mean and conditional covariance matrix may be time-varying.

Throughout this section we assume that $\boldsymbol{\mu}_t$ follows a stationary VAR(p) model with $\boldsymbol{\mu}_t = \boldsymbol{\mu} + \sum_{\ell=1}^{p} \boldsymbol{\Phi}_\ell(\boldsymbol{Y}_{t-\ell} - \boldsymbol{\mu})$, where p is a non-negative integer, $\boldsymbol{\mu}$ is the $d \times 1$ unconditional mean vector, and the $\boldsymbol{\Phi}_\ell$ are $d \times d$ coefficient matrices, respectively. Recall, the residual series considered in Fig. 14.13 were from a VAR model with $p = 1$.

The relationship between the innovation process and the volatility process is defined by

$$\boldsymbol{a}_t = \boldsymbol{\Sigma}_t^{1/2}\boldsymbol{\epsilon}_t, \quad \boldsymbol{\epsilon}_t \overset{iid}{\sim} F(\boldsymbol{0}, \boldsymbol{I}_d), \tag{14.25}$$

in which $\boldsymbol{\Sigma}_t^{1/2}$ is a symmetric *matrix square-root* of $\boldsymbol{\Sigma}_t$, such that $\boldsymbol{\Sigma}_t^{1/2}\boldsymbol{\Sigma}_t^{1/2} = \boldsymbol{\Sigma}_t$. The iid white noise $\boldsymbol{\epsilon}_t$ are *standardized* innovations from a multivariate distribution F with mean zero and a covariance matrix equal to the identity. The models detailed below describe dynamic evolutions for the volatility matrix $\boldsymbol{\Sigma}_t$.

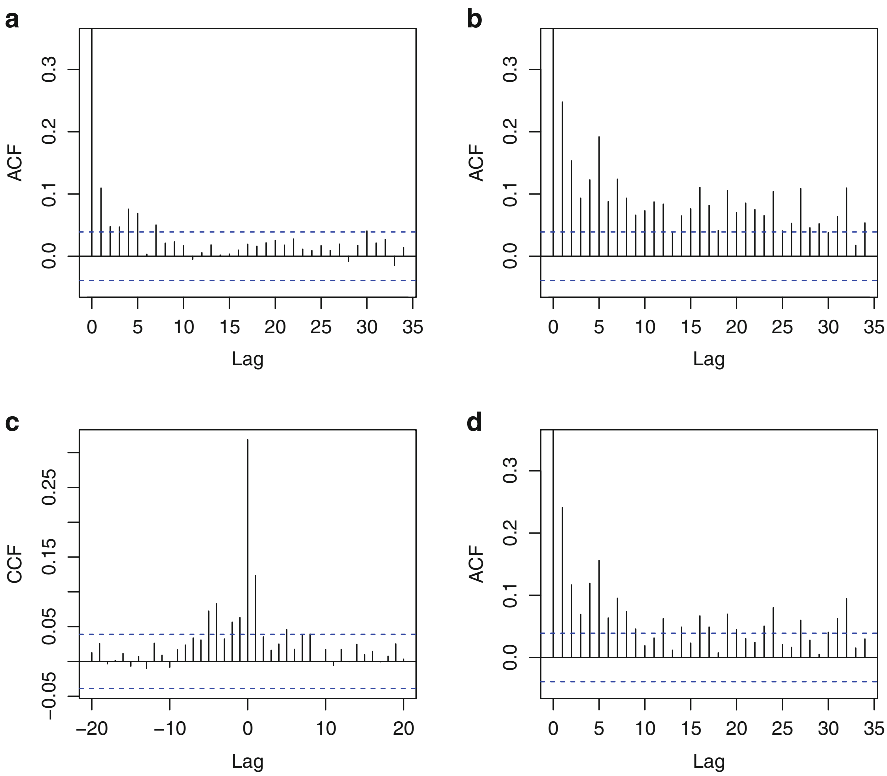

Fig. 14.13. ACFs of squared residuals from a VAR(1) model for (a) IBM and (b) CRSP; (c) CCF between the squared residuals; (d) ACF for the product of the residuals.

14.14.3 Exponentially Weighted Moving Average (EWMA) Model

The simplest matrix generalization of a univariate volatility model is the exponentially weighted moving average (EWMA) model. It is indexed by a single parameter $\lambda \in (0, 1)$, and is defined by the recursion

$$\begin{aligned} \boldsymbol{\Sigma}_t &= (1-\lambda)\boldsymbol{a}_{t-1}\boldsymbol{a}'_{t-1} + \lambda\boldsymbol{\Sigma}_{t-1} \\ &= (1-\lambda)\sum_{\ell=1}^{\infty}\lambda^{\ell-1}\boldsymbol{a}_{t-\ell}\boldsymbol{a}'_{t-\ell}. \end{aligned} \tag{14.26}$$

When the recursion in (14.26) is initialized with a positive-definite (p.d.) matrix the sequence remains p.d. This single parameter model is simple to estimate regardless of the dimension, with large values of λ indicating high persistence in the volatility process. However, the dynamics can be too restrictive in practice, since the component-wise evolutions all have the same discounting factor (i.e., persistence parameter) λ.

Figure 14.14 shows the in-sample fitted EWMA model for $\widehat{\boldsymbol{a}}_t$ assuming a multivariate standard normal distribution for $\boldsymbol{\epsilon}_t$ and using conditional maximum likelihood estimation. The estimated conditional standard deviations are shown in (a) and (d), and the conditional covariances and implied conditional correlations are shown in (b) and (c), respectively. The persistence parameter λ was estimated as 0.985. Estimation and Fig. 14.14 were calculated using the following commands in R.

```
source("SDAFE2.R")
EWMA.param = est.ewma(lambda.0=0.95, innov=A)
EWMA.Sigma = sigma.ewma(lambda=EWMA.param$lambda.hat, innov=A)
par(mfrow = c(2,2))
plot(ts(EWMA.Sigma[1,1,]^.5, start = c(1989, 1), frequency = 253),
     type = 'l', xlab = "year", ylab = NULL,
     main = expression(paste("(a) ", hat(sigma)["1,t"])))
plot(ts(EWMA.Sigma[1,2,], start = c(1989, 1), frequency = 253),
     type = 'l', xlab = "year", ylab = NULL,
     main = expression(paste("(b) ", hat(sigma)["12,t"])))
plot(ts(EWMA.Sigma[1,2,]/(sqrt(EWMA.Sigma[1,1,]* EWMA.Sigma[2,2,])),
        start = c(1989, 1), frequency = 253),
     type = 'l', xlab = "year", ylab = NULL,
     main = expression(paste("(c) ", hat(rho)["12,t"])))
points(ts(mvwindow.cor(A[,1],A[,2], win = 126)$correlation,
          start = c(1989, 1), frequency = 253),
       type = 'l', col = 2, lty = 2, lwd=2)
plot(ts(EWMA.Sigma[2,2,]^.5, start = c(1989, 1), frequency = 253),
     type = 'l', xlab = "year", ylab = NULL,
     main = expression(paste("(d) ", hat(sigma)["2,t"])))
EWMA.param$lambda.hat
```

```
[1] 0.985046
```

14.14.4 Orthogonal GARCH Models

Several factor and orthogonal models have been proposed to reduce the number of parameters and parameter constraints by imposing a common dynamic structure on the elements of the volatility matrix. The orthogonal GARCH (O-GARCH) model of Alexander (2001) is among the most popular because of its simplicity. It is assumed that the innovations $\boldsymbol{a}_t$ can be decomposed into orthogonal components $\boldsymbol{z}_t$ via a linear transformation U. This is done in conjunction with principal component analysis (PCA, see Sect. 18.2) as follows. Let $\boldsymbol{O}$ be the matrix of eigenvectors and $\boldsymbol{\Lambda}$ the diagonal matrix of the corresponding eigenvalues of $\boldsymbol{\Sigma}_{\boldsymbol{a}}$. Then, take $U = \boldsymbol{\Lambda}^{-1/2}\boldsymbol{O}'$, and let

$$\boldsymbol{z}_t = U\boldsymbol{a}_t.$$

The components are constructed such that $\mathrm{Cov}(\boldsymbol{z}_t) = \boldsymbol{I}_d$. The sample estimate of $\boldsymbol{\Sigma}_{\boldsymbol{a}}$ is typically used to estimate U.

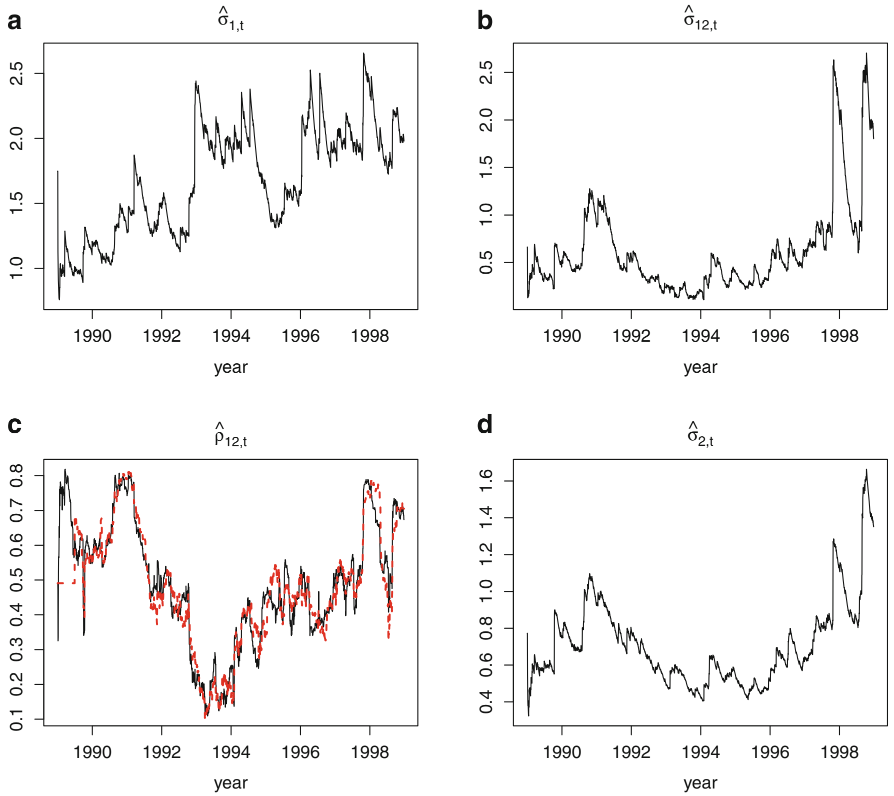

Fig. 14.14. A fitted EWMA model with $\lambda = 0.985$. The red line in (c) is the sample correlation estimate over the previous six months for comparison.

Next, univariate GARCH(1,1) models are individually fit to each orthogonal component to estimate the *conditional* covariance $\boldsymbol{V}_t = \mathrm{Cov}(\boldsymbol{z}_t|\mathcal{F}_{t-1})$. Let

$$\begin{aligned} v_{it}^2 &= \omega_i + \alpha_i z_{i,t-1}^2 + \beta_i v_{i,t-1}^2 \\ \boldsymbol{V}_t &= \mathrm{diag}\{v_{1,t}^2, \ldots, v_{d,t}^2\} \\ \boldsymbol{\Sigma}_t &= U^{-1}\boldsymbol{V}_t U^{-1\prime}. \end{aligned}$$

In summary, a linear transformation U is estimated, using PCA, such that the components of $\boldsymbol{z}_t = U\boldsymbol{a}_t$ have unconditional correlation approximately equal to zero. It is then also *assumed* that the *conditional* correlations of $\boldsymbol{z}_t$ are also zero; however, this is not at all assured to be true. Under this additional stronger assumption, $\boldsymbol{V}_t$, the conditional covariance matrix for $\boldsymbol{z}_t$, is diagonal. For simplicity, univariate models are then fit to model the conditional variance v_{it}^2 for each component of $\boldsymbol{z}_t$.

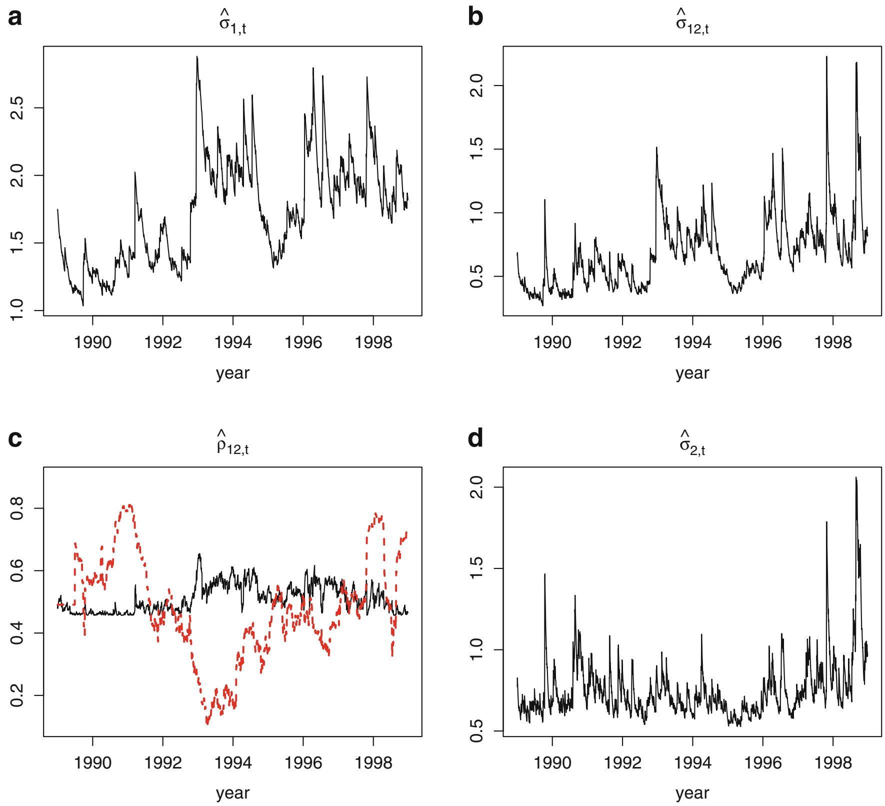

Fig. 14.15. A fitted first order orthogonal GARCH model with $(\omega_1, \alpha_1, \beta_1)' = (0.0038, 0.0212, 0.9758)'$, $(\omega_2, \alpha_2, \beta_2)' = (0.0375, 0.0711, 0.8913)'$, and $U^{-1} = ((1.7278, 0.2706)', (0.2706, 0.7241)')$. The red line in (c) is the sample correlation estimate over the previous six months for comparison.

The main drawback of this model is that the orthogonal components are uncorrelated unconditionally, but they may still be conditionally correlated. The O-GARCH model implicitly assumes the conditional correlations for $\boldsymbol{z}_t$ are zero. Figure 14.15 shows a fitted O-GARCH model for $\widehat{\boldsymbol{a}}_t$ using PCA followed by univariate conditional maximum likelihood estimation. The estimated conditional standard deviations are shown in (a) and (d), and the conditional covariances and conditional correlations are shown in (b) and (c), respectively. The implied conditional correlations do not appear adequate for this fitted model compared to the sample correlation estimate over the previous six months (used as a proxy for the conditional correlation process).

14.14.5 Dynamic Orthogonal Component (DOC) Models

To properly apply univariate modeling after estimating a linear transformation in the spirit of the O-GARCH model above, the resulting component processes must not only be orthogonal contemporaneously, the conditional correlations must also be zero. Additionally, the lagged cross-correlations for the squared components must also be zero. In Matteson and Tsay (2011), if the components of a time series $\boldsymbol{s}_t$ satisfy these conditions, then they are called dynamic orthogonal components (DOCs) in volatility.

Let $\boldsymbol{s}_t = (s_{1,t}, \ldots, s_{d,t})'$ denote a vector time series of DOCs. Without loss of generality, $\boldsymbol{s}_t$ is assumed to be standardized such that $\mathrm{E}(s_{i,t}) = 0$ and $\mathrm{Var}(s_{i,t}) = 1$ for $i = 1, \ldots, d$. A Ljung-Box type statistic, defined below, is used to test for the existence of DOCs in volatility. Including lag zero in the test implies that the pairwise product processes among stationary DOCs $s_{i,t}s_{j,t}$ has zero serial correlation since the Cauchy-Schwarz inequality gives

$$|\mathrm{Cov}(s_{i,t}s_{j,t}, s_{i,t-h}s_{j,t-h})| \leq \mathrm{Var}(s_{i,t}s_{j,t}) = \mathrm{E}(s_i^2 s_j^2), \tag{14.27}$$

and $E(s_{i,t}s_{j,t}) = E(s_{i,t-h}s_{j,t-h}) = 0$ by the assumption of a DOC model.

Let $\rho_{s_i^2,s_j^2}(h) = \mathrm{Corr}(s_{i,t}^2, s_{j,t-h}^2)$. The joint lag-$K$ null and alternative hypotheses to test for the existence of DOCs in volatility are

$$\begin{aligned} H_0 &: \rho_{s_i^2,s_j^2}(h) = 0 \text{ for all } i \neq j,\ h = 0, \ldots, K \\ H_A &: \rho_{s_i^2,s_j^2}(h) \neq 0 \text{ for some } i \neq j,\ h = 0, \ldots, K. \end{aligned}$$

The corresponding Ljung-Box type test statistic is

$$Q_d^0(\boldsymbol{s}^2; K) = n \sum_{i<j} \rho_{s_i^2,s_j^2}(0)^2 + n(n+2) \sum_{h=1}^{K} \sum_{i \neq j} \rho_{s_i^2,s_j^2}(h)^2/(n-h). \tag{14.28}$$

Under H_0, $Q_d^0(\boldsymbol{s}^2; K)$ is asymptotically distributed as Chi-squared with $d(d-1)/2 + Kd(d-1)$ degrees of freedom. The null hypothesis is rejected for a large value of Q_d^0. When H_0 is rejected, one must seek an alternative modeling procedure.

As expected from Fig. 14.13, the DOCs in volatility hypothesis is rejected for the VAR(1) residuals. The test statistic is $Q_2^0(\widehat{\boldsymbol{a}}^2, 5) = 356.926$ with a p-value near zero. DOCs in volatility is also rejected for the principal components used in the O-GARCH model, the test statistic is $Q_2^0(\boldsymbol{z}^2, 5) = 135.492$ with a p-value near zero. Starting with the uncorrelated principal components $\boldsymbol{z}_t$, Matteson and Tsay (2011) propose estimating an orthogonal matrix $\boldsymbol{W}$ such that the components $\boldsymbol{s}_t = \boldsymbol{W}\boldsymbol{z}_t$ are as close to DOCs in volatility as possible. This is done by minimizing a reweighted version of the Ljung-Box type test statistic (14.28), with respect to the separating matrix $\boldsymbol{W}$. The null hypothesis of DOCs in volatility is accepted for the estimated components $\boldsymbol{s}_t$, with $Q_2^0(\boldsymbol{s}^2, 5) = 7.845$ which has a p-value approximately equal to 0.727.

After DOCs are identified, a univariate volatility model is considered for each process $v_{i,t}^2 = \mathrm{Var}(s_{i,t}|\mathcal{F}_{t-1})$. For example, the following model was fit

$$\begin{aligned}
\boldsymbol{a}_t &= \boldsymbol{M}\boldsymbol{s}_t = \boldsymbol{M}\boldsymbol{V}_t^{1/2}\boldsymbol{\epsilon}_t,\\
\boldsymbol{V}_t &= \mathrm{diag}\{v_{1,t}^2, \ldots, v_{d,t}^2\}, \quad \epsilon_{it} \overset{iid}{\sim} t_{\nu_i}(0,1)\\
v_{i,t}^2 &= \omega_i + \alpha_i s_{i,t-1}^2 + \beta_i v_{i,t-1}^2\\
\boldsymbol{\Sigma}_t &= \boldsymbol{M}\boldsymbol{V}_t\boldsymbol{M}',
\end{aligned}$$

in which $t_{\nu_i}(0,1)$ denotes the standardized Student-t distribution with tail-index ν_i. Each $\boldsymbol{\Sigma}_t$ is positive-definite if $v_{i,t}^2 > 0$ for all components. The fundamental motivation is that empirically the dynamics of $\boldsymbol{a}_t$ can often be well approximated by an invertible linear combination of DOCs $\boldsymbol{a}_t = \boldsymbol{M}\boldsymbol{s}_t$, in which $\boldsymbol{M} = U^{-1}\boldsymbol{W}'$ by definition.

In summary, U is estimated by PCA to uncorrelate $\boldsymbol{a}_t$, $\boldsymbol{W}$ is estimated to minimize a reweighted version of (14.28) defined above (giving more weight to lower lags). The matrices U and $\boldsymbol{W}$ are combined to estimate DOCs s_t, of which univariate volatility modeling may then be appropriately applied. This approach allows modeling of a d-dimensional multivariate volatility process with d univariate volatility models, while greatly reducing both the number of parameters and the computational cost of estimation, and at the same time maintaining adequate empirical performance.

Figure 14.16 shows a fitted DOCs in volatility GARCH model for $\widehat{\boldsymbol{a}}_t$ using generalized decorrelation followed by univariate conditional maximum likelihood estimation. The estimated conditional standard deviations are shown in (a) and (d), and the conditional covariances and implied correlations are shown in (b) and (c), respectively. Unlike the O-GARCH fit, the implied conditional correlations appear adequate compared to the rolling estimator. Estimation of the O-GARCH and DOC models and Figs. 14.15 and 14.16 were calculated using the following commands in R.

```
source("SDAFE2.R")
DOC.fit = doc.garch(E = A, L = 4., c = 2.25, theta.ini = NULL)

par(mfrow = c(2,2)) # O-GARCH
plot(ts(DOC.fit$Sigma.pca[1,1,]^.5, start=c(1989,1), frequency=253),
     type = 'l', xlab = "year", ylab = NULL,
     main = expression(paste("(a) ", hat(sigma)["1,t"])))
plot(ts(DOC.fit$Sigma.pca[2,1,], start=c(1989,1), frequency=253),
     type = 'l', xlab = "year", ylab = NULL,
     main = expression(paste("(b) ", hat(sigma)["12,t"])))
plot(ts(DOC.fit$Sigma.pca[2,1,]/(sqrt(DOC.fit$Sigma.pca[1,1,]*
                                      DOC.fit$Sigma.pca[2,2,])),
        start=c(1989,1), frequency=253),
     type = 'l', xlab = "year", ylab = NULL, ylim = c(0.1,0.9),
     main = expression(paste("(c) ", hat(rho)["12,t"])))
points(ts(mvwindow.cor(A[,1],A[,2], win = 126)$correlation,
          start = c(1989, 1), frequency = 253),
```

```
      type = 'l', col = 2, lty = 2, lwd = 2)
plot(ts(DOC.fit$Sigma.pca[2,2,]^.5, start=c(1989,1), frequency=253),
     type = 'l', xlab = "year", ylab = NULL,
     main = expression(paste("(d) ", hat(sigma)["2,t"])))
```

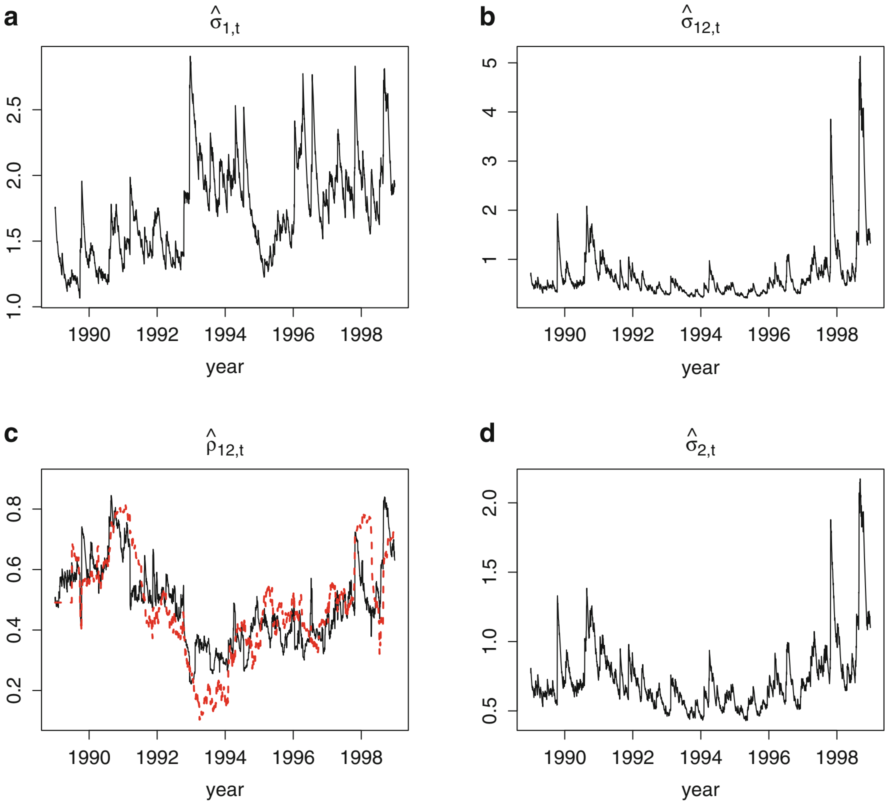

Fig. 14.16. A fitted first order DOCs in volatility GARCH model with $(\omega_1, \alpha_1, \beta_1, \nu_1)' = (0.0049, 0.0256, 0.9703, 4.3131)'$, $(\omega_2, \alpha_2, \beta_2, \nu_2)' = (0.0091, 0.0475, 0.9446, 5.0297)'$, and $\boldsymbol{M} = ((1.5350, 0.838)', (0.0103, 0.7730)')$. The red line in (c) is the sample correlation estimate over the previous six months for comparison.

```
par(mfrow = c(2,2)) # DOCs in volatility
plot(ts(DOC.fit$Sigma.doc[1,1,]^.5, start=c(1989,1), frequency=253),
     type = 'l', xlab = "year", ylab = NULL,
     main = expression(paste("(a) ", hat(sigma)["1,t"])))
plot(ts(DOC.fit$Sigma.doc[2,1,], start=c(1989,1), frequency=253),
     type = 'l', xlab = "year", ylab = NULL,
     main = expression(paste("(b) ", hat(sigma)["12,t"])))
plot(ts(DOC.fit$Sigma.doc[2,1,]/(sqrt(DOC.fit$Sigma.doc[1,1,]*
```

```
                                         DOC.fit$Sigma.doc[2,2,])),
        start=c(1989,1), frequency=253),
     type = 'l', xlab = "year", ylab = NULL, ylim = c(0.1,0.9),
     main = expression(paste("(c) ", hat(rho)["12,t"])))
points(ts(mvwindow.cor(A[,1],A[,2], win = 126)$correlation,
          start=c(1989,1), frequency=253),
       type = 'l', col = 2, lty = 2,lwd=2)
plot(ts(DOC.fit$Sigma.doc[2,2,]^.5, start=c(1989,1), frequency=253),
     type = 'l', xlab = "year", ylab = NULL,
     main = expression(paste("(d) ", hat(sigma)["2,t"])))
```

```
DOC.fit$coef.pca
```

```
           omega     alpha1     beta1
[1,] 0.003845283 0.02118369 0.9758129
[2,] 0.037473820 0.07101731 0.8913321
```

```
DOC.fit$coef.doc
```

```
           omega     alpha1     beta1    shape
[1,] 0.004874403 0.02560464 0.9702966 4.313164
[2,] 0.009092705 0.04740792 0.9446408 5.030019
```

```
DOC.fit$W.hat
```

```
          [,1]       [,2]
[1,] 0.9412834 -0.3376174
[2,] 0.3376174  0.9412834
```

```
DOC.fit$U.hat
```

```
           [,1]       [,2]
[1,]  0.6147515 -0.2297417
[2,] -0.2297417  1.4669516
```

```
DOC.fit$M.hat
```

```
           [,1]      [,2]
[1,] 1.53499088 0.8380397
[2,] 0.01024854 0.7729063
```

```
solve(DOC.fit$U.hat)
```

```
          [,1]      [,2]
[1,] 1.7277983 0.2705934
[2,] 0.2705934 0.7240638
```

14.14.6 Dynamic Conditional Correlation (DCC) Models

Nonlinear combinations of univariate volatility models have been proposed to allow for time-varying correlations, a feature that is prevalent in many financial applications. Both Tse and Tsui (2002) and Engle (2002) generalize the constant correlation model of Bollerslev (1990) to allow for such dynamic conditional correlations (DCC).

Analogously to the GARCH(1,1) model, the first order form of the DCC model in Engle (2002) may be represented by the following equations

$$
\begin{aligned}
\sigma_{i,t}^2 &= \omega_i + \alpha_i a_{i,t-1}^2 + \beta_i \sigma_{i,t-1}^2, \\
\boldsymbol{D}_t &= \operatorname{diag}\{\sigma_{1,t}, \dots, \sigma_{d,t}\}, \\
\boldsymbol{\varepsilon}_t &= \boldsymbol{D}_t^{-1} \boldsymbol{a}_t, \\
\boldsymbol{Q}_t &= (1-\lambda)\boldsymbol{\varepsilon}_{t-1}\boldsymbol{\varepsilon}_{t-1}' + \lambda \boldsymbol{Q}_{t-1}, \\
\boldsymbol{R}_t &= \operatorname{diag}\{\boldsymbol{Q}_t\}^{-\frac{1}{2}} \boldsymbol{Q}_t \operatorname{diag}\{\boldsymbol{Q}_t\}^{-\frac{1}{2}}, \\
\boldsymbol{\Sigma}_t &= \boldsymbol{D}_t \boldsymbol{R}_t \boldsymbol{D}_t.
\end{aligned}
$$

The main idea is to first model the conditional variance of each individual series σ_{it}^2 with a univariate volatility model, estimate the *scaled* innovations $\boldsymbol{\varepsilon}_t$ (not to be confused with *standardized* innovations $\boldsymbol{\epsilon}_t$) from these models, then focus on modeling the conditional correlation matrix $\boldsymbol{R}_t$. These are then combined at each time point t to estimate the volatility matrix $\boldsymbol{\Sigma}_t$. The recursion $\boldsymbol{Q}_t$ is an EWMA model applied to the scaled innovations $\boldsymbol{\varepsilon}_t$. It is indexed by a single parameter $\lambda \in (0,1)$. The matrix $\boldsymbol{Q}_t$ needs to be rescaled to form a proper correlation matrix $\boldsymbol{R}_t$ with the value 1 for all elements on the main diagonal.

The DCC model parameters can be estimated consistently in two stages using quasi-maximum likelihood. First, a univariate GARCH$(1,1)$ model is fit to each series to estimate σ_{it}^2. Then, given $\boldsymbol{\varepsilon}_t$, λ is estimated by maximizing the components of the quasi-likelihood that only depend on the correlations. This is justified since the squared residuals do not depend on the correlation parameters.

In the form above, the variance components only condition on their own individual lagged returns and not the joint returns. Also, the dynamics for each of the conditional correlations are constrained to have equal persistence parameters, similar to the EWMA model. An explicit parameterization of the conditional correlation matrix $\boldsymbol{R}_t$, with flexible dynamics, is just as difficult to estimate in high dimensions as $\boldsymbol{\Sigma}_t$ itself. Figure 14.17 shows a fitted DCC model for $\widehat{\boldsymbol{a}}_t$ using quasi-maximum likelihood estimation. The estimated conditional standard deviations are shown in (a) and (d), and the conditional covariances and conditional correlations are shown in (b) and (c), respectively. Estimation and Fig. 14.17 were calculated using the following commands in R.

```
source("SDAFE2.R")
DCCe.fit = fit.DCCe(theta.0=0.95, innov=A)
DCCe.fit$coef
```

```
          omega     alpha1     beta1
[1,] 0.07435095 0.05528162 0.9231251
[2,] 0.02064808 0.08341755 0.8822517
```

```
DCCe.fit$lambda
```

```
[1] 0.9876297
```

```
par(mfrow = c(2,2))
plot(ts(DCCe.fit$Sigma.t[1,1,]^.5, start=c(1989, 1), frequency=253),
     type = 'l', xlab = "year", ylab = NULL,
     main = expression(paste("(a) ", hat(sigma)["1,t"])))
plot(ts(DCCe.fit$Sigma.t[2,1,], start=c(1989, 1), frequency=253),
     type = 'l', xlab = "year", ylab = NULL,
     main = expression(paste("(b) ", hat(sigma)["12,t"])))
plot(ts(DCCe.fit$R.t[2,1,], start=c(1989, 1), frequency=253),
     type = 'l', xlab = "year", ylab = NULL,
     main = expression(paste("(c) ", hat(rho)["12,t"])))
points(ts(mvwindow.cor(A[,1],A[,2], win = 126)$correlation,
          start=c(1989, 1), frequency=253),
       type = 'l', col = 2, lty = 2, lwd=2)
plot(ts(DCCe.fit$Sigma.t[2,2,]^.5, start=c(1989, 1), frequency=253),
     type = 'l', xlab = "year", ylab = NULL,
     main = expression(paste("(d) ", hat(sigma)["2,t"])))
```

14.14.7 Model Checking

For a fitted volatility sequence $\widehat{\boldsymbol{\Sigma}}_t$, the *standardized* residuals are defined as

$$\widehat{\boldsymbol{\epsilon}}_t = \widehat{\boldsymbol{\Sigma}}_t^{-1/2}\boldsymbol{a}_t, \tag{14.29}$$

in which $\widehat{\boldsymbol{\Sigma}}_t^{-1/2}$ denotes the inverse of the matrix $\widehat{\boldsymbol{\Sigma}}_t^{1/2}$. To verify the adequacy of a fitted volatility model, lagged cross-correlations of the squared standardized residuals should be zero. The product process $\widehat{\epsilon}_{it}\widehat{\epsilon}_{jt}$ should also have no serial correlation. Additional diagnostic checks for time series are considered in Li (2003). Since the standardized residuals are estimated and not observed, all p-values given in this section are only approximate.

To check the first condition we can apply a multivariate Ljung-Box test to the squared standardized residuals. For the EWMA model, $Q_2(\widehat{\boldsymbol{\epsilon}}_t^2, 5) = 26.40$ with a p-value of 0.153, implying no significant serial correlation. For the O-GARCH model, $Q_2(\widehat{\boldsymbol{\epsilon}}_t^2, 5) = 30.77$ with a p-value 0.058. In this case, there is some minor evidence of serial correlation. For the DOC in volatility model, $Q_2(\widehat{\boldsymbol{\epsilon}}_t^2, 5) = 18.68$ with a p-value 0.543, implying no significant serial correlation. For the DCC model, $Q_2(\widehat{\boldsymbol{\epsilon}}_t^2, 5) = 10.54$ with a p-value 0.957, implying no significant serial correlation.

```
n = dim(A)[1] ; d = dim(A)[2]
stdResid.EWMA = matrix(0,n,d)
stdResid.PCA = matrix(0,n,d)
stdResid.DOC = matrix(0,n,d)
stdResid.DCCe = matrix(0,n,d)
for(t in 1:n){
 stdResid.EWMA[t,] = A[t,] %*% matrix.sqrt.inv(EWMA.Sigma[,,t])
 stdResid.PCA[t,] = A[t,] %*% matrix.sqrt.inv(DOC.fit$Sigma.pca[,,t])
 stdResid.DOC[t,] = A[t,] %*% matrix.sqrt.inv(DOC.fit$Sigma.doc[,,t])
```

```
  stdResid.DCCe[t,] = A[t,] %*% matrix.sqrt.inv(DCCe.fit$Sigma.t[,,t])
}
mLjungBox(stdResid.EWMA^2, lag=5)
mLjungBox(stdResid.PCA^2, lag=5)
mLjungBox(stdResid.DOC^2, lag=5)
mLjungBox(stdResid.DCCe^2, lag=5)
```

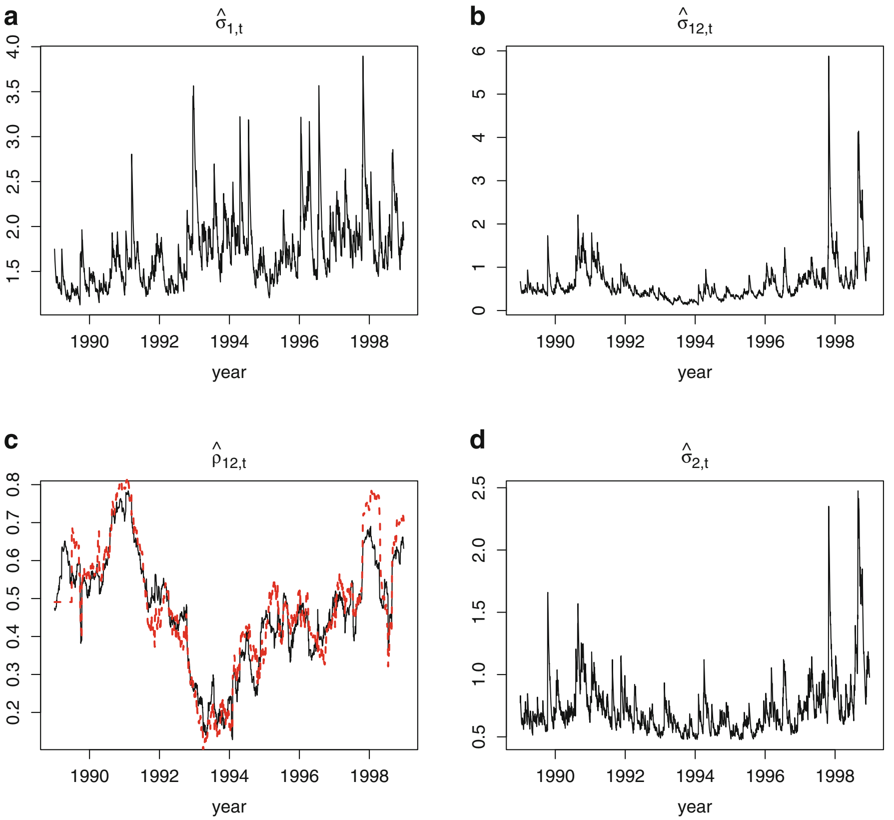

Fig. 14.17. A fitted first order DCC model with $(\omega_1, \alpha_1, \beta_1)' = (0.0741, 0.0552, 0.9233)'$, $(\omega_2, \alpha_2, \beta_2)' = (0.0206, 0.0834, 0.8823)'$, and $\lambda = 0.9876$. The red line in (c) is the sample correlation estimate over the previous six months for comparison.

The multivariate Ljung-Box test for the squared standardized residuals is not sensitive to misspecification of the conditional correlation structure. To check this condition, we apply univariate Ljung-Box tests to the product of each pair of standardized residuals. For the EWMA model, $Q(\widehat{\epsilon}_{1t}\widehat{\epsilon}_{2t}, 5) = 16.45$ with a p-value of 0.006. This model has not adequately accounted for the time-varying conditional correlation. For the O-GARCH model, $Q(\widehat{\epsilon}_{1t}\widehat{\epsilon}_{2t}, 5) = 63.09$ with a p-value near zero. This model also fails to account for the observed time-varying conditional correlation. For the DOC in volatility model, $Q(\widehat{\epsilon}_{1t}\widehat{\epsilon}_{2t}, 5) = 9.07$ with a p-value of 0.106, implying no significant serial correlation. For the DCC model, $Q(\widehat{\epsilon}_{1t}\widehat{\epsilon}_{2t}, 5) = 8.37$ with a p-value of 0.137, also implying no significant serial correlation.

```
mLjungBox(stdResid.EWMA[,1] * stdResid.EWMA[,2], lag=5)
mLjungBox(stdResid.PCA[,1] * stdResid.PCA[,2], lag=5)
mLjungBox(stdResid.DOC[,1] * stdResid.DOC[,2], lag=5)
mLjungBox(stdResid.DCCe[,1] * stdResid.DCCe[,2], lag=5)
```

14.15 Bibliographic Notes

Modeling nonconstant conditional variances in regression is treated in depth in the book by Carroll and Ruppert (1988).

There is a vast literature on GARCH processes beginning with Engle (1982), where ARCH models were introduced. Hamilton (1994), Enders (2004), Pindyck and Rubinfeld (1998), Gourieroux and Jasiak (2001), Alexander (2001), and Tsay (2005) have chapters on GARCH models. There are many review articles, including Bollerslev (1986), Bera and Higgins (1993), Bollerslev, Engle, and Nelson (1994), and Bollerslev, Chou, and Kroner (1992). Jarrow (1998) and Rossi (1996) contain a number of papers on volatility in financial markets. Duan (1995), Ritchken and Trevor (1999), Heston and Nandi (2000), Hsieh and Ritchken (2000), Duan and Simonato (2001), and many other authors study the effects of GARCH errors on options pricing, and Bollerslev, Engle, and Wooldridge (1988) use GARCH models in the CAPM.

For a thorough review of multivariate GARCH modeling see Bauwens, Laurent, and Rombouts (2006), and Silvennoinen and Teräsvirta (2009).

14.16 R Lab

14.16.1 Fitting GARCH Models

Run the following code to load the data set `TbGdpPi.csv`, which has three variables: the 91-day T-bill rate, the log of real GDP, and the inflation rate. In this lab you will use only the T-bill rate.

```
TbGdpPi = read.csv("TbGdpPi.csv", header=TRUE)
#  r = the 91-day treasury bill rate
#  y = the log of real GDP
#  pi = the inflation rate
TbGdpPi = ts(TbGdpPi, start = 1955, freq = 4)
Tbill = TbGdpPi[,1]
Tbill.diff = diff(Tbill)
```

Problem 1 *Plot both* `Tbill` *and* `Tbill.diff`*. Use both time series and ACF plots. Also, perform ADF and KPSS tests on both series. Which series do you think are stationary? Why? What types of heteroskedasticity can you see in the* `Tbill.diff` *series?*

In the following code, the variable `Tbill` can be used if you believe that series is stationary. Otherwise, replace `Tbill` by `Tbill.diff`. This code will fit an ARMA+GARCH model to the series.

```
library(rugarch)
arma.garch.norm = ugarchspec(mean.model=list(armaOrder=c(1,0)),
                             variance.model=list(garchOrder=c(1,1)))
Tbill.arma.garch.norm = ugarchfit(data=Tbill, spec=arma.garch.norm)
show(Tbill.arma.garch.norm)
```

Problem 2 *(a) Which ARMA+GARCH model is being fit? Write down the model using the same parameter names as in the* `R` *output.*
(b) What are the estimates of each of the parameters in the model?

Next, plot the residuals (ordinary or raw) and standardized residuals in various ways using the code below. The standardized residuals are best for checking the model, but the residuals are useful to see if there are GARCH effects in the series.

```
res = ts(residuals(Tbill.arma.garch.norm, standardize=FALSE),
         start = 1955, freq = 4)
res.std = ts(residuals(Tbill.arma.garch.norm, standardize=TRUE),
             start = 1955, freq = 4)
par(mfrow=c(2,3))
plot(res)
acf(res)
acf(res^2)
plot(res.std)
acf(res.std)
acf(res.std^2)
```

Problem 3 *(a) Describe what is plotted by* `acf(res)`*. What, if anything, does the plot tell you about the fit of the model?*
(b) Describe what is plotted by `acf(res^2)`*. What, if anything, does the plot tell you about the fit of the model?*
(c) Describe what is plotted by `acf(res_std^2)`*. What, if anything, does the plot tell you about the fit of the model?*
(d) Is there anything noteworthy in the figure produced by the command `plot(res.std)`*?*

Problem 4 *Now find an ARMA+GARCH model for the series* `diff.log.Tbill`*, which we will define as* `diff(log(Tbill))`*. Do you see any advantages of working with the differences of the logarithms of the T-bill rate, rather than with the difference of* `Tbill` *as was done earlier?*

14.16.2 The GARCH-in-Mean (GARCH-M) Model

A GARCH-in-Mean or *GARCH-M* model takes the form

$$\begin{aligned} Y_t &= \mu + \delta\sigma_t + a_t \\ a_t &= \sigma_t \epsilon_t \\ \sigma_t^2 &= \omega + \alpha a_{t-1}^2 + \beta\sigma_{t-1}^2 \end{aligned}$$

in which $\epsilon_t \overset{iid}{\sim} (0,1)$. The GARCH-M model directly incorporates volatility as a regression variable. The parameter δ represents the *risk premium*, or reward for additional risk. Modern portfolio theory dictates that increased volatility leads to increased risk, requiring larger expected returns. The presence of volatility as a statistically significant predictor of returns is one of the primary contributors to serial correlation in historic return series. The data set `GPRO.csv()` contains the adjusted daily closing price of GoPro stock from June 26, 2014 to January 28, 2015.

Run the following R commands to fit a GARCH-M model to the GoPro stock returns.

```
library(rugarch)
GPRO = read.table("GPRO.csv")
garchm = ugarchspec(mean.model=list(armaOrder=c(0,0),
                                    archm=T,archpow=1),
                    variance.model=list(garchOrder=c(1,1)))
GPRO.garchm = ugarchfit(garchm, data=GPRO)
show(GPRO.garchm)
```

Problem 5 *Write out the fitted model. The parameter δ is equal to* `archm` *in the* R *output.*

Problem 6 *Test the one-sided hypothesis that $\delta > 0$ verses the alternative that $\delta = 0$. Is the risk premium significant?*

14.16.3 Fitting Multivariate GARCH Models

Run the following code to again load the data set `TbGdpPi.csv`, which has three variables: the 91-day T-bill rate, the log of real GDP, and the inflation rate. In this lab you will now use the first and third series after taking first differences.

```
TbGdpPi = read.csv("TbGdpPi.csv", header=TRUE)
TbPi.diff = ts(apply(TbGdpPi[,-2],2,diff), start=c(1955,2), freq=4)
plot(TbPi.diff)
acf(TbPi.diff^2)
source("SDAFE2.R")
mLjungBox(TbPi.diff^2, lag=8)
```

Problem 7 *Does the joint series exhibit conditional heteroskedasticity? Why?*

Now fit and plot a EWMA model with the following R commands.

```
EWMA.param = est.ewma(lambda.0=0.95, innov=TbPi.diff)
EWMA.param$lambda.hat
EWMA.Sigma=sigma.ewma(lambda=EWMA.param$lambda.hat,innov=TbPi.diff)
par(mfrow = c(2,2))
plot(ts(EWMA.Sigma[1,1,]^.5, start = c(1955, 2), frequency = 4),
     type = 'l', xlab = "year", ylab = NULL,
     main = expression(paste("(a) ", hat(sigma)["1,t"])))
plot(ts(EWMA.Sigma[1,2,], start = c(1955, 2), frequency = 4),
     type = 'l', xlab = "year", ylab = NULL,
     main = expression(paste("(b) ", hat(sigma)["12,t"])))
plot(ts(EWMA.Sigma[1,2,]/(sqrt(EWMA.Sigma[1,1,]* EWMA.Sigma[2,2,])),
        start = c(1955, 2), frequency = 4),
     type = 'l', xlab = "year", ylab = NULL,
     main = expression(paste("(c) ", hat(rho)["12,t"])))
plot(ts(EWMA.Sigma[2,2,]^.5, start = c(1955, 2), frequency = 4),
     type = 'l', xlab = "year", ylab = NULL,
     main = expression(paste("(d) ", hat(sigma)["2,t"])))
```

Problem 8 *What is the estimated persistence parameter λ?*

Now estimate standardized residuals and check whether they exhibit any conditional heteroskedasticity

```
n = dim(TbPi.diff)[1]
d = dim(TbPi.diff)[2]
stdResid.EWMA = matrix(0,n,d)
for(t in 1:n){
  stdResid.EWMA[t,] = TbPi.diff[t,] %*% matrix.sqrt.inv
  (EWMA.Sigma[,,t])
}
mLjungBox(stdResid.EWMA^2, lag=8)
```

Problem 9 *Based on the output of the Ljung-Box test for the squared standardized residuals, is the EWMA model adequate?*

Run the following command in R to determine whether the joint series are DOCs in volatility.

```
DOC.test(TbPi.diff^2, 8)
```

Problem 10 *Is the null hypothesis rejected? Based on this conclusion, how should the conditional heteroskedasticity in the bivariate series be modeled, jointly or separately?*

14.17 Exercises

1. Let Z have an $N(0,1)$ distribution. Show that
$$E(|Z|) = \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} |z| e^{-z^2/2} dz = 2 \int_0^{\infty} \frac{1}{\sqrt{2\pi}} z e^{-z^2/2} dz = \sqrt{\frac{2}{\pi}}.$$
Hint: $\frac{d}{dz} e^{-z^2/2} = -z e^{-z^2/2}$.
2. Suppose that $f_X(x) = 1/4$ if $|x| < 1$ and $f_X(x) = 1/(4x^2)$ if $|x| \geq 1$. Show that
$$\int_{-\infty}^{\infty} f_X(x) dx = 1,$$
so that f_X really is a density, but that
$$\int_{-\infty}^{0} x f_X(x) dx = -\infty$$
and
$$\int_0^{\infty} x f_X(x) dx = \infty,$$
so that a random variable with this density does not have an expected value.
3. Suppose that ϵ_t is an i.i.d. WN$(0,1)$ process, that
$$a_t = \epsilon_t \sqrt{1 + 0.35 a_{t-1}^2},$$
and that
$$y_t = 3 + 0.72 y_{t-1} + a_t.$$
   (a) Find the mean of y_t.
   (b) Find the variance of y_t.
   (c) Find the autocorrelation function of y_t.
   (d) Find the autocorrelation function of a_t^2.
4. Let y_t be the AR(1)+ARCH(1) model
$$a_t = \epsilon_t \sqrt{\omega + \alpha a_{t-1}^2},$$
$$(y_t - \mu) = \phi(y_{t-1} - \mu) + a_t,$$
where ϵ_t is i.i.d. WN(0,1). Suppose that $\mu = 0.4$, $\phi = 0.45$, $\omega = 1$, and $\alpha_1 = 0.3$.

(a) Find $E(y_2|y_1 = 1, y_0 = 0.2)$.
(b) Find $\text{Var}(y_2|y_1 = 1, y_0 = 0.2)$.

5. Suppose that ϵ_t is white noise with mean 0 and variance 1, that $a_t = \epsilon_t \sqrt{7 + a_{t-1}^2/2}$, and that $Y_t = 2 + 0.67Y_{t-1} + a_t$.
(a) What is the mean of Y_t?
(b) What is the ACF of Y_t?
(c) What is the ACF of a_t?
(d) What is the ACF of a_t^2?

6. Let Y_t be a stock's return in time period t and let X_t be the inflation rate during this time period. Assume the model

$$Y_t = \beta_0 + \beta_1 X_t + \delta\sigma_t + a_t, \tag{14.30}$$

where

$$a_t = \epsilon_t \sqrt{1 + 0.5a_{t-1}^2}. \tag{14.31}$$

Here the ϵ_t are independent $N(0, 1)$ random variables. Model (14.30)–(14.31) is called a *GARCH-in-mean* model or a GARCH-M model.
Assume that $\beta_0 = 0.06$, $\beta_1 = 0.35$, and $\delta = 0.22$.
(a) What is $E(Y_t|X_t = 0.1 \text{ and } a_{t-1} = 0.6)$?
(b) What is $\text{Var}(Y_t|X_t = 0.1 \text{ and } a_{t-1} = 0.6)$?
(c) Is the conditional distribution of Y_t given X_t and a_{t-1} normal? Why or why not?
(d) Is the marginal distribution of Y_t normal? Why or why not?

7. Suppose that $\epsilon_1, \epsilon_2, \ldots$ is a Gaussian white noise process with mean 0 and variance 1, and a_t and y_t are stationary processes such that

$$a_t = \sigma_t \epsilon_t \quad \text{where} \quad \sigma_t^2 = 2 + 0.3a_{t-1}^2,$$

and

$$y_t = 2 + 0.6y_{t-1} + a_t.$$

(a) What type of process is a_t?
(b) What type of process is y_t?
(c) Is a_t Gaussian? If not, does it have heavy or lighter tails than a Gaussian distribution?
(d) What is the ACF of a_t?
(e) What is the ACF of a_t^2?
(f) What is the ACF of y_t?

8. On Black Monday, the return on the S&P 500 was −22.8 %. Ouch! This exercise attempts to answer the question, "what was the conditional probability of a return this small or smaller on Black Monday?" "Conditional" means given the information available the previous trading day. Run the following `R` code:

```
library(rugarch)
library(Ecdat)
data(SP500,package="Ecdat")
returnBlMon = SP500$r500[1805] ; returnBlMon
x = SP500$r500[(1804-2*253+1):1804]
ts.plot(c(x,returnBlMon))
spec = ugarchspec(mean.model=list(armaOrder=c(1,0)),
                  variance.model=list(garchOrder=c(1,1)),
                  distribution.model = "std")
fit = ugarchfit(data=x, spec=spec)
dfhat = coef(fit)[6]
forecast = ugarchforecast(fit, data=x, n.ahead=1)
```

The S&P 500 returns are in the data set `SP500` in the `Ecdat` package. The returns are the variable `r500` (this is the only variable in this data set). Black Monday is the 1805th return in this data set. This code fits an AR(1)+GARCH(1,1) model to the last two years of data before Black Monday, assuming 253 trading days/year. The conditional distribution of the white noise is the t-distribution (called `"std"` in `ugarchspec()`). The code also plots the returns during these two years and on Black Monday. From the plot you can see that Black Monday was highly unusual. The parameter estimates are in `coef(fit)` and the sixth parameter is the degrees of freedom of the t-distribution. The `ugarchforecast()` function is used to predict one-step ahead, that is, to predict the return on Black Monday; the input variable `n.ahead` specifies how many days ahead to forecast, so `n.ahead=5` would forecast the next five days. The object `forecast` will contain `fitted(forecast)`, which is the conditional expected return on Black Monday, and `sigma(forecast)`, which is the conditional standard deviation of the return on Black Monday.

(a) Use the information above to calculate the conditional probability of a return less than or equal to -0.228 on Black Monday.

(b) Compute and plot the standardized residuals. Also plot the ACF of the standardized residuals and their squares. Include all three plots with your work. Do the standardized residuals indicate that the AR(1)+GARCH(1,1) model fits adequately?

(c) Would an AR(1)+ARCH(1) model provide an adequate fit?

(d) Does an AR(1) model with a Gaussian conditional distribution provide an adequate fit? Use the `arima()` function to fit the AR(1) model. This function only allows a Gaussian conditional distribution.

9. This problem uses monthly observations of the two-month yield, that is, Y_T with T equal to two months, in the data set `Irates` in the `Ecdat` package. The rates are log-transformed to stabilize the variance. To fit a GARCH model to the changes in the log rates, run the following `R` code.

```
library(rugarch)
library(Ecdat)
data(Irates)
```

```
r = as.numeric(log(Irates[,2]))
n = length(r)
lagr = r[1:(n-1)]
diffr = r[2:n] - lagr
spec = ugarchspec(mean.model=list(armaOrder=c(1,0)),
                  variance.model=list(garchOrder=c(1,1)),
                  distribution.model = "std")
fit = ugarchfit(data=diffr, spec=spec)
plot(fit, which="all")
```

(a) What model is being fit to the changes in `r`? Describe the model in detail.
(b) What are the estimates of the parameters of the model?
(c) What is the estimated ACF of Δr_t?
(d) What is the estimated ACF of a_t?
(e) What is the estimated ACF of a_t^2?

10. Consider the daily log returns on the S&P 500 index (GSPC). Begin by running the following commands in R, then answer the questions below for the series `y`.

```
library(rugarch)
library(quantmod)
getSymbols("^GSPC", from="2005-01-01", to="2014-12-31")
head(GSPC)
sp500 = xts( diff( log( GSPC[,6] ) )[-1] )
plot(sp500)
y = as.numeric(sp500)
```

(a) Is there any serial correlation in the log returns of S&P 500 index? Why?
(b) Is there any ARCH effect (evidence of conditional heteroskedasticity) in the log returns of S&P 500 index? Why?
(c) Specify and fit an ARCH model to the log returns of S&P 500 index. Write down the fitted model.
(d) Is your fitted ARCH model stationary? Why?
(e) Fit a GARCH(1,1) model for the log returns on the S&P 500 index using the Gaussian distribution for the innovations. Write down the fitted model.
(f) Perform model checking to ensure that the model is adequate using 20 lags in a Ljung-Box test of the standardized residuals and the squared standardized residuals.
(g) Is the fitted GARCH model stationary? Why?
(h) Make a Normal quantile plot for the standardized residuals. Use `qqnorm()` and `qqline()` in R. Is the Gaussian distribution appropriate for the standardized innovations?
(i) Plot the fitted conditional standard deviation process $\hat{\sigma}_t$ and comment.

(j) Calculate the 1–10 step ahead forecasts from the end of the series for both the process y_t and the conditional variance using the `ugarchforecast()` function.

References

Alexander, C. (2001) *Market Models: A Guide to Financial Data Analysis*, Wiley, Chichester.

Bauwens, L., Laurent, S., and Rombouts, J. V. (2006) Multivariate GARCH models: a survey. *Journal of Applied Econometrics*, **21**(1), 79–109.

Bera, A. K., and Higgins, M. L. (1993) A survey of Arch models. *Journal of Economic Surveys*, **7**, 305–366. [Reprinted in Jarrow (1998).]

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

Bollerslev, T. (1990) Modelling the coherence in short-run nominal exchange rates: a multivariate generalized ARCH model. *The Review of Economics and Statistics*, **72**(3), 498–505.

Bollerslev, T., Chou, R. Y., and Kroner, K. F. (1992) ARCH modelling in finance. *Journal of Econometrics*, **52**, 5–59. [Reprinted in Jarrow (1998)]

Bollerslev, T., Engle, R. F., and Nelson, D. B. (1994) ARCH models, In *Handbook of Econometrics, Vol IV*, Engle, R.F., and McFadden, D.L., Elsevier, Amsterdam.

Bollerslev, T., Engle, R. F., and Wooldridge, J. M. (1988) A capital asset pricing model with time-varying covariances. *Journal of Political Economy*, **96**, 116–131.

Carroll, R. J., and Ruppert, D. (1988) *Transformation and Weighting in Regression*, Chapman & Hall, New York.

Duan, J.-C. (1995) The GARCH option pricing model. *Mathematical Finance*, **5**, 13–32. [Reprinted in Jarrow (1998).]

Duan, J-C., and Simonato, J. G. (2001) American option pricing under GARCH by a Markov chain approximation. *Journal of Economic Dynamics and Control*, **25**, 1689–1718.

Enders, W. (2004) *Applied Econometric Time Series*, 2nd ed., Wiley, New York.

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

Engle, R. F. (2002) Dynamic conditional correlation: A simple class of multivariate generalized autoregressive conditional heteroskedasticity models. *Journal of Business & Economic Statistics*, **20**(3), 339–350.

Fisher, T.J., and Gallagher, C.M. (2012) New weighted portmanteau statistics for time series goodness of fit testing. *Journal of the American Statistical Association*, **107**(498), 777–787.

Gourieroux, C. and Jasiak, J. (2001) *Financial Econometrics*, Princeton University Press, Princeton, NJ.

Hamilton, J. D. (1994) *Time Series Analysis*, Princeton University Press, Princeton, NJ.

Heston, S. and Nandi, S. (2000) A closed form GARCH option pricing model. *The Review of Financial Studies*, **13**, 585–625.

Hsieh, K. C. and Ritchken, P. (2000) An empirical comparison of GARCH option pricing models. working paper.

Jarrow, R. (1998) *Volatility: New Estimation Techniques for Pricing Derivatives*, Risk Books, London. (This is a collection of articles, many on GARCH models or on stochastic volatility models, which are related to GARCH models.)

Li, W. K. (2003) *Diagnostic checks in time series*, CRC Press.

Matteson, D. S. and Tsay, R. S. (2011) Dynamic orthogonal components for multivariate time series. *Journal of the American Statistical Association*, **106**(496), 1450–1463.

Palm, F.C. (1996) GARCH models of volatility. *Handbook of Statistics*, **14**, 209–240.

Palma, W. and Zevallos, M. (2004). Analysis of the correlation structure of square time series. *Journal of Time Series Analysis*, **25**(4), 529–550.

Pindyck, R. S. and Rubinfeld, D. L. (1998) *Econometric Models and Economic Forecasts*, Irwin/McGraw Hill, Boston.

Ritchken, P. and Trevor, R. (1999) Pricing options under generalized GARCH and stochastic volatility processes. *Journal of Finance*, **54**, 377–402.

Rossi, P. E. (1996) *Modelling Stock Market Volatility*, Academic Press, San Diego.

Silvennoinen, A. and Teräsvirta, T (2009) Multivariate GARCH models. In *Handbook of Financial Time Series*, 201–229, Springer, Berlin.

Tsay, R. S. (2005) *Analysis of Financial Time Series*, 2nd ed., Wiley, New York.

Tse, Y. K. and Tsui, A. K. C. (2002) A multivariate generalized autoregressive conditional heteroscedasticity model with time-varying correlations. *Journal of Business & Economic Statistics*, **20**(3), 351–362.