18

GARCH Models

18.1 Introduction

As seen in earlier chapters, financial markets data often exhibit volatility clustering, where time series show periods of high volatility and periods of low volatility; see, for example, Figure 18.1. 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 Chapter 9, 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 AutoRegressive Conditional Heteroscedasticity. 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 ARCH(1) model, which is the simplest GARCH model and similar to an AR(1) model. Then we look at ARCH(p) models that are analogous to AR(p) models. Finally, we look at GARCH (Generalized ARCH) models that model conditional variances much as the conditional expectation is modeled by an ARMA model.

D. Ruppert, *Statistics and Data Analysis for Financial Engineering*, Springer Texts in Statistics, DOI 10.1007/978-1-4419-7787-8_18,

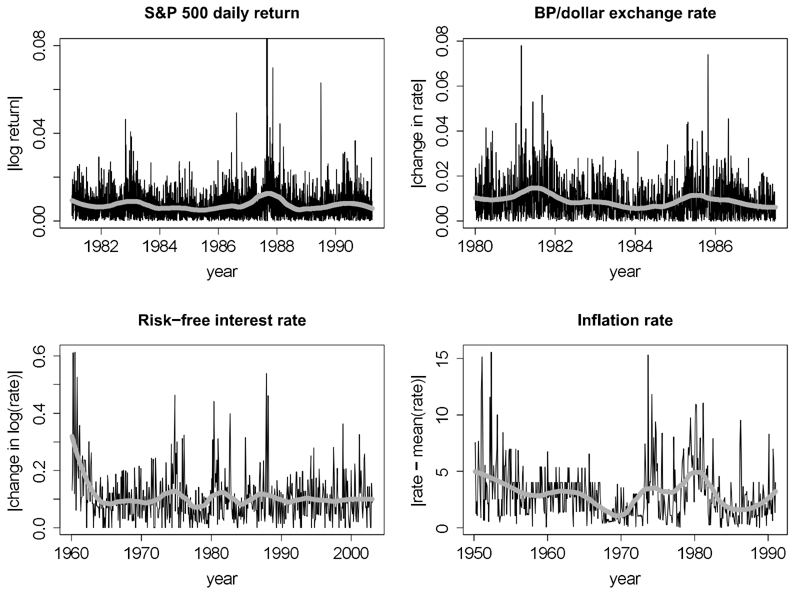

Fig. 18.1. *Examples of financial markets and economic data with time-varying volatility: (a) absolute values of S&P* 500 *log returns; (b) absolute values of changes in the BP/dollar exchange rate; (c) absolute values of changes in the log of the risk-free interest rate; (d) absolute deviations of the inflation rate from its mean. Loess (see Section* 21.2*) smooths have been added.*

18.2 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

$$Y_t = f(X_{1,t}, \ldots, X_{p,t}) + \epsilon_t, \tag{18.1}$$

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_ϵ^2. The function f is the conditional expectation of Y_t given $X_{1,t}, \ldots, X_{p,t}$. Moreover, the conditional variance of Y_t is σ_ϵ^2.

Equation (18.1) 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

$$Y_t = f(X_{1,t}, \ldots, X_{p,t}) + \sigma(X_{1,t}, \ldots, X_{p,t})\, \epsilon_t, \tag{18.2}$$

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.

18.3 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 independent white noise with a possibly nonnormal distribution, such as, a standardized t-distribution. Then

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

and

$$\mathrm{Var}(\epsilon_t | \epsilon_{t-1}, \ldots) = 1. \tag{18.3}$$

Property (18.3) is called *conditional homoskedasticity*.

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

$$a_t = \sqrt{\omega + \alpha_1 a_{t-1}^2}\,\epsilon_t, \tag{18.4}$$

which is a special case of (18.2) with f equal to 0 and σ equal to $\sqrt{\omega + \alpha_1 a_{t-1}^2}$. We require that $\omega > 0$ and $\alpha_1 \geq 0$ so that $\alpha_0 + \alpha_1 a_{t-1}^2 > 0$. It is also required that $\alpha_1 < 1$ in order for a_t to be stationary with a finite variance. Equation (18.4) can be written as

$$a_t^2 = (\omega + \alpha_1 a_{t-1}^2)\,\epsilon_t^2,$$

which is very much like 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.

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 $E(\epsilon_t^2) = \mathrm{Var}(\epsilon_t) = 1$,

$$E(a_t | a_{t-1}, \ldots) = 0, \tag{18.5}$$

and

$$
\begin{aligned}
\sigma_t^2 &= E\left\{\left(\omega + \alpha_1 a_{t-1}^2\right)\epsilon_t^2 | a_{t-1}, a_{t-2}, \ldots\right\} \\
&= (\omega + \alpha_1 a_{t-1}^2) E\left\{\epsilon_t^2 | a_{t-1}, a_{t-2}, \ldots\right\} \\
&= \alpha_0 + \alpha_1 a_{t-1}^2. \qquad (18.6)
\end{aligned}
$$

Equation (18.6) 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 deviation that makes σ_{t+1}^2 large so that a_{t+1} tends to be large 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 $\alpha_1 < 1$, 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 (18.6), which give us

$$\gamma_a(0) = \omega + \alpha_1 \gamma_a(0).$$

This equation has a positive solution if $\alpha_1 < 1$:

$$\gamma_a(0) = \omega/(1 - \alpha_1).$$

If $\alpha_1 = 1$, then $\gamma_a(0)$ is infinite, but a_t is stationary nonetheless and is called an integrated GARCH model (I-GARCH) process.

Straightforward calculations using (18.5) show that the ACF of a_t is

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

In fact, any process such that 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 the GARCH processes, where 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 uncorrelated, the process a_t^2 has a more interesting ACF: if $\alpha_1 < 1$, then

$$\rho_{a^2}(h) = \alpha_1^{|h|}, \quad \forall\ h.$$

If $\alpha_1 \geq 1$, then a_t^2 either is nonstationary or has an infinite variance, so it does not have an ACF.

Example 18.1. A simulated ARCH(1) process

A simulated ARCH(1) process is shown in Figure 18.2. Panel (a) shows the i.i.d. white noise process, ϵ_t, (b) shows $\sigma_t = \sqrt{1 + 0.95 a_{t-1}^2}$, the conditional standard deviation process, (c) shows $a_t = \sigma_t \epsilon_t$, the ARCH(1) process. As discussed in the next 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 variance of a_t is $\gamma_a(0) = 1/(1 - 0.95) = 20$, so the standard deviation is $\sqrt{20} = 4.47$. Panels (e)–(h) are ACF plots of the ARCH and AR/ARCH processes and squared processes. Notice that for the ARCH process, the process is uncorrelated but the squared process has correlation. The processes were all started at 0 and simulated for 100 observations. The first 10 observations were treated as a burn-in period and discarded.

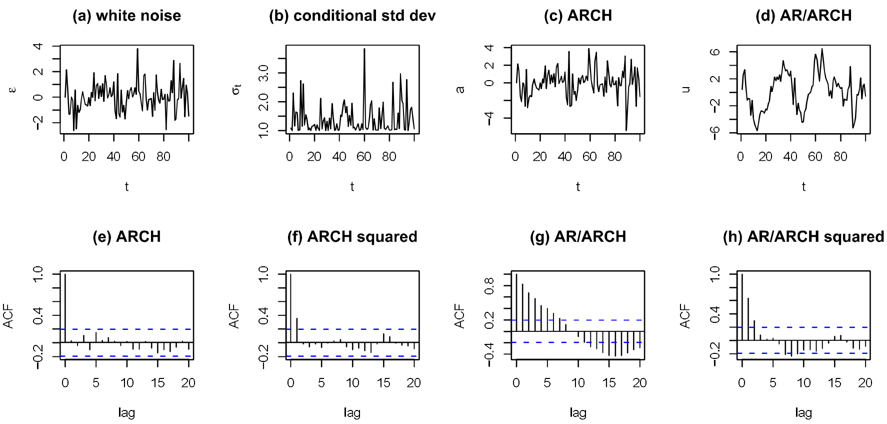

Fig. 18.2. *Simulation of* 100 *observations from an ARCH(*1*) process and an AR(*1*)/ARCH(*1*) process. The parameters are* $\omega = 1$, $\alpha_1 = 0.95$, $\mu = 0.1$, *and* $\phi = 0.8$.

□

18.4 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 Section 18.6. In this section we combine an AR(1) model with an ARCH(1) model.

Let a_t be an ARCH(1) process so that $a_t = \sqrt{\omega + \alpha_1 a_{t-1}^2}\epsilon_t$, where ϵ_t is i.i.d. $N(0,1)$, and suppose that

$$u_t - \mu = \phi(u_{t-1} - \mu) + a_t.$$

The process u_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, a_t has the same ACF as independent white noise and therefore u_t has the same ACF as an AR(1) process with independent white noise:

$$\rho_u(h) = \phi^{|h|} \quad \forall\ h.$$

Moreover, a_t^2 has the ARCH(1) ACF:

$$\rho_{a^2}(h) = \alpha_1^{|h|} \quad \forall\ h.$$

We need to assume that both $|\phi| < 1$ and $\alpha_1 < 1$ in order for u to be stationary with a finite variance. Of course, $\omega > 0$ and $\alpha_1 \geq 0$ are also assumed.

The process u_t is such that its conditional mean and variance, given the past, are both nonconstan, so a wide variety of time series can be modeled.

Example 18.2. Simulated AR(1)/ARCH(1) process

A simulation of an AR(1)/ARCH(1) process is shown in panel (d) of Figure 18.2 and the ACFs of the process and the squared process are in panels (g) and (h). Notice that both ACFs show autocorrelation.

□

18.5 ARCH(p) Models

As before, let ϵ_t be Gaussian white noise with unit variance. Then a_t is an ARCH(q) 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(q) 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 is the same as the ACF of an AR(q) process; see Section 18.9.

18.6 ARIMA(p_A, d, q_A)/GARCH(p_G, q_G) Models

A deficiency of ARCH(q) 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 Figure 18.2(b). GARCH models permit a wider range of behavior, in particular, more persistent volatility. The GARCH(p, q) model is

$$a_t = \sigma_t \epsilon_t,$$

where

$$\sigma_t = \sqrt{\omega + \sum_{i=1}^{p} \alpha_i a_{t-i}^2 + \sum_{i=1}^{q} \beta_i \sigma_{t-i}^2}. \tag{18.7}$$

Because past values of the σ_t process are fed back into the present value, the conditional standard deviation can exhibit more persistent periods of high or low volatility than seen in an ARCH process. The process a_t is uncorrelated with a stationary mean and variance and a_t^2 has an ACF like an ARMA process (see Section 18.9). 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_G, q_G) and uses a_t as the noise term in an ARIMA(p_A, d, q_A) model. The subscripts on p and q distinguish between the GARCH (G) and ARIMA (A) parameters. We will call such a model an ARIMA(p_A, d, q_A)/GARCH(p_G, q_G) model.

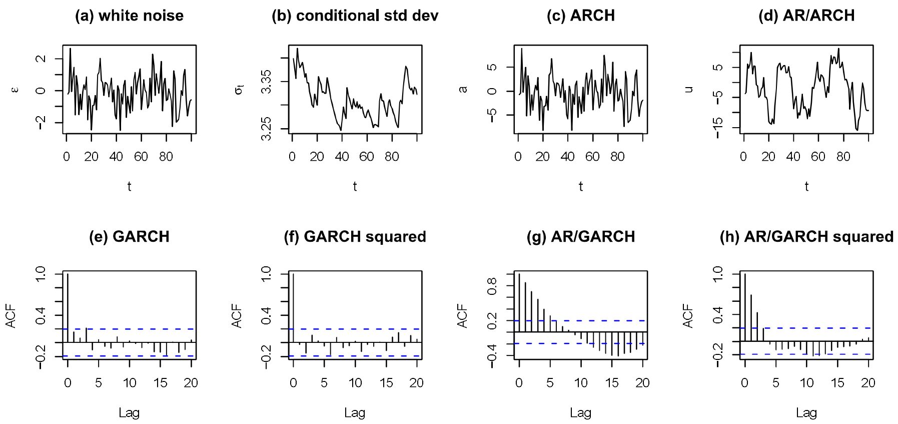

Fig. 18.3. *Simulation of GARCH(1, 1) and AR(1)/GARCH(1, 1) processes. The parameters are $\omega = 1$, $\alpha_1 = 0.08$, $\beta_1 = 0.9$, and $\phi = 0.8$.*

Figure 18.3 is a simulation of 100 observations from a GARCH(1,1) process and from a AR(1)/GARCH(1,1) process. The GARCH parameters are $\omega = 1$, $\alpha_1 = 0.08$, and $\beta_1 = 0.9$. The large value of β_1 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 Figure 18.2.

18.6.1 Residuals for ARIMA(p_A, d, q_A)/GARCH(p_G, q_G) Models

When one fits an ARIMA(p_A, d, q_A)/GARCH(p_G, q_G) 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 divided by its 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.

The $\widehat{a}_t$ are the residuals of the ARIMA process and are used when forecasting by the methods in Section 9.12.

18.7 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 Section 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 markets 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.

18.8 Fitting ARMA/GARCH Models

Example 18.3. AR(1)/GARCH(1,1) model fit to BMW returns

This example uses the BMW daily log returns. An AR(1)/GARCH(1,1) model was fit to these returns using R's `garchFit` function in the `fGarch` package. Although garchFit allows the white noise to have a nonGaussian distribution, in this example we specified Gaussian white noise (the default). The results include

```
Call:  garchFit(formula = ~arma(1, 0) + garch(1, 1), data = bmw,
     cond.dist = "norm")

Mean and Variance Equation:
 data ~ arma(1, 0) + garch(1, 1)
 [data = bmw]

Conditional Distribution: norm

Coefficient(s):
        mu         ar1       omega      alpha1       beta1
4.0092e-04  9.8596e-02  8.9043e-06  1.0210e-01  8.5944e-01

Std. Errors: based on Hessian

Error Analysis:
        Estimate  Std. Error  t value Pr(>|t|)
mu     4.009e-04   1.579e-04    2.539   0.0111 *
ar1    9.860e-02   1.431e-02    6.888 5.65e-12 ***
omega  8.904e-06   1.449e-06    6.145 7.97e-10 ***
alpha1 1.021e-01   1.135e-02    8.994  < 2e-16 ***
beta1  8.594e-01   1.581e-02   54.348  < 2e-16 ***
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1   1

Log Likelihood: 17757    normalized:  2.89

Information Criterion Statistics:
  AIC   BIC   SIC  HQIC
-5.78 -5.77 -5.78 -5.77
```

In the output, ϕ is denoted by ar1, the mean is mean, and ω is called omega. Note that $\widehat{\phi} = 0.0986$ and is statistically significant, implying that this is a small amount of positive autocorrelation. Both α_1 and β_1 are highly significant and $\widehat{\beta}_1 = 0.859$, which implies rather persistent volatility clustering. There are two additional information criteria reported, SIC (Schwarz's information criterion) and HQIC (Hannan–Quinn information criterion). These are less widely used compared to AIC and BIC and will not be discussed here.[1]

[1] To make matters even more confusing, some authors use SIC as a synonym for BIC, since BIC is due to Schwarz. Also, the term SBIC (Schwarz's Bayesian information criterion) is used in the literature, sometimes as a synonym for BIC and SIC and sometimes as a third criterion. Moreover, BIC does not mean the same thing to all authors. We will not step any further into this quagmire. For-

In the output from `garchFit`, the normalized log-likelihood is the log-likelihood divided by n. The AIC and BIC values have also 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 output also included the following tests applied to the standardized residuals and squared residuals:

```
Standardised Residuals Tests:
                                Statistic p-Value
 Jarque-Bera Test   R    Chi^2  11378     0
 Ljung-Box Test     R    Q(10)  15.2      0.126
 Ljung-Box Test     R    Q(15)  20.1      0.168
 Ljung-Box Test     R    Q(20)  30.5      0.0614
 Ljung-Box Test     R^2  Q(10)  5.03      0.889
 Ljung-Box Test     R^2  Q(15)  7.54      0.94
 Ljung-Box Test     R^2  Q(20)  9.28      0.98
 LM Arch Test       R    TR^2   6.03      0.914
```

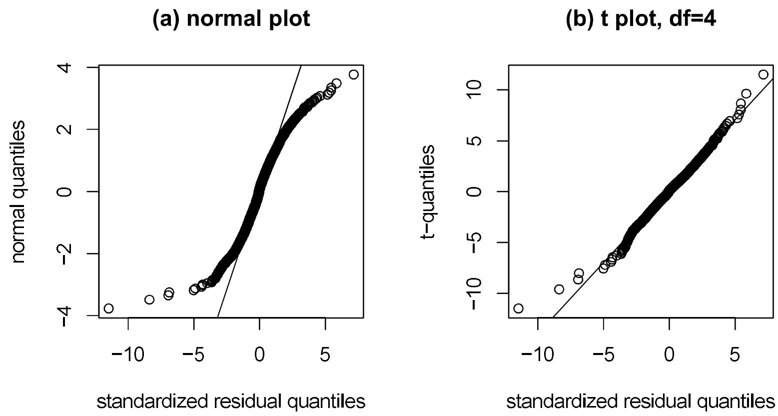

Fig. 18.4. *QQ plots of standardized residuals from an AR(1)/GARCH(1,1) fit to daily BMW log returns. The reference lines go through the first and third quartiles.*

The Jarque–Bera test of normality strongly rejects the null hypothesis that the white noise innovation process $\{\epsilon_t\}$ is Gaussian. Figure 18.4 shows two QQ plots of the standardized residuals, a normal plot and a t-plot with 4 df.

tunately, the various versions of BIC, SIC, and SBIC are similar. In this book, BIC is always defined by (5.30) and `garchFit` uses this definition of BIC as well.

The latter 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-model would be suitable for the white noise.

The Ljung–Box tests with an `R` in the second column are applied to the residuals (here `R` = residuals, not the `R` software), while the Ljung–Box tests with `R^2` are applied to the squared residuals. None of the tests is significant, which indicates that the model fits the data well, except for the nonnormality of the $\{\epsilon_t\}$ noted earlier. The nonsignificant `LM Arch Test` indicates the same.

A t-distribution was fit to the standardized residuals by maximum likelihood using `R`'s `fitdistr` function. The MLE of the degrees-of-freedom parameter was 4.1. This confirms the good fit by this distribution seen in Figure 18.4. The AR(1)/GARCH(1,1) model was refit assuming t-distributed errors, so `cond.dist = "std"`, with the following results:

```
Call:
 garchFit(formula = ~arma(1, 1) + garch(1, 1), data = bmw,
    cond.dist = "std")

Mean and Variance Equation:
 data ~ arma(1, 1) + garch(1, 1) [data = bmw]

Conditional Distribution: std

Coefficient(s):
         mu          ar1          ma1        omega       alpha1
 1.7358e-04  -2.9869e-01   3.6896e-01   6.0525e-06   9.2924e-02
        beta1       shape
   8.8688e-01 4.0461e+00

Std. Errors: based on Hessian

Error Analysis:
         Estimate  Std. Error  t value Pr(>|t|)
mu      1.736e-04   1.855e-04    0.936  0.34929
ar1    -2.987e-01   1.370e-01   -2.180  0.02924 *
ma1     3.690e-01   1.345e-01    2.743  0.00608 **
omega   6.052e-06   1.344e-06    4.502 6.72e-06 ***
alpha1  9.292e-02   1.312e-02    7.080 1.44e-12 ***
beta1   8.869e-01   1.542e-02   57.529  < 2e-16 ***
shape   4.046e+00   2.315e-01   17.480  < 2e-16 ***
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1   1

Log Likelihood:
 18159    normalized:  2.9547

Standardised Residuals Tests:
                                Statistic p-Value
```

```
 Jarque-Bera Test   R    Chi^2  13355     0
 Shapiro-Wilk Test  R    W      NA        NA
 Ljung-Box Test     R    Q(10)  21.933    0.015452
 Ljung-Box Test     R    Q(15)  26.501    0.033077
 Ljung-Box Test     R    Q(20)  36.79     0.012400
 Ljung-Box Test     R^2  Q(10)  5.8285    0.82946
 Ljung-Box Test     R^2  Q(15)  8.0907    0.9201
 Ljung-Box Test     R^2  Q(20)  10.733    0.95285
 LM Arch Test       R    TR^2   7.009     0.85701

Information Criterion Statistics:
    AIC     BIC     SIC    HQIC
-5.9071 -5.8994 -5.9071 -5.9044
```

The 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.

□

18.9 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

$$E_{t-1}(a_t^2) = \sigma_t^2 = \omega + \alpha_1 a_{t-1}^2 + \beta_1 \sigma_{t-1}^2, \tag{18.8}$$

where E_{t-1} is the conditional expectation given the information set at time $t-1$. Define $\eta_t = a_t^2 - \sigma_t^2$. Since $E_{t-1}(\eta_t) = E_{t-1}(a_t^2) - \sigma_t^2 = 0$, 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

$$\sigma_t^2 = \omega + (\alpha_1 + \beta_1) a_{t-1}^2 - \beta_1 \eta_{t-1} \tag{18.9}$$

and therefore

$$a_t^2 = \sigma_t^2 + \eta_t = \omega + (\alpha_1 + \beta_1) a_{t-1}^2 - \beta_1 \eta_{t-1} + \eta_t. \tag{18.10}$$

Assume that $\alpha_1 + \beta_1 < 1$. If $\mu = \omega / \{1 - (\alpha_1 + \beta_1)\}$, then

$$a_t^2 - \mu = (\alpha_1 + \beta_1)(a_{t-1}^2 - \mu) + \beta_1 \eta_{t-1} + \eta_t. \tag{18.11}$$

From (18.11) one sees that a_t^2 is an ARMA(1,1) process with mean μ. Using the notation of (9.25), the AR(1) coefficient is $\phi_1 = \alpha_1 + \beta_1$ and the MA(1) coefficient is $\theta_1 = -\beta_1$.

For the general case, assume that σ_t follows (18.7) so that

$$\sigma_t^2 = \omega + \sum_{i=1}^{p} \alpha_i a_{t-i}^2 + \sum_{i=1}^{q} \beta_i \sigma_{t-i}^2 . \tag{18.12}$$

Assume also that $p \le q$—this assumption causes no loss of generality because, if $q > p$, then we can increase p to equal q by defining $\alpha_i = 0$ for $i = p{+}1, \ldots, q$. Define $\mu = \omega/\{1 - \sum_{i=1}^{p}(\alpha_i + \beta_i)\}$. Straightforward algebra similar to the GARCH(1,1) case shows that

$$a_t^2 - \mu = \sum_{i=1}^{p} (\alpha_i + \beta_i)(a_{t-i}^2 - \mu) - \sum_{i=1}^{q} \beta_i \eta_{t-i} + \eta_t, \tag{18.13}$$

so that a_t^2 is an ARMA(p, q) process with mean μ. As a byproduct of these calculations, we obtain a necessary condition for a_t to be stationary:

$$\sum_{i=1}^{p} (\alpha_i + \beta_i) < 1. \tag{18.14}$$

18.10 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 (9.31) and (9.32). After some algebra, one finds that

$$\rho_{a^2}(1) = \frac{\alpha_1(1 - \alpha_1\beta_1 - \beta_1^2)}{1 - 2\alpha_1\beta_1 - \beta_1^2} \tag{18.15}$$

and

$$\rho_{a^2}(k) = (\alpha_1 + \beta_1)^{k-1} \rho_{a^2}(1), \quad k \ge 2. \tag{18.16}$$

By (18.15), there are infinitely many values of (α_1, β_1) with the same value of $\rho_{a^2}(1)$. By (18.16), a higher value of $\alpha_1 + \beta_1$ means a slower decay of ρ_{a^2} after the first lag. This behavior is illustrated in Figure 18.5, 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 $\alpha_1 + \beta_1$ and the ACF decays very slowly. The dotted curve is a pure AR(1) process and has the most rapid decay.

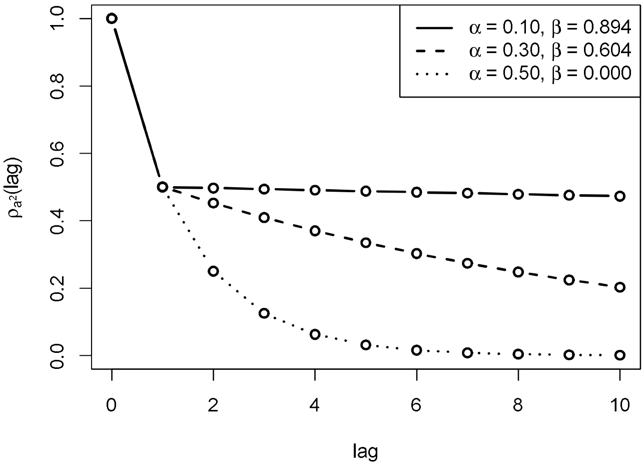

Fig. 18.5. *ACFs of three GARCH(1,1) processes with* $\rho_{a^2}(1) = 0.5$.

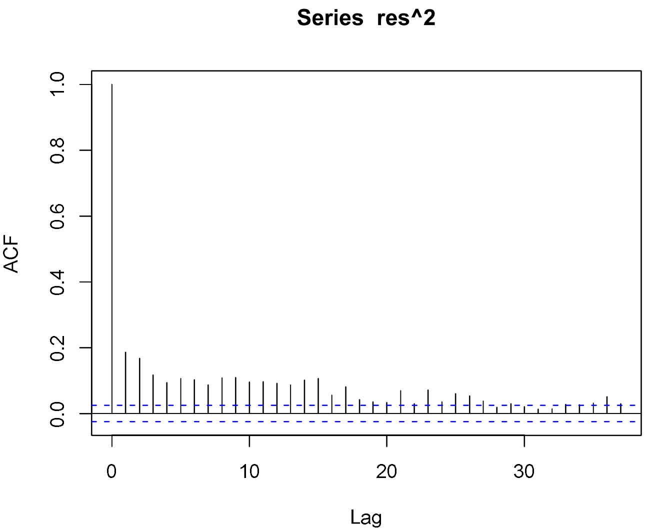

Fig. 18.6. *ACF of the squared residuals from an AR(1) fit to the BMW log returns.*

In Example 18.3, an AR(1)/GARCH(1,1) model was fit to the BMW daily log returns. The GARCH parameters were estimated to be $\widehat{\alpha}_1 = 0.10$ and $\widehat{\beta}_1 = 0.86$. By (18.15) the $\widehat{\rho}_{a^2}(1) = 0.197$ for this process and the high value of $\widehat{\beta}_1$ suggests slow decay. The sample ACF of the squared residuals [from an AR(1) model] is plotted in Figure 18.6. 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.

18.11 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 *leverage effect*. Standard GARCH models, that is, the models given by (18.7), 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

$$\sigma_t^\delta = \omega + \sum_{i=1}^{p} \alpha_i (|a_{t-1}| - \gamma_i a_{t-1})^\delta + \sum_{j=1}^{q} \beta_j \sigma_{t-j}^\delta, \tag{18.17}$$

where $\delta > 0$ and $-1 < \gamma_j < 1$, $j = 1, \ldots, p$. Note that $\delta = 2$ and $\gamma_1 = \cdots = \gamma_p = 0$ 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 18.7 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 18.4. AR(1)/APARCH(1,1) fit to BMW returns

In this example, an AR(1)/APARCH(1,1) model with t-distributed errors is fit to the BMW log returns. The output from `garchFit` is below. The

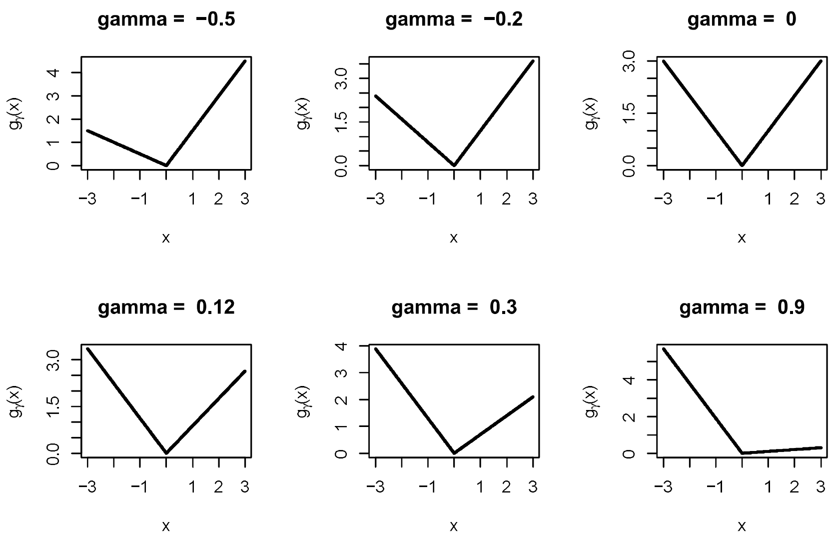

Fig. 18.7. *Plots of $g_\gamma(x)$ for various values of γ.*

estimate of δ is 1.46 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 0.0045, 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 Figure 18.7 with $\gamma = 0.12$. The leverage might not be of practical importance.

```
Call:
 garchFit(formula = ~arma(1, 0) + aparch(1, 1), data = bmw,
    cond.dist = "std",  include.delta = T)

Mean and Variance Equation:
 data ~ arma(1, 0) + aparch(1, 1)
 [data = bmw]

Conditional Distribution:
 std

Coefficient(s):
        mu          ar1        omega       alpha1       gamma1
4.1696e-05   6.3761e-02   5.4746e-05   1.0050e-01   1.1998e-01


       beta1      delta        shape
  8.9817e-011.4585e+00   4.0665e+00
```

```
Std. Errors:
 based on Hessian

Error Analysis:
        Estimate  Std. Error  t value Pr(>|t|)
mu     4.170e-05   1.377e-04    0.303  0.76208
ar1    6.376e-02   1.237e-02    5.155 2.53e-07 ***
omega  5.475e-05   1.230e-05    4.452 8.50e-06 ***
alpha1 1.005e-01   1.275e-02    7.881 3.33e-15 ***
gamma1 1.200e-01   4.498e-02    2.668  0.00764 **
beta1  8.982e-01   1.357e-02   66.171  < 2e-16 ***
delta  1.459e+00   1.434e-01   10.169  < 2e-16 ***
shape  4.066e+00   2.344e-01   17.348  < 2e-16 ***
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1   1

Log Likelihood:
 18166    normalized:  2.9557

Description:
 Sat Dec 06 09:11:54 2008 by user: DavidR


Standardised Residuals Tests:
                                Statistic p-Value
 Jarque-Bera Test   R    Chi^2  10267     0
 Shapiro-Wilk Test  R    W      NA        NA
 Ljung-Box Test     R    Q(10)  24.076    0.0074015
 Ljung-Box Test     R    Q(15)  28.868    0.016726
 Ljung-Box Test     R    Q(20)  38.111    0.0085838
 Ljung-Box Test     R^2  Q(10)  8.083     0.62072
 Ljung-Box Test     R^2  Q(15)  9.8609    0.8284
 Ljung-Box Test     R^2  Q(20)  13.061    0.87474
 LM Arch Test       R    TR^2   9.8951    0.62516

Information Criterion Statistics:
    AIC     BIC     SIC    HQIC
-5.9088 -5.9001 -5.9088 -5.9058
```

As mentioned earlier, in the output from `garchFit`, the normalized log-likelihood is the log-likelihood divided by n. The AIC and BIC values have also been normalized by dividing by n, though this is not noted in the output.

The normalized BIC for this model (-5.9001) is very nearly the same as the normalized BIC for the GARCH model with t-distributed errors (-5.8994), but after multiplying by $n = 6146$, the difference in the BIC values is 4.30. The difference between the two normalized AIC values, -5.9088 and -5.9071, is even larger, 10.4, 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.

`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.

□

18.12 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 Section 14.1. The model there was

$$Y_i = \beta_0 + \beta_1 X_{i,1} + \cdots + \beta_p X_{i,p} + \epsilon_i, \tag{18.18}$$

where

$$(1 - \phi_1 B - \cdots - \phi_p B^p)(\epsilon_t - \mu) = (1 + \theta_1 B + \ldots + \theta_q B^q)u_t, \tag{18.19}$$

and $\{u_t\}$ is i.i.d. white noise. This model is good as far as it goes, but it does not accommodate volatility clustering, which is often found in the residuals. Therefore, we will now assume that, instead of being i.i.d. white noise, $\{u_t\}$ is a GARCH process so that

$$u_t = \sigma_t v_t, \tag{18.20}$$

where

$$\sigma_t = \sqrt{\omega + \sum_{i=1}^{p} \alpha_i u_{t-i}^2 + \sum_{i=1}^{q} \beta_i \sigma_{t-i}^2}, \tag{18.21}$$

and $\{v_t\}$ is i.i.d. white noise. The model given by (18.18)–(18.21) is a *linear regression model with ARMA/GARCH disturbances*.

Some software can fit the linear regression model with ARMA/GARCH disturbances in one step. If such software is not available, then a three-step estimation method is the following:

1. estimate the parameters in (18.18) by ordinary least-squares;
2. fit model (18.19)–(18.21) to the ordinary least-squares residuals;
3. reestimate the parameters in (18.18) by weighted least-squares with weights equal to the reciprocals of the conditional variances from step 2.

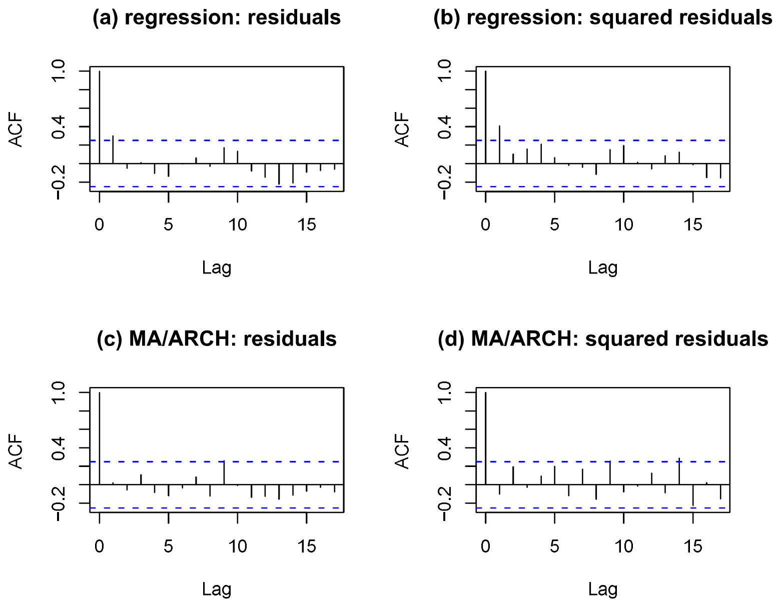

Fig. 18.8. *(a) ACF of the externally studentized residuals from a linear model and (b) their squared values. (c) ACF of the residuals from an MA(1)/ARCH(1) fit to the regression residuals and (d) their squared values.*

Example 18.5. Regression analysis with ARMA/GARCH errors of the Nelson–Plosser data

In Example 12.9, we saw that a parsimonious model for the yearly log returns on the stock index used `diff(log(ip))` and `diff(bnd)` as predictors. Figure 18.8 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 the residuals and certainly a GARCH effect. `R`'s `auto.arima` selected an ARIMA(0,0,1) model for the residuals.

Next an MA(1)/ARCH(1) model was fit to the regression model's raw residuals with the following results:

```
Call:
 garchFit(formula = ~arma(0, 1) + garch(1, 0),
    data = residuals(fit_lm2))

Mean and Variance Equation:
 data ~ arma(0, 1) + garch(1, 0)
 [data = residuals(fit_lm2)]
```

```
Conditional Distribution: norm

Error Analysis:
         Estimate  Std. Error   t value Pr(>|t|)
mu     -2.527e-17   2.685e-02 -9.41e-16  1.00000
ma1     3.280e-01   1.602e-01     2.048  0.04059 *
omega   1.400e-02   4.403e-03     3.180  0.00147 **
alpha1  2.457e-01   2.317e-01     1.060  0.28897
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1   1

Log Likelihood:
 36    normalized:  0.59

Standardised Residuals Tests:
                                Statistic p-Value
 Jarque-Bera Test   R    Chi^2  0.72      0.7
 Shapiro-Wilk Test  R    W      0.99      0.89
 Ljung-Box Test     R    Q(10)  14        0.18
 Ljung-Box Test     R    Q(15)  25        0.054
 Ljung-Box Test     R    Q(20)  28        0.12
 Ljung-Box Test     R^2  Q(10)  11        0.35
 Ljung-Box Test     R^2  Q(15)  18        0.26
 Ljung-Box Test     R^2  Q(20)  25        0.21
 LM Arch Test       R    TR^2   11        0.5

Information Criterion Statistics:
 AIC  BIC  SIC HQIC
-1.0 -0.9 -1.1 -1.0
```

ACF plots of the standardized residuals from the MA(1)/ARCH(1) model are in Figure 18.8(c) and (d). One sees essentially no short-term autocorrelation in the ARMA/GARCH standardized residuals or squared standardized residuals, which indicates that the ARMA/GARCH model fits 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.

Next, 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.

```
Call:
lm(formula = diff(log(sp)) ~ diff(log(ip)) + diff(bnd),
  data = new_np, weights = 1/nelploss.garch@sigma.t^2)

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)     0.0281     0.0202    1.39   0.1685
diff(log(ip))   0.5785     0.1672    3.46   0.0010 **
```

```
diff(bnd)      -0.1172     0.0580   -2.02   0.0480 *
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1   1

Residual standard error: 1.1 on 58 degrees of freedom
Multiple R-squared: 0.246,      Adjusted R-squared: 0.22
F-statistic: 9.46 on 2 and 58 DF,  p-value: 0.000278
```

There are no striking differences between these results and the unweighted fit in Example 12.9. The main reason for using the GARCH model for the residuals would be in providing more accurate prediction intervals if the model were to be used for forecasting; see Section 18.13.

□

18.13 Forecasting ARMA/GARCH Processes

Forecasting ARMA/GARCH processes is in one way similar to forecasting ARMA processes—the forecasts 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 a 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 k-steps ahead forecasting are $\mu \pm z_{\alpha/2}\sigma_{n+k|n}$ where $\sigma_{n+k|n}$ is the conditional standard deviation of Y_{n+k} given the information available at time n. As k increases, $\sigma_{n+k|n}$ 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 18.6. Forecasting BMW log returns

In this example, we will return to the BMW log returns used in several earlier examples. We have seen in Example 18.3 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.

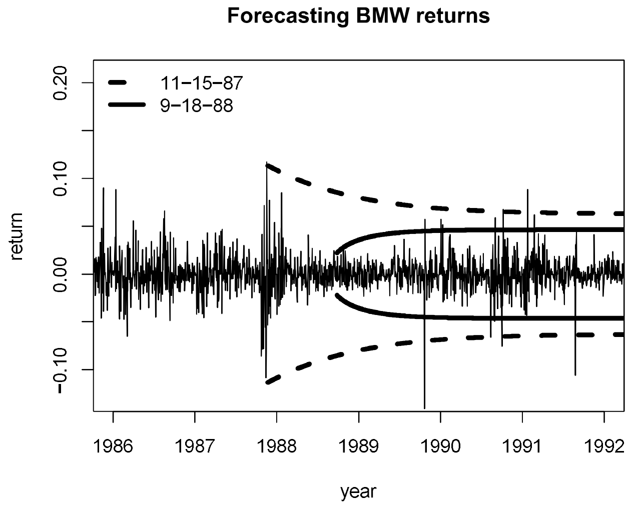

Fig. 18.9. *Prediction limits for forecasting BMW log returns at two time origins.*

Figure 18.9 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. In this example, they do not quite converge to each other because the estimates of σ differ between the two time origins. $\square$

18.14 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.

18.15 References

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

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., and Engle, R. F. (1993) Common persistence in conditional variances. *Econometrica*, **61**, 167–186.

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., and Ng, V. (1993) Measuring and testing the impact of news on volatility. *Journal of Finance*, **4**, 47–59.

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.)

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.

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

18.16 R Lab

18.16.1 Fitting GARCH Models

Run the following code to load the data set `Tbrate`, 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.

```
data(Tbrate,package="Ecdat")
library(tseries)
library(fGarch)
#  r = the 91-day treasury bill rate
#  y = the log of real GDP
#  pi = the inflation rate
Tbill = Tbrate[,1]
Del.Tbill = diff(Tbill)
```

Problem 1 *Plot both* `Tbill` *and* `Del.Tbill`*. 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* `Del.Tbill` *series?*

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

```
garch.model.Tbill = garchFit(formula= ~arma(1,0)  + garch(1,0),Tbill)
summary(garch.model.Tbill)
garch.model.Tbill@fit$matcoef
```

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 = residuals(garch.model.Tbill)
res_std = res / garch.model.Tbill@sigma.t
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) What is contained in the the variable `garch.model.Tbill@sigma.t`*?*
(e) Is there anything noteworthy in the plot produced by the code `plot(res_std)`*?*

Problem 4 *Now find an ARMA/GARCH model for the series* `del.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?*

18.17 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 a WN$(0, 1)$ process, that

$$a_t = \epsilon_t \sqrt{1 + 0.35a_{t-1}^2},$$

and that

$$u_t = 3 + 0.72u_{t-1} + a_t.$$

(a) Find the mean of u_t.
(b) Find the variance of u_t.
(c) Find the autocorrelation function of u_t.
(d) Find the autocorrelation function of a_t^2.

4. Let u_t be the AR(1)/ARCH(1) model

$$a_t = \epsilon_t \sqrt{\omega + \alpha_1 \, a_{t-1}^2},$$
$$(u_t - \mu) = \phi(u_{t-1} - \mu) + a_t,$$

where ϵ_t is WN(0,1). Suppose that $\mu = 0.4$, $\phi = 0.45$, $\omega = 1$, and $\alpha_1 = 0.3$.
(a) Find $E(u_2|u_1 = 1, u_0 = 0.2)$.
(b) Find $\text{Var}(u_2|u_1 = 1, u_0 = 0.2)$.

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

$$Y_t = \beta_0 + \beta_1 X_t + \delta\sigma_t + a_t, \tag{18.22}$$

where

$$a_t = \epsilon_t \sqrt{1 + 0.5 a_{t-1}^2}. \tag{18.23}$$

Here the ϵ_t are independent $N(0,1)$ random variables. Model (18.22)–(18.23) 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 u_t are stationary processes such that

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

and

$$u_t = 2 + 0.6 u_{t-1} + a_t.$$

(a) What type of process is a_t?
(b) What type of process is u_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 u_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(Ecdat)
library(fGarch)
data(SP500,package="Ecdat")
returnBlMon = SP500$r500[1805]
x = SP500$r500[(1804-2*253+1):1804]
plot(c(x,returnBlMon))
results = garchFit(~arma(1,0)+garch(1,1),data=x,cond.dist="std")
dfhat = as.numeric(results@fit$par[6])
forecast = predict(results,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 `garchFit`). 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 `results@fit$par` and the sixth parameter is the degrees of freedom of the t-distribution. The `predict` 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 `meanForecast`, which is the conditional expected return on Black Monday, `meanError`, which you should ignore, and `standardDeviation`, 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? (Warning: If you apply the function `summary` to an `fGarch` object, the AIC value reported has been normalized by division by the sample size. You need to multiply by the sample size to get AIC.)
(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(fGarch)
library(Ecdat)
data(Irates)
r = as.numeric(log(Irates[,2]))
n = length(r)
lagr = r[1:(n-1)]
diffr = r[2:n] - lagr
garchFit(~arma(1,0)+garch(1,1),data=diffr, cond.dist = "std")
```

(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?

19

Risk Management

19.1 The Need for Risk Management

The financial world has always been risky, and financial innovations such as the development of derivatives markets and the packaging of mortgages have now made risk management more important than ever but also more difficult.

There are many different types of risk. *Market risk* is due to changes in prices. *Credit risk* is the danger that a counterparty does not meet contractual obligations, for example, that interest or principal on a bond is not paid. *Liquidity risk* is the potential extra cost of liquidating a position because buyers are difficult to locate. *Operational risk* is due to fraud, mismanagement, human errors, and similar problems.

Early attempts to measure risk such as duration analysis, discussed in Section 3.8.1 and used to estimate the market risk of fixed income securities, were somewhat primitive and of only limited applicability. In contrast, value-at-risk (VaR) and expected shortfall (ES) are widely used because they can be applied to all types of risks and securities, including complex portfolios.

VaR uses two parameters, the time horizon and the confidence level, which are denoted by T and $1 - \alpha$, respectively. Given these, the VaR is a bound such that the loss over the horizon is less than this bound with probability equal to the confidence coefficient. For example, if the horizon is one week, the confidence coefficient is 99% (so $\alpha = 0.01$), and the VaR is \$5 million, then there is only a 1% chance of a loss exceeding \$5 million over the next week. We sometimes use the notation $\mathrm{VaR}(\alpha)$ or $\mathrm{Var}(\alpha, T)$ to indicate the dependence of VaR on α or on both α and the horizon T. Usually, $\mathrm{VaR}(\alpha)$ is used with T being understood.

If $\mathcal{L}$ is the loss over the holding period T, then $\mathrm{VaR}(\alpha)$ is the αth upper quantile of $\mathcal{L}$. Equivalently, if $\mathcal{R} = -\mathcal{L}$ is the revenue, then $\mathrm{VaR}(\alpha)$ is minus the αth quantile of $\mathcal{R}$. For continuous loss distributions, $\mathrm{VaR}(\alpha)$ solves

$$P\{\mathcal{L} > \mathrm{VaR}(\alpha)\} = P\{\mathcal{L} \geq \mathrm{VaR}(\alpha)\} = \alpha, \tag{19.1}$$

and for any loss distribution, continuous or not,

D. Ruppert, *Statistics and Data Analysis for Financial Engineering*, Springer Texts in Statistics,
DOI 10.1007/978-1-4419-7787-8_19,