

CHAPTER 12

TIME SERIES MODELS OF HETEROSCEDASTICITY

The models discussed so far concern the conditional mean structure of time series data. However, more recently, there has been much work on modeling the conditional variance structure of time series data—mainly motivated by the needs for financial modeling. Let $\{Y_t\}$ be a time series of interest. The conditional variance of Y_t given the past Y values, Y_{t-1}, Y_{t-2}, \dots , measures the uncertainty in the deviation of Y_t from its conditional mean $E(Y_t | Y_{t-1}, Y_{t-2}, \dots)$. If $\{Y_t\}$ follows some ARIMA model, the (one-step-ahead) conditional variance is always equal to the noise variance for any present and past values of the process. Indeed, the constancy of the conditional variance is true for predictions of any fixed number of steps ahead for an ARIMA process. In practice, the (one-step-ahead) conditional variance may vary with the current and past values of the process, and, as such, the conditional variance is itself a random process, often referred to as the conditional variance process. For example, daily returns of stocks are often observed to have larger conditional variance following a period of violent price movement than a relatively stable period. The development of models for the conditional variance process with which we can predict the variability of future values based on current and past data is the main concern of the present chapter. In contrast, the ARIMA models studied in earlier chapters focus on how to predict the conditional mean of future values based on current and past data.

In finance, the conditional variance of the return of a financial asset is often adopted as a measure of the risk of the asset. This is a key component in the mathematical theory of pricing a financial asset and the *VaR* (Value at Risk) calculations; see, for example, Tsay (2005). In an efficient market, the expected return (conditional mean) should be zero, and hence the return series should be white noise. Such series have the simplest autocorrelation structure. Thus, for ease of exposition, we shall assume in the first few sections of this chapter that the data are returns of some financial asset and are white noise; that is, serially uncorrelated data. By doing so, we can concentrate initially on studying how to model the conditional variance structure of a time series. By the end of the chapter, we discuss some simple schemes for simultaneously modeling the conditional mean and conditional variance structure by combining an ARIMA model with a model of conditional heteroscedasticity.

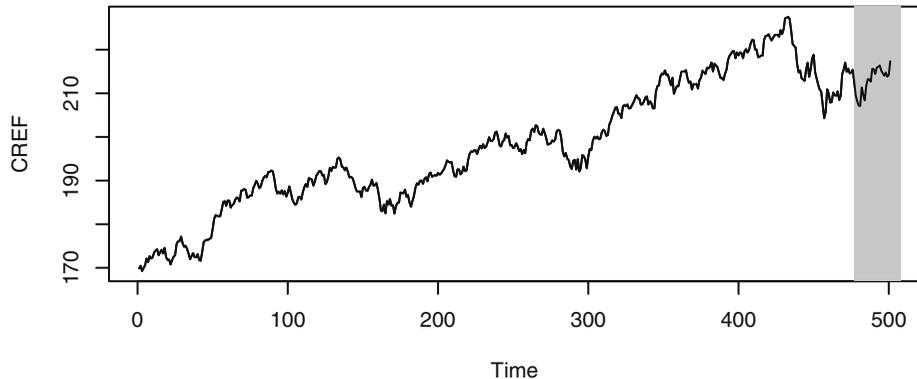
12.1 Some Common Features of Financial Time Series

As an example of financial time series, we consider the daily values of a unit of the CREF stock fund over the period from August 26, 2004 to August 15, 2006. The CREF stock fund is a fund of several thousand stocks and is not openly traded in the stock market.[†] Since stocks are not traded over weekends or on holidays, only on so-called *trading days*, the CREF data do not change over weekends and holidays. For simplicity, we will analyze the data as if they were equally spaced. Exhibit 12.1 shows the time series plot of the CREF data. It shows a generally increasing trend with a hint of higher variability with higher level of the stock value. Let $\{p_t\}$ be the time series of, say, the daily price of some financial asset. The (continuously compounded) return on the t th day is defined as

$$r_t = \log(p_t) - \log(p_{t-1}) \quad (12.1.1)$$

Sometimes the returns are then multiplied by 100 so that they can be interpreted as percentage changes in the price. The multiplication may also reduce numerical errors as the raw returns could be very small numbers and render large rounding errors in some calculations.

Exhibit 12.1 Daily CREF Stock Values: August 26, 2004 to August 15, 2006



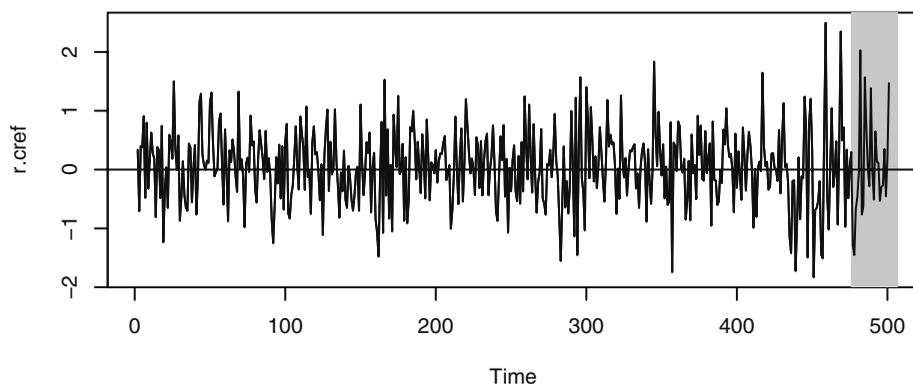
```
> win.graph(width=4.875,height=2.5,pointsize=8)
> data(CREF); plot(CREF)
```

Exhibit 12.2 plots the CREF return series (sample size = 500). The plot shows that the returns were more volatile over some time periods and became very volatile toward the end of the study period. This observation may be more clearly seen by plotting the time sequence plot of the absolute or squared returns; see Exercise 12.1, page 316.

[†] CREF stands for College Retirement Equities Fund—a group of stock and bond funds crucial to many college faculty retirement plans.

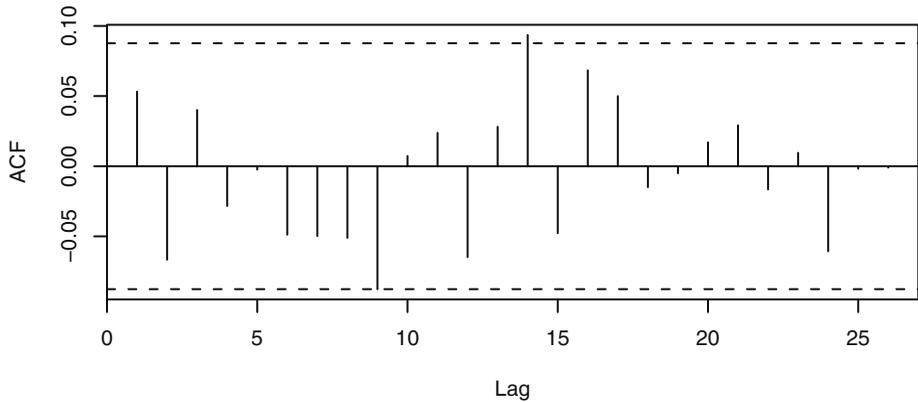
These results might be triggered by the instability in the Middle East due to a war in southern Lebanon from July 12 to August 14, 2006, the period that is shaded in gray in Exhibits 12.1 and 12.2. This pattern of alternating quiet and volatile periods of substantial duration is referred to as *volatility clustering* in the literature. Volatility in a time series refers to the phenomenon where the conditional variance of the time series varies over time. The study of the dynamical pattern in the volatility of a time series (that is, the conditional variance process of the time series) constitutes the main subject of this chapter.

Exhibit 12.2 Daily CREF Stock Returns: August 26, 2004 to August 15, 2006

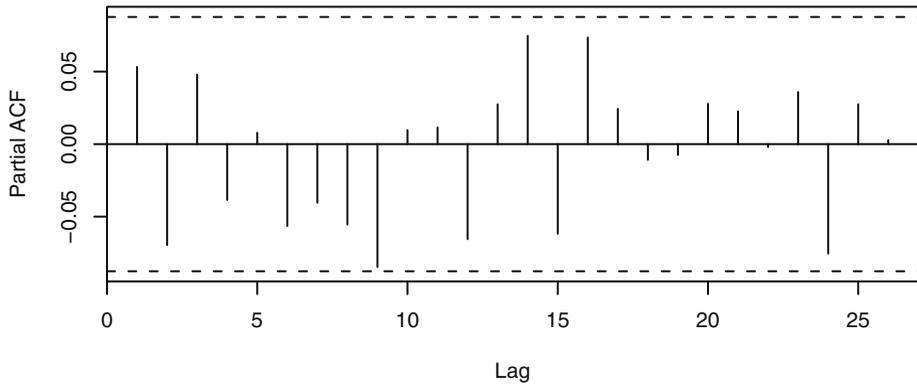


```
> r.cref=diff(log(CREF))*100
> plot(r.cref); abline(h=0)
```

The sample ACF and PACF of the daily CREF returns (multiplied by 100), shown in Exhibits 12.3 and 12.4, suggest that the returns have little serial correlation at all. The sample EACF (not shown) also suggests that a white noise model is appropriate for these data. The average CREF return equals 0.0493 with a standard error of 0.02885. Thus the mean of the return process is not statistically significantly different from zero. This is expected based on the efficient-market hypothesis alluded to in the introduction to this chapter.

Exhibit 12.3 Sample ACF of Daily CREF Returns: 8/26/04 to 8/15/06

```
> acf(r.cref)
```

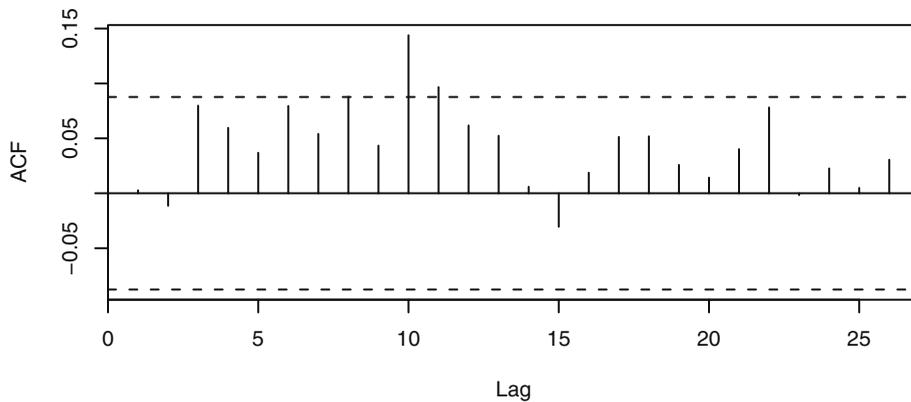
Exhibit 12.4 Sample PACF of Daily CREF Returns: 8/26/04 to 8/15/06

```
> pacf(r.cref)
```

However, the volatility clustering observed in the CREF return data gives us a hint that they may not be independently and identically distributed—otherwise the variance would be constant over time. This is the first occasion in our study of time series models where we need to distinguish between series values being uncorrelated and series values being independent. If series values are truly independent, then nonlinear instantaneous transformations such as taking logarithms, absolute values, or squaring preserves independence. However, the same is not true of correlation, as correlation is only a measure of linear dependence. Higher-order serial dependence structure in data can be explored by studying the autocorrelation structure of the absolute returns (of lesser sampling vari-

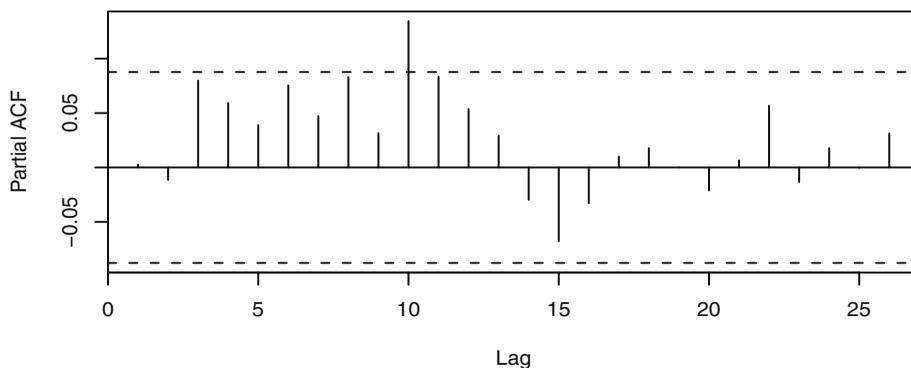
ability with less mathematical tractability) or that of the squared returns (of greater sampling variability but with more manageability in terms of statistical theory). If the returns are independently and identically distributed, then so are the absolute returns (as are the squared returns), and hence they will be white noise as well. Hence, if the absolute or squared returns admit some significant autocorrelations, then these autocorrelations furnish some evidence against the hypothesis that the returns are independently and identically distributed. Indeed, the sample ACF and PACF of the absolute returns and those of the squared returns in Exhibits 12.5 through 12.8 display some significant autocorrelations and hence provide some evidence that the daily CREF returns are not independently and identically distributed.

Exhibit 12.5 Sample ACF of the Absolute Daily CREF Returns

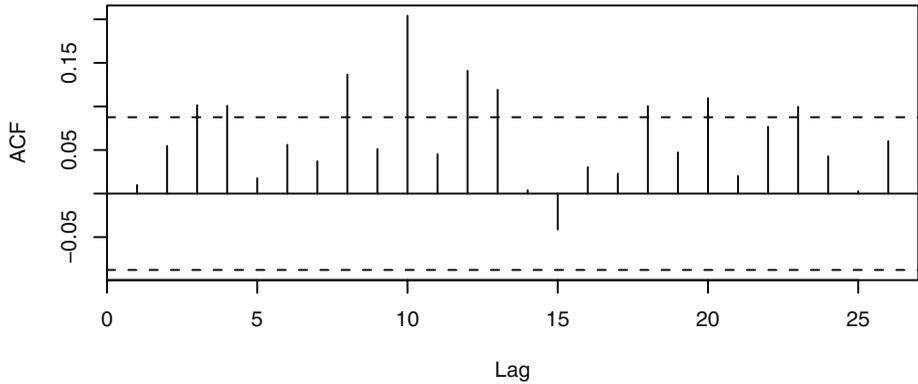


```
> acf(abs(r.cref))
```

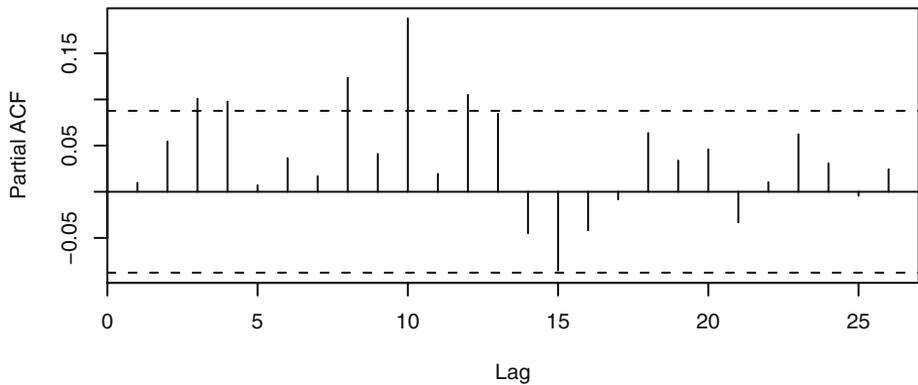
Exhibit 12.6 Sample PACF of the Absolute Daily CREF Returns



```
> pacf(abs(r.cref))
```

Exhibit 12.7 Sample ACF of the Squared Daily CREF Returns

```
> acf(r.cref^2)
```

Exhibit 12.8 Sample PACF of the Squared Daily CREF Returns

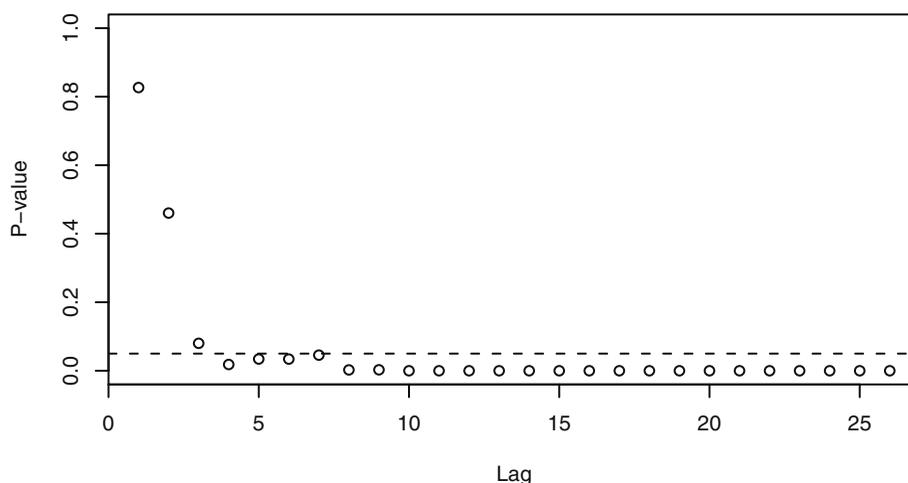
```
> pacf(r.cref^2)
```

These visual tools are often supplemented by formally testing whether the squared data are autocorrelated using the Box-Ljung test. Because no model fitting is required, the degrees of freedom of the approximating chi-square distribution for the Box-Ljung statistic equals the number of correlations used in the test. Hence, if we use m autocorrelations of the squared data in the test, the test statistic is approximately chi-square distributed with m degrees of freedom, if there is no ARCH. This approach can be extended to the case when the conditional mean of the process is non-zero and if an ARMA model is adequate in describing the autocorrelation structure of the data. In which case, the first m autocorrelations of the squared residuals from this model can be used to test for the presence of ARCH. The corresponding Box-Ljung statistic will have a

chi-square distribution with m degrees of freedom under the assumption of no ARCH effect, see McLeod and Li (1983) and Li(2004). Below, we shall refer to the test for ARCH effects using the Box-Ljung statistic with the squared residuals or data as the McLeod- Li test.

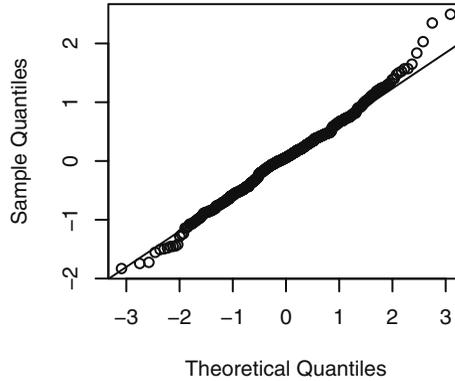
In practice, it is useful to apply the McLeod-Li test for ARCH using a number of lags and plot the p -values of the test. Exhibit 12.9 shows that the McLeod-Li tests are all significant at the 5% significance level when more than 3 lags are included in the test. This is broadly consistent with the visual pattern in Exhibit 12.7 and formally shows strong evidence for ARCH in this data.

Exhibit 12.9 McLeod-Li Test Statistics for Daily CREF Returns



```
> win.graph(width=4.875, height=3,pointsize=8)
> McLeod.Li.test (y=r.cref)
```

The distributional shape of the CREF returns can be explored by constructing a QQ normal scores plot—see Exhibit 12.10. The QQ plot suggests that the distribution of returns may have a tail thicker than that of a normal distribution and may be somewhat skewed to the right. Indeed, the Shapiro-Wilk test statistic for testing normality equals 0.9932 with p -value equal to 0.024, and hence we reject the normality hypothesis at the usual significance levels.

Exhibit 12.10 QQ Normal Plot of Daily CREF Returns

```
> win.graph(width=2.5,height=2.5,pointsize=8)
> qqnorm(r.cref); qqline(r.cref)
```

The *skewness* of a random variable, say Y , is defined by $E(Y-\mu)^3/\sigma^3$, where μ and σ are the mean and standard deviation of Y , respectively. It can be estimated by the sample skewness

$$g_1 = \frac{\sum_{i=1}^n (Y_i - \bar{Y})^3}{(n\hat{\sigma}^3)} \quad (12.1.2)$$

where $\hat{\sigma}^2 = \Sigma(Y_i - \bar{Y})^2/n$ is the sample variance. The sample skewness of the CREF returns equals 0.116. The thickness of the tail of a distribution relative to that of a normal distribution is often measured by the (excess) *kurtosis*, defined as $E(Y - \mu)^4/\sigma^4 - 3$. For normal distributions, the kurtosis is always equal to zero. A distribution with positive kurtosis is called a *heavy-tailed* distribution, whereas it is called *light-tailed* if its kurtosis is negative. The kurtosis can be estimated by the sample kurtosis

$$g_2 = \frac{\sum_{i=1}^n (Y_i - \bar{Y})^4}{(n\hat{\sigma}^4)} - 3 \quad (12.1.3)$$

The sample kurtosis of the CREF returns equals 0.6274. An alternative definition of kurtosis modifies the formula and uses $E(r_t - \mu)^4/\sigma^4$; that is, it does not subtract three from the ratio. We shall always use the former definition for kurtosis.

Another test for normality is the Jarque-Bera test, which is based on the fact that a normal distribution has zero skewness and zero kurtosis. Assuming independently and identically distributed data Y_1, Y_2, \dots, Y_n , the Jarque-Bera test statistic is defined as

$$JB = \frac{ng_1^2}{6} + \frac{ng_2^2}{24} \quad (12.1.4)$$

where g_1 is the sample skewness and g_2 is the sample kurtosis. Under the null hypothesis of normality, the Jarque-Bera test statistic is approximately distributed as χ^2 with two degrees of freedom. In fact, under the normality assumption, each summand defining the Jarque-Bera statistic is approximately χ^2 with 1 degree of freedom. The Jarque-Bera test rejects the normality assumption if the test statistic is too large. For the CREF returns, $JB = 500 \times 0.116^2/6 + 500 \times 0.6274^2/24 = 1.12 + 8.20 = 9.32$ with a p -value equal to 0.011. Recall that the upper 5 percentage point of a χ^2 distribution with unit degree of freedom equals 3.84. Hence, the data appear not to be skewed but do have a relatively heavy tail. In particular, the normality assumption is inconsistent with the CREF return data—a conclusion that is also consistent with the finding of the Shapiro-Wilk test.

In summary, the CREF return data are found to be serially uncorrelated but admit a higher-order dependence structure, namely volatility clustering, and a heavy-tailed distribution. It is commonly observed that such characteristics are rather prevalent among financial time series data. The GARCH models introduced in the next sections attempt to provide a framework for modeling and analyzing time series that display some of these characteristics.

12.2 The ARCH(1) Model

Engle (1982) first proposed the autoregressive conditional heteroscedasticity (ARCH) model for modeling the changing variance of a time series. As discussed in the previous section, the return series of a financial asset, say $\{r_t\}$, is often a serially uncorrelated sequence with zero mean, even as it exhibits volatility clustering. This suggests that the conditional variance of r_t given past returns is not constant. The conditional variance, also referred to as the *conditional volatility*, of r_t will be denoted by $\sigma_{t|t-1}^2$, with the subscript $t-1$ signifying that the conditioning is upon returns through time $t-1$. When r_t is available, the squared return r_t^2 provides an unbiased estimator of $\sigma_{t|t-1}^2$. A series of large squared returns may foretell a relatively volatile period. Conversely, a series of small squared returns may foretell a relatively quiet period. The ARCH model is formally a regression model with the conditional volatility as the response variable and the past lags of the squared return as the covariates. For example, the ARCH(1) model assumes that the return series $\{r_t\}$ is generated as follows:

$$r_t = \sigma_{t|t-1} \varepsilon_t \quad (12.2.1)$$

$$\sigma_{t|t-1}^2 = \omega + \alpha r_{t-1}^2 \quad (12.2.2)$$

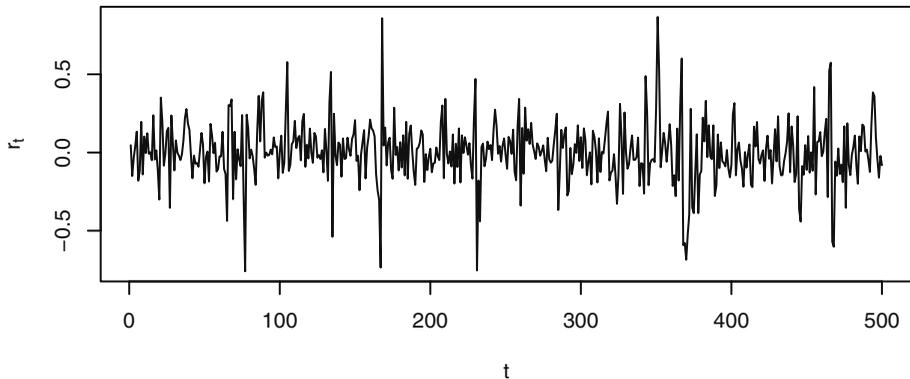
where α and ω are unknown parameters, $\{\varepsilon_t\}$ is a sequence of *independently and identically distributed* random variables each with zero mean and unit variance (also known as the *innovations*), and ε_t is independent of r_{t-j} , $j = 1, 2, \dots$. The innovation ε_t is presumed to have unit variance so that the conditional variance of r_t equals $\sigma_{t|t-1}^2$. This follows from

$$\begin{aligned}
E(r_t^2 | r_{t-j}, j = 1, 2, \dots) &= E(\sigma_{t|t-1}^2 \varepsilon_t^2 | r_{t-j}, j = 1, 2, \dots) \\
&= \sigma_{t|t-1}^2 E(\varepsilon_t^2 | r_{t-j}, j = 1, 2, \dots) \\
&= \sigma_{t|t-1}^2 E(\varepsilon_t^2) \\
&= \sigma_{t|t-1}^2
\end{aligned} \tag{12.2.3}$$

The second equality follows because $\sigma_{t|t-1}$ is known given the past returns, the third equality holds because ε_t is independent of past returns, and the last equality results from the assumption that the variance of ε_t equals 1.

Exhibit 12.11 shows the time series plot of a simulated series of size 500 from an ARCH(1) model with $\omega = 0.01$ and $\alpha = 0.9$. Volatility clustering is evident in the data as larger fluctuations cluster together, although the series is able to recover from large fluctuations quickly because of the very short memory in the conditional variance process.[†]

Exhibit 12.11 Simulated ARCH(1) Model with $\omega = 0.01$ and $\alpha_1 = 0.9$



```

> set.seed(1235678); library(tseries)
> garch01.sim=garch.sim(alpha=c(.01,.9),n=500)
> plot(garch01.sim,type='l',ylab=expression(r[t]), xlab='t')

```

While the ARCH model resembles a regression model, the fact that the conditional variance is not directly observable (and hence is called a latent variable) introduces some subtlety in the use of ARCH models in data analysis. For example, it is not obvious how to explore the regression relationship graphically. To do so, it is pertinent to replace the conditional variance by some observable in Equation (12.2.2). Let

[†] The R package named `tseries` is required for this chapter. We assume that the reader has downloaded and installed it.

$$\eta_t = r_t^2 - \sigma_{t|t-1}^2 \quad (12.2.4)$$

It can be verified that $\{\eta_t\}$ is a serially uncorrelated series with zero mean. Moreover, η_t is uncorrelated with past returns. Substituting $\sigma_{t|t-1}^2 = r_t^2 - \eta_t$ into Equation (12.2.2) it is obvious that

$$r_t^2 = \omega + \alpha r_{t-1}^2 + \eta_t \quad (12.2.5)$$

Thus, the squared return series satisfies an AR(1) model under the assumption of an ARCH(1) model for the return series! Based on this useful observation, an ARCH(1) model may be specified if an AR(1) specification for the squared returns is warranted by techniques learned from earlier chapters.

Besides its value in terms of data analysis, the deduced AR(1) model for the squared returns can be exploited to gain theoretical insights on the parameterization of the ARCH model. For example, because the squared returns must be nonnegative, it makes sense to always restrict the parameters ω and α to be nonnegative. Also, if the return series is stationary with variance σ^2 , then taking expectation on both sides of Equation (12.2.5) yields

$$\sigma^2 = \omega + \alpha \sigma^2 \quad (12.2.6)$$

That is, $\sigma^2 = \omega / (1 - \alpha)$ and hence $0 \leq \alpha < 1$. Indeed, it can be shown (Ling and McAleer, 2002) that the condition $0 \leq \alpha < 1$ is necessary and sufficient for the (weak) stationarity of the ARCH(1) model. At first sight, it seems that the concepts of stationarity and conditional heteroscedasticity may be incompatible. However, recall that weak stationarity of a process requires that the mean of the process be constant and the covariance of the process at any two epochs be finite and identical whenever the lags of the two epochs are the same. In particular, the variance is constant for a weakly stationary process. The condition $0 \leq \alpha < 1$ implies that there exists an initial distribution for r_0 such that r_t defined by Equations (12.2.1) and (12.2.2) for $t \geq 1$ is weakly stationary in the sense above. It is interesting to observe that weak stationarity does not preclude the possibility of a nonconstant conditional variance process, as is the case for the ARCH(1) model! It can be checked that the ARCH(1) process is white noise. Hence, it is an example of a white noise that admits a nonconstant conditional variance process as defined by Equation (12.2.2) that varies with the lag one of the squared process.

A satisfying feature of the ARCH(1) model is that, even if the innovation η_t has a normal distribution, the stationary distribution of an ARCH(1) model with $1 > \alpha > 0$ has fat tails; that is, its kurtosis, $E(r_t^4) / \sigma^4 - 3$, is greater than zero. (Recall that the kurtosis of a normal distribution is always equal to 0, and a distribution with positive kurtosis is said to be fat-tailed, while one with a negative kurtosis is called a light-tailed distribution.) To see the validity of this claim, consider the case where the $\{\varepsilon_t\}$ are independently and identically distributed as standard normal variables. Raising both sides of Equation (12.2.1) on page 285 to the fourth power and taking expectations gives

$$\begin{aligned}
E(r_t^4) &= E[E(\sigma_{t|t-1}^4 \varepsilon_t^4 | r_{t-j}, j = 1, 2, 3, \dots)] \\
&= E[\sigma_{t|t-1}^4 E(\varepsilon_t^4 | r_{t-j}, j = 1, 2, 3, \dots)] \\
&= E[\sigma_{t|t-1}^4 E(\varepsilon_t^4)] \\
&= 3E(\sigma_{t|t-1}^4)
\end{aligned} \tag{12.2.7}$$

The first equality follows from the iterated-expectation formula, which, in the simple case of two random variables X, Y , states that $E[E(X|Y)] = E(X)$. [See Equation (9.E.5) on page 218 for a review.] The second equality results from the fact that $\sigma_{t|t-1}$ is known given past returns. The third equality is a result of the independence between ε_t and past returns, and the final equality follows from the normality assumption. It remains to calculate $E(\sigma_{t|t-1}^4)$. Now, it is unclear whether the preceding expectation exists as a finite number. For the moment, assume it does and, assuming stationarity, let it be denoted by τ . Below, we shall derive a condition for this assumption to be valid. Raising both sides of Equation (12.2.2) to the second power and taking expectation yields

$$\tau = \omega^2 + 2\omega\alpha\sigma^2 + \alpha^2 3\tau \tag{12.2.8}$$

which implies

$$\tau = \frac{\omega^2 + 2\omega\alpha\sigma^2}{1 - 3\alpha^2} \tag{12.2.9}$$

This equality shows that a necessary (and, in fact, also sufficient) condition for the finiteness of τ is that $0 \leq \alpha < 1/\sqrt{3}$, in which case the ARCH(1) process has finite fourth moment. Incidentally, this shows that a stationary ARCH(1) model need not have finite fourth moments. The existence of finite higher moments will further restrict the parameter range—a feature also shared by higher-order analogues of the ARCH model and its variants. Returning to the calculation of the kurtosis of an ARCH(1) process, it can be shown by tedious algebra that Equation (12.2.1) implies that $\tau > \sigma^4$ and hence $E(r_t^4) > 3\sigma^4$. Thus the kurtosis of a stationary ARCH(1) process is greater than zero. This verifies our earlier statement that an ARCH(1) process has fat tails even with normal innovations. In other words, the fat tail is a result of the volatility clustering as specified by Equation (12.2.2).

A main use of the ARCH model is to predict the future conditional variances. For example, one might be interested in forecasting the h -step-ahead conditional variance

$$\sigma_{t+h|t}^2 = E(r_{t+h}^2 | r_t, r_{t-1}, \dots) \tag{12.2.10}$$

For $h = 1$, the ARCH(1) model implies that

$$\sigma_{t+1|t}^2 = \omega + \alpha r_t^2 = (1 - \alpha)\sigma^2 + \alpha r_t^2 \tag{12.2.11}$$

which is a weighted average of the long-run variance and the current squared return. Similarly, using the iterated expectation formula, we have

$$\begin{aligned}
\sigma_{t+h|t}^2 &= E(r_{t+h}^2 | r_p, r_{t-1}, \dots) \\
&= E[E(\sigma_{t+h|t+h-1}^2 \varepsilon_{t+h}^2 | r_{t+h-1}, r_{t+h-2}, \dots) | r_p, r_{t-1}, \dots] \\
&= E[\sigma_{t+h|t+h-1}^2 E(\varepsilon_{t+h}^2) | r_p, r_{t-1}, \dots] \\
&= E(\sigma_{t+h|t+h-1}^2 | r_p, r_{t-1}, \dots) \\
&= \omega + \alpha E(r_{t+h-1}^2 | r_p, r_{t-1}, \dots) \\
&= \omega + \alpha \sigma_{t+h-1|t}^2
\end{aligned} \tag{12.2.12}$$

where we adopt the convention that $\sigma_{t+h|t}^2 = r_{t+h}^2$ for $h < 0$. The formula above provides a recursive recipe for computing the h -step-ahead conditional variance.

12.3 GARCH Models

The forecasting formulas derived in the previous section show both the strengths and weaknesses of an ARCH(1) model, as the forecasting of the future conditional variances only involves the most recent squared return. In practice, one may expect that the accuracy of forecasting may improve by including all past squared returns with lesser weight for more distant volatilities. One approach is to include further lagged squared returns in the model. The ARCH(q) model, proposed by Engle (1982), generalizes Equation (12.2.2) on page 285, by specifying that

$$\sigma_{t|t-1}^2 = \omega + \alpha_1 r_{t-1}^2 + \alpha_2 r_{t-2}^2 + \dots + \alpha_q r_{t-q}^2 \tag{12.3.1}$$

Here, q is referred to as the ARCH order. Another approach, proposed by Bollerslev (1986) and Taylor (1986), introduces p lags of the conditional variance in the model, where p is referred to as the GARCH order. The combined model is called the generalized autoregressive conditional heteroscedasticity, GARCH(p, q), model.

$$\begin{aligned}
\sigma_{t|t-1}^2 &= \omega + \beta_1 \sigma_{t-1|t-2}^2 + \dots + \beta_p \sigma_{t-p|t-p-1}^2 + \alpha_1 r_{t-1}^2 \\
&\quad + \alpha_2 r_{t-2}^2 + \dots + \alpha_q r_{t-q}^2
\end{aligned} \tag{12.3.2}$$

In terms of the backshift B notation, the model can be expressed as

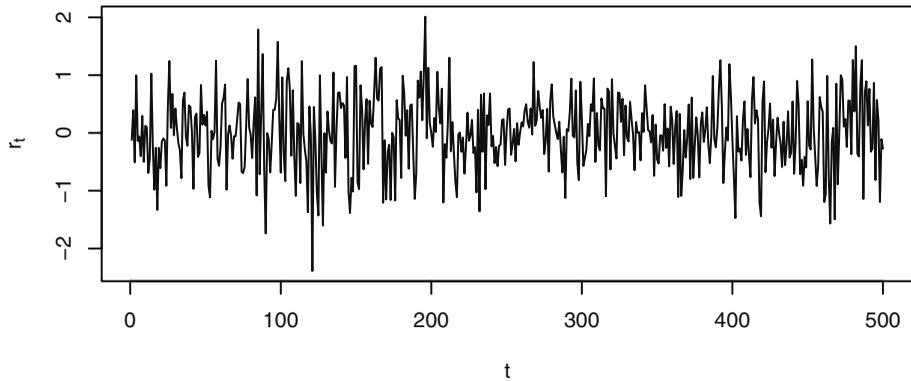
$$(1 - \beta_1 B - \dots - \beta_p B^p) \sigma_{t|t-1}^2 = \omega + (\alpha_1 B + \dots + \alpha_q B^q) r_t^2 \tag{12.3.3}$$

We note that in some of the literature, the notation GARCH(p, q) is written as GARCH(q, p); that is, the orders are switched. It can be rather confusing but true that the two different sets of conventions are used in different software! A reader must find out which convention is used by the software on hand before fitting or interpreting a GARCH model.

Because conditional variances must be nonnegative, the coefficients in a GARCH model are often constrained to be nonnegative. However, the nonnegative parameter constraints are not necessary for a GARCH model to have nonnegative conditional variances with probability 1; see Nelson and Cao (1992) and Tsai and Chan (2006). Allowing the parameter values to be negative may increase the dynamical patterns that can be captured by the GARCH model. We shall return to this issue later. Henceforth, within this section, we shall assume the nonnegative constraint for the GARCH parameters.

Exhibit 12.12 shows the time series plot of a time series, of size 500, simulated from a GARCH(1,1) model with standard normal innovations and parameter values $\omega = 0.02$, $\alpha = 0.05$, and $\beta = 0.9$. Volatility clustering is evident in the plot, as large (small) fluctuations are usually succeeded by large (small) fluctuations. Moreover, the inclusion of the lag 1 of the conditional variance in the model successfully enhances the memory in the volatility.

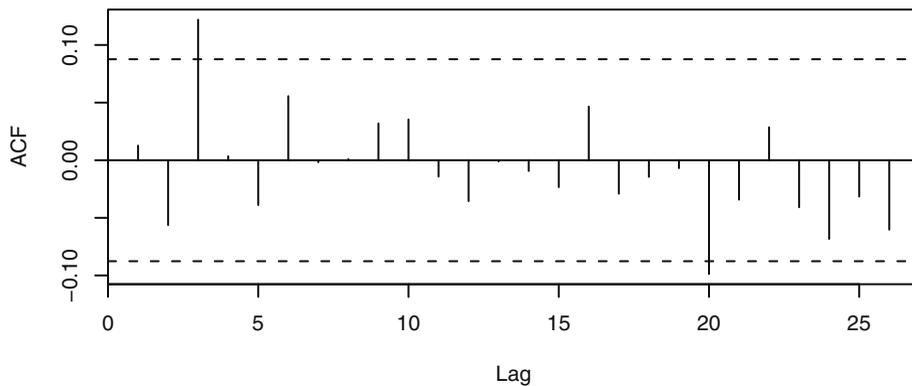
Exhibit 12.12 Simulated GARCH(1,1) Process



```
> set.seed(1234567)
> garch11.sim=garch.sim(alpha=c(0.02,0.05),beta=.9,n=500)
> plot(garch11.sim,type='l',ylab=expression(r[t]), xlab='t')
```

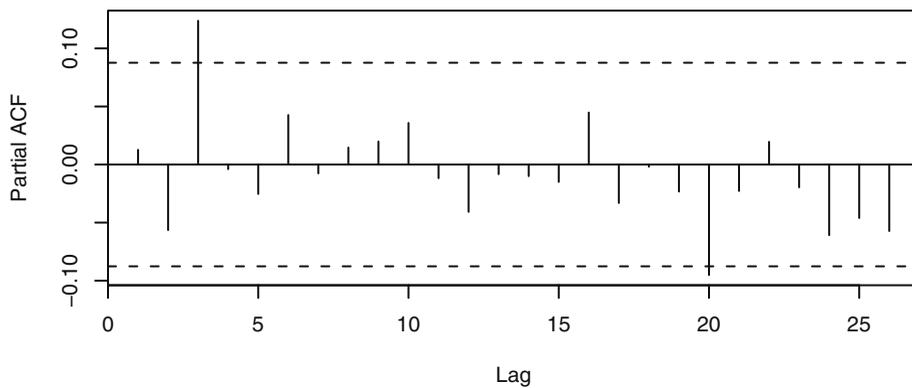
Except for lags 3 and 20, which are mildly significant, the sample ACF and PACF of the simulated data, shown in Exhibits 12.13 and 12.14, do not show significant correlations. Hence, the simulated process seems to be basically serially uncorrelated as it is.

Exhibit 12.13 Sample ACF of Simulated GARCH(1,1) Process



```
> acf(garch11.sim)
```

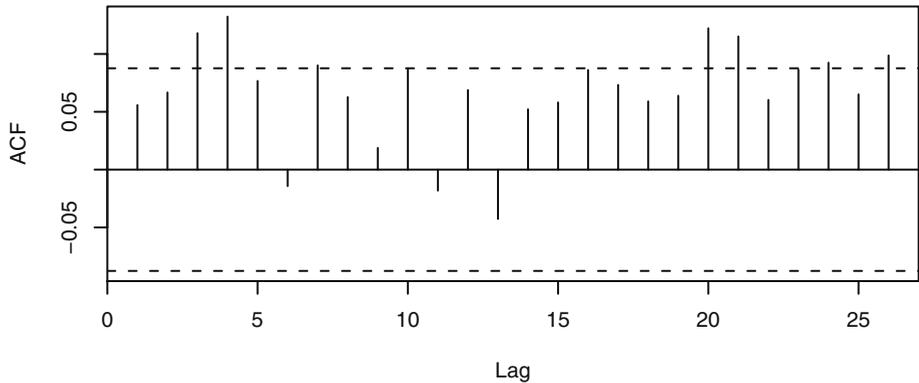
Exhibit 12.14 Sample PACF of Simulated GARCH(1,1) Process



```
> pacf(garch11.sim)
```

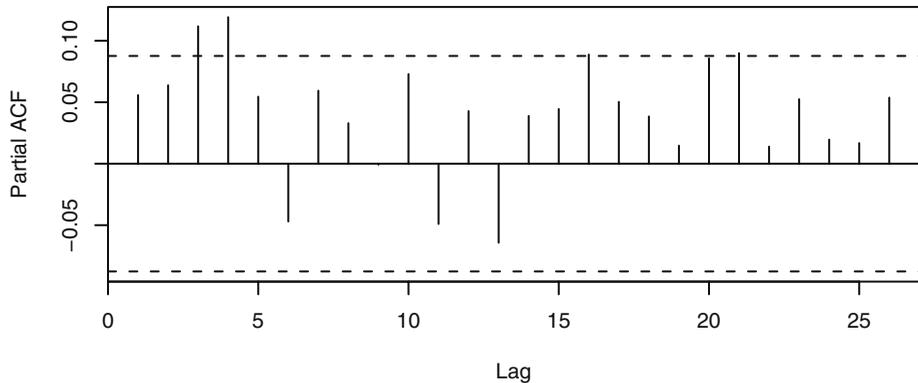
Exhibits 12.15 through 12.18 show the sample ACF and PACF of the absolute values and the squares of the simulated data.

Exhibit 12.15 Sample ACF of the Absolute Values of the Simulated GARCH(1,1) Process



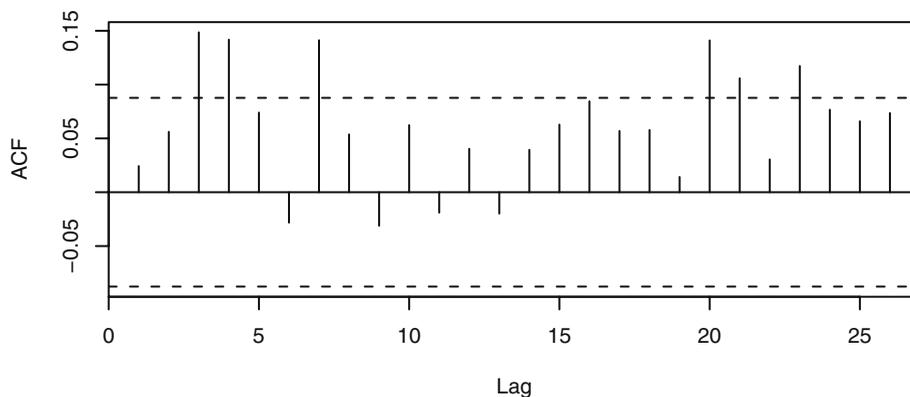
```
> acf(abs(garch11.sim))
```

Exhibit 12.16 Sample PACF of the Absolute Values of the Simulated GARCH(1,1) Process

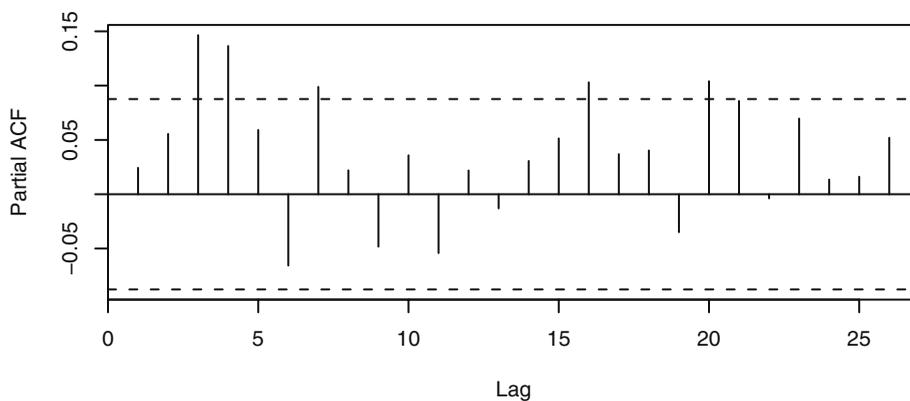


```
> pacf(abs(garch11.sim))
```

These plots indicate the existence of significant autocorrelation patterns in the absolute and squared data and indicate that the simulated process is in fact serially dependent. Interestingly, the lag 1 autocorrelations are not significant in any of these last four plots.

Exhibit 12.17 Sample ACF of the Squared Values of the Simulated GARCH(1,1) Process


```
> acf(garch11.sim^2)
```

Exhibit 12.18 Sample PACF of the Squared Values of the Simulated GARCH(1,1) Process


```
> pacf(garch11.sim^2)
```

For model identification of the GARCH orders, it is again advantageous to express the model for the conditional variances in terms of the squared returns. Recall the definition $\eta_t = r_t^2 - \sigma_{t|t-1}^2$. Similar to the ARCH(1) model, we can show that $\{\eta_t\}$ is a serially uncorrelated sequence. Moreover, η_t is uncorrelated with past squared returns. Substituting the expression $\sigma_{t|t-1}^2 = r_t^2 - \eta_t$ into Equation (12.3.2) yields

$$r_t^2 = \omega + (\beta_1 + \alpha_1)r_{t-1}^2 + \cdots + (\beta_{\max(p,q)} + \alpha_{\max(p,q)})r_{t-\max(p,q)}^2 + \eta_t - \beta_1\eta_{t-1} - \cdots - \beta_p\eta_{t-p} \quad (12.3.4)$$

where $\beta_k = 0$ for all integers $k > p$ and $\alpha_k = 0$ for $k > q$. This shows that the GARCH(p, q) model for the return series implies that the model for the squared returns is an ARMA($\max(p, q), p$) model. Thus, we can apply the model identification techniques for ARMA models to the squared return series to identify p and $\max(p, q)$. Notice that if q is smaller than p , it will be masked in the model identification. In such cases, we can first fit a GARCH(p, p) model and then estimate q by examining the significance of the resulting ARCH coefficient estimates.

As an illustration, Exhibit 12.19 shows the sample EACF of the squared values from the simulated GARCH(1,1) series.

Exhibit 12.19 Sample EACF for the Squared Simulated GARCH(1,1) Series

AR/MA	0	1	2	3	4	5	6	7	8	9	10	11	12	13
0	o	o	x	x	o	o	x	o	o	o	o	o	o	o
1	x	o	o	o	x	o	x	x	o	o	o	o	o	o
2	x	o	o	o	o	o	x	o	o	o	o	o	o	o
3	x	x	x	o	o	x	o	o	o	o	o	o	o	o
4	x	x	o	x	x	o	o	o	o	o	o	o	o	o
5	x	o	x	x	o	o	o	o	o	o	o	o	o	o
6	x	o	x	x	o	x	o	o	o	o	o	o	o	o
7	x	x	x	x	x	x	o	o	o	o	o	o	o	o

```
> eacf((garch11.sim)^2)
```

The pattern in the EACF table is not very clear, although an ARMA(2,2) model seems to be suggested. The fuzziness of the signal in the EACF table is likely caused by the larger sampling variability when we deal with higher moments. Shin and Kang (2001) argued that, to a first-order approximation, a power transformation preserves the theoretical autocorrelation function and hence the order of a stationary ARMA process. Their result suggests that the GARCH order may also be identified by studying the absolute returns. Indeed, the sample EACF table for the absolute returns, shown in Exhibit 12.20, more convincingly suggests an ARMA(1,1) model, and therefore a GARCH(1,1) model for the original data, although there is also a hint of a GARCH(2,2) model.

Exhibit 12.20 Sample EACF for Absolute Simulated GARCH(1,1) Series

AR/MA	0	1	2	3	4	5	6	7	8	9	10	11	12	13
0	o	o	x	x	o	o	x	o	o	o	o	o	o	o
1	x	o	o	o	x	o	o	o	o	o	o	o	o	o
2	x	x	o	o	o	o	o	o	o	o	o	o	o	o
3	x	x	o	o	o	x	o	o	o	o	o	o	o	o
4	x	x	o	x	o	x	o	o	o	o	o	o	o	o
5	x	o	x	x	x	o	o	o	o	o	o	o	o	o
6	x	o	x	x	x	x	o	o	o	o	o	o	o	o
7	x	x	x	x	x	o	x	o	o	o	o	o	o	o

```
> eacf(abs(garch11.sim))
```

For the absolute CREF daily return data, the sample EACF table is reported in Exhibit 12.21, which suggests a GARCH(1,1) model. The corresponding EACF table for the squared CREF returns (not shown) is, however, less clear and may suggest a GARCH(2,2) model.

Exhibit 12.21 Sample EACF for the Absolute Daily CREF Returns

AR/MA	0	1	2	3	4	5	6	7	8	9	10	11	12	13
0	o	o	o	o	o	o	o	o	o	x	x	o	o	o
1	x	o	o	o	o	o	o	o	o	o	o	o	o	o
2	x	o	o	o	o	o	o	o	o	o	o	o	o	o
3	x	o	x	o	o	o	o	o	o	o	o	o	o	o
4	x	o	x	o	o	o	o	o	o	o	o	o	o	o
5	x	x	x	x	o	o	o	o	o	o	o	o	o	o
6	x	x	x	x	o	o	o	o	o	o	o	o	o	o
7	x	x	x	x	o	o	o	o	o	o	o	o	o	o

```
> eacf(abs(r.cref))
```

Furthermore, the parameter estimates of the fitted ARMA model for the absolute data may yield initial estimates for maximum likelihood estimation of the GARCH model. For example, Exhibit 12.22 reports the estimated parameters of the fitted ARMA(1,1) model for the absolute simulated GARCH(1,1) process.

Exhibit 12.22 Parameter Estimates with ARMA(1,1) Model for the Absolute Simulated GARCH(1,1) Series

Coefficient	ar1	ma1	Intercept
Estimate	0.9821	-0.9445	0.5077
s.e.	0.0134	0.0220	0.0499

```
> arima(abs(garch11.sim), order=c(1, 0, 1))
```

Using Equation (12.3.4), it can be seen that β is estimated by 0.9445, α is estimated by $0.9821 - 0.9445 = 0.03763$, and ω can be estimated as the variance of the original data times the estimate of $1 - \alpha - \beta$, which equals 0.0073. Amazingly, these estimates turn out to be quite close to the maximum likelihood estimates reported in the next section!

We now derive the condition for a GARCH model to be weakly stationary. Assume for the moment that the return process is weakly stationary. Taking expectations on both sides of Equation (12.3.4) gives an equation for the unconditional variance σ^2

$$\sigma^2 = \omega + \sigma^2 \sum_{i=1}^{\max(p,q)} (\beta_i + \alpha_i) \quad (12.3.5)$$

so that

$$\sigma^2 = \frac{\omega}{1 - \sum_{i=1}^{\max(p,q)} (\beta_i + \alpha_i)} \quad (12.3.6)$$

which is finite if

$$\sum_{i=1}^{\max(p,q)} (\beta_i + \alpha_i) < 1 \quad (12.3.7)$$

This condition can be shown to be necessary and sufficient for the weak stationarity of a GARCH(p, q) model. (Recall that we have implicitly assumed that $\alpha_1 \geq 0, \dots, \alpha_p \geq 0$, and $\beta_1 \geq 0, \dots, \beta_q \geq 0$.) Henceforth, we assume $p = q$ for ease of notation.

As in the case of an ARCH(1) model, finiteness of higher moments of the GARCH model requires further stringent conditions on the coefficients; see Ling and McAleer (2002). Also, the stationary distribution of a GARCH model is generally fat-tailed even if the innovations are normal.

In terms of forecasting the h -step-ahead conditional variance $\sigma_{t+h|t}^2$, we can repeat the arguments used in the preceding section to derive the recursive formula that for $h > p$

$$\sigma_{t+h|t}^2 = \omega + \sum_{i=1}^p (\alpha_i + \beta_i) \sigma_{t+h-i|t}^2 \quad (12.3.8)$$

More generally, for arbitrary $h \geq 1$, the formula is more complex, as

$$\sigma_{t+h|t}^2 = \omega + \sum_{i=1}^p \alpha_i \sigma_{t+h-i|t}^2 + \sum_{i=1}^p \beta_i \hat{\sigma}_{t+h-i|t+h-i-1}^2 \quad (12.3.9)$$

where

$$\sigma_{t+h|t}^2 = r_{t+h}^2 \text{ for } h < 0 \quad (12.3.10)$$

and

$$\hat{\sigma}_{t+h-i|t+h-i-1}^2 = \begin{cases} \sigma_{t+h-i|t}^2 & \text{for } h-i-1 > 0 \\ \sigma_{t+h-i|t+h-i-1}^2 & \text{otherwise} \end{cases} \quad (12.3.11)$$

The computation of the conditional variances may be best illustrated using the GARCH(1,1) model. Suppose that there are n observations r_1, r_2, \dots, r_n and

$$\sigma_{t|t-1}^2 = \omega + \alpha_1 r_{t-1}^2 + \beta_1 \sigma_{t-1|t-2}^2 \quad (12.3.12)$$

To compute the conditional variances for $2 \leq t \leq n$, we need to set the initial value $\sigma_{1|0}^2$. This may be set to the stationary unconditional variance $\sigma^2 = \omega / (1 - \alpha_1 - \beta_1)$ under the stationarity assumption or simply as r_1^2 . Thereafter, we can compute $\sigma_{t|t-1}^2$ by the formula defining the GARCH model. It is interesting to observe that

$$\sigma_{t|t-1}^2 = (1 - \alpha_1 - \beta_1) \sigma^2 + \alpha_1 r_{t-1}^2 + \beta_1 \sigma_{t-1|t-2}^2 \quad (12.3.13)$$

so that the estimate of the one-step-ahead conditional volatility is a weighted average of the long-run variance, the current squared return, and the current estimate of the conditional volatility. Further, the MA(∞) representation of the conditional variance implies that

$$\sigma_{t|t-1}^2 = \sigma^2 + \alpha_1 (r_{t-1}^2 + \beta_1 r_{t-2}^2 + \beta_1^2 r_{t-3}^2 + \beta_1^3 r_{t-4}^2 + \dots) \quad (12.3.14)$$

an infinite moving average of past squared returns. The formula shows that the squared returns in the distant past receive exponentially diminishing weights. In contrast, simple moving averages of the squared returns are sometimes used to estimate the conditional variance. These, however, suffer much larger bias.

If $\alpha_1 + \beta_1 = 1$, then the GARCH(1,1) model is nonstationary and instead is called an IGARCH(1,1) model with the letter I standing for integrated. In such a case, we shall drop the subscript from the notation and let $\alpha = 1 - \beta$. Suppose that $\omega = 0$. Then

$$\sigma_{t|t-1}^2 = (1 - \beta)(r_{t-1}^2 + \beta r_{t-2}^2 + \beta^2 r_{t-3}^2 + \beta^3 r_{t-4}^2 + \dots), \quad (12.3.15)$$

an exponentially weighted average of the past squared returns. The famed Riskmetrics software in finance employs the IGARCH(1,1) model with $\beta = 0.94$ for estimating conditional variances; see Andersen et al. (2006).

12.4 Maximum Likelihood Estimation

The likelihood function of a GARCH model can be readily derived for the case of normal innovations. We illustrate the computation for the case of a stationary GARCH(1,1) model. Extension to the general case is straightforward. Given the parameters ω , α , and β , the conditional variances can be computed recursively by the formula

$$\sigma_{t|t-1}^2 = \omega + \alpha r_{t-1}^2 + \beta \sigma_{t-1|t-2}^2 \quad (12.4.1)$$

for $t \geq 2$, with the initial value, $\sigma_1^2|0$, set under the stationarity assumption as the stationary unconditional variance $\sigma^2 = \omega/(1 - \alpha - \beta)$. We use the conditional pdf

$$f(r_t|r_{t-1}, \dots, r_1) = \frac{1}{\sqrt{2\pi\sigma_{t|t-1}^2}} \exp[-r_t^2/(2\sigma_{t|t-1}^2)] \quad (12.4.2)$$

and the joint pdf

$$f(r_n, \dots, r_1) = f(r_{n-1}, \dots, r_1)f(r_n|r_{n-1}, \dots, r_1) \quad (12.4.3)$$

Iterating this last formula and taking logs gives the following formula for the log-likelihood function:

$$L(\omega, \alpha, \beta) = -\frac{n}{2}\log(2\pi) - \frac{1}{2}\sum_{i=1}^n \left\{ \log(\sigma_{i-1|i-2}^2) + r_i^2/\sigma_{i-1|i-2}^2 \right\} \quad (12.4.4)$$

There is no closed-form solution for the maximum likelihood estimators of ω , α , and β , but they can be computed by maximizing the log-likelihood function numerically. The maximum likelihood estimators can be shown to be approximately normally distributed with the true parameter values as their means. Their covariances may be collected into a matrix denoted by Λ , which can be obtained as follows. Let

$$\theta = \begin{bmatrix} \omega \\ \alpha \\ \beta \end{bmatrix} \quad (12.4.5)$$

be the vector of parameters. Write the i th component of θ as θ_i so that $\theta_1 = \omega$, $\theta_2 = \alpha$, and $\theta_3 = \beta$. The diagonal elements of Λ are the approximate variances of the estimators, whereas the off-diagonal elements are their approximate covariances. So, the first diagonal element of Λ is the approximate variance of $\hat{\omega}$, the (1,2)th element of Λ is the approximate covariance between $\hat{\omega}$ and $\hat{\alpha}$, and so forth. We now outline the computation of Λ . Readers not interested in the mathematical details may skip the rest of this paragraph. The 3×3 matrix Λ is approximately equal to the inverse matrix of the 3×3 matrix whose (i, j) th element equals

$$\frac{1}{2} \sum_{t=1}^n \frac{1}{\sigma_{t|t-1}^4} \frac{\partial \sigma_{t|t-1}^2}{\partial \theta_i} \frac{\partial \sigma_{t|t-1}^2}{\partial \theta_j} \quad (12.4.6)$$

The partial derivatives in this expression can be obtained recursively by differentiating Equation (12.4.1). For example, differentiating both sides of Equation (12.4.1) with respect to ω yields the recursive formula

$$\frac{\partial \sigma_{t|t-1}^2}{\partial \omega} = 1 + \beta \frac{\partial \sigma_{t-1|t-2}^2}{\partial \omega} \quad (12.4.7)$$

Other partial derivatives can be computed similarly.

Recall that, in the previous section, the simulated GARCH(1,1) series was identified to be either a GARCH(1,1) model or a GARCH(2,2) model. The model fit of the GARCH(2,2) model is reported in Exhibit 12.23, where the estimate of ω is denoted by a0, that of α_1 by a1, that of β_1 by b1, and so forth. Note that none of the coefficients is significant, although a2 is close to being significant. The model fit for the GARCH(1,1) model is given in Exhibit 12.24.

Exhibit 12.23 Estimates for GARCH(2,2) Model of a Simulated GARCH(1,1) Series

Coefficient	Estimate	Std. Error	t-value	Pr(> t)
a0	1.835e-02	1.515e-02	1.211	0.2257
a1	4.09e-15	4.723e-02	8.7e-14	1.0000
a2	1.136e-01	5.855e-02	1.940	0.0524
b1	3.369e-01	3.696e-01	0.911	0.3621
b2	5.100e-01	3.575e-01	1.426	0.1538

```
> g1=garch(garch11.sim,order=c(2,2))
> summary(g1)
```

Exhibit 12.24 Estimates for GARCH(1,1) Model of a Simulated GARCH(1,1) Series

Coefficient	Estimate	Std. Error	t-value	Pr(> t)
a0	0.007575	0.007590	0.998	0.3183
a1	0.047184	0.022308	2.115	0.0344
b1	0.935377	0.035839	26.100	< 0.0001

```
> g2=garch(garch11.sim,order=c(1,1))
> summary(g2)
```

Now all coefficient estimates (except a_0) are significant. The AIC of the fitted GARCH(2,2) model is 961.0, while that of the fitted GARCH(1,1) model is 958.0, and thus the GARCH(1,1) model provides a better fit to the data. (Here, AIC is defined as minus two times the log-likelihood of the fitted model plus twice the number of parameters. As in the case of ARIMA models, a smaller AIC is preferable.) A 95% confidence interval for a parameter is given (approximately) by the estimate ± 1.96 times its standard error. So, an approximate 95% confidence interval for ω equals $(-0.0073, 0.022)$, that of α_1 equals $(0.00345, 0.0909)$, and that of β_1 equals $(0.865, 1.01)$. These all contain their true values of 0.02, 0.05, and 0.9, respectively. Note that the standard error of b_1 is 0.0358. Since the standard error is approximately proportional to $1/\sqrt{n}$, the standard error of b_1 is expected to be about 0.0566 (0.0462) if the sample size n is 200 (300). Indeed, fitting the GARCH(1,1) model to the first 200 simulated data, b_1 was found to equal 0.0603 with standard error equal to 50.39! When the sample size was increased to 300, b_1 became 0.935 with standard error equal to 0.0449. This example illustrates that fitting a GARCH model generally requires a large sample size for the theoretical sampling distribution to be valid and useful; see Shephard (1996, p. 10) for a relevant discussion.

For the CREF return data, we earlier identified either a GARCH(1,1) or GARCH(2,2) model. The AIC of the fitted GARCH(1,1) model is 969.6, whereas that of the GARCH(2,2) model is 970.3. Hence the GARCH(1,1) model provides a marginally better fit to the data. Maximum likelihood estimates of the fitted GARCH(1,1) model are reported in Exhibit 12.25.

Exhibit 12.25 Maximum Likelihood Estimates of the GARCH(1,1) Model for the CREF Stock Returns

Parameter	Estimate [†]	Std. Error	t-value	Pr(> t)
a0	0.01633	0.01237	1.320	0.1869
a1	0.04414	0.02097	2.105	0.0353
b1	0.91704	0.04570	20.066	< 0.0001

[†] As remarked earlier, the analysis depends on the scale of measurement. In particular, a GARCH(1,1) model based on the raw CREF stock returns yields estimates $a_0 = 0.00000511$, $a_1 = 0.0941$, and $b_1 = 0.789$.

```
> ml=garch(x=r.cref,order=c(1,1))
> summary(ml)
```

Note that the long-term variance of the GARCH(1,1) model is estimated to be

$$\hat{\omega}/(1 - \hat{\alpha} - \hat{\beta}) = 0.01633/(1 - 0.04414 - 0.91704) = 0.4206 \quad (12.4.8)$$

which is very close to the sample variance of 0.4161.

In practice, the innovations need not be normally distributed. In fact, many financial time series appear to have nonnormal innovations. Nonetheless, we can proceed to esti-

mate the GARCH model by pretending that the innovations are normal. The resulting likelihood function is called the Gaussian likelihood, and estimators maximizing the Gaussian likelihood are called the *quasi-maximum likelihood estimators* (QMLEs). It can be shown that, under some mild regularity conditions, including stationarity, the quasi-maximum likelihood estimators are approximately normal, centered at the true parameter values, and their covariance matrix equals $[(\kappa + 2)/2]\Lambda$, where κ is the (excess) kurtosis of the innovations and Λ is the covariance matrix assuming the innovations are normally distributed—see the discussion above for the normal case. Note that the heavy-tailedness of the innovations will inflate the covariance matrix and hence result in less reliable parameter estimates. In the case where the innovations are deemed nonnormal, this result suggests a simple way to adjust the standard errors of the quasi-maximum likelihood estimates by multiplying the standard errors of the Gaussian likelihood estimates from a routine that assumes normal innovations by $\sqrt{(\kappa + 2)/2}$, where κ can be substituted with the sample kurtosis of the standardized residuals that are defined below. It should be noted that one disadvantage of QMLE is that the AIC is not strictly applicable.

Let the estimated conditional standard deviation be denoted by $\hat{\sigma}_{t|t-1}$. The standardized residuals are then defined as

$$\hat{\varepsilon}_t = r_t / \hat{\sigma}_{t|t-1} \quad (12.4.9)$$

The standardized residuals from the fitted model are proxies for the innovations and can be examined to cast light on the distributional form of the innovations. Once a (parameterized) distribution for the innovations is specified, for example a t -distribution, the corresponding likelihood function can be derived and optimized to obtain maximum likelihood estimators; see Tsay (2005) for details. The price of not correctly specifying the distributional form of the innovation is a loss in efficiency of estimation, although, with large datasets, the computational convenience of the Gaussian likelihood approach may outweigh the loss of estimation efficiency.

12.5 Model Diagnostics

Before we accept a fitted model and interpret its findings, it is essential to check whether the model is correctly specified, that is, whether the model assumptions are supported by the data. If some key model assumptions seem to be violated, then a new model should be specified; fitted, and checked again until a model is found that provides an adequate fit to the data. Recall that the standardized residuals are defined as

$$\hat{\varepsilon}_t = r_t / \hat{\sigma}_{t|t-1} \quad (12.5.1)$$

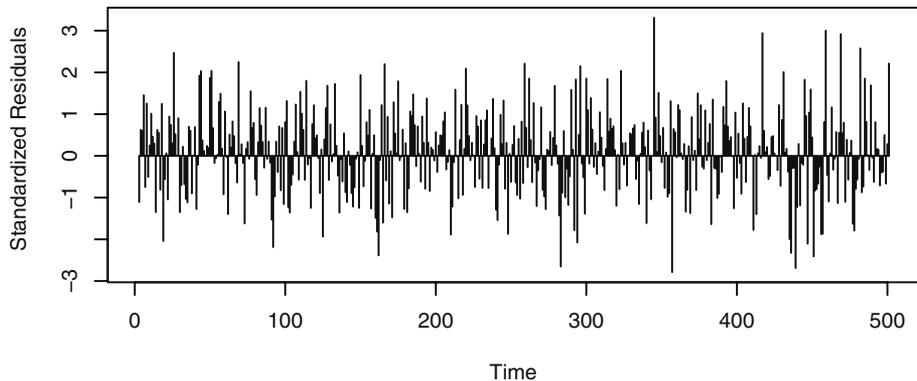
which are approximately independently and identically distributed if the model is correctly specified. As in the case of model diagnostics for ARIMA models, the standardized residuals are very useful for checking the model specification. The normality assumption of the innovations can be explored by plotting the QQ normal scores plot. Deviations from a straight line pattern in the QQ plot furnish evidence against normality and may provide clues on the distributional form of the innovations. The Shapiro-Wilk

test and the Jarque-Bera test are helpful for formally testing the normality of the innovations.[†]

For the GARCH(1,1) model fitted to the simulated GARCH(1,1) process, the sample skewness and kurtosis of the standardized residuals equal -0.0882 and -0.104 , respectively. Moreover, both the Shapiro-Wilk test and the Jarque-Bera test suggest that the standardized residuals are normal.

For the GARCH(1,1) model fitted to the CREF return data, the standardized residuals are plotted in Exhibit 12.26. There is some tendency for the residuals to be larger in magnitude towards the end of the study period, perhaps suggesting that there is some residual pattern in the volatility. The QQ plot of the standardized residuals is shown in Exhibit 12.27. The QQ plot shows a largely straight-line pattern. The skewness and the kurtosis of the standardized residuals are 0.0341 and 0.205 , respectively. The p -value of the Jarque-Bera test equals 0.58 and that of the Shapiro-Wilk test is 0.34 . Hence, the normality assumption cannot be rejected.

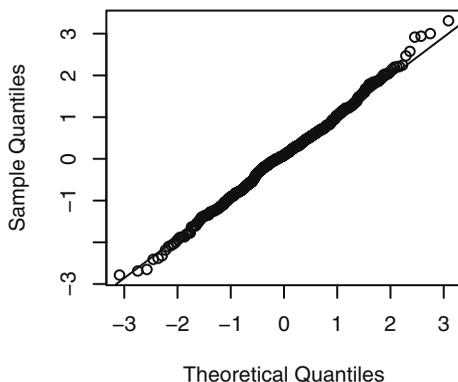
Exhibit 12.26 Standardized Residuals from the Fitted GARCH(1,1) Model of Daily CREF Returns



```
> plot(residuals(m1), type='h', ylab='Standardized Residuals')
```

[†] Chen and Kuan (2006) have shown that the Jarque-Bera test with the residuals from a GARCH model is no longer approximately chi-square distributed under the null hypothesis of normal innovations. Their simulation results suggest that, in such cases, the Jarque-Bera test tends to be liberal; that is, it rejects the normality hypothesis more often than its nominal significance level. The authors have proposed a modification of the Jarque-Bera test that retains the chi-square null distribution approximately. Similarly, it can be expected that the Shapiro-Wilk test may require modification when it is applied to residuals from a GARCH model, although the problem seems open.

Exhibit 12.27 QQ Normal Scores Plot of Standardized Residuals from the Fitted GARCH(1,1) Model of Daily CREF Returns



```
> win.graph(width=2.5,height=2.5,pointsize=8)
> qqnorm(residuals(m1)); qqline(residuals(m1))
```

If the GARCH model is correctly specified, then the standardized residuals $\{\hat{\varepsilon}_t\}$ should be close to independently and identically distributed. The independently and identically distributed assumption of the innovations can be checked by examining their sample acf. Recall that the portmanteau statistic equals

$$n \sum_{k=1}^m \hat{\rho}_k^2$$

where $\hat{\rho}_k$ is the lag k autocorrelation of the standardized residuals and n is the sample size. (Recall that the same statistic is also known as the Box-Pierce statistic and, in a modified version, the Ljung-Box statistic.) Furthermore, it can be shown that the test statistic is approximately χ^2 distributed with m degrees of freedom under the null hypothesis that the model is correctly specified. This result relies on the fact that the sample autocorrelations of nonzero lags from an independently and identically distributed sequence are approximately independent and normally distributed with zero mean and variance $1/n$, and this result holds approximately also for the sample autocorrelations of the standardized residuals if the data are truly generated by a GARCH model of the same orders as those of the fitted model. However, the portmanteau test does not have strong power against uncorrelated and yet serially dependent innovations. In fact, we start out with the assumption that the return data are uncorrelated, so the preceding test is of little interest.

More useful tests may be devised by studying the autocorrelation structure of the absolute standardized residuals or the squared standardized residuals. Let the lag k autocorrelation of the absolute standardized residuals be denoted by $\hat{\rho}_{k,1}$ and that of the squared standardized residuals by $\hat{\rho}_{k,2}$. Unfortunately, the approximate χ^2 distribution with m degrees of freedom for the corresponding portmanteau statistics based on $\hat{\rho}_{k,1}$ ($\hat{\rho}_{k,2}$) is no longer valid, the reason being that the estimation of the unknown param-

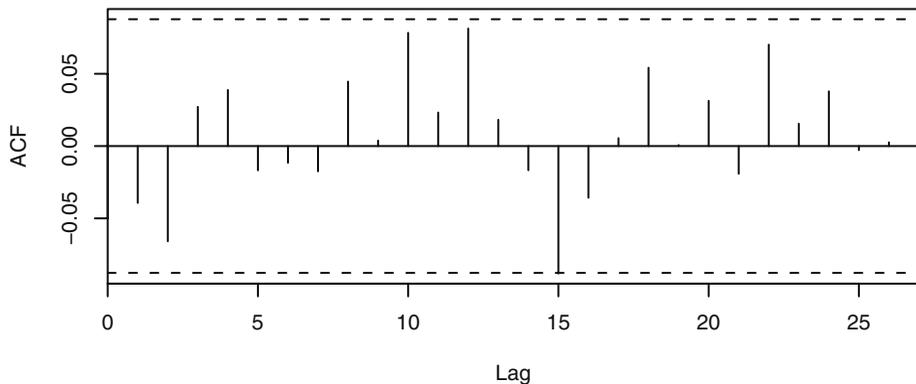
ters induces a nonnegligible effect on the tests. Li and Mak (1994) showed that the χ^2 approximate distribution may be preserved by replacing the sum of squared autocorrelations by a quadratic form in the autocorrelations; see also Li (2003). For the absolute standardized residuals, the test statistic takes the form

$$n \sum_{i=1}^m \sum_{j=1}^m q_{i,j} \hat{\rho}_{i,1} \hat{\rho}_{j,1} \quad (12.5.2)$$

We shall call this modified test statistic the *generalized portmanteau test statistic*. However, the q 's depend on m , the number of lags, and they are specific to the underlying true model and so must be estimated from the data. For the squared residuals, the q 's take different values. See Appendix I on page 318 for the formulas for the q 's.

We illustrate the generalized portmanteau test with the CREF data. Exhibit 12.28, plots the sample ACF of the squared standardized residuals from the fitted GARCH(1,1) model. The (individual) critical limits in the figure are based on the $1/n$ nominal variance under the assumption of independently and identically distributed data. As discussed above, this nominal value could be very different from the actual variance of the autocorrelations of the squared residuals even when the model is correctly specified. Nonetheless, the general impression from the figure is that the squared residuals are serially uncorrelated.

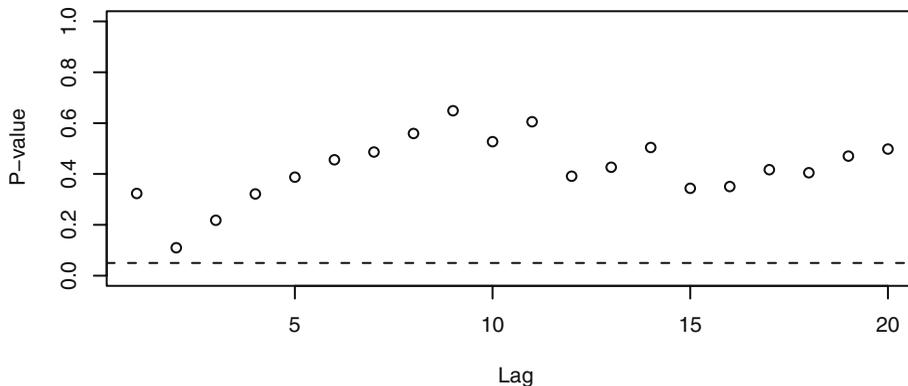
Exhibit 12.28 Sample ACF of Squared Standardized Residuals from the GARCH(1,1) Model of the Daily CREF Returns



```
> acf(residuals(m1)^2,na.action=na.omit)
```

Exhibit 12.29 displays the p -values of the generalized portmanteau tests with the squared standardized residuals from the fitted GARCH(1,1) model of the CREF data for $m = 1$ to 20. All p -values are higher than 5%, suggesting that the squared residuals are uncorrelated over time, and hence the standardized residuals may be independent.

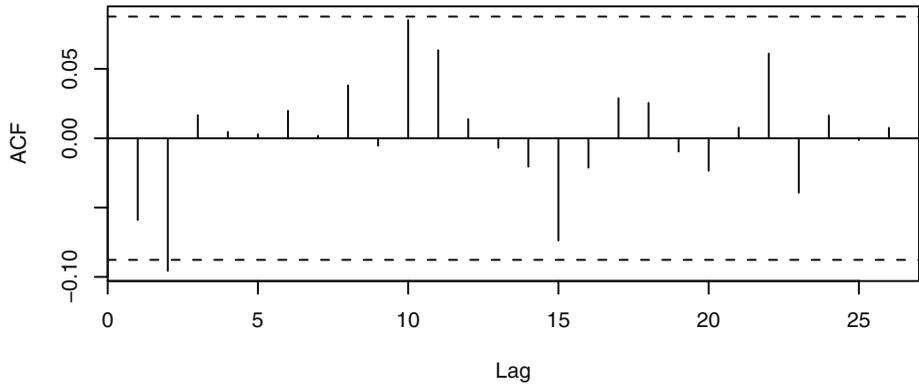
Exhibit 12.29 Generalized Portmanteau Test p -Values for the Squared Standardized Residuals for the GARCH(1,1) Model of the Daily CREF Returns



```
> gBox(m1, method='squared')
```

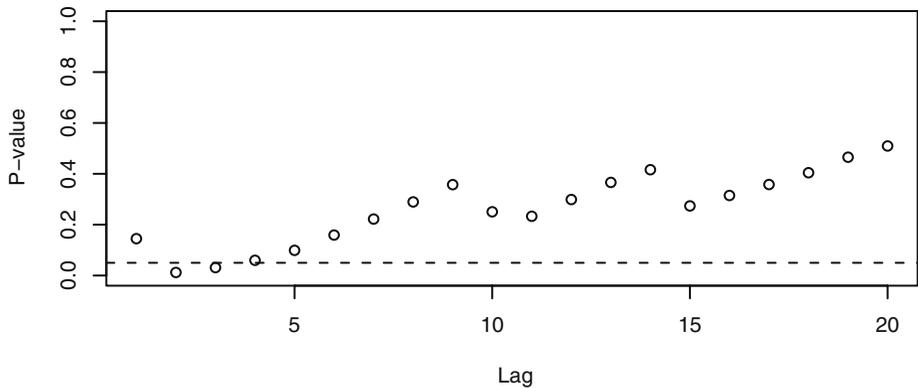
We repeated checking the model using the absolute standardized residuals—see Exhibits 12.30 and 12.31. The lag 2 autocorrelation of the absolute residuals is significant according to the nominal critical limits shown. Furthermore, the generalized portmanteau tests are significant when $m = 2$ and 3 and marginally not significant at $m = 4$. The sample EACF table (not shown) of the absolute standardized residuals suggests an AR(2) model for the absolute residuals and hence points to the possibility that the CREF returns may be identified as a GARCH(1,2) process. However, the fitted GARCH(1,2) model to the CREF data did not improve the fit, as its AIC was 978.2—much higher than 969.6, that of the GARCH(1,1) model. Therefore, we conclude that the fitted GARCH(1,1) model provides a good fit to the CREF data.

Exhibit 12.30 Sample ACF of the Absolute Standardized Residuals from the GARCH(1,1) Model for the Daily CREF Returns



```
> acf(abs(residuals(m1)), na.action=na.omit)
```

Exhibit 12.31 Generalized Portmanteau Test p -Values for the Absolute Standardized Residuals for the GARCH(1,1) Model of the Daily CREF Returns



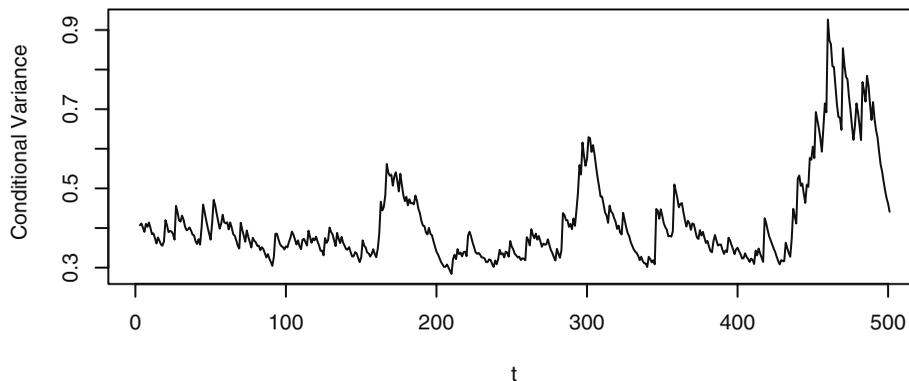
```
> gBox(m1, method='absolute')
```

Given that the GARCH(1,1) model provides a good fit to the CREF data, we may use it to forecast the future conditional variances. Exhibit 12.32 shows the within-sample estimates of the conditional variances, which capture several periods of high volatility, especially the one at the end of the study period. At the final time point, the squared return equals 2.159, and the conditional variance is estimated to be 0.4411. These values combined with Equations (12.3.8) and (12.3.9) can be used to compute the forecasts of future conditional variances. For example, the one-step-ahead forecast of the conditional variance equals $0.01633 + 0.04414 \cdot 2.159 + 0.91704 \cdot 0.4411 = 0.5161$. The two-step forecast of the conditional variance equals $0.01633 + 0.04414 \cdot 0.5161 +$

$0.91704 \times 0.5161 = 0.5124$, and so forth, with the longer lead forecasts eventually approaching 0.42066, the long-run variance of the model. The conditional variances may be useful for pricing financial assets through the Black-Scholes formula and calculation of the value at risk (*VaR*); see Tsay (2005) and Andersen et al. (2006).

It is interesting to note that the need for incorporating ARCH in the data is also supported by the McLeod-Li test applied to the residuals of the AR(1) + outlier model; see Exhibit (12.9), page 283.

Exhibit 12.32 Estimated Conditional Variances of the Daily CREF Returns



```
> plot((fitted(m1)[,1])^2,type='l',ylab='Conditional Variance',
      xlab='t')
```

12.6 Conditions for the Nonnegativity of the Conditional Variances

Because the conditional variance $\sigma_{i|t-1}^2$ must be nonnegative, the GARCH parameters are often constrained to be nonnegative. However, the nonnegativity parameter constraints need not be necessary for the nonnegativity of the conditional variances. This issue was first explored by Nelson and Cao (1992) and more recently by Tsai and Chan (2006). To better understand the problem, first consider the case of an ARCH(q) model. Then the conditional variance is given by the formula

$$\sigma_{i|t-1}^2 = \omega + \alpha_1 r_{t-1}^2 + \alpha_2 r_{t-2}^2 + \cdots + \alpha_q r_{t-q}^2 \quad (12.6.1)$$

Assume that q consecutive returns can take on any arbitrary set of real numbers. If one of the α 's is negative, say $\alpha_1 < 0$, then $\sigma_{i|t-1}^2$ will be negative if r_{t-1}^2 is sufficiently large and the other r 's are sufficiently close to zero. Hence, it is clear that all α 's must be nonnegative for the conditional variances to be nonnegative. Similarly, by letting the returns be close to zero, it can be seen that ω must be nonnegative—otherwise the conditional variance may become negative. Thus, it is clear that for an ARCH model, the

non-negativity of all ARCH coefficients is necessary and sufficient for the conditional variances σ_{it-1}^2 to be always nonnegative.

The corresponding problem for a GARCH(p, q) model can be studied by expressing the GARCH model as an infinite-order ARCH model. The conditional variance process $\{\sigma_{it-1}^2\}$ is an ARMA(p, q) model with the squared returns playing the role of the noise process. Recall that an ARMA(p, q) model can be expressed as an MA(∞) model if all the roots of the AR characteristic polynomial lie outside the unit circle. Hence, assuming that all the roots of $1 - \beta_1 x - \beta_2 x^2 - \dots - \beta_p x^p = 0$ have magnitude greater than 1, the conditional variances satisfy the equation

$$\sigma_{it-1}^2 = \omega^* + \psi_1 r_{t-1}^2 + \psi_2 r_{t-2}^2 + \dots \quad (12.6.2)$$

where

$$\omega^* = \omega / \left(1 - \sum_{i=1}^p \beta_i \right) \quad (12.6.3)$$

It can be similarly shown that the conditional variances are all nonnegative if and only if $\omega^* \geq 0$ and $\psi_j \geq 0$ for all integers $j \geq 1$. The coefficients in the ARCH(∞) representation relate to the parameters of the GARCH model through the equality

$$\frac{\alpha_1 B + \dots + \alpha_q B^q}{1 - \beta_1 B - \dots - \beta_p B^p} = \psi_1 B + \psi_2 B^2 + \dots \quad (12.6.4)$$

If $p = 1$, then it can be easily checked that $\psi_k = \beta_1 \psi_{k-1}$ for $k > q$. Thus, $\psi_j \geq 0$ for all $j \geq 1$ if and only if $\beta_1 \geq 0$ and $\psi_1 \geq 0, \dots, \psi_q \geq 0$. For higher GARCH order, the situation is more complex. Let $\lambda_j, 1 \leq j \leq p$, be the roots of the characteristic equation

$$1 - \beta_1 x - \dots - \beta_p x^p = 0 \quad (12.6.5)$$

With no loss of generality, we can and shall henceforth assume the convention that

$$|\lambda_1| \leq |\lambda_2| \leq \dots \leq |\lambda_p| \quad (12.6.6)$$

Let $i = \sqrt{-1}$ and $\bar{\lambda}$ denote the complex conjugate of λ , $B(x) = 1 - \beta_1 x - \dots - \beta_p x^p$, and $B^{(1)}$ be the first derivative of B . We then have the following result.

Result 1: Consider a GARCH(p, q) model where $p \geq 2$. Assume A1, that all the roots of the equation

$$1 - \beta_1 x - \beta_2 x^2 - \dots - \beta_p x^p = 0 \quad (12.6.7)$$

have magnitude greater than 1, and A2, that none of these roots satisfy the equation

$$\alpha_1 x + \dots + \alpha_q x^q = 0 \quad (12.6.8)$$

Then the following hold:

- (a) $\omega^* \geq 0$ if and only if $\omega \geq 0$.

(b) Assuming the roots $\lambda_1, \dots, \lambda_p$ are distinct, and $|\lambda_1| < |\lambda_2|$, then the conditions given in Equation (12.6.9) are necessary and sufficient for $\psi_k \geq 0$ for all positive integers k :

$$\left. \begin{array}{l} \lambda_1 \text{ is real and } \lambda_1 > 1 \\ \alpha(\lambda_1) > 0 \\ \psi_k \geq 0 \text{ for } k = 1, \dots, k^* \end{array} \right\} \quad (12.6.9)$$

where k^* is the smallest integer greater than or equal to

$$r_j = -\frac{\alpha(\lambda_j)}{B^{(1)}(\lambda_j)}, \quad \text{for } 1 \leq j \leq p, \quad \text{and} \quad r^* = \max_{2 \leq j \leq p} (|r_j|) \quad (12.6.10)$$

$$\frac{\log(r_1) - \log[(p-1)r^*]}{\log(|\lambda_1|) - \log(|\lambda_2|)},$$

For $p = 2$, the k^* defined in Result 1 can be shown to be $q + 1$; see Theorem 2 of Nelson and Cao (1992). If the k^* defined in Equations (12.6.10) is a negative number, then it can be seen from the proof given in Tsai and Chan (2006) that $\psi_k \geq 0$ for all positive k .

Tsai and Chan (2006) have also derived some more readily verifiable conditions for the conditional variances to be always nonnegative.

Result 2: Let the assumptions of Result 1 be satisfied. Then the following hold:

(a) For a GARCH($p, 1$) model, if λ_j is real and $\lambda_j > 1$, for $j = 1, \dots, p$, and $\alpha_1 \geq 0$, then $\psi_k \geq 0$ for all positive integers k .

(b) For a GARCH($p, 1$) model, if $\psi_k \geq 0$ for all positive integers k , then $\alpha_1 \geq 0$,

$$\sum_{j=1}^p \lambda_j^{-1} \geq 0, \quad \lambda_1 \text{ is real, and } \lambda_1 > 1.$$

(c) For a GARCH(3,1) model, $\psi_k \geq 0$ for all positive integers k if and only if $\alpha_1 \geq 0$ and either of the following cases hold:

Case 1. All the λ_j 's are real numbers, $\lambda_1 > 1$, and $\lambda_1^{-1} + \lambda_2^{-1} + \lambda_3^{-1} \geq 0$.

Case 2. $\lambda_1 > 1$, and $\lambda_2 = \bar{\lambda}_3 = |\lambda_2|e^{i\theta} = a + bi$, where a and b are real numbers, $b > 0$, and $0 < \theta < \pi$:

Case 2.1. $\theta = 2\pi/r$ for some integer $r \geq 3$, and $1 < \lambda_1 \leq |\lambda_2|$.

Case 2.2. $\theta \notin \{2\pi/r \mid r = 3, 4, \dots\}$, and $|\lambda_2|/\lambda_1 \geq x_0 > 1$, where x_0 is the largest real root of $f_{n,\theta}(x) = 0$, and

$$f_{n,\theta}(x) = x^{n+2} - x \frac{\sin[(n+2)\theta]}{\sin\theta} + \frac{\sin[(n+1)\theta]}{\sin\theta} \quad (12.6.11)$$

where n is the smallest positive integer such that $\sin((n+1)\theta) < 0$ and $\sin((n+2)\theta) > 0$.

- (d) For a GARCH(3,1) model, if $\lambda_2 = \bar{\lambda}_3 = |\lambda_2|e^{i\theta} = a + bi$, where a and b are real numbers, $b > 0$, and $a \geq \lambda_1 > 1$, then $\psi_k \geq 0$ for all positive integers k .
- (e) For a GARCH(4,1) model, if the λ_j 's are real for $1 \leq j \leq 4$, then a necessary and sufficient condition for $\{\psi_i\}_{i=0}^{\infty}$ to be nonnegative is that $\alpha_1 \geq 0$, $\lambda_1^{-1} + \lambda_2^{-1} + \lambda_3^{-1} + \lambda_4^{-1} \geq 0$, and $\lambda_1 > 1$.

Note that x_0 is the only real root of Equation (12.6.11) that is greater than or equal to 1. Also, Tsai and Chan (2006) proved that if the ARCH coefficients (α 's) of a GARCH(p, q) model are all nonnegative, the model has nonnegative conditional variances if the nonnegativity property holds for the associated GARCH($p, 1$) models with a nonnegative α_1 coefficient.

12.7 Some Extensions of the GARCH Model

The GARCH model may be generalized in several directions. First, the GARCH model assumes that the conditional mean of the time series is zero. Even for financial time series, this strong assumption need not always hold. In the more general case, the conditional mean structure may be modeled by some ARMA(u, v) model, with the white noise term of the ARMA model modeled by some GARCH(p, q) model. Specifically, let $\{Y_t\}$ be a time series given by (now we switch to using the notation Y_t to denote a general time series)

$$\left. \begin{aligned} Y_t &= \phi_1 Y_{t-1} + \cdots + \phi_u Y_{t-u} + \theta_0 + e_t + \theta_1 e_{t-1} + \cdots + \theta_v e_{t-v} \\ e_t &= \sigma_{t|t-1} \varepsilon_t \\ \sigma_{t|t-1}^2 &= \omega + \alpha_1 e_{t-1}^2 + \cdots + \alpha_q e_{t-q}^2 + \beta_1 \sigma_{t-1|t-2}^2 + \cdots + \beta_p \sigma_{t-p|t-p-1}^2 \end{aligned} \right\} \quad (12.7.1)$$

and where we have used the plus convention in the MA parts of the model. The ARMA orders can be identified based on the time series $\{Y_t\}$, whereas the GARCH orders may be identified based on the squared residuals from the fitted ARMA model. Once the orders are identified, full maximum likelihood estimation for the ARMA + GARCH model can be carried out by maximizing the likelihood function as defined in Equation (12.4.4) on page 298 but with r_t there replaced by e_t that are recursively computed according to Equation (12.7.1). The maximum likelihood estimators of the ARMA parameters are approximately independent of their GARCH counterparts if the innovations ε_t have a symmetric distribution (for example, a normal or t -distribution) and their standard errors are approximately given by those in the pure ARMA case. Likewise, the GARCH parameter estimators enjoy distributional results similar to those for the pure GARCH case. However, the ARMA estimators and the GARCH estimators are correlated if the innovations have a skewed distribution. In the next section, we illustrate the ARMA + GARCH model with the daily exchange rates of the U.S. dollar to the Hong Kong dollar.

Another direction of generalization concerns nonlinearity in the volatility process. For financial data, this is motivated by a possible asymmetric market response that may,

for example, react more strongly to a negative return than a positive return of the same magnitude. The idea can be simply illustrated in the setting of an ARCH(1) model, where the asymmetry can be modeled by specifying that

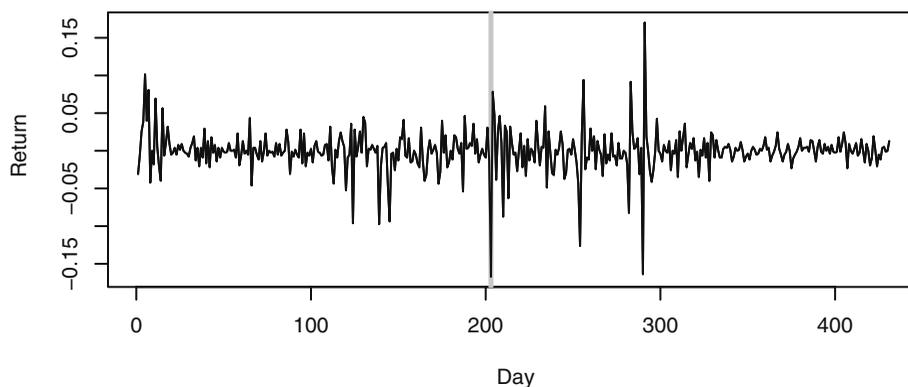
$$\sigma_{t|t-1}^2 = \omega + \alpha e_{t-1}^2 + \gamma \min(e_{t-1}, 0)^2 \quad (12.7.2)$$

Such a model is known as a GJR model—a variant of which allows the threshold to be unknown and other than 0. See Tsay (2005) for other useful extensions of the GARCH models.

12.8 Another Example: The Daily USD/HKD Exchange Rates

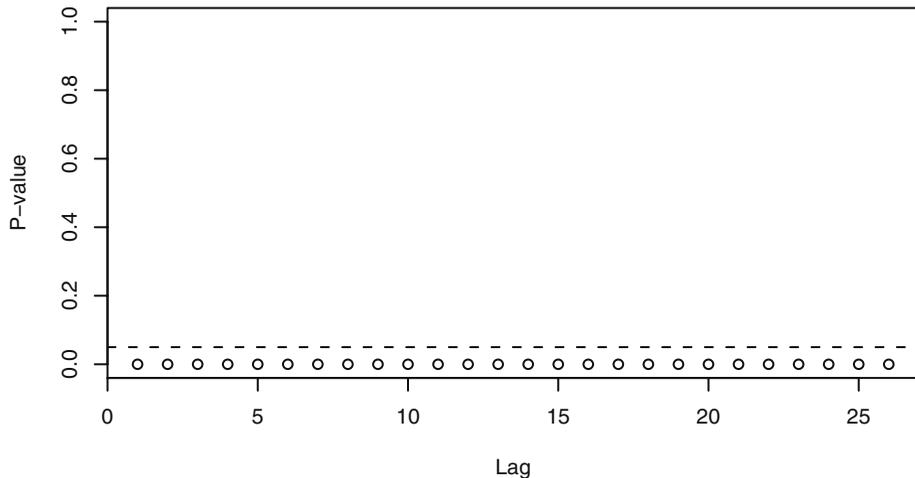
As an illustration for the ARIMA + GARCH model, we consider the daily USD/HKD (U.S. dollar to Hong Kong dollar) exchange rate from January 1, 2005 to March 7, 2006, altogether 431 days of data. The returns of the daily exchange rates are shown in Exhibit 12.33 and appear to be stationary, although volatility clustering is evident in the plot.

Exhibit 12.33 Daily Returns of USD/HKD Exchange Rate: 1/1/05–3/7/06



```
> data(usd.hkd)
> plot(ts(usd.hkd$hkrate, freq=1), type='l', xlab='Day',
       ylab='Return')
```

It is interesting to note that the need for incorporating ARCH in the data is also supported by the McLeod-Li test applied to the residuals of the AR(1) + outlier model; see below for further discussion of the additive outlier. Exhibit 12.34 shows that the tests are all significant when the number of lags of the autocorrelations of the squared residuals ranges from 1 to 26, displaying strong evidence of conditional heteroscedascity.

Exhibit 12.34 McLeod-Li Test Statistics for the USD/HKD Exchange Rate

```
> attach(usd.hkd)
> McLeod.Li.test(arima(hkrate,order=c(1,0,0),
  xreg=data.frame(outlier1)))
```

An AR(1) + GARCH(3,1) model was fitted to the (raw) return data with an additive outlier one day after July 22, 2005, the date when China revalued the yuan by 2.1% and adopted a floating-rate system for it. The outlier is shaded in gray in Exhibit 12.33. The intercept term in the conditional mean function was found to be insignificantly different from zero and hence is omitted from the model. Thus we take the returns to have zero mean unconditionally. The fitted model has an AIC = -2070.9, being smallest among various competing (weakly) stationary models—see Exhibit 12.35. Interestingly, for lower GARCH orders ($p \leq 2$), the fitted models are nonstationary, but the fitted models are largely stationary when the GARCH order is higher than 2. As the data appear to be stationary, we choose the AR(1) + GARCH(3,1) model as the final model.

The AR + GARCH models partially reported in Exhibit 12.35 were fitted using the Proc Autoreg routine in the SAS software.[†] We used the default option of imposing that the Nelson-Cao inequality constraints for the GARCH conditional variance process be nonnegative. However, the inequality constraints so imposed are only necessary and sufficient for the nonnegativity of the conditional variances of a GARCH(p, q) model for $p \leq 2$. For higher-order GARCH models, Proc Autoreg imposes the constraints that (1) $\psi_k \geq 0$, $1 \leq k \leq \max(q - 1, p) + 1$ and (2) the nonnegativity of the in-sample conditional variances; see the SAS 9.1.3 Help and Documentation manual. Hence, higher-order GARCH models estimated by Proc Autoreg with the Nelson-Cao option need not have nonnegative conditional variances with probability one.

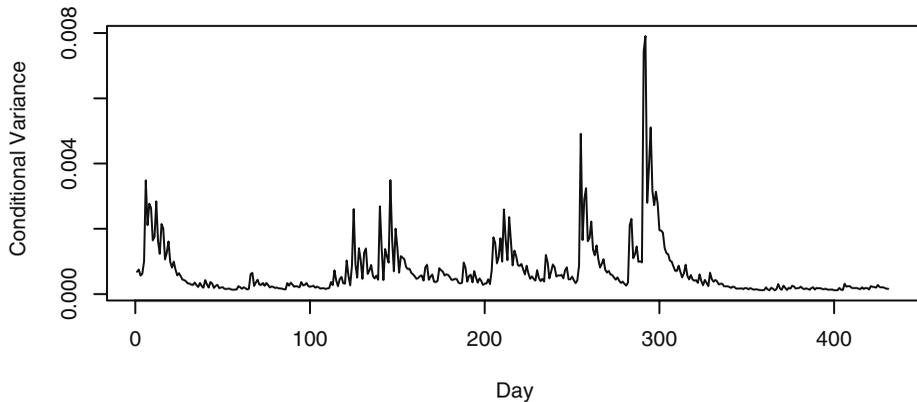
[†] Proc Autoreg of SAS has the option of imposing the Nelson-Cao inequality constraint in the GARCH model, hence it is used here.

Exhibit 12.35 AIC Values for Various Fitted Models for the Daily Returns of the USD/HKD Exchange Rate

AR order	GARCH order (p)	ARCH order (q)	AIC	Stationarity
0	3	1	-1915.3	nonstationary
1	1	1	-2054.3	nonstationary
1	1	2	-2072.5	nonstationary
1	1	3	-2051.0	nonstationary
1	2	1	-2062.2	nonstationary
1	2	2	-2070.5	nonstationary
1	2	3	-2059.2	nonstationary
1	3	1	-2070.9	stationary
1	3	2	-2064.8	stationary
1	3	3	-2062.8	stationary
1	4	1	-2061.7	nonstationary
1	4	2	-2054.8	stationary
1	4	3	-2062.4	stationary
2	3	1	-2066.6	stationary

For the Hong Kong exchange rate data, the fitted model from Proc Autoreg is listed in Exhibit 12.37 with the estimated conditional variances shown in Exhibit 12.36. Note that the GARCH2 (β_2) coefficient estimate is negative.

Exhibit 12.36 Estimated Conditional Variances of the Daily Returns of USD/HKD Exchange Rate from the Fitted AR(1) + GARCH(3,1) Model



```
> plot(ts(usd.hkd$v, freq=1), type='l', xlab='Day',
       ylab='Conditional Variance')
```

Since both the intercept and the ARCH coefficient are positive, we can apply part (c) of Result 2 to check whether or not the conditional variance process defined by the fitted model is always nonnegative. The characteristic equation $1 - \beta_1 x - \beta_2 x^2 - \beta_3 x^3 = 0$ admits three roots equal to 1.153728 and $-0.483294 \pm 1.221474i$. Thus $\lambda_1 = 1.153728$ and $|\lambda_2|/\lambda_1 = 1.138579$. Based on numerical computations, n in Equation (12.6.11) turns out to be 2 and Equation (12.6.11) has one real root equal to 1.1385751 which is strictly less than $1.138579 = |\lambda_2|/\lambda_1$. Hence, we can conclude that the fitted model always results in nonnegative conditional variances.

Exhibit 12.37 Fitted AR(1) + ARCH(3,1) Model for Daily Returns of USD/HKD Exchange Rate

Coefficient	Estimate	Std. error	t-ratio	p-value
AR1	0.1635	0.005892	21.29	0.0022
ARCH0 (ω)	2.374×10^{-5}	6.93×10^{-6}	3.42	0.0006
ARCH1 (α_1)	0.2521	0.0277	9.09	< 0.0001
GARCH1 (β_1)	0.3066	0.0637	4.81	< 0.0001
GARCH2 (β_2)	-0.09400	0.0391	-2.41	0.0161
GARCH3 (β_3)	0.5023	0.0305	16.50	< 0.0001
Outlier	-0.1255	0.00589	-21.29	< 0.0001

```

> SAS code: data hkex; infile 'hkrate.dat'; input hkrate;
  outlier1=0;
  day+1; if day=203 then outlier1=1;
  proc autoreg data=hkex;
    model hkrate=outlier1 /noint nlag=1 garch=(p=3,q=1)
    maxiter=200 archtest;
    /*hetero outlier /link=linear;*/
    output out=a cev=v residual=r;
  run;

```

12.9 Summary

This chapter began with a brief description of some terms and issues associated with financial time series. Autoregressive conditional heteroscedasticity (ARCH) models were then introduced in an attempt to model the changing variance of a time series. The ARCH model of order 1 was thoroughly explored from identification through parameter estimation and prediction. These models were then generalized to the generalized autoregressive conditional heteroscedasticity, GARCH(p, q), model. The GARCH models were also thoroughly explored with respect to identification, maximum likelihood estimation, prediction, and model diagnostics. Examples with both simulated and real time series data were used to illustrate the ideas.

EXERCISES

- 12.1** Display the time sequence plot of the absolute returns for the CREF data. Repeat the plot with the squared returns. Comment on the volatility patterns observed in these plots. (The data are in file named CREF.)
- 12.2** Plot the time sequence plot of the absolute returns for the USD/HKD exchange rate data. Repeat the plot with the squared returns. Comment on the volatility patterns observed in these plots. (The data are in the file named usd.hkd.)
- 12.3** Use the definition $\eta_t = r_t^2 - \sigma_{t|t-1}^2$ [Equation (12.2.4) on page 287] and show that $\{\eta_t\}$ is a serially uncorrelated sequence. Show also that η_t is uncorrelated with past squared returns, that is, show that $\text{Corr}(\eta_t, r_{t-k}^2) = 0$ for $k > 0$.
- 12.4** Substituting $\sigma_{t|t-1}^2 = r_t^2 - \eta_t$ into Equation (12.2.2) on page 285 show the algebra that leads to Equation (12.2.5) on page 287.
- 12.5** Verify Equation (12.2.8) on page 288.
- 12.6** Without doing any theoretical calculations, order the kurtosis values of the following four distributions in ascending order: the t -distribution with 10 DF, the t -distribution with 30 DF, the uniform distribution on $[-1, 1]$, and the normal distribution with mean 0 and variance 4. Explain your answer.
- 12.7** Simulate a GARCH(1,1) process with $\alpha = 0.1$ and $\beta = 0.8$ and of length 500. Plot the time series and inspect its sample ACF, PACF, and EACF. Are the data consistent with the assumption of white noise?
- Square the data and identify a GARCH model for the raw data based on the sample ACF, PACF, and EACF of the squared data.
 - Identify a GARCH model for the raw data based on the sample ACF, PACF and EACF of the absolute data. Discuss and reconcile any discrepancy between the tentative model identified with the squared data and that with the absolute data.
 - Perform the McLeod-Li test on your simulated series. What do you conclude?
 - Repeat the exercise but now using only the first 200 simulated data. Discuss your findings.
- 12.8** The file cref.bond contains the daily price of the CREF bond fund from August 26, 2004 to August, 15, 2006. These data are available only on trading days, but proceed to analyze the data as if they were sampled regularly.
- Display the time sequence plot of the daily bond price data and comment on the main features in the data.
 - Compute the daily bond returns by log-transforming the data and then computing the first differences of the transformed data. Plot the daily bond returns, and comment on the result.
 - Perform the McLeod-Li test on the returns series. What do you conclude?
 - Show that the returns of the CREF bond price series appear to be independently and identically distributed and not just serially uncorrelated; that is, there is no discernible volatility clustering.

- 12.9** The daily returns of Google stock from August 20, 2004 to September 13, 2006 are stored in the file named `google`.
- (a) Display the time sequence plot for the return data and show that the data are essentially uncorrelated over time.
 - (b) Compute the mean of the Google daily returns. Does it appear to be significantly different from 0?
 - (c) Perform the McLeod-Li test on the Google daily returns series. What do you conclude?
 - (d) Identify a GARCH model for the Google daily return data. Estimate the identified model and perform model diagnostics with the fitted model.
 - (e) Draw and comment on the time sequence plot of the estimated conditional variances.
 - (f) Plot the QQ normal plot for the standardized residuals from the fitted model. Do the residuals appear to be normal? Discuss the effects of the normality on the model fit, for example, regarding the computation of the confidence interval.
 - (g) Construct a 95% confidence interval for b_1 .
 - (h) What are the stationary mean and variance according to the fitted GARCH model? Compare them with those of the data.
 - (i) Based on the GARCH model, construct the 95% prediction intervals for h -step-ahead forecast, for $h = 1, 2, \dots, 5$.
- 12.10** In Exercise 11.21 on page 276, we investigated the existence of outliers with the logarithms of monthly oil prices within the framework of an IMA(1,1) model. Here, we explore the effects of “outliers” on the GARCH specification. The data are in the file named `oil.price`.
- (a) Based on the sample ACF, PACF, and EACF of the absolute and squared residuals from the fitted IMA(1,1) model (without outlier adjustment), show that a GARCH(1,1) model may be appropriate for the residuals.
 - (b) Fit an IMA(1,1) + GARCH(1,1) model to the logarithms of monthly oil prices.
 - (c) Draw the time sequence plot for the standardized residuals from the fitted IMA(1,1) + GARCH(1,1) model. Are there any outliers?
 - (d) For the log oil prices, fit an IMA(1,1) model with two IOs at $t = 2$ and $t = 56$ and an AO at $t = 8$. Show that the residuals from the IMA plus outlier model appear to be independently and identically distributed and not just serially uncorrelated; that is, there is no discernible volatility clustering.
 - (e) Between the outlier and the GARCH model, which one do you think is more appropriate for the oil price data? Explain your answer.

Appendix I: Formulas for the Generalized Portmanteau Tests

We first present the formula for $Q = (q_{i,j})$ for the case where the portmanteau test is based on the squared standardized residuals. Readers may consult Li and Mak (1994) for proofs of the formulas. Let θ denote the vector of GARCH parameters. For example, for a GARCH(1,1) model,

$$\theta = \begin{bmatrix} \omega \\ \alpha \\ \beta \end{bmatrix} \quad (12.I.1)$$

Write the i th component of θ as θ_i so that $\theta_1 = \omega$, $\theta_2 = \alpha$, and $\theta_3 = \beta$ for the GARCH(1,1) model. In the general case, let $k = p + q + 1$ be the number of GARCH parameters. Let J be an $m \times k$ matrix whose (i,j) th element equals

$$\frac{1}{n} \sum_{t=i+1}^n \frac{1}{\sigma_{t|t-1}^2} \frac{\partial \sigma_{t|t-1}^2}{\partial \theta_j} (\varepsilon_{t-i}^2 - 1) \quad (12.I.2)$$

and Λ be the $k \times k$ covariance matrix of the approximate normal distribution of the maximum likelihood estimator of θ for the model assuming normal innovations; see Section 12.4. Let $Q = (q_{i,j})$ be the matrix of the q 's appearing in the quadratic form of the generalized portmanteau test. It can be shown that the matrix Q equals

$$\left[I - \frac{1}{2(\kappa + 2)} J \Lambda J^T \right]^{-1} \quad (12.I.3)$$

where I is the $m \times m$ identity matrix, κ is the (excess) kurtosis of the innovations, J^T is the transpose of J , and the superscript -1 denotes the matrix inverse.

Next, we present the formulas for the case where the tests are computed based on the absolute standardized residuals. In this case, the (i,j) th element of the J matrix equals

$$\frac{1}{n} \sum_t \frac{1}{\sigma_{t|t-1}^2} \frac{\partial \sigma_{t|t-1}^2}{\partial \theta_j} (|\varepsilon_{t-i}| - \tau) \quad (12.I.4)$$

where $\tau = E(|\varepsilon_t|)$, and Q equals

$$\left[I - \frac{-(\kappa + 2)\tau^2/8 + \tau(v - \tau)}{(1 - \tau^2)^2} J \Lambda J^T \right]^{-1} \quad (12.I.5)$$

with $v = E(|\varepsilon_t^3|)$.