CHAPTER 3

Conditional Heteroscedastic Models

The objective of this chapter is to study some methods and econometric models available in the literature for modeling the volatility of an asset return. The models are referred to as conditional heteroscedastic models.

Volatility is an important factor in options trading. Here volatility means the conditional standard deviation of the underlying asset return. Consider, for example, the price of a European *call option*, which is a contract giving its holder the right, but not the obligation, to buy a fixed number of shares of a specified common stock at a fixed price on a given date. The fixed price is called the *strike price* and is commonly denoted by K. The given date is called the expiration date. The important time duration here is the time to expiration (measured in years), and we denote it by ℓ. The well-known Black–Scholes option pricing formula states that the price of such a call option is

$$c_t = P_t\Phi(x) - Ke^{-r\ell}\Phi(x - \sigma_t\sqrt{\ell}), \quad \text{and} \quad x = \frac{\ln(P_t/K) + r\ell}{\sigma_t\sqrt{\ell}} + \frac{1}{2}\sigma_t\sqrt{\ell}, \tag{3.1}$$

where P_t is the current price of the underlying stock, r is the continuously compounded risk-free interest rate, σ_t is the annualized conditional standard deviation of the log return of the specified stock, and $\Phi(x)$ is the cumulative distribution function of the standard normal random variable evaluated at x. A derivation of the formula is given in Chapter 6. The formula has several nice interpretations, but it suffices to say here that the conditional standard deviation of the log return of the underlying stock plays an important role. This volatility evolves over time. If the holder can exercise her right any time on or before the expiration date, then the option is called an *American call option*.

Volatility has many other financial applications. As discussed in Chapter 7, volatility modeling provides a simple approach to calculating value at risk of a financial position in risk management. It plays an important role in asset allocation

Analysis of Financial Time Series, Third Edition, By Ruey S. Tsay

under the mean-variance framework. Furthermore, modeling the volatility of a time series can improve the efficiency in parameter estimation and the accuracy in interval forecast. Finally, the volatility index of a market has recently become a financial instrument. The VIX volatility index compiled by the Chicago Board of Option Exchange (CBOE) started to trade in futures on March 26, 2004.

The univariate volatility models discussed in this chapter include the autoregressive conditional heteroscedastic (ARCH) model of Engle (1982), the generalized ARCH (GARCH) model of Bollerslev (1986), the exponential GARCH (EGARCH) model of Nelson (1991), the threshold GARCH (TGARCH) model of Glosten, Jagannathan, and Runkle (1993) and Zakoian (1994), the conditional heteroscedastic autoregressive moving-average (CHARMA) model of Tsay (1987), the random coefficient autoregressive (RCA) model of Nicholls and Quinn (1982), and the stochastic volatility (SV) models of Melino and Turnbull (1990), Taylor (1994), Harvey, Ruiz, and Shephard (1994), and Jacquier, Polson, and Rossi (1994). We also discuss advantages and weaknesses of each volatility model and show some applications of the models. Multivariate volatility models, including those with time-varying correlations, are discussed in Chapter 10. The chapter also discusses some alternative approaches to volatility modeling in Section 3.15, including use of daily high and low prices of an asset.

3.1 CHARACTERISTICS OF VOLATILITY

A special feature of stock volatility is that it is not directly observable. For example, consider the daily log returns of IBM stock. The daily volatility is not directly observable from the return data because there is only one observation in a trading day. If intraday data of the stock, such as 10-minute returns, are available, then one can estimate the daily volatility. See Section 3.15. The accuracy of such an estimate deserves a careful study, however. For example, stock volatility consists of intraday volatility and overnight volatility with the latter denoting variation between trading days. The high-frequency intraday returns contain only very limited information about the overnight volatility. The unobservability of volatility makes it difficult to evaluate the forecasting performance of conditional heteroscedastic models. We discuss this issue later.

In options markets, if one accepts the idea that the prices are governed by an econometric model such as the Black–Scholes formula, then one can use the price to obtain the "implied" volatility. Yet this approach is often criticized for using a specific model, which is based on some assumptions that might not hold in practice. For instance, from the observed prices of a European call option, one can use the Black–Scholes formula in Eq. (3.1) to deduce the conditional standard deviation σ_t. The resulting value of σ_t is called the *implied volatility* of the underlying stock. However, this implied volatility is derived under the assumption that the price of the underlying asset follows a geometric Brownian motion. It might be different from the actual volatility. Experience shows that implied volatility of an asset return tends to be larger than that obtained by using a GARCH type of volatility model.

This might be due to the risk premium for volatility or to the way daily returns are calculated. The VIX of CBOE is an implied volatility.

Although volatility is not directly observable, it has some characteristics that are commonly seen in asset returns. First, there exist volatility clusters (i.e., volatility may be high for certain time periods and low for other periods). Second, volatility evolves over time in a continuous manner—that is, volatility jumps are rare. Third, volatility does not diverge to infinity—that is, volatility varies within some fixed range. Statistically speaking, this means that volatility is often stationary. Fourth, volatility seems to react differently to a big price increase or a big price drop, referred to as the *leverage* effect. These properties play an important role in the development of volatility models. Some volatility models were proposed specifically to correct the weaknesses of the existing ones for their inability to capture the characteristics mentioned earlier. For example, the EGARCH model was developed to capture the asymmetry in volatility induced by big "positive" and "negative" asset returns.

3.2 STRUCTURE OF A MODEL

Let r_t be the log return of an asset at time index t. The basic idea behind volatility study is that the series $\{r_t\}$ is either serially uncorrelated or with minor lower order serial correlations, but it is a dependent series. For illustration, consider the monthly log stock returns of Intel Corporation from January 1973 to December 2008 shown in Figure 3.1. Figure 3.2(a) shows the sample ACF of the log return series, which suggests no significant serial correlations except for a minor one at lag 7. Figure 3.2(c) shows the sample ACF of the absolute log returns (i.e., $|r_t|$), whereas Figure 3.2(b) shows the sample ACF of the squared log returns r_t^2. These two plots clearly suggest that the monthly log returns are not serially independent. Combining the three plots, it seems that the log returns are indeed serially uncorrelated but dependent. Volatility models attempt to capture such dependence in the return series.

To put the volatility models in proper perspective, it is informative to consider the conditional mean and variance of r_t given F_{t-1}; that is,

$$\mu_t = E(r_t|F_{t-1}), \qquad \sigma_t^2 = \mathrm{Var}(r_t|F_{t-1}) = E[(r_t - \mu_t)^2|F_{t-1}], \tag{3.2}$$

where F_{t-1} denotes the information set available at time $t-1$. Typically, F_{t-1} consists of all linear functions of the past returns. As shown by the empirical examples of Chapter 2 and Figure 3.2, serial dependence of a stock return series r_t is weak if it exists at all. Therefore, the equation for μ_t in (3.2) should be simple, and we assume that r_t follows a simple time series model such as a stationary ARMA(p, q) model with some explanatory variables. In other words, we entertain the model

$$r_t = \mu_t + a_t, \quad \mu_t = \sum_{i=1}^{p} \phi_i y_{t-i} - \sum_{i=1}^{q} \theta_i a_{t-i}, \quad y_t = r_t - \phi_0 - \sum_{i=1}^{k} \beta_i x_{it}, \tag{3.3}$$

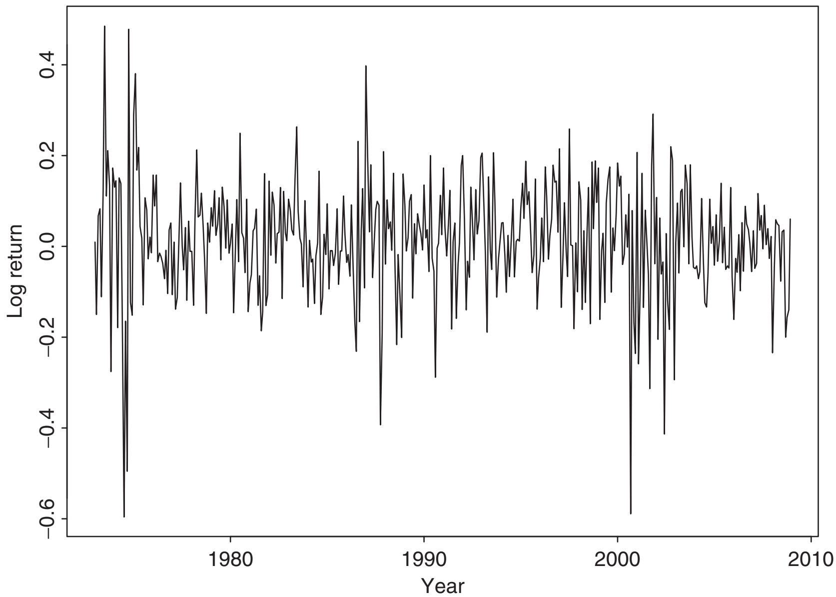

Figure 3.1 Time plot of monthly log returns of Intel stock from January 1973 to December 2008.

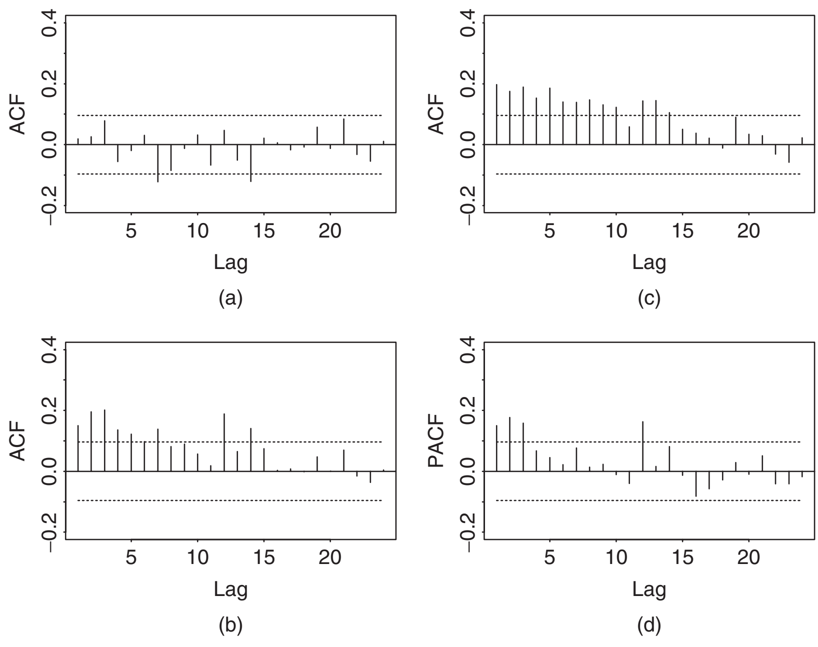

Figure 3.2 Sample ACF and PACF of various functions of monthly log stock returns of Intel Corporation from January 1973 to December 2008: (a) ACF of the log returns, (b) ACF of the squared log returns, (c) ACF of the absolute log returns, and (d) PACF of the squared log returns.

for r_t, where k, p, and q are nonnegative integers, and x_{it} are explanatory variables. Here y_t is simply a notation representing the adjusted return series after removing the effect of explanatory variables.

Model (3.3) illustrates a possible financial application of the regression model with time series errors of Chapter 2. The order (p, q) of an ARMA model may depend on the frequency of the return series. For example, daily returns of a market index often show some minor serial correlations, but monthly returns of the index may not contain any significant serial correlation. The explanatory variables x_t in Eq. (3.3) are flexible. For example, a dummy variable can be used for the Mondays to study the effect of the weekend on daily stock returns. In the capital asset pricing model (CAPM), the mean equation of r_t can be written as $r_t = \phi_0 + \beta r_{m,t} + a_t$, where $r_{m,t}$ denotes the market return.

Combining Eqs. (3.2) and (3.3), we have

$$\sigma_t^2 = \text{Var}(r_t|F_{t-1}) = \text{Var}(a_t|F_{t-1}). \tag{3.4}$$

The conditional heteroscedastic models of this chapter are concerned with the evolution of σ_t^2. The manner under which σ_t^2 evolves over time distinguishes one volatility model from another.

Conditional heteroscedastic models can be classified into two general categories. Those in the first category use an exact function to govern the evolution of σ_t^2, whereas those in the second category use a stochastic equation to describe σ_t^2. The GARCH model belongs to the first category, whereas the stochastic volatility model is in the second category.

Throughout the book, a_t is referred to as the *shock* or *innovation* of an asset return at time t and σ_t is the positive square root of σ_t^2. The model for μ_t in Eq. (3.3) is referred to as the *mean* equation for r_t and the model for σ_t^2 is the *volatility* equation for r_t. Therefore, modeling conditional heteroscedasticity amounts to augmenting a dynamic equation, which governs the time evolution of the conditional variance of the asset return, to a time series model.

3.3 MODEL BUILDING

Building a volatility model for an asset return series consists of four steps:

1. Specify a mean equation by testing for serial dependence in the data and, if necessary, building an econometric model (e.g., an ARMA model) for the return series to remove any linear dependence.
2. Use the residuals of the mean equation to test for ARCH effects.
3. Specify a volatility model if ARCH effects are statistically significant, and perform a joint estimation of the mean and volatility equations.
4. Check the fitted model carefully and refine it if necessary.

For most asset return series, the serial correlations are weak, if any. Thus, building a mean equation amounts to removing the sample mean from the data if the sample mean is significantly different from zero. For some daily return series, a simple AR model might be needed. In some cases, the mean equation may employ some explanatory variables such as an indicator variable for weekend or January effects.

In what follows, we use R (both with and without OX) and S-Plus in empirical illustrations. Other software packages (e.g., Eviews, SCA, and RATS) can also be used.

3.3.1 Testing for ARCH Effect

For ease in notation, let $a_t = r_t - \mu_t$ be the residuals of the mean equation. The squared series a_t^2 is then used to check for conditional heteroscedasticity, which is also known as the *ARCH* effects. Two tests are available. The first test is to apply the usual Ljung–Box statistics $Q(m)$ to the $\{a_t^2\}$ series; see McLeod and Li (1983). The null hypothesis is that the first m lags of ACF of the a_t^2 series are zero. The second test for conditional heteroscedasticity is the Lagrange multiplier test of Engle (1982). This test is equivalent to the usual F statistic for testing $\alpha_i = 0$ $(i = 1, \ldots, m)$ in the linear regression

$$a_t^2 = \alpha_0 + \alpha_1 a_{t-1}^2 + \cdots + \alpha_m a_{t-m}^2 + e_t, \qquad t = m+1, \ldots, T,$$

where e_t denotes the error term, m is a prespecified positive integer, and T is the sample size. Specifically, the null hypothesis is $H_0 : \alpha_1 = \cdots = \alpha_m = 0$. Let $\text{SSR}_0 = \sum_{t=m+1}^{T}(a_t^2 - \bar{\omega})^2$, where $\bar{\omega} = (1/T)\sum_{t=1}^{T} a_t^2$ is the sample mean of a_t^2, and $\text{SSR}_1 = \sum_{t=m+1}^{T} \hat{e}_t^2$, where $\hat{e}_t$ is the least-squares residual of the prior linear regression. Then we have

$$F = \frac{(\text{SSR}_0 - \text{SSR}_1)/m}{\text{SSR}_1/(T - 2m - 1)},$$

which is asymptotically distributed as a chi-squared distribution with m degrees of freedom under the null hypothesis. The decision rule is to reject the null hypothesis if $F > \chi_m^2(\alpha)$, where $\chi_m^2(\alpha)$ is the upper $100(1-\alpha)$th percentile of χ_m^2, or the p value of F is less than α, type-I error.

To demonstrate, we consider the monthly log stock returns of Intel Corporation from 1973 to 2008; see Example 3.1. The series does not have significant serial correlations so that it can be directly used to test for the ARCH effect. Indeed, the $Q(m)$ statistics of the return series give $Q(12) = 18.26$ with a p value of 0.11, confirming no serial correlations in the data. On the other hand, the Lagrange multiplier test shows strong ARCH effects with test statistic $F \approx 53.62$, the p value of which is close to zero. The Ljung–Box statistics of the a_t series also shows strong ARCH effects with $Q(12) = 89.85$, the p value of which is close to zero.

S-Plus Demonstration

Denote the return series by `intc`. Note that the command `archTest` applies directly to the a_t series, not to a_t^2.

```
> da=read.table("m-intc7308.txt",header=T)
> intc=log(da[,2]+1)
> autocorTest(intc,lag=12)
Test for Autocorrelation: Ljung-Box
Null Hypothesis: no autocorrelation

Test Statistics:
Test Stat 18.2635  p.value  0.1079

Dist. under Null: chi-square with 12 degrees of freedom
   Total Observ.: 432

> archTest(intc,lag=12)
Test for ARCH Effects: LM Test
Null Hypothesis: no ARCH effects

Test Statistics:
Test Stat 53.6197  p.value  0.0000

Dist. under Null: chi-square with 12 degrees of freedom
```

R Demonstration

```
> da=read.table("m-intc7308.txt",header=T)
> intc=log(da[,2]+1)
> Box.test(intc,lag=12,type='Ljung')
        Box-Ljung test

data:  intc
X-squared = 18.2635, df = 12, p-value = 0.1079

> at=intc-mean(intc)
> Box.test(at^2,lag=12,type='Ljung')
        Box-Ljung test

data:  at^2
X-squared = 89.8509, df = 12, p-value = 5.274e-14
```

3.4 THE ARCH MODEL

The first model that provides a systematic framework for volatility modeling is the ARCH model of Engle (1982). The basic idea of ARCH models is that (a) the shock a_t of an asset return is serially uncorrelated, but dependent, and

(b) the dependence of a_t can be described by a simple quadratic function of its lagged values. Specifically, an ARCH(m) model assumes that

$$a_t = \sigma_t \epsilon_t, \qquad \sigma_t^2 = \alpha_0 + \alpha_1 a_{t-1}^2 + \cdots + \alpha_m a_{t-m}^2, \tag{3.5}$$

where $\{\epsilon_t\}$ is a sequence of independent and identically distributed (iid) random variables with mean zero and variance 1, $\alpha_0 > 0$, and $\alpha_i \geq 0$ for $i > 0$. The coefficients α_i must satisfy some regularity conditions to ensure that the unconditional variance of a_t is finite. In practice, ϵ_t is often assumed to follow the standard normal or a standardized Student-t or a generalized error distribution.

From the structure of the model, it is seen that large past squared shocks $\{a_{t-i}^2\}_{i=1}^m$ imply a large conditional variance σ_t^2 for the innovation a_t. Consequently, a_t tends to assume a large value (in modulus). This means that, under the ARCH framework, large shocks tend to be followed by another large shock. Here I use the word *tend* because a large variance does not necessarily produce a large realization. It only says that the probability of obtaining a large variate is greater than that of a smaller variance. This feature is similar to the volatility clusterings observed in asset returns.

The ARCH effect also occurs in other financial time series. Figure 3.3 shows the time plots of (a) the percentage changes in Deutsche mark/U.S. dollar exchange rate measured in 10-minute intervals from June 5, 1989, to June 19, 1989, for 2488 observations, and (b) the squared series of the percentage changes. Big percentage

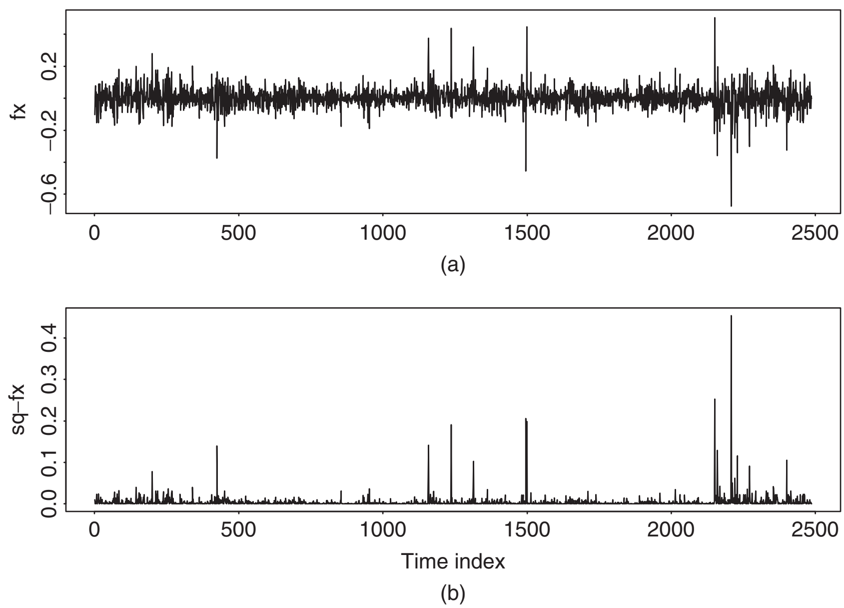

Figure 3.3 (a) Time plot of 10-minute returns of exchange rate between Deutsche mark and U.S. dollar from June 5, 1989, to June 19, 1989, and (b) the squared returns.

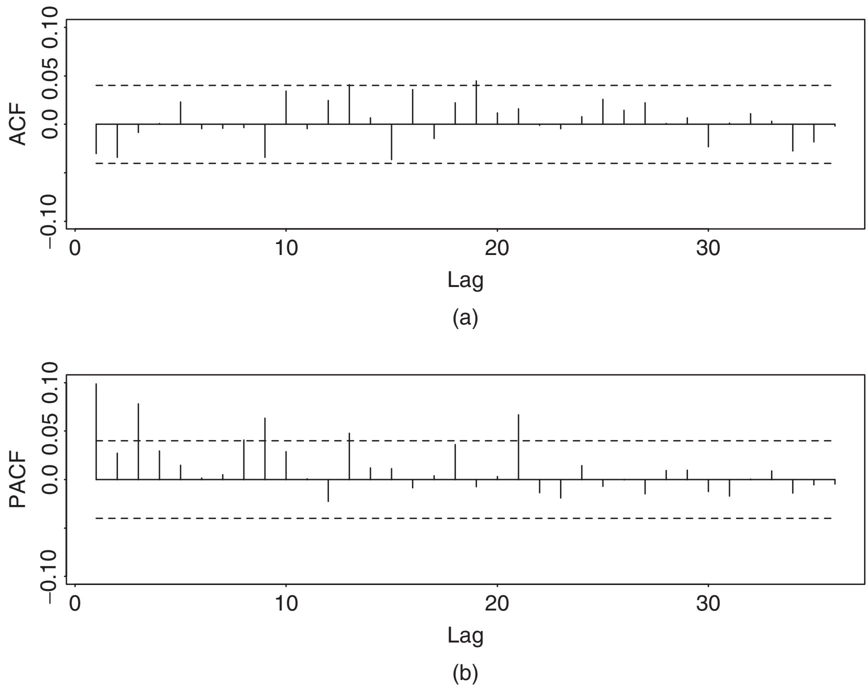

Figure 3.4 (a) Sample autocorrelation function of return series of mark/dollar exchange rate and (b) sample partial autocorrelation function of squared returns.

changes occurred occasionally, but there were certain stable periods. Figure 3.4(a) shows the sample ACF of the percentage change series. Clearly, the series has no serial correlation. Figure 3.4(b) shows the sample PACF of the squared series of percentage change. It is seen that there are some big spikes in the PACF. Such spikes suggest that the percentage changes are not serially independent and have some ARCH effects.

Remark. Some authors use h_t to denote the conditional variance in Eq. (3.5). In this case, the shock becomes $a_t = \sqrt{h_t}\epsilon_t$. □

3.4.1 Properties of ARCH Models

To understand the ARCH models, it pays to carefully study the ARCH(1) model

$$a_t = \sigma_t \epsilon_t, \qquad \sigma_t^2 = \alpha_0 + \alpha_1 a_{t-1}^2,$$

where $\alpha_0 > 0$ and $\alpha_1 \geq 0$. First, the unconditional mean of a_t remains zero because

$$E(a_t) = E[E(a_t|F_{t-1})] = E[\sigma_t E(\epsilon_t)] = 0.$$

Second, the unconditional variance of a_t can be obtained as

$$\begin{aligned}\mathrm{Var}(a_t) &= E(a_t^2) = E[E(a_t^2|F_{t-1})]\\ &= E(\alpha_0 + \alpha_1 a_{t-1}^2) = \alpha_0 + \alpha_1 E(a_{t-1}^2).\end{aligned}$$

Because a_t is a stationary process with $E(a_t) = 0$, $\mathrm{Var}(a_t) = \mathrm{Var}(a_{t-1}) = E(a_{t-1}^2)$. Therefore, we have $\mathrm{Var}(a_t) = \alpha_0 + \alpha_1 \mathrm{Var}(a_t)$ and $\mathrm{Var}(a_t) = \alpha_0/(1-\alpha_1)$. Since the variance of a_t must be positive, we require $0 \le \alpha_1 < 1$. Third, in some applications, we need higher order moments of a_t to exist and, hence, α_1 must also satisfy some additional constraints. For instance, to study its tail behavior, we require that the fourth moment of a_t is finite. Under the normality assumption of ϵ_t in Eq. (3.5), we have

$$E(a_t^4|F_{t-1}) = 3[E(a_t^2|F_{t-1})]^2 = 3(\alpha_0 + \alpha_1 a_{t-1}^2)^2.$$

Therefore,

$$E(a_t^4) = E[E(a_t^4|F_{t-1})] = 3E(\alpha_0 + \alpha_1 a_{t-1}^2)^2 = 3E\left(\alpha_0^2 + 2\alpha_0\alpha_1 a_{t-1}^2 + \alpha_1^2 a_{t-1}^4\right).$$

If a_t is fourth-order stationary with $m_4 = E(a_t^4)$, then we have

$$\begin{aligned}m_4 &= 3[\alpha_0^2 + 2\alpha_0\alpha_1 \mathrm{Var}(a_t) + \alpha_1^2 m_4]\\ &= 3\alpha_0^2\left(1 + 2\frac{\alpha_1}{1-\alpha_1}\right) + 3\alpha_1^2 m_4.\end{aligned}$$

Consequently,

$$m_4 = \frac{3\alpha_0^2(1+\alpha_1)}{(1-\alpha_1)(1-3\alpha_1^2)}.$$

This result has two important implications: (a) since the fourth moment of a_t is positive, we see that α_1 must also satisfy the condition $1 - 3\alpha_1^2 > 0$; that is, $0 \le \alpha_1^2 < \frac{1}{3}$; and (b) the unconditional kurtosis of a_t is

$$\frac{E(a_t^4)}{[\mathrm{Var}(a_t)]^2} = 3\frac{\alpha_0^2(1+\alpha_1)}{(1-\alpha_1)(1-3\alpha_1^2)} \times \frac{(1-\alpha_1)^2}{\alpha_0^2} = 3\frac{1-\alpha_1^2}{1-3\alpha_1^2} > 3.$$

Thus, the excess kurtosis of a_t is positive and the tail distribution of a_t is heavier than that of a normal distribution. In other words, the shock a_t of a conditional Gaussian ARCH(1) model is more likely than a Gaussian white noise series to produce "outliers." This is in agreement with the empirical finding that "outliers" appear more often in asset returns than that implied by an iid sequence of normal random variates.

These properties continue to hold for general ARCH models, but the formulas become more complicated for higher order ARCH models. The condition $\alpha_i \ge 0$ in

Eq. (3.5) can be relaxed. It is a condition to ensure that the conditional variance σ_t^2 is positive for all t. In fact, a natural way to achieve positiveness of the conditional variance is to rewrite an ARCH(m) model as

$$a_t = \sigma_t \epsilon_t, \qquad \sigma_t^2 = \alpha_0 + A'_{m,t-1} \Omega A_{m,t-1}, \tag{3.6}$$

where $A_{m,t-1} = (a_{t-1}, \ldots, a_{t-m})'$ and Ω is an $m \times m$ nonnegative definite matrix. The ARCH(m) model in Eq. (3.5) requires Ω to be diagonal. Thus, Engle's model uses a parsimonious approach to approximate a quadratic function. A simple way to achieve Eq. (3.6) is to employ a random-coefficient model for a_t; see the CHARMA and RCA models discussed later.

3.4.2 Weaknesses of ARCH Models

The advantages of ARCH models include properties discussed in the previous section. The model also has some weaknesses:

1. The model assumes that positive and negative shocks have the same effects on volatility because it depends on the square of the previous shocks. In practice, it is well known that the price of a financial asset responds differently to positive and negative shocks.
2. The ARCH model is rather restrictive. For instance, α_1^2 of an ARCH(1) model must be in the interval $[0, \frac{1}{3}]$ if the series has a finite fourth moment. The constraint becomes complicated for higher order ARCH models. In practice, it limits the ability of ARCH models with Gaussian innovations to capture excess kurtosis.
3. The ARCH model does not provide any new insight for understanding the source of variations of a financial time series. It merely provides a mechanical way to describe the behavior of the conditional variance. It gives no indication about what causes such behavior to occur.
4. ARCH models are likely to overpredict the volatility because they respond slowly to large isolated shocks to the return series.

3.4.3 Building an ARCH Model

Among volatility models, specifying an ARCH model is relatively easy. Details are given below.

Order Determination

If an ARCH effect is found to be significant, one can use the PACF of a_t^2 to determine the ARCH order. Using PACF of a_t^2 to select the ARCH order can be justified as follows. From the model in Eq. (3.5), we have

$$\sigma_t^2 = \alpha_0 + \alpha_1 a_{t-1}^2 + \cdots + \alpha_m a_{t-m}^2.$$

For a given sample, a_t^2 is an unbiased estimate of σ_t^2. Therefore, we expect that a_t^2 is linearly related to $a_{t-1}^2, \ldots, a_{t-m}^2$ in a manner similar to that of an autoregressive model of order m. Note that a single a_t^2 is generally not an efficient estimate of σ_t^2, but it can serve as an approximation that could be informative in specifying the order m.

Alternatively, define $\eta_t = a_t^2 - \sigma_t^2$. It can be shown that $\{\eta_t\}$ is an uncorrelated series with mean 0. The ARCH model then becomes

$$a_t^2 = \alpha_0 + \alpha_1 a_{t-1}^2 + \cdots + \alpha_m a_{t-m}^2 + \eta_t,$$

which is in the form of an AR(m) model for a_t^2, except that $\{\eta_t\}$ is not an iid series. From Chapter 2, PACF of a_t^2 is a useful tool to determine the order m. Because $\{\eta_t\}$ are not identically distributed, the least-squares estimates of the prior model are consistent but not efficient. The PACF of a_t^2 may not be effective when the sample size is small.

Estimation

Several likelihood functions are commonly used in ARCH estimation, depending on the distributional assumption of ϵ_t. Under the normality assumption, the likelihood function of an ARCH(m) model is

$$\begin{aligned} f(a_1, \ldots, a_T|\boldsymbol{\alpha}) &= f(a_T|F_{T-1})f(a_{T-1}|F_{T-2}) \cdots f(a_{m+1}|F_m)f(a_1, \cdots, a_m|\boldsymbol{\alpha}) \\ &= \prod_{t=m+1}^{T} \frac{1}{\sqrt{2\pi\sigma_t^2}} \exp\left(-\frac{a_t^2}{2\sigma_t^2}\right) \times f(a_1, \ldots, a_m|\boldsymbol{\alpha}), \end{aligned}$$

where $\boldsymbol{\alpha} = (\alpha_0, \alpha_1, \ldots, \alpha_m)'$ and $f(a_1, \ldots, a_m|\boldsymbol{\alpha})$ is the joint probability density function of $a_1, \ldots, a_m$. Since the exact form of $f(a_1, \ldots, a_m|\boldsymbol{\alpha})$ is complicated, it is commonly dropped from the prior likelihood function, especially when the sample size is sufficiently large. This results in using the conditional-likelihood function

$$f(a_{m+1}, \ldots, a_T|\boldsymbol{\alpha}, a_1, \ldots, a_m) = \prod_{t=m+1}^{T} \frac{1}{\sqrt{2\pi\sigma_t^2}} \exp\left(-\frac{a_t^2}{2\sigma_t^2}\right),$$

where σ_t^2 can be evaluated recursively. We refer to estimates obtained by maximizing the prior likelihood function as the conditional maximum-likelihood estimates (MLEs) under normality.

Maximizing the conditional-likelihood function is equivalent to maximizing its logarithm, which is easier to handle. The conditional log-likelihood function is

$$\ell(a_{m+1}, \ldots, a_T|\boldsymbol{\alpha}, a_1, \ldots, a_m) = \sum_{t=m+1}^{T} \left[-\frac{1}{2}\ln(2\pi) - \frac{1}{2}\ln(\sigma_t^2) - \frac{1}{2}\frac{a_t^2}{\sigma_t^2}\right].$$

Since the first term $\ln(2\pi)$ does not involve any parameters, the log-likelihood function becomes

$$\ell(a_{m+1}, \ldots, a_T|\boldsymbol{\alpha}, a_1, \ldots, a_m) = -\sum_{t=m+1}^{T}\left[\frac{1}{2}\ln(\sigma_t^2) + \frac{1}{2}\frac{a_t^2}{\sigma_t^2}\right],$$

where $\sigma_t^2 = \alpha_0 + \alpha_1 a_{t-1}^2 + \cdots + \alpha_m a_{t-m}^2$ can be evaluated recursively.

In some applications, it is more appropriate to assume that ϵ_t follows a heavy-tailed distribution such as a standardized Student-t distribution. Let x_v be a Student-t distribution with v degrees of freedom. Then $\mathrm{Var}(x_v) = v/(v-2)$ for $v > 2$, and we use $\epsilon_t = x_v/\sqrt{v/(v-2)}$. The probability density function of ϵ_t is

$$f(\epsilon_t|v) = \frac{\Gamma[(v+1)/2]}{\Gamma(v/2)\sqrt{(v-2)\pi}}\left(1 + \frac{\epsilon_t^2}{v-2}\right)^{-(v+1)/2}, \qquad v > 2, \tag{3.7}$$

where $\Gamma(x)$ is the usual gamma function (i.e., $\Gamma(x) = \int_0^\infty y^{x-1}e^{-y}\,dy$). Using $a_t = \sigma_t\epsilon_t$, we obtain the conditional-likelihood function of a_t as

$$f(a_{m+1}, \ldots, a_T|\boldsymbol{\alpha}, A_m) = \prod_{t=m+1}^{T}\frac{\Gamma[(v+1)/2]}{\Gamma(v/2)\sqrt{(v-2)\pi}}\frac{1}{\sigma_t}\left[1 + \frac{a_t^2}{(v-2)\sigma_t^2}\right]^{-(v+1)/2},$$

where $v > 2$ and $A_m = (a_1, a_2, \ldots, a_m)$. We refer to the estimates that maximize the prior likelihood function as the conditional MLEs under t distribution. The degrees of freedom of the t distribution can be specified a priori or estimated jointly with other parameters. A value between 4 and 8 is often used if it is prespecified.

If the degrees of freedom v of the Student-t distribution is prespecified, then the conditional log-likelihood function is

$$\ell(a_{m+1}, \ldots, a_T|\boldsymbol{\alpha}, A_m) = -\sum_{t=m+1}^{T}\left[\frac{v+1}{2}\ln\left(1 + \frac{a_t^2}{(v-2)\sigma_t^2}\right) + \frac{1}{2}\ln(\sigma_t^2)\right]. \tag{3.8}$$

If one wishes to estimate v jointly with other parameters, then the log-likelihood function becomes

$$\begin{aligned}\ell(a_{m+1}, \ldots, a_T|\boldsymbol{\alpha}, v, A_m) = (T-m)\Bigg\{&\ln\left[\Gamma\left(\frac{v+1}{2}\right)\right] - \ln\left[\Gamma\left(\frac{v}{2}\right)\right] \\ &- 0.5\ln[(v-2)\pi]\Bigg\} + \ell(a_{m+1}, \ldots, a_T|\boldsymbol{\alpha}, A_m),\end{aligned}$$

where the second term is given in Eq. (3.8).

Besides fat tails, empirical distributions of asset returns may also be skewed. To handle this additional characteristic of asset returns, the Student-t distribution has been modified to become a skew-Student-t distribution. There are multiple versions of skew-Student-t distribution, but we shall adopt the approach of Fernández and Steel (1998), which can introduce skewness into any continuous unimodal and symmetric (with respect to 0) univariate distribution. Specifically, for the innovation ϵ_t of an ARCH process, Lambert and Laurent (2001) apply the Fernández and Steel method to the standardized Student-t distribution in Eq. (3.7) to obtain a standardized skew-Student-t distribution. The resulting probability density function is

$$g(\epsilon_t|\xi, v) = \begin{cases} \dfrac{2}{\xi + \frac{1}{\xi}} \varrho f[\xi(\varrho\epsilon_t + \overline{\omega})|v] & \text{if } \epsilon_t < -\overline{\omega}/\varrho, \\ \dfrac{2}{\xi + \frac{1}{\xi}} \varrho f[(\varrho\epsilon_t + \overline{\omega})/\xi|v] & \text{if } \epsilon_t \geq -\overline{\omega}/\varrho, \end{cases} \tag{3.9}$$

where $f(\cdot)$ is the probability density function (pdf) of the standardized Student-t distribution in Eq. (3.7), ξ is the skewness parameter, $v > 2$ is the degrees of freedom, and the parameters ϱ and $\overline{\omega}$ are given below:

$$\varpi = \frac{\Gamma[(v-1)/2]\sqrt{v-2}}{\sqrt{\pi}\Gamma(v/2)}(\xi - \frac{1}{\xi}), \qquad \varrho^2 = (\xi^2 + \frac{1}{\xi^2} - 1) - \varpi^2.$$

In Eq. (3.9), ξ^2 is equal to the ratio of probability masses above and below the mode of the distribution and, hence, it is a measure of the skewness.

Finally, ϵ_t may assume a generalized error distribution (GED) with probability density function

$$f(x) = \frac{v \exp(-\frac{1}{2}|x/\lambda|^v)}{\lambda 2^{(1+1/v)}\Gamma(1/v)}, \qquad -\infty < x < \infty, \qquad 0 < v \leq \infty, \tag{3.10}$$

where $\Gamma(\cdot)$ is the gamma function and $\lambda = [2^{(-2/v)}\Gamma(1/v)/\Gamma(3/v)]^{1/2}$. This distribution reduces to a Gaussian distribution if $v = 2$, and it has heavy tails when $v < 2$. The conditional log-likelihood function $\ell(a_{m+1}, \ldots, a_T|\boldsymbol{\alpha}, \boldsymbol{A}_m)$ can easily be obtained.

Remark. Skew Student-t, skew normal, and skew GED distributions are available in the `fGarch` package of `Rmetrics`. The commands are `sstd`, `snorm`, and `sged`, respectively. See the R demonstration below for an example. □

Model Checking

For a properly specified ARCH model, the standardized residuals

$$\tilde{a}_t = \frac{a_t}{\sigma_t}$$

form a sequence of iid random variables. Therefore, one can check the adequacy of a fitted ARCH model by examining the series $\{\tilde{a}_t\}$. In particular, the Ljung–Box statistics of $\tilde{a}_t$ can be used to check the adequacy of the mean equation and that of $\tilde{a}_t^2$ can be used to test the validity of the volatility equation. The skewness, kurtosis, and quantile-to-quantile plot (i.e., QQ plot) of $\{\tilde{a}_t\}$ can be used to check the validity of the distribution assumption. Many residual plots are available in S-Plus for model checking.

Forecasting

Forecasts of the ARCH model in Eq. (3.5) can be obtained recursively as those of an AR model. Consider an ARCH(m) model. At the forecast origin h, the 1-step-ahead forecast of σ_{h+1}^2 is

$$\sigma_h^2(1) = \alpha_0 + \alpha_1 a_h^2 + \cdots + \alpha_m a_{h+1-m}^2.$$

The 2-step-ahead forecast is

$$\sigma_h^2(2) = \alpha_0 + \alpha_1 \sigma_h^2(1) + \alpha_2 a_h^2 + \cdots + \alpha_m a_{h+2-m}^2,$$

and the ℓ-step-ahead forecast for $\sigma_{h+\ell}^2$ is

$$\sigma_h^2(\ell) = \alpha_0 + \sum_{i=1}^{m} \alpha_i \sigma_h^2(\ell - i), \tag{3.11}$$

where $\sigma_h^2(\ell - i) = a_{h+\ell-i}^2$ if $\ell - i \le 0$.

3.4.4 Some Examples

In this section, we illustrate ARCH modeling by considering two examples.

Example 3.1. We first apply the modeling procedure to build a simple ARCH model for the monthly log returns of Intel stock. The sample ACF and PACF of the squared returns in Figure 3.2 clearly show the existence of conditional heteroscedasticity. This is confirmed by the ARCH test shown in Section 3.3.1, and we proceed to identify the order of an ARCH model. The sample PACF in Figure 3.2(d) indicates that an ARCH(3) model might be appropriate. Consequently, we specify the model

$$r_t = \mu + a_t, \qquad a_t = \sigma_t \epsilon_t, \qquad \sigma_t^2 = \alpha_0 + \alpha_1 a_{t-1}^2 + \alpha_2 a_{t-2}^2 + \alpha_3 a_{t-3}^2$$

for the monthly log returns of Intel stock. Assuming that ϵ_t are iid standard normal, we obtain the fitted model

$$r_t = 0.0122 + a_t, \qquad \sigma_t^2 = 0.0106 + 0.2131 a_{t-1}^2 + 0.0770 a_{t-2}^2 + 0.0599 a_{t-3}^2,$$

where the standard errors of the parameters are 0.0057, 0.0010, 0.0757, 0.0480, and 0.0688, respectively; see the output below. While the estimates meet the general requirement of an ARCH(3) model, the estimates of α_2 and α_3 appear to be statistically nonsignificant at the 5% level. Therefore, the model can be simplified.

S-Plus Demonstration

The following output has been edited and % marks explanation:

```
> module(finmetrics)
> da=read.table("m-intc7308.txt",header=T)
> intc=log(da[,2]+1)
> arch3.fit=garch(intc~1,~garch(3,0))
> summary(arch3.fit)
garch(formula.mean = intc ~ 1, formula.var = ~ garch(3, 0))

Mean Equation: structure(.Data = intc ~  1, class = "for-
mula")
Conditional Variance Equation:structure(.Data=~garch(3,0),..)
Conditional Distribution:  gaussian
--------------------------------------------------------------
Estimated Coefficients:
--------------------------------------------------------------
            Value Std.Error t value Pr(>|t|)
      C 0.01216 0.0056986  2.1341 0.033402
      A 0.01058 0.0009643 10.9739 0.000000
ARCH(1) 0.21307 0.0756708  2.8157 0.005093
ARCH(2) 0.07698 0.0480170  1.6032 0.109638
ARCH(3) 0.05988 0.0688081  0.8703 0.384628
--------------------------------------------------------------
> arch1=garch(intc~1,~garch(1,0))
> summary(arch1)
garch(formula.mean = intc ~ 1, formula.var = ~ garch(1, 0))

Conditional Distribution:  gaussian
--------------------------------------------------------------
Estimated Coefficients:
--------------------------------------------------------------
            Value Std.Error t value  Pr(>|t|)
      C 0.01261 0.0052624   2.397 1.695e-02
      A 0.01113 0.0009971  11.164 0.000e+00
ARCH(1) 0.35602 0.0761267   4.677 3.912e-06
--------------------------------------------------------------
AIC(3) = -570.0179, BIC(3) = -557.8126

Ljung-Box test for standardized residuals:
--------------------------------------------------------------
Statistic P-value Chi^2-d.f.
     14.26  0.2844         12

Ljung-Box test for squared standardized residuals:
--------------------------------------------------------------
Statistic  P-value Chi^2-d.f.
     32.11 0.001329         12
> stres=arch1$residuals/arch1$sigma.t %standardized residuals
```

```
> autocorTest(stres,lag=10)
Test for Autocorrelation: Ljung-Box

Null Hypothesis: no autocorrelation
Test Statistics:
Test Stat 12.6386,  p.value  0.2446

Dist. under Null: chi-square with 10 degrees of freedom
> archTest(stres,lag=10)
Test for ARCH Effects: LM Test

Null Hypothesis: no ARCH effects
Test Statistics:
Test Stat 14.7481,  p.value  0.1415

Dist. under Null: chi-square with 10 degrees of freedom
> arch1$asymp.sd  %Obtain unconditional standard error

[1] 0.1314698
> plot(arch1)  % Obtain various plots, including the
               % fitted volatility series.
```

Dropping the two nonsignificant parameters, we obtain the model

$$r_t = 0.0126 + a_t, \qquad \sigma_t^2 = 0.0111 + 0.3560a_{t-1}^2, \tag{3.12}$$

where the standard errors of the parameters are 0.0053, 0.0010, and 0.0761, respectively. All the estimates are highly significant. Figure 3.5 shows the standardized residuals $\{\tilde{a}_t\}$ and the sample ACF of some functions of $\{\tilde{a}_t\}$. The Ljung–Box statistics of standardized residuals give $Q(10) = 12.64$ with a p value of 0.24 and those of $\{\tilde{a}_t^2\}$ give $Q(10) = 14.75$ with a p value of 0.14. See the output. Consequently, the ARCH(1) model in Eq. (3.12) is adequate for describing the conditional heteroscedasticity of the data at the 5% significance level.

The ARCH(1) model in Eq. (3.12) has some interesting properties. First, the expected monthly log return for Intel stock is about 1.26%, which is remarkable, especially since the data span includes the period after the Internet bubble. Second, $\hat{\alpha}_1^2 = 0.356^2 < \frac{1}{3}$ so that the unconditional fourth moment of the monthly log return of Intel stock exists. Third, the unconditional standard deviation of r_t is $\sqrt{0.0111/(1-0.356)} \approx 0.1315$. Finally, the ARCH(1) model can be used to predict the monthly volatility of Intel stock returns.

***t* Innovation**

For comparison, we also fit an ARCH(1) model with Student-t innovations to the series. The resulting model is

$$r_t = 0.0169 + a_t, \qquad \sigma_t^2 = 0.0120 + 0.2845a_{t-1}^2, \tag{3.13}$$

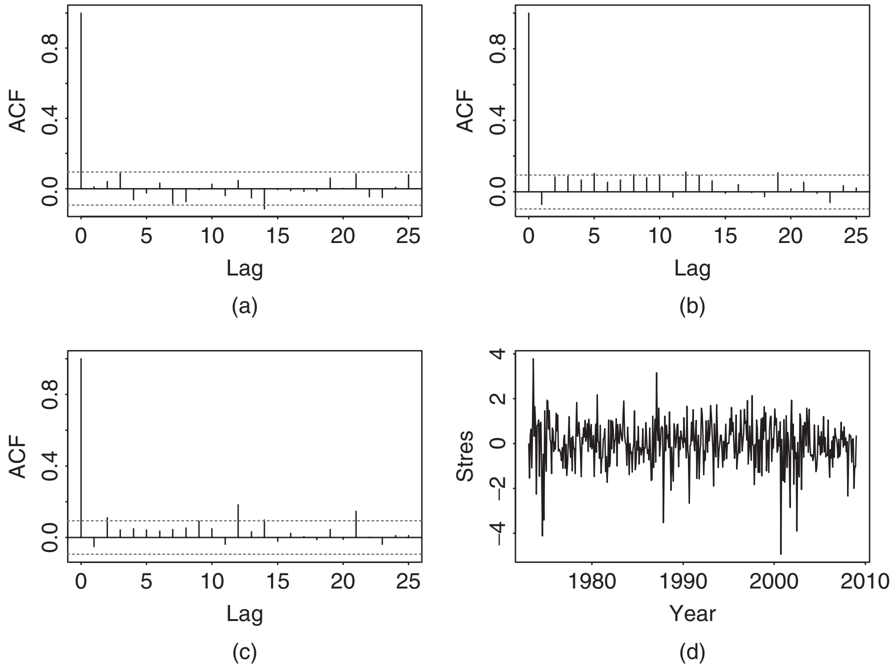

Figure 3.5 Model checking statistics of Gaussian ARCH(1) model in Eq. (3.12) for monthly log returns of Intel stock from January 1973 to December 2008: Parts (a), (b), and (c) show sample ACF of standardized residuals, their squared series, and absolute series, respectively; part (d) is time plot of standardized residuals.

where the standard errors of the parameters are 0.0053, 0.0017, and 0.1120, respectively. The estimated degrees of freedom is 6.01 with standard error 1.50. All the estimates are significant at the 5% level. The unconditional standard deviation of a_t is $\sqrt{0.0120/(1-0.2845)} \approx 0.1295$, which is close to that obtained under normality. The Ljung–Box statistics of the standardized residuals give $Q(12) = 14.88$ with a p value of 0.25, confirming that the mean equation is adequate. However, the Ljung–Box statistics for the squared standardized residuals show $Q(12) = 35.42$ with a p value of 0.0004. The volatility equation is inadequate at the 1% level. Further analysis shows that $Q(10) = 15.90$ with a p value of 0.10 for the squared standardized residuals. The inadequancy of the volatility equation is due to a large lag-12 ACF ($\rho_{12} = 0.188$) of the squared standardized residuals.

Comparing models (3.12) and (3.13), we see that (a) using a heavy-tailed distribution for ϵ_t reduces the ARCH coefficient, and (b) the difference between the two models is small for this particular instance. Finally, a more appropriate conditional heteroscedastic model for the monthly log returns of Intel stock is a GARCH(1,1) model, which is discussed in the next section.

S-Plus Demonstration
Note the following output with t innovations:

```
> arch1t=garch(intc~1,~garch(1,0),cond.dist="t")
> summary(arch1t)
Call:
garch(formula.mean=intc~1,formula.var=~garch(1,0),
   cond.dist="t")

Mean Equation: structure(.Data = intc ~ 1, class = "formula")
Cond. Variance Equation:structure(.Data=~ garch(1,0), ...)
Cond. Distribution:  t
 with estimated parameter 6.012769 and standard error 1.502179
--------------------------------------------------------------
Estimated Coefficients:
--------------------------------------------------------------
           Value Std.Error t value  Pr(>|t|)
      C 0.01688  0.005288   3.193 1.512e-03
      A 0.01195  0.001667   7.169 3.345e-12
ARCH(1) 0.28445  0.111998   2.540 1.145e-02
--------------------------------------------------------------
AIC(4) = -597.3379, BIC(4) = -581.0642

Ljung-Box test for standardized residuals:
--------------------------------------------------------------
Statistic P-value Chi^2-d.f.
     14.88  0.2482         12

Ljung-Box test for squared standardized residuals:
--------------------------------------------------------------
Statistic   P-value Chi^2-d.f.
     35.42 0.0004014         12
```

Remark. In S-Plus, the command `garch` allows for several conditional distributions. They are specified by `cond.dist = ''t'' or ''ged''`. The default is Gaussian. The R output is given below. The estimates are close to those of S-Plus. □

R Demonstration
The following output uses the `fGarch` package with command `garchFit` and % denotes explanation:

```
> da=read.table("m-intc7308.txt",header=T)
> library(fGarch)   % Load the package
> intc=log(da[,2]+1)
> m1=garchFit(intc~garch(1,0),data=intc,trace=F)
> summary(m1)  % Obtain results
```

```
Title:
 GARCH Modelling
Call:
garchFit(formula=intc~garch(1,0), data=intc, trace=F)

Mean and Variance Equation: data ~ garch(1, 0) [data = intc]
Conditional Distribution: norm

Coefficient(s):
      mu     omega    alpha1
0.012637  0.011195  0.379492

Std. Errors:
 based on Hessian

Error Analysis:
         Estimate  Std. Error  t value Pr(>|t|)
mu       0.012637    0.005428    2.328  0.01990 *
omega    0.011195    0.001239    9.034  < 2e-16 ***
alpha1   0.379492    0.115534    3.285  0.00102 **
---
Log Likelihood:
 288.0589    normalized:  0.6668031

Standardised Residuals Tests:   %Model checking
                                Statistic p-Value
 Jarque-Bera Test   R    Chi^2  137.919   0
 Shapiro-Wilk Test  R    W      0.9679255 4.025172e-08
 Ljung-Box Test     R    Q(10)  12.54002  0.2505382
 Ljung-Box Test     R    Q(15)  21.33508  0.1264607
 Ljung-Box Test     R    Q(20)  23.19679  0.2792354
 Ljung-Box Test     R^2  Q(10)  16.0159   0.09917815
 Ljung-Box Test     R^2  Q(15)  36.08022  0.001721296
 Ljung-Box Test     R^2  Q(20)  37.43683  0.01036728
 LM Arch Test       R    TR^2   26.57744  0.008884587

Information Criterion Statistics:
      AIC       BIC       SIC      HQIC
-1.319717 -1.291464 -1.319813 -1.308563

> predict(m1,5)  % Obtain 1 to 5-step predictions
  meanForecast meanError standardDeviation
1   0.01263656 0.1278609         0.1098306
2   0.01263656 0.1278609         0.1255897
3   0.01263656 0.1278609         0.1310751
4   0.01263656 0.1278609         0.1330976
5   0.01263656 0.1278609         0.1338571
```

```
   % The next command fits a GARCH(1,1) model
> m2=garchFit(intc~garch(1,1),data=intc,trace=F)
> summary(m2) % output edited.
Coefficient(s):
        mu       omega      alpha1       beta1
0.01073352  0.00095445  0.08741989  0.85118414

Error Analysis:
        Estimate  Std. Error  t value Pr(>|t|)
mu     0.0107335   0.0055289    1.941  0.05222 .
omega  0.0009544   0.0003989    2.392  0.01674 *
alpha1 0.0874199   0.0269810    3.240  0.00120 **
beta1  0.8511841   0.0393702   21.620  < 2e-16 ***
---
Standardised Residuals Tests:
                                Statistic p-Value
 Jarque-Bera Test   R    Chi^2  165.5740  0
 Shapiro-Wilk Test  R    W      0.9712087 1.626824e-07
 Ljung-Box Test     R    Q(10)  8.267633  0.6027128
 Ljung-Box Test     R    Q(15)  14.42612  0.4934871
 Ljung-Box Test     R    Q(20)  15.13331  0.7687297
 Ljung-Box Test     R^2  Q(10)  0.9891848 0.9998363
 Ljung-Box Test     R^2  Q(15)  11.36596  0.7262473
 Ljung-Box Test     R^2  Q(20)  12.68143  0.8906302
 LM Arch Test       R    TR^2   10.70199  0.5546164

 % The next command fits an ARCH(1) model with Student-t dist.
> m3=garchFit(intc~garch(1,0),data=intc,trace=F,
   cond.dist='std')
> summary(m3) % Output shortened.

Call:
 garchFit(formula=intc~garch(1,0), data=intc, cond.dist="std",
    trace = F)
Mean and Variance Equation: data ~ garch(1, 0) [data = intc]
Conditional Distribution:  std   % Student-t distribution

Coefficient(s):
      mu     omega    alpha1     shape
0.016731  0.011939  0.285320  6.015195

Error Analysis:
        Estimate  Std. Error  t value Pr(>|t|)
mu      0.016731    0.005302    3.155 0.001603 **
omega   0.011939    0.001603    7.449  9.4e-14 ***
alpha1  0.285320    0.110607    2.580 0.009892 **
shape   6.015195    1.562620    3.849 0.000118 ***
% Degrees of freedom
```

```
% The next command fits an ARCH(1) model with skew
%Student-t dist.
> m4=garchFit(intc~garch(1,0),data=intc,cond.dist='sstd',
   trace=F)
% Next, fit an ARMA(1,0)+GARCH(1,1) model with
% Gaussian noises.
> m5=garchFit(intc~arma(1,0)+garch(1,1),data=intc,trace=F)
```

R Demonstration

The following output was generated with Ox and G@RCH4.2 package and % denotes explanation:

```
> source("garchoxfit_R.txt")
% In G@RCH package, an ARCH(1) model is specified as
% GARCH(0,1).
> m1=garchOxFit(formula.mean=~arma(0,0),
   formula.var=~garch(0,1), series=intc)
% ** SPECIFICATIONS **
Dependent variable : X
Mean Equation : ARMA (0, 0) model.
No regressor in the mean
Variance Equation : GARCH (0, 1) model.
No regressor in the variance
The distribution is a Gauss distribution.

Maximum Likelihood Estimation(Std.Errors based on 2nd deriv.)
                Coefficient  Std.Error  t-value  t-prob
Cst(M)             0.012630  0.0054130    2.333  0.0201
Cst(V)             0.011129  0.0012355    9.007  0.0000
ARCH(Alpha1)       0.387223    0.11688    3.313  0.0010

% ** TESTS **
Q-Statistics on Standardized Residuals
  Q(10)=12.4952 [0.2532785],  Q(20)=23.1210 [0.2828934]
H0: No serial correlation ==> Accept H0 when prob. is High.
------------
Q-Statistics on Squared Standardized Residuals
  --> P-values adjusted by 1 degree(s) of freedom
  Q(10)=15.7849 [0.0715122], Q( 20)=37.0238 [0.0078807]
------------
ARCH 1-10 test:   F(10,410)=   1.4423 [0.1592]
------------
% Apply Student-t distribution
> m2=garchOxFit(formula.mean=~arma(0,0),
   formula.var=~garch(0,1),
  series=intc,cond.dist="t")
% ** SPECIFICATIONS **
```

```
Dependent variable : X
Mean Equation : ARMA (0, 0) model.
No regressor in the mean
Variance Equation : GARCH (0, 1) model.
No regressor in the variance
The distribution is a Student distribution, with 6.02272 df.

Maximum Likelihood Estimation(Std.Errors based on 2nd deriv.)
                Coefficient  Std.Error  t-value  t-prob
Cst(M)             0.016702  0.0052934    3.155  0.0017
Cst(V)             0.011870  0.0015969    7.433  0.0000
ARCH(Alpha1)       0.292318    0.11223    2.605  0.0095
Student(DF)        6.022723     1.5663    3.845  0.0001
** TESTS **
Q-Statistics on Standardized Residuals
  Q(10)=13.0837 [0.2190281],  Q(20)=24.0724 [0.2392436]
------------
Q-Statistics on Squared Standardized Residuals
  --> P-values adjusted by 1 degree(s) of freedom
  Q(10)=18.6982 [0.0278845], Q( 20)=41.7182 [0.0019343]
```

Example 3.2. Consider the percentage changes of the exchange rate between mark and dollar in 10-minute intervals. The data are shown in Figure 3.3(a). As shown in Figure 3.4(a), the series has no serial correlations. However, the sample PACF of the squared series a_t^2 shows some big spikes, especially at lags 1 and 3. There are some large PACF at higher lags, but the lower order lags tend to be more important. Following the procedure discussed in the previous section, we specify an ARCH(3) model for the series. Using the conditional Gaussian likelihood function, we obtain the fitted model $r_t = 0.0018 + \sigma_t\epsilon_t$ and

$$\sigma_t^2 = 0.22 \times 10^{-2} + 0.322a_{t-1}^2 + 0.074a_{t-2}^2 + 0.093a_{t-3}^2,$$

where all the estimates in the volatility equation are statistically significant at the 5% significant level, and the standard errors of the parameters are 0.47×10^{-6}, 0.017, 0.016, and 0.014, respectively. Model checking, using the standardized residual $\tilde{a}_t$, indicates that the model is adequate.

3.5 THE GARCH MODEL

Although the ARCH model is simple, it often requires many parameters to adequately describe the volatility process of an asset return. For instance, consider the monthly excess returns of S&P 500 index of Example 3.3. An ARCH(9) model is needed for the volatility process. Some alternative model must be sought. Bollerslev (1986) proposes a useful extension known as the generalized ARCH (GARCH) model. For a log return series r_t, let $a_t = r_t - \mu_t$ be the innovation at time t. Then

a_t follows a GARCH(m, s) model if

$$a_t = \sigma_t \epsilon_t, \qquad \sigma_t^2 = \alpha_0 + \sum_{i=1}^{m} \alpha_i a_{t-i}^2 + \sum_{j=1}^{s} \beta_j \sigma_{t-j}^2, \tag{3.14}$$

where again $\{\epsilon_t\}$ is a sequence of iid random variables with mean 0 and variance 1.0, $\alpha_0 > 0$, $\alpha_i \geq 0$, $\beta_j \geq 0$, and $\sum_{i=1}^{\max(m,s)}(\alpha_i + \beta_i) < 1$. Here it is understood that $\alpha_i = 0$ for $i > m$ and $\beta_j = 0$ for $j > s$. The latter constraint on $\alpha_i + \beta_i$ implies that the unconditional variance of a_t is finite, whereas its conditional variance σ_t^2 evolves over time. As before, ϵ_t is often assumed to follow a standard normal or standardized Student-t distribution or generalized error distribution. Equation (3.14) reduces to a pure ARCH(m) model if $s = 0$. The α_i and β_j are referred to as ARCH and GARCH parameters, respectively.

To understand properties of GARCH models, it is informative to use the following representation. Let $\eta_t = a_t^2 - \sigma_t^2$ so that $\sigma_t^2 = a_t^2 - \eta_t$. By plugging $\sigma_{t-i}^2 = a_{t-i}^2 - \eta_{t-i}$ $(i = 0, \ldots, s)$ into Eq. (3.14), we can rewrite the GARCH model as

$$a_t^2 = \alpha_0 + \sum_{i=1}^{\max(m,s)} (\alpha_i + \beta_i) a_{t-i}^2 + \eta_t - \sum_{j=1}^{s} \beta_j \eta_{t-j}. \tag{3.15}$$

It is easy to check that $\{\eta_t\}$ is a martingale difference series [i.e., $E(\eta_t) = 0$ and $\text{cov}(\eta_t, \eta_{t-j}) = 0$ for $j \geq 1$]. However, $\{\eta_t\}$ in general is not an iid sequence. Equation (3.15) is an ARMA form for the squared series a_t^2. Thus, a GARCH model can be regarded as an application of the ARMA idea to the squared series a_t^2. Using the unconditional mean of an ARMA model, we have

$$E(a_t^2) = \frac{\alpha_0}{1 - \sum_{i=1}^{\max(m,s)}(\alpha_i + \beta_i)}$$

provided that the denominator of the prior fraction is positive.

The strengths and weaknesses of GARCH models can easily be seen by focusing on the simplest GARCH(1,1) model with

$$\sigma_t^2 = \alpha_0 + \alpha_1 a_{t-1}^2 + \beta_1 \sigma_{t-1}^2, \qquad 0 \leq \alpha_1, \beta_1 \leq 1, (\alpha_1 + \beta_1) < 1. \tag{3.16}$$

First, a large a_{t-1}^2 or σ_{t-1}^2 gives rise to a large σ_t^2. This means that a large a_{t-1}^2 tends to be followed by another large a_t^2, generating, again, the well-known behavior of volatility clustering in financial time series. Second, it can be shown that if $1 - 2\alpha_1^2 - (\alpha_1 + \beta_1)^2 > 0$, then

$$\frac{E(a_t^4)}{[E(a_t^2)]^2} = \frac{3[1 - (\alpha_1 + \beta_1)^2]}{1 - (\alpha_1 + \beta_1)^2 - 2\alpha_1^2} > 3.$$

Consequently, similar to ARCH models, the tail distribution of a GARCH(1,1) process is heavier than that of a normal distribution. Third, the model provides a simple parametric function that can be used to describe the volatility evolution.

Forecasts of a GARCH model can be obtained using methods similar to those of an ARMA model. Consider the GARCH(1,1) model in Eq. (3.16) and assume that the forecast origin is h. For 1-step-ahead forecast, we have

$$\sigma_{h+1}^2 = \alpha_0 + \alpha_1 a_h^2 + \beta_1 \sigma_h^2,$$

where a_h and σ_h^2 are known at the time index h. Therefore, the 1-step-ahead forecast is

$$\sigma_h^2(1) = \alpha_0 + \alpha_1 a_h^2 + \beta_1 \sigma_h^2.$$

For multistep-ahead forecasts, we use $a_t^2 = \sigma_t^2 \epsilon_t^2$ and rewrite the volatility equation in Eq. (3.16) as

$$\sigma_{t+1}^2 = \alpha_0 + (\alpha_1 + \beta_1)\sigma_t^2 + \alpha_1 \sigma_t^2(\epsilon_t^2 - 1).$$

When $t = h + 1$, the equation becomes

$$\sigma_{h+2}^2 = \alpha_0 + (\alpha_1 + \beta_1)\sigma_{h+1}^2 + \alpha_1 \sigma_{h+1}^2(\epsilon_{h+1}^2 - 1).$$

Since $E(\epsilon_{h+1}^2 - 1|F_h) = 0$, the 2-step-ahead volatility forecast at the forecast origin h satisfies the equation

$$\sigma_h^2(2) = \alpha_0 + (\alpha_1 + \beta_1)\sigma_h^2(1).$$

In general, we have

$$\sigma_h^2(\ell) = \alpha_0 + (\alpha_1 + \beta_1)\sigma_h^2(\ell - 1), \qquad \ell > 1. \tag{3.17}$$

This result is exactly the same as that of an ARMA(1,1) model with AR polynomial $1 - (\alpha_1 + \beta_1)B$. By repeated substitutions in Eq. (3.17), we obtain that the ℓ-step-ahead forecast can be written as

$$\sigma_h^2(\ell) = \frac{\alpha_0[1 - (\alpha_1 + \beta_1)^{\ell-1}]}{1 - \alpha_1 - \beta_1} + (\alpha_1 + \beta_1)^{\ell-1}\sigma_h^2(1).$$

Therefore,

$$\sigma_h^2(\ell) \to \frac{\alpha_0}{1 - \alpha_1 - \beta_1}, \qquad \text{as} \quad \ell \to \infty$$

provided that $\alpha_1 + \beta_1 < 1$. Consequently, the multistep-ahead volatility forecasts of a GARCH(1,1) model converge to the unconditional variance of a_t as the forecast horizon increases to infinity provided that $\mathrm{Var}(a_t)$ exists.

The literature on GARCH models is enormous; see Bollerslev, Chou, and Kroner (1992), Bollerslev, Engle, and Nelson (1994), and the references therein. The model encounters the same weaknesses as the ARCH model. For instance, it responds equally to positive and negative shocks. In addition, recent empirical studies of high-frequency financial time series indicate that the tail behavior of GARCH models remains too short even with standardized Student-t innovations. For further information about kurtosis of GARCH models, see Section 3.16.

3.5.1 An Illustrative Example

The modeling procedure of ARCH models can also be used to build a GARCH model. However, specifying the order of a GARCH model is not easy. Only lower order GARCH models are used in most applications, say, GARCH(1,1), GARCH(2,1), and GARCH(1,2) models. The conditional maximum-likelihood method continues to apply provided that the starting values of the volatility $\{\sigma_t^2\}$ are assumed to be known. Consider, for instance, a GARCH(1,1) model. If σ_1^2 is treated as fixed, then σ_t^2 can be computed recursively for a GARCH(1,1) model. In some applications, the sample variance of a_t serves as a good starting value of σ_1^2. The fitted model can be checked by using the standardized residual $\tilde{a}_t = a_t/\sigma_t$ and its squared process.

Example 3.3. In this example, we consider the monthly excess returns of S&P 500 index starting from 1926 for 792 observations. The series is shown in Figure 3.6. Denote the excess return series by r_t. Figure 3.7 shows the sample ACF

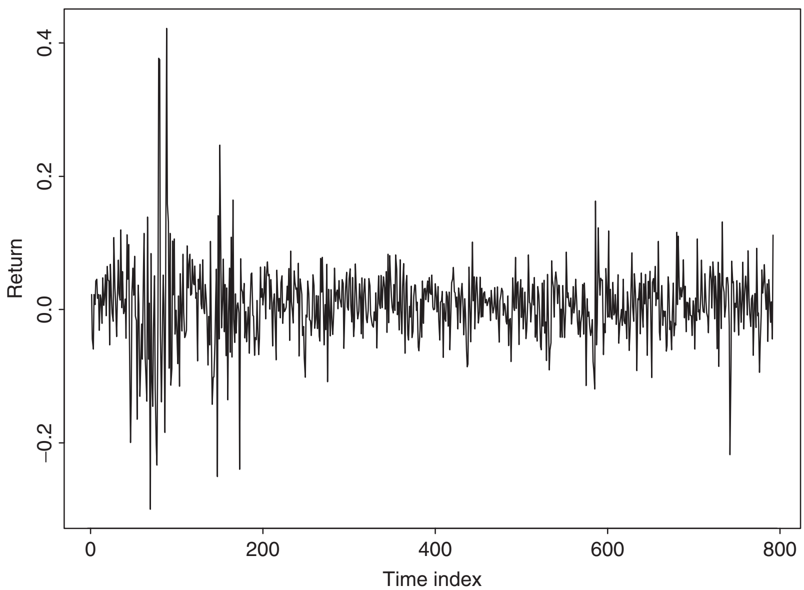

Figure 3.6 Time series plot of monthly excess returns of S&P 500 index from 1926 to 1991.

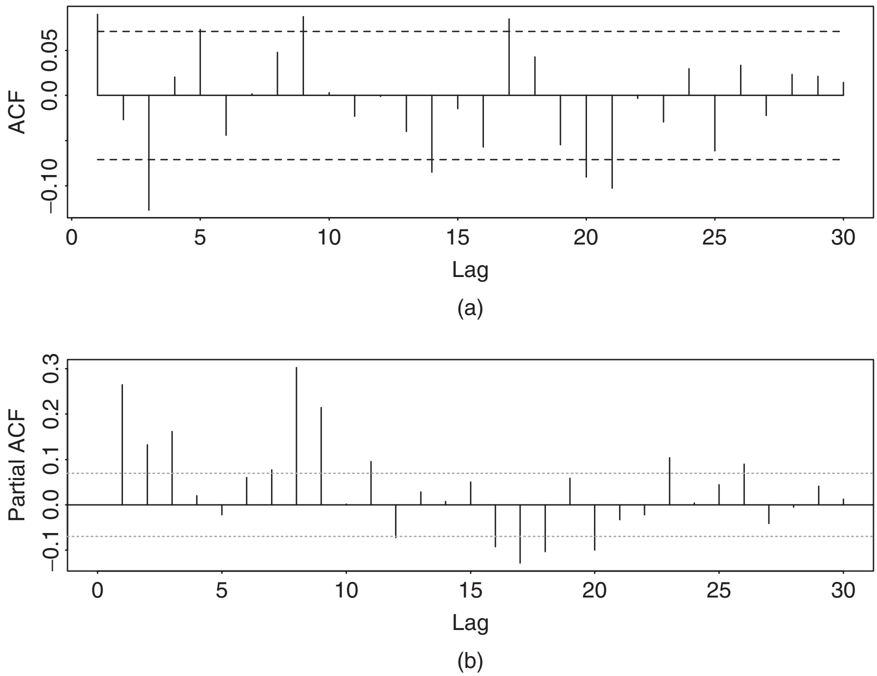

Figure 3.7 (a) Sample ACF of monthly excess returns of S&P 500 index and (b) sample PACF of squared monthly excess returns. Sample period is from 1926 to 1991.

of r_t and the sample PACF of r_t^2. The r_t series has some serial correlations at lags 1 and 3, but the key feature is that the PACF of r_t^2 shows strong linear dependence. If an MA(3) model is entertained, we obtain

$$r_t = 0.0062 + a_t + 0.0944a_{t-1} - 0.1407a_{t-3}, \qquad \hat{\sigma}_a = 0.0576$$

for the series, where all of the coefficients are significant at the 5% level. However, for simplicity, we use instead an AR(3) model

$$r_t = \phi_1 r_{t-1} + \phi_2 r_{t-2} + \phi_3 r_{t-3} + \beta_0 + a_t.$$

The fitted AR(3) model, under the normality assumption, is

$$r_t = 0.088r_{t-1} - 0.023r_{t-2} - 0.123r_{t-3} + 0.0066 + a_t, \quad \hat{\sigma}_a^2 = 0.00333. \quad (3.18)$$

For the GARCH effects, we use the GARCH(1,1) model

$$a_t = \sigma_t \epsilon_t, \qquad \sigma_t^2 = \alpha_0 + \beta_1 \sigma_{t-1}^2 + \alpha_1 a_{t-1}^2.$$

A joint estimation of the AR(3)–GARCH(1,1) model gives

$$r_t = 0.0078 + 0.032r_{t-1} - 0.029r_{t-2} - 0.008r_{t-3} + a_t,$$
$$\sigma_t^2 = 0.000084 + 0.1213a_{t-1}^2 + 0.8523\sigma_{t-1}^2.$$

From the volatility equation, the implied unconditional variance of a_t is

$$\frac{0.000084}{1 - 0.8523 - 0.1213} = 0.00317,$$

which is close to that of Eq. (3.18). However, t ratios of the parameters in the mean equation suggest that all three AR coefficients are insignificant at the 5% level. Therefore, we refine the model by dropping all AR parameters. The refined model is

$$r_t = 0.0076 + a_t, \qquad \sigma_t^2 = 0.000086 + 0.1216a_{t-1}^2 + 0.8511\sigma_{t-1}^2. \tag{3.19}$$

The standard error of the constant in the mean equation is 0.0015, whereas those of the parameters in the volatility equation are 0.000024, 0.0197, and 0.0190, respectively. The unconditional variance of a_t is $0.000086/(1 - 0.8511 - 0.1216) = 0.00314$. This is a simple stationary GARCH(1,1) model. Figure 3.8 shows the estimated volatility process, σ_t, and the standardized shocks $\tilde{a}_t = a_t/\sigma_t$ for the

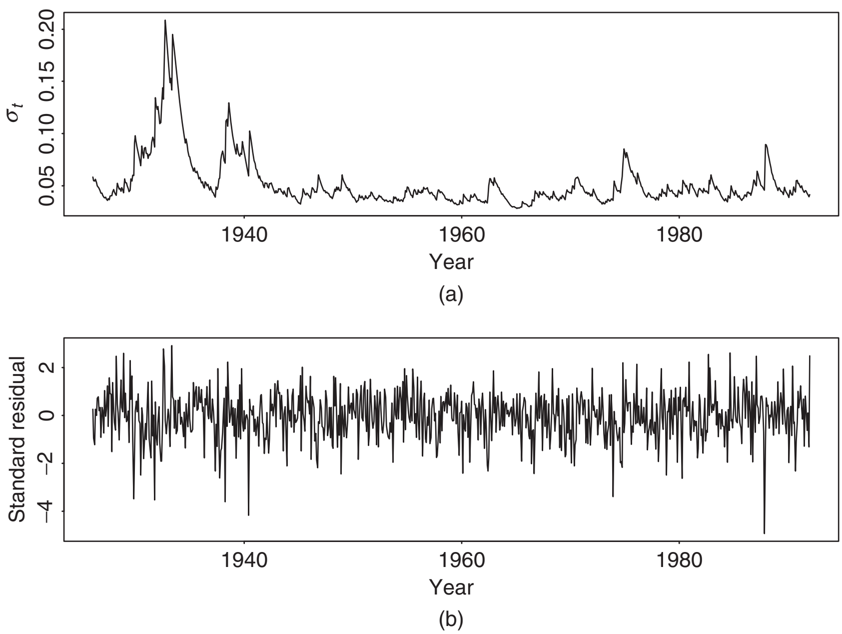

Figure 3.8 (a) Time series plot of estimated volatility (σ_t) for monthly excess returns of S&P 500 index and (b) standardized shocks of monthly excess returns of S&P 500 index. Both plots are based on GARCH(1,1) model in Eq. (3.19).

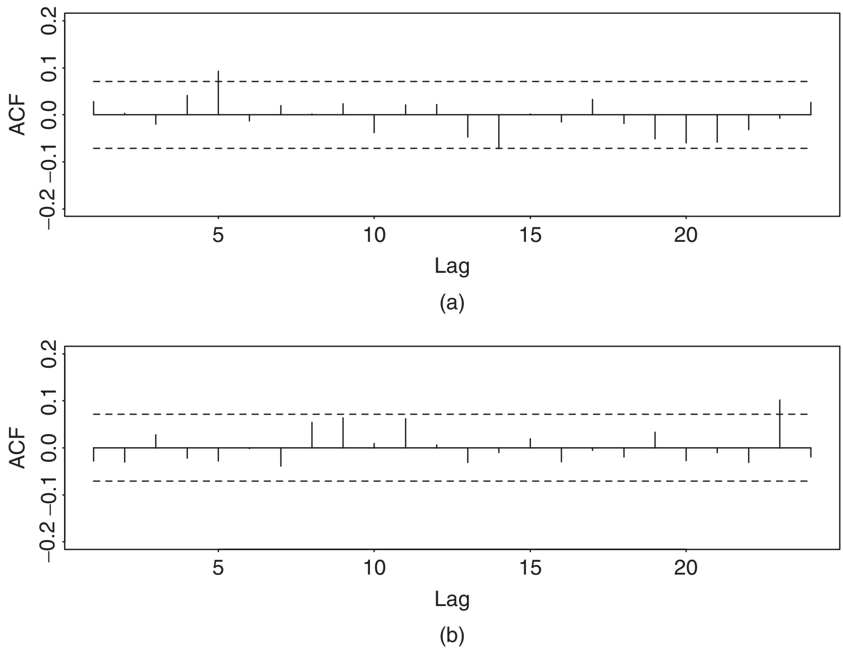

Figure 3.9 Model checking of GARCH(1,1) model in Eq. (3.19) for monthly excess returns of S&P 500 index: (a) Sample ACF of standardized residuals and (b) sample ACF of the squared standardized residuals.

GARCH(1,1) model in Eq. (3.19). The $\tilde{a}_t$ series looks like a white noise process. Figure 3.9 provides the sample ACF of the standardized residuals $\tilde{a}_t$ and the squared process $\tilde{a}_t^2$. These ACFs fail to suggest any significant serial correlations or conditional heteroscedasticity in the standardized residual series. More specifically, we have $Q(12) = 11.99(0.45)$ and $Q(24) = 28.52(0.24)$ for $\tilde{a}_t$, and $Q(12) = 13.11(0.36)$ and $Q(24) = 26.45(0.33)$ for $\tilde{a}_t^2$, where the number in parentheses is the p value of the test statistic. Thus, the model appears to be adequate in describing the linear dependence in the return and volatility series. Note that the fitted model shows $\hat{\alpha}_1 + \hat{\beta}_1 = 0.9772$, which is close to 1. This phenomenon is commonly observed in practice and it leads to imposing the constraint $\alpha_1 + \beta_1 = 1$ in a GARCH(1,1) model, resulting in an integrated GARCH (or IGARCH) model; see Section 3.6.

Finally, to forecast the volatility of monthly excess returns of the S&P 500 index, we can use the volatility equation in Eq. (3.19). For instance, at the forecast origin h, we have $\sigma_{h+1}^2 = 0.000086 + 0.1216a_h^2 + 0.8511\sigma_h^2$. The 1-step-ahead forecast is then

$$\sigma_h^2(1) = 0.000086 + 0.1216a_h^2 + 0.8511\sigma_h^2,$$

TABLE 3.1 Volatility Forecasts for Monthly Excess Returns of S&P 500 Index[a]

Horizon	1	2	3	4	5	∞
Return	0.0076	0.0076	0.0076	0.0076	0.0076	0.0076
Volatility	0.0536	0.0537	0.0537	0.0538	0.0538	0.0560

[a] The forecast origin is $h = 792$, which corresponds to December 1991. Here volatility denotes conditional standard deviation.

where a_h is the residual of the mean equation at time h and σ_h is obtained from the volatility equation. The starting value σ_0^2 is fixed at either zero or the unconditional variance of a_t. For multistep-ahead forecasts, we use the recursive formula in Eq. (3.17). Table 3.1 shows some mean and volatility forecasts for the monthly excess return of the S&P 500 index with forecast origin $h = 792$ based on the GARCH(1,1) model in Eq. (3.19).

Some S-Plus Commands Used in Example 3.3.

```
> fit=garch(sp~ar(3),~garch(1,1))
> summary(fit)
> fit=garch(sp~1,~garch(1,1))
> summary(fit)
> names(fit)
 [1] "residuals" "sigma.t"    "df.residual" "coef" "model"
 [6] "cond.dist" "likelihood" "opt.index"   "cov"
     "prediction"
[11] "call"      "asymp.sd"   "series"
>
> stdresi=fit$residuals/fit$sigma.t
> autocorTest(stdresi,lag=24)
> autocorTest(stdresi^2,lag=24)
> predict(fit,5)
```

Note that in the prior commands the volatility series σ_t is stored in `fit$sigma.t` and the residual series of the returns in `fit$residuals`.

t Innovation

Assuming that ϵ_t follows a standardized Student-t distribution with 5 degrees of freedom, we reestimate the GARCH(1,1) model and obtain

$$r_t = 0.0085 + a_t, \qquad \sigma_t^2 = 0.00012 + 0.1121 a_{t-1}^2 + 0.8432 \sigma_{t-1}^2, \tag{3.20}$$

where the standard errors of the parameters are 0.0015, 0.51×10^{-4}, 0.0296, and 0.0371, respectively. This model is essentially an IGARCH(1,1) model as $\hat{\alpha}_1 + \hat{\beta}_1 \approx 0.95$, which is close to 1. The Ljung–Box statistics of the standardized residuals give $Q(10) = 11.38$ with a p value of 0.33 and those of the $\{\tilde{a}_t^2\}$ series give $Q(10) = 10.48$ with a p value of 0.40. Thus, the fitted GARCH(1,1) model with Student-t distribution is adequate.

S-Plus Commands Used

```
> fit1 = garch(sp~1,~garch(1,1),cond.dist='t',cond.par=5,
+ cond.est=F)
> summary(fit1)
> stresi=fit1$residuals/fit1$sigma.t
> autocorTest(stresi,lag=10)
> autocorTest(stresi^2,lag=10)
```

Estimation of Degrees of Freedom

If we further extend the GARCH(1,1) model by estimating the degrees of freedom of the Student-t distribution used, we obtain the model

$$r_t = 0.0085 + a_t, \qquad \sigma_t^2 = 0.00012 + 0.1121a_{t-1}^2 + 0.8432\sigma_{t-1}^2, \tag{3.21}$$

where the estimated degrees of freedom is 7.02. Standard errors of the estimates in Eq. (3.21) are close to those in Eq. (3.20). The standard error of the estimated degrees of freedom is 1.78. Consequently, we cannot reject the hypothesis of using a standardized Student-t distribution with 5 degrees of freedom at the 5% significance level.

S-Plus Commands Used

```
> fit2 = garch(sp~1,~garch(1,1),cond.dist='t')
> summary(fit2)
```

R Commands Used in Example 3.3

```
> library(fGarch)
> sp5=scan(file='sp500.txt') % Load data
> plot(sp5,type='l')
% Below, fit an AR(3)+GARCH(1,1) model.
> m1=garchFit(~arma(3,0)+garch(1,1),data=sp5,trace=F)
> summary(m1)
% Below, fit a GARCH(1,1) model with Student-t distribution.
> m2=garchFit(~garch(1,1),data=sp5,trace=F,cond.dist="std")
> summary(m2)
% Obtain standardized residuals.
> stresi=residuals(m2,standardize=T)
> plot(stresi,type='l')
> Box.test(stresi,10,type='Ljung')
> predict(m2,5)
```

3.5.2 Forecasting Evaluation

Since the volatility of an asset return is not directly observable, comparing the forecasting performance of different volatility models is a challenge to data analysts. In the literature, some researchers use out-of-sample forecasts and compare the volatility forecasts $\sigma_h^2(\ell)$ with the shock $a_{h+\ell}^2$ in the forecasting sample to assess the forecasting performance of a volatility model. This approach often finds a low correlation coefficient between $a_{h+\ell}^2$ and $\sigma_h^2(\ell)$, that is, low R^2. However, such a finding is not surprising because $a_{h+\ell}^2$ alone is not an adequate measure of the volatility at time index $h+\ell$. Consider the 1-step-ahead forecasts. From a statistical point of view, $E(a_{h+1}^2|F_h) = \sigma_{h+1}^2$ so that a_{h+1}^2 is a consistent estimate of σ_{h+1}^2. But it is not an accurate estimate of σ_{h+1}^2 because a single observation of a random variable with a known mean value cannot provide an accurate estimate of its variance. Consequently, such an approach to evaluate forecasting performance of volatility models is strictly speaking not proper. For more information concerning forecasting evaluation of GARCH models, readers are referred to Andersen and Bollerslev (1998).

3.5.3 A Two-Pass Estimation Method

Based on Eq. (3.15), a two-pass estimation method can be used to estimate GARCH models. First, ignoring any ARCH effects, one estimates the mean equation of a return series using the methods discussed in Chapter 2 (e.g., maximum-likelihood method). Denote the residual series by a_t. Second, treating $\{a_t^2\}$ as an observed time series, one applies the maximum-likelihood method to estimate parameters of Eq. (3.15). Denote the AR and MA coefficient estimates by $\hat{\phi}_i$ and $\hat{\theta}_i$. The GARCH estimates are obtained as $\hat{\beta}_i = \hat{\theta}_i$ and $\hat{\alpha}_i = \hat{\phi}_i - \hat{\theta}_i$. Obviously, such estimates are approximations to the true parameters and their statistical properties have not been rigorously investigated. However, limited experience shows that this simple approach often provides good approximations, especially when the sample size is moderate or large. For instance, consider the monthly excess return series of the S&P 500 index of Example 3.3. Using the conditional MLE method in SCA, we obtain the model

$$r_t = 0.0061 + a_t, \qquad a_t^2 = 0.00014 + 0.9583a_{t-1}^2 + \eta_t - 0.8456\eta_{t-1},$$

where all estimates are significantly different from zero at the 5% level. From the estimates, we have $\hat{\beta}_1 = 0.8456$ and $\hat{\alpha}_1 = 0.9583 - 0.8456 = 0.1127$. These approximate estimates are very close to those in Eq. (3.19) or (3.21). Furthermore, the fitted volatility series of the two-pass method is very close to that of Figure 3.8(a).

3.6 THE INTEGRATED GARCH MODEL

If the AR polynomial of the GARCH representation in Eq. (3.15) has a unit root, then we have an IGARCH model. Thus, IGARCH models are unit-root GARCH models. Similar to ARIMA models, a key feature of IGARCH models is that the impact of past squared shocks $\eta_{t-i} = a_{t-i}^2 - \sigma_{t-i}^2$ for $i > 0$ on a_t^2 is persistent.

An IGARCH(1,1) model can be written as

$$a_t = \sigma_t \epsilon_t, \qquad \sigma_t^2 = \alpha_0 + \beta_1 \sigma_{t-1}^2 + (1 - \beta_1) a_{t-1}^2,$$

where $\{\epsilon_t\}$ is defined as before and $1 > \beta_1 > 0$. For the monthly excess returns of the S&P 500 index, an estimated IGARCH(1,1) model is

$$\begin{aligned} r_t &= 0.0067 + a_t, \qquad a_t = \sigma_t \epsilon_t, \\ \sigma_t^2 &= 0.000119 + 0.8059 \sigma_{t-1}^2 + 0.1941 a_{t-1}^2, \end{aligned}$$

where the standard errors of the estimates in the volatility equation are 0.0017, 0.000013, and 0.0144, respectively. The parameter estimates are close to those of the GARCH(1,1) model shown before, but there is a major difference between the two models. The unconditional variance of a_t, hence that of r_t, is not defined under the above IGARCH(1,1) model. This seems hard to justify for an excess return series. From a theoretical point of view, the IGARCH phenomenon might be caused by occasional level shifts in volatility. The actual cause of persistence in volatility deserves a careful investigation.

When $\alpha_1 + \beta_1 = 1$, repeated substitutions in Eq. (3.17) give

$$\sigma_h^2(\ell) = \sigma_h^2(1) + (\ell - 1)\alpha_0, \qquad \ell \geq 1, \tag{3.22}$$

where h is the forecast origin. Consequently, the effect of $\sigma_h^2(1)$ on future volatilities is also persistent, and the volatility forecasts form a straight line with slope α_0. Nelson (1990) studies some probability properties of the volatility process σ_t^2 under an IGARCH model. The process σ_t^2 is a martingale for which some nice results are available in the literature. Under certain conditions, the volatility process is strictly stationary but not weakly stationary because it does not have the first two moments.

The case of $\alpha_0 = 0$ is of particular interest in studying the IGARCH(1,1) model. In this case, the volatility forecasts are simply $\sigma_h^2(1)$ for all forecast horizons; see Eq. (3.22). This special IGARCH(1,1) model is the volatility model used in RiskMetrics, which is an approach for calculating value at risk; see Chapter 7. The model is also an exponential smoothing model for the $\{a_t^2\}$ series. To see this, rewrite the model as

$$\begin{aligned} \sigma_t^2 &= (1 - \beta_1) a_{t-1}^2 + \beta_1 \sigma_{t-1}^2 \\ &= (1 - \beta_1) a_{t-1}^2 + \beta_1 [(1 - \beta) a_{t-2}^2 + \beta_1 \sigma_{t-2}^2] \\ &= (1 - \beta_1) a_{t-1}^2 + (1 - \beta_1) \beta_1 a_{t-2}^2 + \beta_1^2 \sigma_{t-2}^2. \end{aligned}$$

By repeated substitutions, we have

$$\sigma_t^2 = (1 - \beta_1)(a_{t-1}^2 + \beta_1 a_{t-2}^2 + \beta_1^2 a_{t-3}^3 + \cdots),$$

which is the well-known exponential smoothing formation with β_1 being the discounting factor. Exponential smoothing methods can thus be used to estimate such an IGARCH(1,1) model.

3.7 THE GARCH-M MODEL

In finance, the return of a security may depend on its volatility. To model such a phenomenon, one may consider the GARCH-M model, where M stands for GARCH *in the mean*. A simple GARCH(1,1)-M model can be written as

$$\begin{aligned} r_t &= \mu + c\sigma_t^2 + a_t, \qquad a_t = \sigma_t\epsilon_t, \\ \sigma_t^2 &= \alpha_0 + \alpha_1 a_{t-1}^2 + \beta_1\sigma_{t-1}^2, \end{aligned} \tag{3.23}$$

where μ and c are constants. The parameter c is called the risk premium parameter. A positive c indicates that the return is positively related to its volatility. Other specifications of risk premium have also been used in the literature, including $r_t = \mu + c\sigma_t + a_t$ and $r_t = \mu + c\ \ln(\sigma_t^2) + a_t$.

The formulation of the GARCH-M model in Eq. (3.23) implies that there are serial correlations in the return series r_t. These serial correlations are introduced by those in the volatility process $\{\sigma_t^2\}$. The existence of risk premium is, therefore, another reason that some historical stock returns have serial correlations.

For illustration, we consider a GARCH(1,1)-M model with Gaussian innovations for the monthly excess returns of the S&P 500 index from January 1926 to December 1991. The fitted model is

$$r_t = 0.0055 + 1.09\sigma_t^2 + a_t, \qquad \sigma_t^2 = 8.76 \times 10^{-5} + 0.123a_{t-1}^2 + 0.849\sigma_{t-1}^2,$$

where the standard errors for the two parameters in the mean equation are 0.0023 and 0.818, respectively, and those for the parameters in the volatility equation are $2.51\times^{-5}$, 0.0205, and 0.0196, respectively. The estimated risk premium for the index return is positive but is not statistically significant at the 5% level. Here the result is obtained using S-Plus. Other forms of GARCH-M specification in S-Plus are given in Table 3.2. The idea of risk premium applies to other GARCH models.

TABLE 3.2 GARCH-M Models Allowed in S-Plus: Mean Equation Is $r_t = \mu + cg(\sigma_t) + a_t$

$g(\sigma_t)$	Command
σ_t^2	var.in.mean
σ_t	sd.in.mean
$\ln(\sigma_t^2)$	logvar.in.mean

S-Plus Demonstration

```
> sp.fit = garch(sp~1+var.in.mean,~garch(1,1))
> summary(sp.fit)
```

3.8 THE EXPONENTIAL GARCH MODEL

To overcome some weaknesses of the GARCH model in handling financial time series, Nelson (1991) proposes the exponential GARCH (EGARCH) model. In particular, to allow for asymmetric effects between positive and negative asset returns, he considered the weighted innovation

$$g(\epsilon_t) = \theta\epsilon_t + \gamma[|\epsilon_t| - E(|\epsilon_t|)], \tag{3.24}$$

where θ and γ are real constants. Both ϵ_t and $|\epsilon_t| - E(|\epsilon_t|)$ are zero-mean iid sequences with continuous distributions. Therefore, $E[g(\epsilon_t)] = 0$. The asymmetry of $g(\epsilon_t)$ can easily be seen by rewriting it as

$$g(\epsilon_t) = \begin{cases} (\theta + \gamma)\epsilon_t - \gamma E(|\epsilon_t|) & \text{if } \epsilon_t \geq 0, \\ (\theta - \gamma)\epsilon_t - \gamma E(|\epsilon_t|) & \text{if } \epsilon_t < 0. \end{cases}$$

Remark. For the standard Gaussian random variable ϵ_t, $E(|\epsilon_t|) = \sqrt{2/\pi}$. For the standardized Student-t distribution in Eq. (3.7), we have

$$E(|\epsilon_t|) = \frac{2\sqrt{v-2}\,\Gamma[(v+1)/2]}{(v-1)\Gamma(v/2)\sqrt{\pi}}.$$

□

An EGARCH(m, s) model can be written as

$$a_t = \sigma_t\epsilon_t, \qquad \ln(\sigma_t^2) = \alpha_0 + \frac{1 + \beta_1 B + \cdots + \beta_{s-1}B^{s-1}}{1 - \alpha_1 B - \cdots - \alpha_m B^m} g(\epsilon_{t-1}), \tag{3.25}$$

where α_0 is a constant, B is the back-shift (or lag) operator such that $Bg(\epsilon_t) = g(\epsilon_{t-1})$, and $1 + \beta_1 B + \cdots + \beta_{s-1}B^{s-1}$ and $1 - \alpha_1 B - \cdots - \alpha_m B^m$ are polynomials with zeros outside the unit circle and have no common factors. By outside the unit circle we mean that absolute values of the zeros are greater than 1. Again, Eq. (3.25) uses the usual ARMA parameterization to describe the evolution of the conditional variance of a_t. Based on this representation, some properties of the EGARCH model can be obtained in a similar manner as those of the GARCH model. For instance, the unconditional mean of $\ln(\sigma_t^2)$ is α_0. However, the model differs from the GARCH model in several ways. First, it uses logged conditional variance to relax the positiveness constraint of model coefficients. Second, the use of $g(\epsilon_t)$ enables the model to respond asymmetrically to positive and negative lagged values of a_t. Some additional properties of the EGARCH model can be found in Nelson (1991).

To better understand the EGARCH model, let us consider the simple model with order (1,1):

$$a_t = \sigma_t \epsilon_t, \qquad (1 - \alpha B)\ln(\sigma_t^2) = (1-\alpha)\alpha_0 + g(\epsilon_{t-1}), \tag{3.26}$$

where the ϵ_t are iid standard normal and the subscript of α_1 is omitted. In this case, $E(|\epsilon_t|) = \sqrt{2/\pi}$ and the model for $\ln(\sigma_t^2)$ becomes

$$(1-\alpha B)\ln(\sigma_t^2) = \begin{cases} \alpha_* + (\gamma+\theta)\epsilon_{t-1} & \text{if } \epsilon_{t-1} \ge 0, \\ \alpha_* + (\gamma-\theta)(-\epsilon_{t-1}) & \text{if } \epsilon_{t-1} < 0, \end{cases} \tag{3.27}$$

where $\alpha_* = (1-\alpha)\alpha_0 - \sqrt{2/\pi}\gamma$. This is a nonlinear function similar to that of the threshold autoregressive (TAR) model of Tong (1978, 1990). It suffices to say that for this simple EGARCH model the conditional variance evolves in a nonlinear manner depending on the sign of a_{t-1}. Specifically, we have

$$\sigma_t^2 = \sigma_{t-1}^{2\alpha}\exp(\alpha_*)\begin{cases} \exp\left[(\gamma+\theta)\dfrac{a_{t-1}}{\sigma_{t-1}}\right] & \text{if } a_{t-1} \ge 0, \\ \exp\left[(\gamma-\theta)\dfrac{|a_{t-1}|}{\sigma_{t-1}}\right] & \text{if } a_{t-1} < 0. \end{cases}$$

The coefficients $(\gamma+\theta)$ and $(\gamma-\theta)$ show the asymmetry in response to positive and negative a_{t-1}. The model is, therefore, nonlinear if $\theta \neq 0$. Since negative shocks tend to have larger impacts, we expect θ to be negative. For higher order EGARCH models, the nonlinearity becomes much more complicated. Cao and Tsay (1992) use nonlinear models, including EGARCH models, to obtain multistep-ahead volatility forecasts. We discuss nonlinearity in financial time series in Chapter 4.

3.8.1 Alternative Model Form

An alternative form for the EGARCH(m, s) model is

$$\ln(\sigma_t^2) = \alpha_0 + \sum_{i=1}^{s}\alpha_i \frac{|a_{t-i}| + \gamma_i a_{t-i}}{\sigma_{t-i}} + \sum_{j=1}^{m}\beta_j \ln(\sigma_{t-j}^2). \tag{3.28}$$

Here a positive a_{t-i} contributes $\alpha_i(1+\gamma_i)|\epsilon_{t-i}|$ to the log volatility, whereas a negative a_{t-i} gives $\alpha_i(1-\gamma_i)|\epsilon_{t-i}|$, where $\epsilon_{t-i} = a_{t-i}/\sigma_{t-i}$. The γ_i parameter thus signifies the leverage effect of a_{t-i}. Again, we expect γ_i to be negative in real applications. This is the model form used in S-Plus.

3.8.2 Illustrative Example

Nelson (1991) applies an EGARCH model to the daily excess returns of the value-weighted market index from the Center for Research in Security Prices from July 1962 to December 1987. The excess returns are obtained by removing monthly Treasury bill returns from the value-weighted index returns, assuming that the Treasury bill return was constant for each calendar day within a given month. There are 6408 observations. Denote the excess return by r_t. The model used is as follows:

$$r_t = \phi_0 + \phi_1 r_{t-1} + c\sigma_t^2 + a_t, \tag{3.29}$$
$$\ln(\sigma_t^2) = \alpha_0 + \ln(1 + wN_t) + \frac{1 + \beta B}{1 - \alpha_1 B - \alpha_2 B^2} g(\epsilon_{t-1}),$$

where σ_t^2 is the conditional variance of a_t given F_{t-1}, N_t is the number of nontrading days between trading days $t-1$ and t, α_0 and w are real parameters, $g(\epsilon_t)$ is defined in Eq. (3.24), and ϵ_t follows a generalized error distribution in Eq. (3.10). Similar to a GARCH-M model, the parameter c in Eq. (3.29) is the risk premium parameter. Table 3.3 gives the parameter estimates and their standard errors of the model. The mean equation of model (3.29) has two features that are of interest. First, it uses an AR(1) model to take care of possible serial correlation in the excess returns. Second, it uses the volatility σ_t^2 as a regressor to account for risk premium. The estimated risk premium is negative, but statistically insignificant.

3.8.3 Second Example

As another illustration, we consider the monthly log returns of IBM stock from January 1926 to December 1997 for 864 observations. An AR(1)–EGARCH(1,1) model is entertained and the fitted model is

$$r_t = 0.0105 + 0.092 r_{t-1} + a_t, \qquad a_t = \sigma_t \epsilon_t, \tag{3.30}$$
$$\ln(\sigma_t^2) = -5.496 + \frac{g(\epsilon_{t-1})}{1 - 0.856B},$$
$$g(\epsilon_{t-1}) = -0.0795\epsilon_{t-1} + 0.2647\left(|\epsilon_{t-1}| - \sqrt{2/\pi}\right), \tag{3.31}$$

TABLE 3.3 Estimated AR(1)–EGARCH(2,2) Model for Daily Excess Returns of Value-Weighted CRSP Market Index: July 1962–December 1987

Parameter	α_0	w	γ	α_1	α_2	β
Estimate	−10.06	0.183	0.156	1.929	−0.929	−0.978
Error	0.346	0.028	0.013	0.015	0.015	0.006
Parameter	θ	ϕ_0	ϕ_1	c	v	
Estimate	−0.118	$3.5 \cdot 10^{-4}$	0.205	−3.361	1.576	
Error	0.009	$9.9 \cdot 10^{-5}$	0.012	2.026	0.032	

where $\{\epsilon_t\}$ is a sequence of independent standard Gaussian random variates. All parameter estimates are statistically significant at the 5% level. For model checking, the Ljung–Box statistics give $Q(10) = 6.31(0.71)$ and $Q(20) = 21.4(0.32)$ for the standardized residual process $\tilde{a}_t = a_t/\sigma_t$ and $Q(10) = 4.13(0.90)$ and $Q(20) = 15.93(0.66)$ for the squared process $\tilde{a}_t^2$, where again the number in parentheses denotes p value. Therefore, there is no serial correlation or conditional heteroscedasticity in the standardized residuals of the fitted model. The prior AR(1)–EGARCH(1,1) model is adequate.

From the estimated volatility equation in (3.31) and using $\sqrt{2/\pi} \approx 0.7979$, we obtain the volatility equation as

$$\ln(\sigma_t^2) = -1.001 + 0.856\ \ln(\sigma_{t-1}^2) + \begin{cases} 0.1852\epsilon_{t-1} & \text{if } \epsilon_{t-1} \geq 0, \\ -0.3442\epsilon_{t-1} & \text{if } \epsilon_{t-1} < 0. \end{cases}$$

Taking antilog transformation, we have

$$\sigma_t^2 = \sigma_{t-1}^{2\times 0.856} e^{-1.001} \times \begin{cases} e^{0.1852\epsilon_{t-1}} & \text{if } \epsilon_{t-1} \geq 0, \\ e^{-0.3442\epsilon_{t-1}} & \text{if } \epsilon_{t-1} < 0. \end{cases}$$

This equation highlights the asymmetric responses in volatility to the past positive and negative shocks under an EGARCH model. For example, for a standardized shock with magnitude 2 (i.e., two standard deviations), we have

$$\frac{\sigma_t^2(\epsilon_{t-1} = -2)}{\sigma_t^2(\epsilon_{t-1} = 2)} = \frac{\exp[-0.3442 \times (-2)]}{\exp(0.1852 \times 2)} = e^{0.318} = 1.374.$$

Therefore, the impact of a negative shock of size 2 standard deviations is about 37.4% higher than that of a positive shock of the same size. This example clearly demonstrates the asymmetric feature of EGARCH models. In general, the bigger the shock, the larger the difference in volatility impact.

Finally, we extend the sample period to include the log returns from 1998 to 2003 so that there are 936 observations and use S-Plus to fit an EGARCH(1,1) model. The results are given below.

S-Plus Demonstration

The following output has been edited:

```
> ibm.egarch=garch(ibmln~1,~egarch(1,1),leverage=T,
+ cond.dist='ged')

> summary(ibm.egarch)
Call:
garch(formula.mean = ibmln ~ 1, formula.var = ~ egarch(1, 1),
      leverage = T,cond.dist = "ged")

Mean Equation: ibmln ~ 1
```

```
Conditional Variance Equation: ~ egarch(1, 1)
Conditional Distribution:  ged
 with estimated parameter 1.5003 and standard error 0.09912
--------------------------------------------------------------
Estimated Coefficients:
--------------------------------------------------------------
             Value Std.Error t value  Pr(>|t|)
       C  0.01181  0.002012   5.870 3.033e-09
       A -0.55680  0.171602  -3.245 6.088e-04
 ARCH(1)  0.22025  0.052824   4.169 1.669e-05
GARCH(1)  0.92910  0.026743  34.742 0.000e+00
  LEV(1) -0.26400  0.126096  -2.094 1.828e-02
--------------------------------------------------------------
Ljung-Box test for standardized residuals:
--------------------------------------------------------------
Statistic P-value Chi^2-d.f.
     17.87  0.1195         12

Ljung-Box test for squared standardized residuals:
--------------------------------------------------------------
Statistic P-value Chi^2-d.f.
     6.723  0.8754         12
```

The fitted GARCH(1,1) model is

$$\begin{aligned} r_t &= 0.0118 + a_t, \qquad a_t = \sigma_t \epsilon_t, \\ \ln(\sigma_t^2) &= -0.557 + 0.220 \frac{|a_{t-1}| - 0.264 a_{t-1}}{\sigma_{t-1}} + 0.929 \ln(\sigma_{t-1}^2), \end{aligned} \tag{3.32}$$

where ϵ_t follows a GED distribution with parameter 1.5. This model is adequate and based on the Ljung–Box statistics of the standardized residual series and its squared process. As expected, the output shows that the estimated leverage effect is negative and is statistically significant at the 5% level with a t ratio of -2.094.

3.8.4 Forecasting Using an EGARCH Model

We use the EGARCH(1,1) model to illustrate multistep-ahead forecasts of EGARCH models, assuming that the model parameters are known and the innovations are standard Gaussian. For such a model, we have

$$\begin{aligned} \ln(\sigma_t^2) &= (1 - \alpha_1)\alpha_0 + \alpha_1 \ln(\sigma_{t-1}^2) + g(\epsilon_{t-1}), \\ g(\epsilon_{t-1}) &= \theta \epsilon_{t-1} + \gamma(|\epsilon_{t-1}| - \sqrt{2/\pi}). \end{aligned}$$

Taking exponentials, the model becomes

$$\begin{aligned} \sigma_t^2 &= \sigma_{t-1}^{2\alpha_1} \exp[(1 - \alpha_1)\alpha_0] \exp[g(\epsilon_{t-1})], \\ g(\epsilon_{t-1}) &= \theta \epsilon_{t-1} + \gamma(|\epsilon_{t-1}| - \sqrt{2/\pi}). \end{aligned} \tag{3.33}$$

Let h be the forecast origin. For the 1-step-ahead forecast, we have

$$\sigma_{h+1}^2 = \sigma_h^{2\alpha_1} \exp[(1-\alpha_1)\alpha_0] \exp[g(\epsilon_h)],$$

where all of the quantities on the right-hand side are known. Thus, the 1-step-ahead volatility forecast at the forecast origin h is simply $\hat{\sigma}_h^2(1) = \sigma_{h+1}^2$ given earlier. For the 2-step-ahead forecast, Eq. (3.33) gives

$$\sigma_{h+2}^2 = \sigma_{h+1}^{2\alpha_1} \exp[(1-\alpha_1)\alpha_0] \exp[g(\epsilon_{h+1})].$$

Taking conditional expectation at time h, we have

$$\hat{\sigma}_h^2(2) = \hat{\sigma}_h^{2\alpha_1}(1) \exp[(1-\alpha_1)\alpha_0] E_h\{\exp[g(\epsilon_{h+1})]\},$$

where E_h denotes a conditional expectation taken at the time origin h. The prior expectation can be obtained as follows:

$$\begin{aligned}
E\{\exp[g(\epsilon)]\} &= \int_{-\infty}^{\infty} \exp[\theta\epsilon + \gamma(|\epsilon| - \sqrt{2/\pi})] f(\epsilon)\, d\epsilon \\
&= \exp\left(-\gamma\sqrt{2/\pi}\right) \left[\int_0^{\infty} e^{(\theta+\gamma)\epsilon} \frac{1}{\sqrt{2\pi}} e^{-\epsilon^2/2}\, d\epsilon \right. \\
&\quad + \left. \int_{-\infty}^{0} e^{(\theta-\gamma)\epsilon} \frac{1}{\sqrt{2\pi}} e^{-\epsilon^2/2}\, d\epsilon \right] \\
&= \exp\left(-\gamma\sqrt{2/\pi}\right) \left[e^{(\theta+\gamma)^2/2}\Phi(\theta+\gamma) + e^{(\theta-\gamma)^2/2}\Phi(\gamma-\theta)\right],
\end{aligned}$$

where $f(\epsilon)$ and $\Phi(x)$ are the probability density function and CDF of the standard normal distribution, respectively. Consequently, the 2-step-ahead volatility forecast is

$$\begin{aligned}
\hat{\sigma}_h^2(2) &= \hat{\sigma}_h^{2\alpha_1}(1) \exp\left[(1-\alpha_1)\alpha_0 - \gamma\sqrt{2/\pi}\right] \\
&\quad \times \left\{\exp[(\theta+\gamma)^2/2]\Phi(\theta+\gamma) + \exp[(\theta-\gamma)^2/2]\Phi(\gamma-\theta)\right\}.
\end{aligned}$$

Repeating the previous procedure, we obtain a recursive formula for a j-step-ahead forecast:

$$\begin{aligned}
\hat{\sigma}_h^2(j) &= \widehat{\sigma_h^{2\alpha_1}}(j-1) \exp(\omega) \\
&\quad \times \left\{\exp[(\theta+\gamma)^2/2]\Phi(\theta+\gamma) + \exp[(\theta-\gamma)^2/2]\Phi(\gamma-\theta)\right\},
\end{aligned}$$

where $\omega = (1-\alpha_1)\alpha_0 - \gamma\sqrt{2/\pi}$. The values of $\Phi(\theta+\gamma)$ and $\Phi(\gamma-\theta)$ can be obtained from most statistical packages. Alternatively, accurate approximations to these values can be obtained by using the method in Appendix B of Chapter 6.

For illustration, consider the AR(1)–EGARCH(1,1) model of the previous section for the monthly log returns of IBM stock, ending December 1997. Using the fitted EGARCH(1,1) model, we can compute the volatility forecasts for the series. At the forecast origin $t = 864$, the forecasts are $\hat{\sigma}^2_{864}(1) = 6.05 \times 10^{-3}$, $\hat{\sigma}^2_{864}(2) = 5.82 \times 10^{-3}$, $\hat{\sigma}^2_{864}(3) = 5.63 \times 10^{-3}$, and $\hat{\sigma}^2_{864}(10) = 4.94 \times 10^{-3}$. These forecasts converge gradually to the sample variance 4.37×10^{-3} of the shock process a_t of Eq. (3.30).

3.9 THE THRESHOLD GARCH MODEL

Another volatility model commonly used to handle leverage effects is the threshold GARCH (or TGARCH) model; see Glosten, Jagannathan, and Runkle (1993) and Zakoian (1994). A TGARCH(m, s) model assumes the form

$$\sigma_t^2 = \alpha_0 + \sum_{i=1}^{s} (\alpha_i + \gamma_i N_{t-i}) a_{t-i}^2 + \sum_{j=1}^{m} \beta_j \sigma_{t-j}^2, \tag{3.34}$$

where N_{t-i} is an indicator for *negative* a_{t-i}, that is,

$$N_{t-i} = \begin{cases} 1 & \text{if } a_{t-i} < 0, \\ 0 & \text{if } a_{t-i} \geq 0, \end{cases}$$

and α_i, γ_i, and β_j are nonnegative parameters satisfying conditions similar to those of GARCH models. From the model, it is seen that a positive a_{t-i} contributes $\alpha_i a_{t-i}^2$ to σ_t^2, whereas a negative a_{t-i} has a larger impact $(\alpha_i + \gamma_i) a_{t-i}^2$ with $\gamma_i > 0$. The model uses zero as its *threshold* to separate the impacts of past shocks. Other threshold values can also be used; see Chapter 4 for the general concept of threshold models. Model (3.34) is also called the GJR model because Glosten et al. (1993) proposed essentially the same model.

For illustration, consider the monthly log returns of IBM stock from 1926 to 2003. The fitted TGARCH(1,1) model with conditional GED innovations is

$$\begin{aligned} r_t &= 0.0121 + a_t, \qquad a_t = \sigma_t \epsilon_t, \\ \sigma_t^2 &= 3.45 \times 10^{-4} + (0.0658 + 0.0843 N_{t-1}) a_{t-1}^2 + 0.8182 \sigma_{t-1}^2, \end{aligned} \tag{3.35}$$

where the estimated parameter of the GED is 1.51 with standard error 0.099. The standard error of the parameter for the mean equation is 0.002 and the standard errors of the parameters in the volatility equation are $1.26\times^{-4}$, 0.0314, 0.0395, and 0.049, respectively. To check the fitted model, we have $Q(12) = 18.34(0.106)$ for the standardized residual $\tilde{a}_t$ and $Q(12) = 5.36$ (0.95) for $\tilde{a}_t^2$. The model is adequate in modeling the first two conditional moments of the log return series. Based on the fitted model, the leverage effect is significant at the 5% level.

S-Plus Commands Used

```
> ibm.tgarch = garch(ibmln~1,~tgarch(1,1),leverage=T,
+ cond.dist='ged')
> summary(ibm.tgarch)
> plot(ibm.tgarch)
```

It is interesting to compare the two models in Eqs. (3.32) and (3.35) for the monthly log returns of IBM stock. Assume that $a_{t-1} = \pm 2\sigma_{t-1}$ so that $\epsilon_{t-1} = \pm 2$. The EGARCH(1,1) model gives

$$\frac{\sigma_t^2(\epsilon_{t-1} = -2)}{\sigma_t^2(\epsilon_{t-1} = 2)} = e^{0.22\times 2\times 0.632} \approx 1.264.$$

On the other hand, ignoring the constant term 0.000345, the TGARCH(1,1) model gives

$$\frac{\sigma_t^2(\epsilon_{t-1} = -2)}{\sigma_t^2(\epsilon_{t-1} = 2)} \approx \frac{[(0.0658 + 0.0843)4 + 0.8182]\sigma_{t-1}^2}{(0.0658 \times 4 + 0.8182)\sigma_{t-1}^2} = 1.312.$$

The two models provide similar leverage effects.

3.10 THE CHARMA MODEL

Many other econometric models have been proposed in the literature to describe the evolution of the conditional variance σ_t^2 in Eq. (3.2). We mention the conditional heteroscedastic ARMA (CHARMA) model that uses random coefficients to produce conditional heteroscedasticity; see Tsay (1987). The CHARMA model is not the same as the ARCH model, but the two models have similar second-order conditional properties. A CHARMA model is defined as

$$r_t = \mu_t + a_t, \qquad a_t = \delta_{1t}a_{t-1} + \delta_{2t}a_{t-2} + \cdots + \delta_{mt}a_{t-m} + \eta_t, \tag{3.36}$$

where $\{\eta_t\}$ is a Gaussian white noise series with mean zero and variance σ_η^2, $\{\boldsymbol{\delta}_t\} = \{(\delta_{1t}, \ldots, \delta_{mt})'\}$ is a sequence of iid random vectors with mean zero and nonnegative definite covariance matrix $\boldsymbol{\Omega}$, and $\{\boldsymbol{\delta}_t\}$ is independent of $\{\eta_t\}$. In this section, we use some basic properties of vector and matrix operations to simplify the presentation. Readers may consult Appendix A of Chapter 8 for a brief review of these properties. For $m > 0$, the model can be written as

$$a_t = \boldsymbol{a}_{t-1}'\boldsymbol{\delta}_t + \eta_t,$$

where $\boldsymbol{a}_{t-1} = (a_{t-1}, \ldots, a_{t-m})'$ is a vector of lagged values of a_t and is available at time $t - 1$. The conditional variance of a_t of the CHARMA model in Eq. (3.36) is then

$$\begin{aligned}\sigma_t^2 &= \sigma_\eta^2 + \boldsymbol{a}'_{t-1}\mathrm{Cov}(\boldsymbol{\delta}_t)\boldsymbol{a}_{t-1}\\ &= \sigma_\eta^2 + (a_{t-1}, \ldots, a_{t-m})\boldsymbol{\Omega}(a_{t-1}, \ldots, a_{t-m})'. \end{aligned} \tag{3.37}$$

Denote the (i, j)th element of $\boldsymbol{\Omega}$ by ω_{ij}. Because the matrix is symmetric, we have $\omega_{ij} = \omega_{ji}$. If $m = 1$, then Eq. (3.37) reduces to $\sigma_t^2 = \sigma_\eta^2 + \omega_{11}a_{t-1}^2$, which is an ARCH(1) model. If $m = 2$, then Eq. (3.37) reduces to

$$\sigma_t^2 = \sigma_\eta^2 + \omega_{11}a_{t-1}^2 + 2\omega_{12}a_{t-1}a_{t-2} + \omega_{22}a_{t-2}^2,$$

which differs from an ARCH(2) model by the cross-product term $a_{t-1}a_{t-2}$. In general, the conditional variance of a CHARMA(m) model is equivalent to that of an ARCH(m) model if $\boldsymbol{\Omega}$ is a diagonal matrix. Because $\boldsymbol{\Omega}$ is a covariance matrix, which is nonnegative definite, and σ_η^2 is a variance, which is positive, we have $\sigma_t^2 \geq \sigma_\eta^2 > 0$ for all t. In other words, the positiveness of σ_t^2 is automatically satisfied under a CHARMA model.

An obvious difference between ARCH and CHARMA models is that the latter use cross products of the lagged values of a_t in the volatility equation. The cross-product terms might be useful in some applications. For example, in modeling an asset return series, cross-product terms denote interactions between previous returns. It is conceivable that stock volatility may depend on such interactions. However, the number of cross-product terms increases rapidly with the order m, and some constraints are needed to keep the model simple. A possible constraint is to use a small number of cross-product terms in a CHARMA model. Another difference between the two models is that higher order properties of CHARMA models are harder to obtain than those of ARCH models because it is in general harder to handle multiple random variables.

For illustration, we employ the CHARMA model

$$r_t = \phi_0 + a_t, \qquad a_t = \delta_{1t}a_{t-1} + \delta_{2t}a_{t-2} + \eta_t$$

for the monthly excess returns of the S&P 500 index used before in GARCH modeling. The fitted model is

$$r_t = 0.00635 + a_t, \qquad \sigma_t^2 = 0.00179 + (a_{t-1}, a_{t-2})\widehat{\boldsymbol{\Omega}}(a_{t-1}, a_{t-2})',$$

where

$$\widehat{\boldsymbol{\Omega}} = \begin{bmatrix} 0.1417(0.0333) & -0.0594(0.0365) \\ -0.0594(0.0365) & 0.3081(0.0340) \end{bmatrix},$$

where the numbers in parentheses are standard errors. The cross-product term of $\widehat{\boldsymbol{\Omega}}$ has a t ratio of -1.63, which is marginally significant at the 10% level. If we refine the model to

$$r_t = \phi_0 + a_t, \qquad a_t = \delta_{1t}a_{t-1} + \delta_{2t}a_{t-2} + \delta_{3t}a_{t-3} + \eta_t,$$

but assume that δ_{3t} is uncorrelated with $(\delta_{1t}, \delta_{2t})$, then we obtain the fitted model

$$r_t = 0.0068 + a_t, \qquad \sigma_t^2 = 0.00136 + (a_{t-1}, a_{t-2}, a_{t-3})\widehat{\mathbf{\Omega}}(a_{t-1}, a_{t-2}, a_{t-3})',$$

where the elements of $\widehat{\mathbf{\Omega}}$ and their standard errors, shown in parentheses, are

$$\widehat{\mathbf{\Omega}} = \begin{bmatrix} 0.1212(0.0355) & -0.0622(0.0283) & 0 \\ -0.0622(0.0283) & 0.1913(0.0254) & 0 \\ 0 & 0 & 0.2988(0.0420) \end{bmatrix}.$$

All of the estimates are now statistically significant at the 5% level. From the model, $a_t = r_t - 0.0068$ is the deviation of the monthly excess return from its average. The fitted CHARMA model shows that there is some interaction effect between the first two lagged deviations. Indeed, the volatility equation can be written approximately as

$$\sigma_t^2 = 0.00136 + 0.12a_{t-1}^2 - 0.12a_{t-1}a_{t-2} + 0.19a_{t-2}^2 + 0.30a_{t-3}^2.$$

The conditional variance is slightly larger when $a_{t-1}a_{t-2}$ is negative.

3.10.1 Effects of Explanatory Variables

The CHARMA model can easily be generalized so that the volatility of r_t may depend on some explanatory variables. Let $\{x_{it}\}_{i=1}^m$ be m explanatory variables available at time t. Consider the model

$$r_t = \mu_t + a_t, \qquad a_t = \sum_{i=1}^m \delta_{it} x_{i,t-1} + \eta_t, \tag{3.38}$$

where $\boldsymbol{\delta}_t = (\delta_{1t}, \ldots, \delta_{mt})'$ and η_t are random vector and variable defined in Eq. (3.36). Then the conditional variance of a_t is

$$\sigma_t^2 = \sigma_\eta^2 + (x_{1,t-1}, \ldots, x_{m,t-1})\mathbf{\Omega}(x_{1,t-1}, \ldots, x_{m,t-1})'.$$

In application, the explanatory variables may include some lagged values of a_t.

3.11 RANDOM COEFFICIENT AUTOREGRESSIVE MODELS

In the literature, the random coefficient autoregressive (RCA) model is introduced to account for variability among different subjects under study, similar to the panel data analysis in econometrics and the hierarchical model in statistics. We classify the RCA model as a conditional heteroscedastic model, but historically it is used to obtain a better description of the conditional mean equation of the process by allowing for the parameters to evolve over time. A time series r_t is said to follow an RCA(p) model if it satisfies

$$r_t = \phi_0 + \sum_{i=1}^{p} (\phi_i + \delta_{it}) r_{t-i} + a_t, \tag{3.39}$$

where p is a positive integer, $\{\boldsymbol{\delta}_t\} = \{(\delta_{1t}, \ldots, \delta_{pt})'\}$ is a sequence of independent random vectors with mean zero and covariance matrix $\boldsymbol{\Omega}_\delta$, and $\{\boldsymbol{\delta}_t\}$ is independent of $\{a_t\}$; see Nicholls and Quinn (1982) for further discussions of the model. The conditional mean and variance of the RCA model in Eq. (3.39) are

$$\mu_t = E(r_t | F_{t-1}) = \phi_0 + \sum_{i=1}^{p} \phi_i r_{t-i},$$

$$\sigma_t^2 = \sigma_a^2 + (r_{t-1}, \ldots, r_{t-p}) \boldsymbol{\Omega}_\delta (r_{t-1}, \ldots, r_{t-p})',$$

which is in the same form as that of a CHARMA model. However, there is a subtle difference between RCA and CHARMA models. For the RCA model, the volatility is a quadratic function of the observed lagged values r_{t-i}. Yet the volatility is a quadratic function of the lagged innovations a_{t-i} in a CHARMA model.

3.12 STOCHASTIC VOLATILITY MODEL

An alternative approach to describe the volatility evolution of a financial time series is to introduce an innovation to the conditional variance equation of a_t; see Melino and Turnbull (1990), Taylor (1994), Harvey, Ruiz, and Shephard (1994), and Jacquier, Polson, and Rossi (1994). The resulting model is referred to as a stochastic volatility (SV) model. Similar to EGARCH models, to ensure positiveness of the conditional variance, SV models use $\ln(\sigma_t^2)$ instead of σ_t^2. A SV model is defined as

$$a_t = \sigma_t \epsilon_t, \qquad (1 - \alpha_1 B - \cdots - \alpha_m B^m) \ln(\sigma_t^2) = \alpha_0 + v_t, \tag{3.40}$$

where the ϵ_t are iid $N(0, 1)$, the v_t are iid $N(0, \sigma_v^2)$, $\{\epsilon_t\}$ and $\{v_t\}$ are independent, α_0 is a constant, and all zeros of the polynomial $1 - \sum_{i=1}^{m} \alpha_i B^i$ are greater than 1 in modulus. Adding the innovation v_t substantially increases the flexibility of the model in describing the evolution of σ_t^2, but it also increases the difficulty in parameter estimation. To estimate an SV model, we need a quasi-likelihood method via Kalman filtering or a Monte Carlo method. Jacquier, Polson, and Rossi (1994) provide some comparison of estimation results between quasi-likelihood and Markov chain Monte Carlo (MCMC) methods. The difficulty in estimating an SV model is understandable because for each shock a_t the model uses two innovations ϵ_t and v_t. We discuss an MCMC method to estimate SV models in Chapter 12. For more discussions on stochastic volatility models, see Taylor (1994).

The appendixes of Jacquier, Polson, and Rossi (1994) provide some properties of the SV model when $m = 1$. For instance, with $m = 1$, we have

$$\ln(\sigma_t^2) \sim N\left(\frac{\alpha_0}{1-\alpha_1}, \frac{\sigma_v^2}{1-\alpha_1^2}\right) \equiv N(\mu_h, \sigma_h^2),$$

and $E(a_t^2) = \exp(\mu_h + \sigma_h^2/2)$, $E(a_t^4) = 3\exp(2\mu_h^2 + 2\sigma_h^2)$, and $\text{corr}(a_t^2, a_{t-i}^2) = [\exp(\sigma_h^2\alpha_1^i) - 1]/[3\exp(\sigma_h^2) - 1]$. Limited experience shows that SV models often provided improvements in model fitting, but their contributions to out-of-sample volatility forecasts received mixed results.

3.13 LONG-MEMORY STOCHASTIC VOLATILITY MODEL

More recently, the SV model is further extended to allow for long memory in volatility, using the idea of fractional difference. As stated in Chapter 2, a time series is a long-memory process if its autocorrelation function decays at a hyperbolic, instead of an exponential, rate as the lag increases. The extension to long-memory models in volatility study is motivated by the fact that the autocorrelation function of the squared or absolute-valued series of an asset return often decays slowly, even though the return series has no serial correlation; see Ding, Granger, and Engle (1993). Figure 3.10 shows the sample ACF of the daily absolute returns for IBM stock and the S&P 500 index from July 3, 1962, to December 31, 2003. These sample ACFs are positive with moderate magnitude but decay slowly.

A simple long-memory stochastic volatility (LMSV) model can be written as

$$a_t = \sigma_t\epsilon_t, \qquad \sigma_t = \sigma\exp(u_t/2), \qquad (1-B)^d u_t = \eta_t, \tag{3.41}$$

where $\sigma > 0$, the ϵ_t are iid $N(0, 1)$, the η_t are iid $N(0, \sigma_\eta^2)$ and independent of ϵ_t, and $0 < d < 0.5$. The feature of long memory stems from the fractional difference $(1 - B)^d$, which implies that the ACF of u_t decays slowly at a hyperbolic, instead of an exponential, rate as the lag increases. For model (3.41), we have

$$\begin{aligned}
\ln(a_t^2) &= \ln(\sigma^2) + u_t + \ln(\epsilon_t^2) \\
&= [\ln(\sigma^2) + E(\ln\epsilon_t^2)] + u_t + [\ln(\epsilon_t^2) - E(\ln\epsilon_t^2)] \\
&\equiv \mu + u_t + e_t.
\end{aligned}$$

Thus, the $\ln(a_t^2)$ series is a Gaussian long-memory signal plus a non-Gaussian white noise; see Breidt, Crato, and de Lima (1998). Estimation of the LMSV model is complicated, but the fractional difference parameter d can be estimated by using either a quasi-maximum-likelihood method or a regression method. Using the log series of squared daily returns for companies in the S&P 500 index, Bollerslev and Jubinski (1999) and Ray and Tsay (2000) found that the median estimate of d is about 0.38. For applications, Ray and Tsay (2000) studied common long-memory

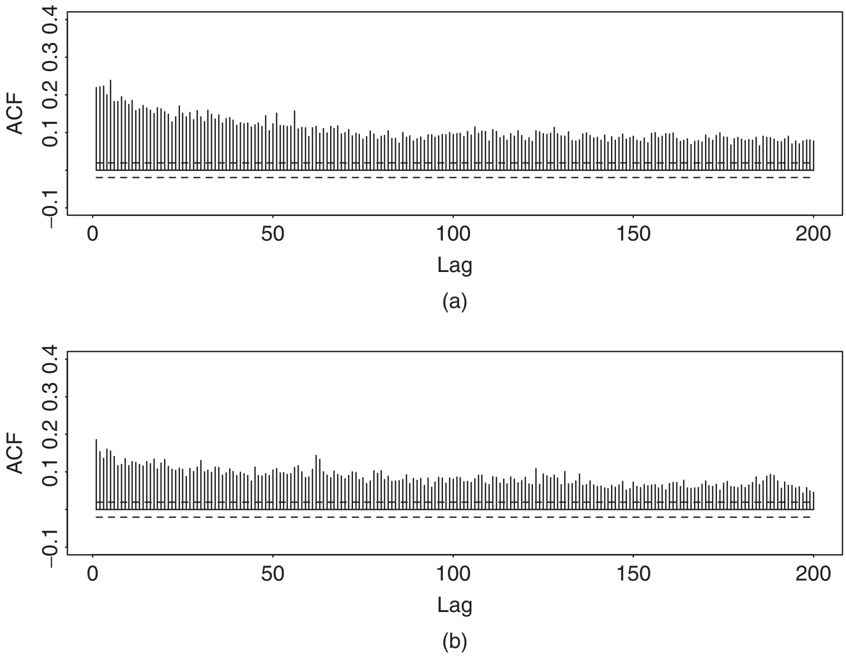

Figure 3.10 Sample ACF of daily absolute log returns for (a) S&P 500 index and (b) IBM stock for period from July 3, 1962, to December 31, 2003. Two horizontal lines denote asymptotic 5% limits.

components in daily stock volatilities of groups of companies classified by various characteristics. They found that companies in the same industrial or business sector tend to have more common long-memory components (e.g., big U.S. national banks and financial institutions).

3.14 APPLICATION

In this section, we apply the volatility models discussed in this chapter to investigate some problems of practical importance. The data used are the monthly log returns of IBM stock and the S&P 500 index from January 1926 to December 1999. There are 888 observations, and the returns are in percentages and include dividends. Figure 3.11 shows the time plots of the two return series. Note that the result of this section was obtained by the RATS program.

Example 3.4. The questions we address here are whether the daily volatility of a stock is lower in the summer and, if so, by how much. Affirmative answers to these two questions have practical implications in stock option pricing. We use the monthly log returns of IBM stock shown in Figure 3.11(a) as an illustrative example.

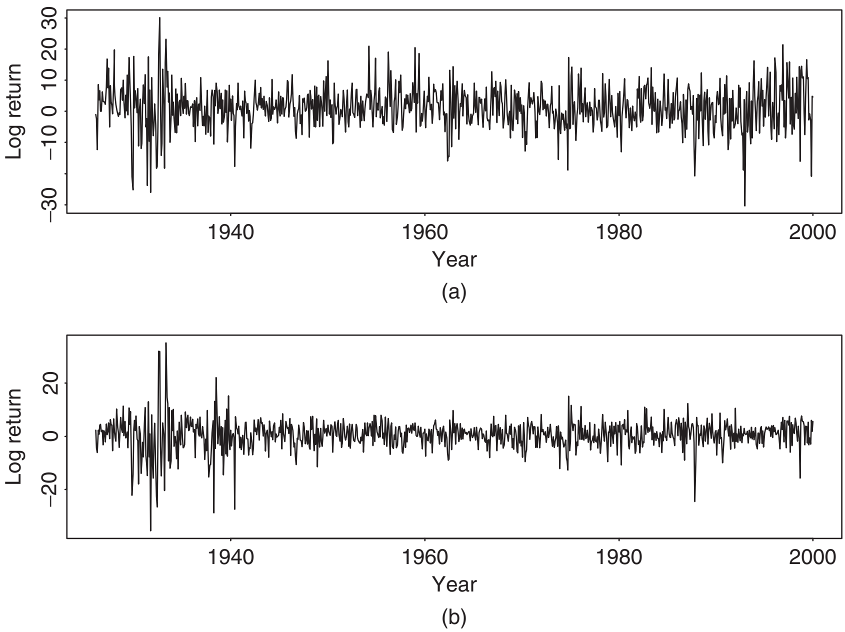

Figure 3.11 Time plots of monthly log returns for (a) IBM stock and (b) S&P 500 index. Sample period is from January 1926 to December 1999. Returns are in percentages and include dividends.

Denote the monthly log return series by r_t. If Gaussian GARCH models are entertained, we obtain the GARCH(1,1) model:

$$
\begin{aligned}
r_t &= 1.23 + 0.099 r_{t-1} + a_t, \qquad a_t = \sigma_t \epsilon_t, \\
\sigma_t^2 &= 3.206 + 0.103 a_{t-1}^2 + 0.825 \sigma_{t-1}^2,
\end{aligned} \tag{3.42}
$$

for the series. The standard errors of the two parameters in the mean equation are 0.222 and 0.037, respectively, whereas those of the parameters in the volatility equation are 0.947, 0.021, and 0.037, respectively. Using the standardized residuals $\tilde{a}_t = a_t/\sigma_t$, we obtain $Q(10) = 7.82(0.553)$ and $Q(20) = 21.22(0.325)$, where the p value is in parentheses. Therefore, there are no serial correlations in the residuals of the mean equation. The Ljung–Box statistics of the $\tilde{a}_t^2$ series show $Q(10) = 2.89(0.98)$ and $Q(20) = 7.26(0.99)$, indicating that the standardized residuals have no conditional heteroscedasticity. The fitted model seems adequate. This model serves as a starting point for further study.

To study the summer effect on stock volatility of an asset, we define an indicator variable

$$
u_t = \begin{cases} 1 & \text{if } t \text{ is June, July, or August,} \\ 0 & \text{otherwise,} \end{cases} \tag{3.43}
$$

and modify the volatility equation to

$$\sigma_t^2 = \alpha_0 + \alpha_1 a_{t-1}^2 + \beta_1 \sigma_{t-1}^2 + u_t(\alpha_{00} + \alpha_{10} a_{t-1}^2 + \beta_{10}\sigma_{t-1}^2).$$

This equation uses two GARCH(1,1) models to describe the volatility of a stock return; one model for the summer months and the other for the remaining months. For the monthly log returns of IBM stock, estimation results show that the estimates of α_{10} and β_{10} are statistically nonsignificant at the 10% level. Therefore, we refine the equation and obtain the model

$$\begin{aligned} r_t &= 1.21 + 0.099 r_{t-1} + a_t, \qquad a_t = \sigma_t \epsilon_t, \\ \sigma_t^2 &= 4.539 + 0.113 a_{t-1}^2 + 0.816\sigma_{t-1}^2 - 5.154 u_t. \end{aligned} \tag{3.44}$$

The standard errors of the parameters in the mean equation are 0.218 and 0.037, respectively, and those of the parameters in the volatility equation are 1.071, 0.022, 0.037, and 1.900, respectively. The Ljung–Box statistics for the standardized residuals $\tilde{a}_t = a_t/\sigma_t$ show $Q(10) = 7.66(0.569)$ and $Q(20) = 21.64(0.302)$. Therefore, there are no serial correlations in the standardized residuals. The Ljung–Box statistics for $\tilde{a}_t^2$ give $Q(10) = 3.38(0.97)$ and $Q(20) = 6.82(0.99)$, indicating no conditional heteroscedasticity in the standardized residuals either. The refined model seems adequate.

Comparing the volatility models in Eqs. (3.42) and (3.44), we obtain the following conclusions. First, because the coefficient -5.514 is significantly different from zero with a p value of 0.0067, the summer effect on stock volatility is statistically significant at the 1% level. Furthermore, the negative sign of the estimate confirms that the volatility of IBM monthly log stock returns is indeed lower during the summer. Second, rewrite the volatility model in Eq. (3.44) as

$$\sigma_t^2 = \begin{cases} -0.615 + 0.113 a_{t-1}^2 + 0.816\sigma_{t-1}^2 & \text{if } t \text{ is June, July,or August,} \\ 4.539 + 0.113 a_{t-1}^2 + 0.816\sigma_{t-1}^2 & \text{otherwise.} \end{cases}$$

The negative constant term $-0.615 = 4.539 - 5.514$ is counterintuitive. However, since the standard errors of 4.539 and 5.514 are relatively large, the estimated difference -0.615 might not be significantly different from zero. To verify the assertion, we refit the model by imposing the constraint that the constant term of the volatility equation is zero for the summer months. This can easily be done by using the equation

$$\sigma_t^2 = \alpha_1 a_{t-1}^2 + \beta_1 \sigma_{t-1}^2 + \gamma(1 - u_t).$$

The fitted model is

$$\begin{aligned} r_t &= 1.21 + 0.099 r_{t-1} + a_t, \qquad a_t = \sigma_t \epsilon_t, \\ \sigma_t^2 &= 0.114 a_{t-1}^2 + 0.811\sigma_{t-1}^2 + 4.552(1 - u_t). \end{aligned} \tag{3.45}$$

The standard errors of the parameters in the mean equation are 0.219 and 0.038, respectively, and those of the parameters in the volatility equation are 0.022, 0.034, and 1.094, respectively. The Ljung–Box statistics of the standardized residuals show $Q(10) = 7.68$ and $Q(20) = 21.67$, and those of the $\tilde{a}_t^2$ series give $Q(10) = 3.17$ and $Q(20) = 6.85$. These test statistics are close to what we had before and are not significant at the 5% level.

The volatility Eq. (3.45) can readily be used to assess the summer effect on the IBM stock volatility. For illustration, based on the model in Eq. (3.45) the medians of a_t^2 and σ_t^2 are 29.4 and 75.1, respectively, for the IBM monthly log returns in 1999. Using these values, we have $\sigma_t^2 = 0.114 \times 29.4 + 0.811 \times 75.1 = 64.3$ for the summer months and $\sigma_t^2 = 68.8$ for the other months. The ratio of the two volatilities is $64.3/68.8 \approx 93\%$. Thus, there is a 7% reduction in the volatility of the monthly log return of IBM stock in the summer months.

Example 3.5. The S&P 500 index is widely used in the derivative markets. As such, modeling its volatility is a subject of intensive study. The question we ask in this example is whether the past returns of individual components of the index contribute to the modeling of the S&P 500 index volatility in the presence of its own returns. A thorough investigation on this topic is beyond the scope of this chapter, but we use the past returns of IBM stock as explanatory variables to address the question.

The data used are shown in Figure 3.11. Denote by r_t the monthly log return series of the S&P 500 index. Using the r_t series and Gaussian GARCH models, we obtain the following special GARCH(2,1) model:

$$r_t = 0.609 + a_t, \quad a_t = \sigma_t \epsilon_t, \quad \sigma_t^2 = 0.717 + 0.147 a_{t-2}^2 + 0.839 \sigma_{t-1}^2. \tag{3.46}$$

The standard error of the constant term in the mean equation is 0.138, and those of the parameters in the volatility equation are 0.214, 0.021, and 0.017, respectively. Based on the standardized residuals $\tilde{a}_t = a_t/\sigma_t$, we have $Q(10) = 11.51(0.32)$ and $Q(20) = 23.71(0.26)$, where the number in parentheses denotes the p value. For the $\tilde{a}_t^2$ series, we have $Q(10) = 9.42(0.49)$ and $Q(20) = 13.01(0.88)$. Therefore, the model seems adequate at the 5% significance level.

Next, we evaluate the contributions, if any, of using the past returns of IBM stock, which is a component of the S&P 500 index, in modeling the index volatility. As a simple illustration, we modify the volatility equation as

$$\sigma_t^2 = \alpha_0 + \alpha_2 a_{t-2}^2 + \beta_1 \sigma_{t-1}^2 + \gamma (x_{t-1} - 1.24)^2,$$

where x_t is the monthly log return of IBM stock and 1.24 is the sample mean of x_t. The fitted model for r_t becomes

$$\begin{aligned} r_t &= 0.616 + a_t, \qquad a_t = \sigma_t \epsilon_t, \\ \sigma_t^2 &= 1.069 + 0.148 a_{t-2}^2 + 0.834 \sigma_{t-1}^2 - 0.007 (x_{t-1} - 1.24)^2. \end{aligned} \tag{3.47}$$

TABLE 3.4 Fitted Volatilities for Monthly Log Returns of S&P 500 Index from July to December 1999 Using Models with and without Past Log Return of IBM Stock

Month	7/99	8/99	9/99	10/99	11/99	12/99
Model (3.46)	26.30	26.01	24.73	21.69	20.71	22.46
Model (3.47)	23.32	23.13	22.46	20.00	19.45	18.27

The standard error of the parameter in the mean equation is 0.139 and the standard errors of the parameters in the volatility equation are 0.271, 0.020, 0.018, and 0.002, respectively. For model checking, we have $Q(10) = 11.39(0.33)$ and $Q(20) = 23.63(0.26)$ for the standardized residuals $\tilde{a}_t = a_t/\sigma_t$ and $Q(10) = 9.35(0.50)$ and $Q(20) = 13.51(0.85)$ for the $\tilde{a}_t^2$ series. Therefore, the model is adequate.

Since the p value for testing $\gamma = 0$ is 0.0039, the contribution of the lag-1 IBM stock return to the S&P 500 index volatility is statistically significant at the 1% level. The negative sign is understandable because it implies that using the lag-1 past return of IBM stock reduces the volatility of the S&P 500 index return. Table 3.4 gives the fitted volatility of the S&P 500 index from July to December of 1999 using models (3.46) and (3.47). From the table, the past value of IBM log stock return indeed contributes to the modeling of the S&P 500 index volatility.

3.15 ALTERNATIVE APPROACHES

In this section, we discuss two alternative methods to volatility modeling.

3.15.1 Use of High-Frequency Data

French, Schwert, and Stambaugh (1987) consider an alternative approach for volatility estimation that uses high-frequency data to calculate volatility of low-frequency returns. In recent years, this approach has attracted substantial interest due to the availability of high-frequency financial data; see Andersen, Bollerslev, Diebold, and Labys (2001a, 2001b).

Suppose that we are interested in the monthly volatility of an asset for which daily returns are available. Let r_t^m be the monthly log return of the asset at month t. Assume that there are n trading days in month t and the daily log returns of the asset in the month are $\{r_{t,i}\}_{i=1}^n$. Using properties of log returns, we have

$$r_t^m = \sum_{i=1}^n r_{t,i}.$$

Assuming that the conditional variance and covariance exist, we have

$$\operatorname{Var}(r_t^m|F_{t-1}) = \sum_{i=1}^n \operatorname{Var}(r_{t,i}|F_{t-1}) + 2\sum_{i<j} \operatorname{Cov}[(r_{t,i}, r_{t,j})|F_{t-1}], \tag{3.48}$$

where F_{t-1} denotes the information available at month $t-1$ (inclusive). The prior equation can be simplified if additional assumptions are made. For example, if we assume that $\{r_{t,i}\}$ is a white noise series, then

$$\mathrm{Var}(r_t^m|F_{t-1}) = n\mathrm{Var}(r_{t,1}),$$

where $\mathrm{Var}(r_{t,1})$ can be estimated from the daily returns $\{r_{t,i}\}_{i=1}^n$ by

$$\hat{\sigma}^2 = \frac{\sum_{i=1}^n (r_{t,i} - \bar{r}_t)^2}{n-1},$$

where $\bar{r}_t$ is the sample mean of the daily log returns in month t [i.e., $\bar{r}_t = (\sum_{i=1}^n r_{t,i})/n$]. The estimated monthly volatility is then

$$\hat{\sigma}_m^2 = \frac{n}{n-1}\sum_{i=1}^n (r_{t,i} - \bar{r}_t)^2. \tag{3.49}$$

If $\{r_{t,i}\}$ follows an MA(1) model, then

$$\mathrm{Var}(r_t^m|F_{t-1}) = n\mathrm{Var}(r_{t,1}) + 2(n-1)\mathrm{Cov}(r_{t,1}, r_{t,2}),$$

which can be estimated by

$$\hat{\sigma}_m^2 = \frac{n}{n-1}\sum_{i=1}^n (r_{t,i} - \bar{r}_t)^2 + 2\sum_{i=1}^{n-1} (r_{t,i} - \bar{r}_t)(r_{t,i+1} - \bar{r}_t). \tag{3.50}$$

The previous approach for volatility estimation is simple, but it encounters several difficulties in practice. First, the model for daily returns $\{r_{t,i}\}$ is unknown. This complicates the estimation of covariances in Eq. (3.48). Second, there are roughly 21 trading days in a month, resulting in a small sample size. The accuracy of the estimates of variance and covariance in Eq. (3.48) might be questionable. The accuracy depends on the dynamic structure of $\{r_{t,i}\}$ and their distribution. If the daily log returns have high excess kurtosis and serial correlations, then the sample estimates $\hat{\sigma}_m^2$ in Eqs. (3.49) and (3.50) may not even be consistent; see Bai, Russell, and Tiao (2004). Further research is needed to make this approach valuable.

Example 3.6. Consider the monthly volatility of the log returns of the S&P 500 index from January 1980 to December 1999. We calculate the volatility by three methods. In the first method, we use daily log returns and Eq. (3.49) (i.e., assuming that the daily log returns form a white noise series). The second method also uses daily returns but assumes an MA(1) model [i.e., using Eq. (3.50)]. The third method applies a GARCH(1,1) model to the monthly returns from January

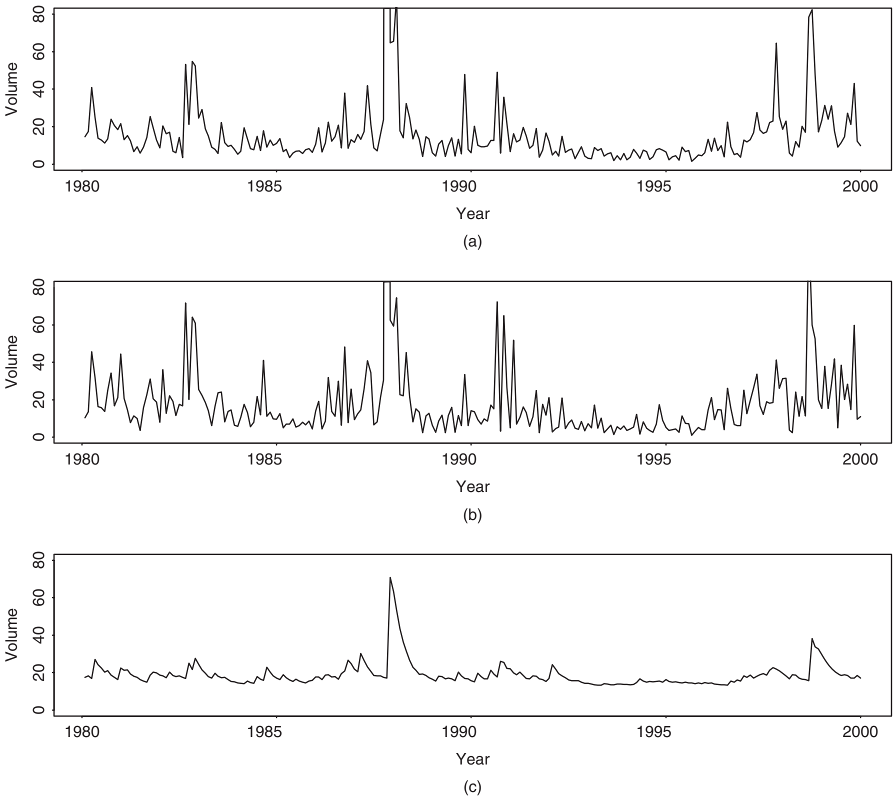

Figure 3.12 Time plots of estimated monthly volatility for log returns of S&P 500 index from January 1980 to December 1999: (a) assumes that daily log returns form a white noise series, (b) assumes that daily log returns follow an MA(1) model, and (c) uses monthly returns from January 1962 to December 1999 and a GARCH(1,1) model.

1962 to December 1999. We use a longer data span to obtain a more accurate estimate of the monthly volatility. The GARCH(1,1) model used is

$$r_t^m = 0.658 + a_t, \qquad a_t = \sigma_t \epsilon_t, \qquad \sigma_t^2 = 3.349 + 0.086 a_{t-1}^2 + 0.735 \sigma_{t-1}^2,$$

where ϵ_t is a standard Gaussian white noise series. Figure 3.12 shows the time plots of the estimated monthly volatility. Clearly the estimated volatilities based on daily returns are much higher than those based on monthly returns and a GARCH(1,1) model. In particular, the estimated volatility for October 1987 was about 680 when daily returns are used. The plots shown were truncated to have the same scale.

In Eq. (3.49), if we further assume that the sample mean $\bar{r}_t$ is zero, then we have $\hat{\sigma}_m^2 \approx \sum_{i=1}^{n} r_{t,i}^2$. In this case, the cumulative sum of squares of daily log returns in a month is used as an estimate of monthly volatility. This concept has been generalized to estimate daily volatility of an asset by using intradaily log returns.

Let r_t be the daily log return of an asset. Suppose that there are n equally spaced intradaily log returns available such that $r_t = \sum_{i=1}^{n} r_{t,i}$. The quantity

$$\mathrm{RV}_t = \sum_{i=1}^{n} r_{t,i}^2,$$

is called the *realized* volatility of r_t; see Andersen et al. (2001a,b). Mathematically, realized volatility is a quadratic variation of r_t, and it assumes that $\{r_{t,i}\}_{i=1}^{n}$ forms an iid sequence with mean zero and finite variance. Limited experience indicates that $\ln(\mathrm{RV}_t)$ often follows approximately a Gaussian ARIMA(0,1,q) model, which can be used to produce forecasts. See demonstration in Section 1.1 for further information.

Advantages of realized volatility include simplicity and making use of intradaily returns. Intuitively, one would like to use as much information as possible by choosing a large n. However, when the time interval between $r_{t,i}$ is small, the returns are subject to the effects of market microstructure, for example, bid–ask bounce, which often results in a biased estimate of the volatility. The problem of choosing an optimal time interval for constructing realized volatility has attracted much research lately. For heavily traded assets in the United States, a time interval of 4–15 minutes is often used. Another problem of using realized volatility for stock returns is that the overnight return, which is the return from the closing price of day $t-1$ to the opening price of t, tends to be substantial. Ignoring overnight returns can seriously underestimate the volatility. On the other hand, our limited experience shows that overnight returns appear to be small for index returns or foreign exchange returns.

In a series of recent articles, Barndorff-Nielsen and Shephard (2004) have used high-frequency returns to study bi-power variations of an asset return and developed some methods to detect jumps in volatility.

3.15.2 Use of Daily Open, High, Low, and Close Prices

For many assets, daily opening, high, low, and closing prices are available. Parkinson (1980), Garman and Klass (1980), Rogers and Satchell (1991), and Yang and Zhang (2000) showed that one can use such information to improve volatility estimation. Figure 3.13 shows a time plot of price versus time for the tth trading day, assuming that time is continuous. For an asset, define the following variables:

- C_t = closing price of the tth trading day.
- O_t = opening price of the tth trading day.
- f = fraction of the day (in interval [0,1]) that trading is closed.
- H_t = highest price of the tth trading period.
- L_t = lowest price of the tth trading period.
- F_{t-1} = public information available at time $t-1$.

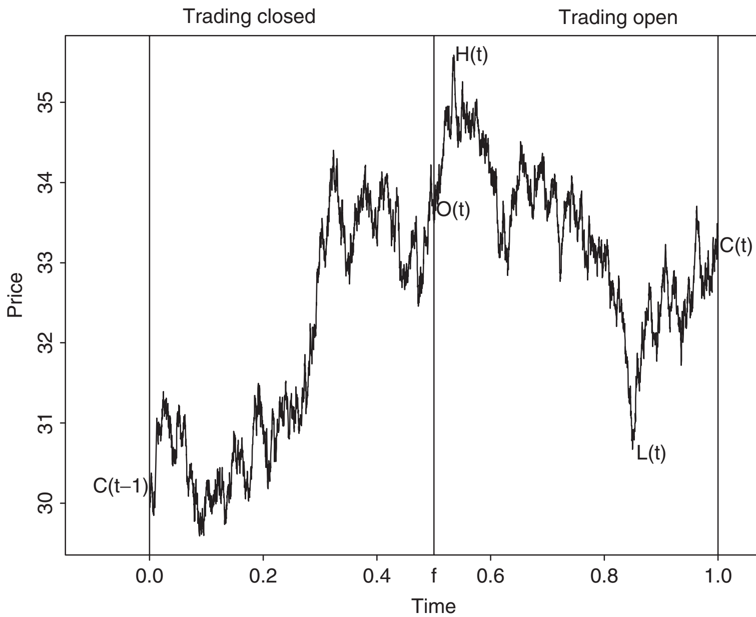

Figure 3.13 Time plot of price over time: scale for price is arbitrary.

The conventional variance (or volatility) is $\sigma_t^2 = E[(C_t - C_{t-1})^2|F_{t-1}]$. Garman and Klass (1980) considered several estimates of σ_t^2 assuming that the price follows a simple diffusion model without drift; see Chapter 6 for more information about stochastic diffusion models. The estimators considered include:

- $\hat{\sigma}_{0,t}^2 = (C_t - C_{t-1})^2$.
- $\hat{\sigma}_{1,t}^2 = \dfrac{(O_t - C_{t-1})^2}{2f} + \dfrac{(C_t - O_t)^2}{2(1-f)}, \qquad 0 < f < 1.$
- $\hat{\sigma}_{2,t}^2 = \dfrac{(H_t - L_t)^2}{4\ln(2)} \approx 0.3607(H_t - L_t)^2.$
- $\hat{\sigma}_{3,t}^2 = 0.17\dfrac{(O_t - C_{t-1})^2}{f} + 0.83\dfrac{(H_t - L_t)^2}{(1-f)4\ln(2)}, \qquad 0 < f < 1.$
- $\hat{\sigma}_{5,t}^2 = 0.5(H_t - L_t)^2 - [2\ln(2) - 1](C_t - O_t)^2$, which is $\approx 0.5(H_t - L_t)^2 - 0.386(C_t - O_t)^2$.
- $\hat{\sigma}_{6,t}^2 = 0.12\dfrac{(O_t - C_{t-1})^2}{f} + 0.88\dfrac{\hat{\sigma}_{5,t}^2}{1-f}, \qquad 0 < f < 1.$

A more precise, but complicated, estimator $\hat{\sigma}_{4,t}^2$ was also considered. However, it is close to $\hat{\sigma}_{5,t}^2$. Defining the efficiency factor of a volatility estimator as

$$\mathrm{Eff}(\hat{\sigma}_{i,t}^2) = \frac{\mathrm{Var}(\hat{\sigma}_{0,t}^2)}{\mathrm{Var}(\hat{\sigma}_{i,t}^2)},$$

Garman and Klass (1980) found that $\mathrm{Eff}(\hat{\sigma}^2_{i,t})$ is approximately 2, 5.2, 6.2, 7.4, and 8.4 for $i = 1$, 2, 3, 5, and 6, respectively, for the simple diffusion model entertained. Note that $\hat{\sigma}^2_{2,t}$ was derived by Parkinson (1980) with $f = 0$.

Turn to log returns. Define the following:

- $o_t = \ln(O_t) - \ln(C_{t-1})$, the normalized open.
- $u_t = \ln(H_t) - \ln(O_t)$, the normalized high.
- $d_t = \ln(L_t) - \ln(O_t)$, the normalized low.
- $c_t = \ln(C_t) - \ln(O_t)$, the normalized close.

Suppose that there are n days of data available and the volatility is constant over the period. Yang and Zhang (2000) recommend the estimate

$$\hat{\sigma}^2_{yz} = \hat{\sigma}^2_o + k\hat{\sigma}^2_c + (1-k)\hat{\sigma}^2_{rs}$$

as a robust estimator of the volatility, where

$$\hat{\sigma}^2_o = \frac{1}{n-1}\sum_{t=1}^{n}(o_t - \bar{o})^2 \quad \text{with} \quad \bar{o} = \frac{1}{n}\sum_{t=1}^{n} o_t,$$

$$\hat{\sigma}^2_c = \frac{1}{n-1}\sum_{t=1}^{n}(c_t - \bar{c})^2 \quad \text{with} \quad \bar{c} = \frac{1}{n}\sum_{t=1}^{n} c_t,$$

$$\hat{\sigma}^2_{rs} = \frac{1}{n}\sum_{t=1}^{n}[u_t(u_t - c_t) + d_t(d_t - c_t)],$$

$$k = \frac{0.34}{1.34 + (n+1)/(n-1)}.$$

The estimate $\hat{\sigma}^2_{rs}$ was proposed by Rogers and Satchell (1991), and the quantity k is chosen to minimize the variance of the estimator of $\hat{\sigma}^2_{yz}$, which is a linear combination of three estimates.

The quantity $H_t - L_t$ is called the *range* of the price in the tth day. This estimator has led to the use of range-based volatility estimates; see, for instance, Alizadeh, Brandt, and Diebold (2002). In practice, stock prices are only observed at discrete time points. As such, the observed daily high is likely lower than H_t and the observed daily low is likely higher than L_t. Consequently, the observed daily price range tends to underestimate the actual range and, hence, may lead to underestimation of volatility. This bias in volatility estimation depends on the trading frequency and tick size of the stocks. For intensively traded stocks, the bias should be negligible. For other stocks, further study is needed to better understand the performance of range-based volatility estimation.

3.16 KURTOSIS OF GARCH MODELS

Uncertainty in volatility estimation is an important issue, but it is often overlooked. To assess the variability of an estimated volatility, one must consider the kurtosis of a volatility model. In this section, we derive the excess kurtosis of a GARCH(1,1) model. The same idea applies to other GARCH models, however. The model considered is

$$a_t = \sigma_t \epsilon_t, \qquad \sigma_t^2 = \alpha_0 + \alpha_1 a_{t-1}^2 + \beta_1 \sigma_{t-1}^2,$$

where $\alpha_0 > 0$, $\alpha_1 \geq 0$, $\beta_1 \geq 0$, $\alpha_1 + \beta_1 < 1$, and $\{\epsilon_t\}$ is an iid sequence satisfying

$$E(\epsilon_t) = 0, \qquad \mathrm{Var}(\epsilon_t) = 1, \qquad E(\epsilon_t^4) = K_\epsilon + 3,$$

where K_ϵ is the excess kurtosis of the innovation ϵ_t. Based on the assumption, we have the following:

- $\mathrm{Var}(a_t) = E(\sigma_t^2) = \alpha_0/[1 - (\alpha_1 + \beta_1)]$.
- $E(a_t^4) = (K_\epsilon + 3)E(\sigma_t^4)$ provided that $E(\sigma_t^4)$ exists.

Taking the square of the volatility model, we have

$$\sigma_t^4 = \alpha_0^2 + \alpha_1^2 a_{t-1}^4 + \beta_1^2 \sigma_{t-1}^4 + 2\alpha_0\alpha_1 a_{t-1}^2 + 2\alpha_0\beta_1\sigma_{t-1}^2 + 2\alpha_1\beta_1\sigma_{t-1}^2 a_{t-1}^2.$$

Taking expectation of the equation and using the two properties mentioned earlier, we obtain

$$E(\sigma_t^4) = \frac{\alpha_0^2(1 + \alpha_1 + \beta_1)}{[1 - (\alpha_1 + \beta_1)][1 - \alpha_1^2(K_\epsilon + 2) - (\alpha_1 + \beta_1)^2]},$$

provided that $1 > \alpha_1 + \beta_1 \geq 0$ and $1 - \alpha_1^2(K_\epsilon + 2) - (\alpha_1 + \beta_1)^2 > 0$. The excess kurtosis of a_t, if it exists, is then

$$\begin{aligned} K_a &= \frac{E(a_t^4)}{[E(a_t^2)]^2} - 3 \\ &= \frac{(K_\epsilon + 3)[1 - (\alpha_1 + \beta_1)^2]}{1 - 2\alpha_1^2 - (\alpha_1 + \beta_1)^2 - K_\epsilon \alpha_1^2} - 3. \end{aligned}$$

This excess kurtosis can be written in an informative expression. First, consider the case that ϵ_t is normally distributed. In this case, $K_\epsilon = 0$, and some algebra shows that

$$K_a^{(g)} = \frac{6\alpha_1^2}{1 - 2\alpha_1^2 - (\alpha_1 + \beta_1)^2},$$

where the superscript (g) is used to denote Gaussian distribution. This result has two important implications: (a) the kurtosis of a_t exists if $1 - 2\alpha_1^2 - (\alpha_1 + \beta_1)^2 > 0$, and (b) if $\alpha_1 = 0$, then $K_a^{(g)} = 0$, meaning that the corresponding GARCH(1,1) model does not have heavy tails.

Second, consider the case that ϵ_t is not Gaussian. Using the prior result, we have

$$\begin{aligned} K_a &= \frac{K_\epsilon - K_\epsilon(\alpha_1 + \beta_1) + 6\alpha_1^2 + 3K_\epsilon\alpha_1^2}{1 - 2\alpha_1^2 - (\alpha_1 + \beta_1)^2 - K_\epsilon\alpha_1^2} \\ &= \frac{K_\epsilon[1 - 2\alpha_1^2 - (\alpha_1 + \beta_1)^2] + 6\alpha_1^2 + 5K_\epsilon\alpha_1^2}{1 - 2\alpha_1^2 - (\alpha_1 + \beta_1)^2 - K_\epsilon\alpha_1^2} \\ &= \frac{K_\epsilon + K_a^{(g)} + \frac{5}{6}K_\epsilon K_a^{(g)}}{1 - \frac{1}{6}K_\epsilon K_a^{(g)}}. \end{aligned}$$

This result was obtained originally by George C. Tiao; see Bai, Russell, and Tiao (2003). It holds for all GARCH models provided that the kurtosis exists. For instance, if $\beta_1 = 0$, then the model reduces to an ARCH(1) model. In this case, it is easy to verify that $K_a^{(g)} = 6\alpha_1^2/(1 - 3\alpha_1^2)$ provided that $1 > 3\alpha_1^2$ and the excess kurtosis of a_t is

$$\begin{aligned} K_a &= \frac{(K_\epsilon + 3)(1 - \alpha_1^2)}{1 - (K_\epsilon + 3)\alpha_1^2} - 3 = \frac{K_\epsilon + 2K_\epsilon\alpha_1^2 + 6\alpha_1^2}{1 - 3\alpha_1^2 - K_\epsilon\alpha_1^2} \\ &= \frac{K_\epsilon(1 - 3\alpha_1^2) + 6\alpha_1^2 + 5K_\epsilon\alpha_1^2}{1 - 3\alpha_1^2 - K_\epsilon\alpha_1^2} \\ &= \frac{K_\epsilon + K_a^{(g)} + \frac{5}{6}K_\epsilon K_a^{(g)}}{1 - \frac{1}{6}K_\epsilon K_a^{(g)}}. \end{aligned}$$

The prior result shows that for a GARCH(1,1) model the coefficient α_1 plays a critical role in determining the tail behavior of a_t. If $\alpha_1 = 0$, then $K_a^{(g)} = 0$ and $K_a = K_\epsilon$. In this case, the tail behavior of a_t is similar to that of the standardized noise ϵ_t. Yet if $\alpha_1 > 0$, then $K_a^{(g)} > 0$ and the a_t process has heavy tails.

For a (standardized) Student-t distribution with v degrees of freedom, we have $E(\epsilon_t^4) = 6/(v - 4) + 3$ if $v > 4$. Therefore, the excess kurtosis of ϵ_t is $K_\epsilon = 6/(v - 4)$ for $v > 4$. This is part of the reason that we used t_5 in the chapter when the degrees of freedom of a t-distribution are prespecified. The excess kurtosis of a_t becomes $K_a = [6 + (v + 1)K_a^{(g)}]/[v - 4 - K_a^{(g)}]$ provided that $1 - 2\alpha_1^2(v - 1)/(v - 4) - (\alpha_1 + \beta_1)^2 > 0$.

APPENDIX: SOME RATS PROGRAMS FOR ESTIMATING VOLATILITY MODELS

The data file used in the illustration is `sp500.txt`, which contains the monthly excess returns of the S&P 500 index with 792 observations. Comments in a RATS program start with *.

A Gaussian GARCH(1,1) Model with a Constant Mean Equation

```
all 0  792:1
open data sp500.txt
data(org=obs) / rt
*** initialize the conditional variance function
set h = 0.0
*** specify the parameters of the model
nonlin mu a0 a1 b1
*** specify the mean equation
frml at = rt(t)-mu
*** specify the volatility equation
frml gvar = a0+a1*at(t-1)**2+b1*h(t-1)
*** specify the log likelihood function
frml garchln = -0.5*log(h(t)=gvar(t))-0.5*at(t)**2/h(t)
*** sample period used in estimation
smpl 2 792
*** initial estimates
compute a0 = 0.01, a1 = 0.1, b1 = 0.5, mu = 0.1
maximize(method=bhhh,recursive,iterations=150) garchln
set fv = gvar(t)
set resid = at(t)/sqrt(fv(t))
set residsq = resid(t)*resid(t)
*** Checking standardized residuals
cor(qstats,number=20,span=10) resid
*** Checking squared standardized residuals
cor(qstats,number=20,span=10) residsq
```

A GARCH(1,1) Model with Student-t Innovation

```
all 0  792:1
open data sp500.txt
data(org=obs) / rt
set h = 0.0
nonlin mu a0 a1 b1 v
frml at = rt(t)-mu
frml gvar = a0+a1*at(t-1)**2+b1*h(t-1)
frml tt = at(t)**2/(h(t)=gvar(t))
frml tln = %LNGAMMA((v+1)/2.)-%LNGAMMA(v/2.)-0.5*log(v-2.)
frml gln = tln-((v+1)/2.)*log(1.0+tt(t)/(v-2.0))-0.5*log(h(t))
smpl  2 792
```

```
compute a0 = 0.01, a1 = 0.1, b1 = 0.5, mu = 0.1, v = 10
maximize(method=bhhh,recursive,iterations=150) gln
set fv = gvar(t)
set resid = at(t)/sqrt(fv(t))
set residsq = resid(t)*resid(t)
cor(qstats,number=20,span=10) resid
cor(qstats,number=20,span=10) residsq
```

An AR(1)−EGARCH(1,1) Model for Monthly Log Returns of IBM Stock

```
all 0  864:1
open data m-ibm.txt
data(org=obs) / rt
set h = 0.0
nonlin c0 p1 th ga a0 a1
frml at = rt(t)-c0-p1*rt(t-1)
frml epsi = at(t)/(sqrt(exp(h(t))))
frml g = th*epsi(t)+ga*(abs(epsi(t))-sqrt(2./% PI))
frml gvar = a1*h(t-1)+(1-a1)*a0+g(t-1)
frml garchln = -0.5*(h(t)=gvar(t))-0.5*epsi(t)**2
smpl 3  864
compute c0 = 0.01, p1 = 0.01, th = 0.1, ga = 0.1
compute a0 = 0.01, a1 = 0.5
maximize(method=bhhh,recursive,iterations=150) garchln
set fv = gvar(t)
set resid = epsi(t)
set residsq = resid(t)*resid(t)
cor(qstats,number=20,span=10) resid
cor(qstats,number=20,span=10) residsq
```

EXERCISES

3.1. Derive multistep-ahead forecasts for a GARCH(1,2) model at the forecast origin h.

3.2. Derive multistep-ahead forecasts for a GARCH(2,1) model at the forecast origin h.

3.3. Suppose that $r_1, \ldots, r_n$ are observations of a return series that follows the AR(1)-GARCH(1,1) model

$$r_t = \mu + \phi_1 r_{t-1} + a_t, \qquad a_t = \sigma_t \epsilon_t, \qquad \sigma_t^2 = \alpha_0 + \alpha_1 a_{t-1}^2 + \beta_1 \sigma_{t-1}^2,$$

where ϵ_t is a standard Gaussian white noise series. Derive the conditional log-likelihood function of the data.

3.4. In the equation in Exercise 3.3, assume that ϵ_t follows a standardized Student-t distribution with v degrees of freedom. Derive the conditional log-likelihood function of the data.

3.5. Consider the monthly simple returns of Intel stock from January 1973 to December 2008 in `m-intc7308.txt`. Transform the returns into log returns. Build a GARCH model for the transformed series and compute 1-step- to 5-step-ahead volatility forecasts at the forecast origin December 2008.

3.6. The file `m-mrk4608.txt` contains monthly simple returns of Merck stock from June 1946 to December 2008. The file has two columns denoting date and simple return. Transform the simple returns to log returns.

(a) Is there any evidence of serial correlations in the log returns? Use autocorrelations and 5% significance level to answer the question. If yes, remove the serial correlations.

(b) Is there any evidence of ARCH effects in the log returns? Use the residual series if there are serial correlations in part (a). Use Ljung–Box statistics for the squared returns (or residuals) with 6 and 12 lags of autocorrelations and 5% significance level to answer the question.

(c) Identify an ARCH model for the data and fit the identified model. Write down the fitted model.

3.7. The file `m-3m4608.txt` contains two columns. They are date and the monthly simple return for 3M stock. Transform the returns to log returns.

(a) Is there any evidence of ARCH effects in the log returns? Use Ljung–Box statistics with 6 and 12 lags of autocorrelations and 5% significance level to answer the question.

(b) Use the PACF of the squared returns to identify an ARCH model. What is the fitted model?

(c) There are 755 data points. Refit the model using the first 750 observations and use the fitted model to predict the volatilities for t from 751 to 755 (the forecast origin is 750).

(d) Build an ARCH-M model for the log return series of 3M stock. Test the hypothesis that the risk premium is zero at the 5% significance level. Draw your conclusion.

(e) Build an EGARCH model for the log return series of 3M stock using the first 750 observations. Use the fitted model to compute 1-step- to 5-step-ahead volatility forecasts at the forecast origin $h = 750$.

3.8. The file `m-gmsp5008.txt` contains the dates and monthly simple returns of General Motors stock and the S&P 500 index from 1950 to 2008.

(a) Build a GARCH model with Gaussian innovations for the log returns of GM stock. Check the model and write down the fitted model.

(b) Build a GARCH-M model with Gaussian innovations for the log returns of GM stock. What is the fitted model?

(c) Build a GARCH model with Student-t distribution for the log returns of GM stock, including estimation of the degrees of freedom. Write down the fitted model. Let v be the degrees of freedom of the Student-t distribution. Test the hypothesis $H_0 : v = 6$ versus $H_a : v \neq 6$, using the 5% significance level.

(d) Build an EGARCH model for the log returns of GM stock. What is the fitted model?

(e) Obtain 1-step- to 6-step-ahead volatility forecasts for all the models obtained. Compare the forecasts.

3.9. Consider the monthly log returns of GM stock in `m-gmsp5008.txt`. Build an adequate TGARCH model for the series. Write down the fitted model and test for the significance of the leverage effect. Obtain 1-step- to 6-steps-ahead volatility forecasts.

3.10. Again, consider the returns in `m-gmsp5008.txt`.

(a) Build a Gaussian GARCH model for the monthly log returns of the S&P 500 index. Check the model carefully.

(b) Is there a summer effect on the volatility of the index return? Use the GARCH model built in part (a) to answer this question.

(c) Are lagged returns of GM stock useful in modeling the index volatility? Again, use the GARCH model of part (a) as a baseline model for comparison.

3.11. The file `d-gmsp9908.txt` contains the daily simple returns of GM stock and the S&P composite index from 1999 to 2008. It has three columns denoting date, GM return, and S&P return.

(a) Compute the daily log returns of GM stock. Is there any evidence of ARCH effects in the log returns? You may use 10 lags of the squared returns and 5% significance level to perform the test.

(b) Compute the PACF of the squared log returns (10 lags).

(c) Specify a GARCH model for the GM log return using a normal distribution for the innovations. Perform model checking and write down the fitted model.

(d) Find an adequate GARCH model for the series but using the generalized error distribution for the innovations. Write down the fitted model.

3.12. Consider the daily simple returns of the S&P composite index in the file `d-gmsp9908.txt`.

(a) Is there any ARCH effect in the simple return series? Use 10 lags of the squared returns and 5% significance level to perform the test.

(b) Build an adequate GARCH model for the simple return series.

(c) Compute 1-step- to 4-step-ahead forecasts of the simple return and its volatility based on the fitted model.

3.13. Again, consider the daily simple returns of GM stock in the file `d-gmsp9908.txt`.

(a) Find an adequate GARCH-M model for the series. Write down the fitted model.

(b) Find an adequate EGARCH model for the series. Is the “leverage” effect significant at the 5% level?

3.14. Revisit the file `d-gmsp9908.txt`. However, we shall investigate the value of using market volatility in modeling volatility of individual stocks. Convert the two simple return series into percentage log return series.

(a) Build an AR(5)–GARCH(1,1) model with generalized error distribution for the log S&P returns. The AR(5) contains only lags 3 and 5. Denote the fitted volatility series by `spvol`.

(b) Estimate a GARCH(1,1) model with `spvol` as an exogenous variable to the log GM return series. Check the adequacy of the model, and write down the fitted model. In S-Plus, the command is

```
fit = garch(gm ~1, ~garch(1,1)+spvol, cond.dist='ged')
```

(c) Discuss the implication of the fitted model.

3.15. Again, consider the percentage daily log returns of GM stock and the S&P 500 index from 1999 to 2008 as before, but we shall investigate whether the volatility of GM stock has any contribution in modeling the S&P index volatility. Follow the steps below to perform the analysis.

(a) Fit a GARCH(1,1) model with generalized error distribution to the percentage log returns of GM stock. Denote the fitted volatility by `gmvol`. Build an adequate GARCH model plus `gmvol` as the exogenous variable for the log S&P return series. Write down the fitted model.

(b) Is the volatility of GM stock returns helpful in modeling the volatility of the S&P index returns? Why?

REFERENCES

Alizadeh, S., Brandt, M., and Diebold, F. X. (2002). Range-based estimation of stochastic volatility models. *Journal of Finance* **57**: 1047–1092.

Andersen, T. G. and Bollerslev, T. (1998). Answering the skeptics: Yes, standard volatility models do provide accurate forecasts. *International Economic Review* **39**: 885–905.

Andersen, T. G., Bollerslev, T., Diebold, F. X., and Labys, P. (2001a). The distribution of realized exchange rate volatility. *Journal of the American Statistical Association* **96**: 42–55.

Andersen, T. G., Bollerslev, T., Diebold, F. X., and Labys, P. (2001b). The distribution of realized stock return volatility. *Journal of Financial Economics* **61**: 43–76.

Bai, X., Russell, J. R., and Tiao, G. C. (2003). Kurtosis of GARCH and stochastic volatility models with non-normal innovations. *Journal of Econometrics* **114**: 349–360.

Bai, X., Russell, J. R., and Tiao, G. C. (2004). Effects of non-normality and dependence on the precision of variance estimates using high-frequency financial data. Revised working paper, Graduate School of Business, University of Chicago.

Barndorff-Nielsen, O. E. and Shephard, N. (2004). Power and bi-power variations with stochastic volatility and jumps (with discussion). *Journal of Financial Econometrics* **2**: 1–48.

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

Bollerslev, T. and Jubinski, D. (1999). Equality trading volume and volatility: Latent information arrivals and common long-run dependencies. *Journal of Business & Economic Statistics* **17**: 9–21.

Bollerslev, T., Chou, R. Y., and Kroner, K. F. (1992). ARCH modeling in finance. *Journal of Econometrics* **52**: 5–59.

Bollerslev, T., Engle, R. F., and Nelson, D. B. (1994). ARCH model. In R. F. Engle and D. C. McFadden (eds.). *Handbook of Econometrics IV*, pp. 2959–3038. Elsevier Science, Amsterdam.

Breidt, F. J., Crato, N., and de Lima, P. (1998). On the detection and estimation of long memory in stochastic volatility. *Journal of Econometrics* **83**: 325–348.

Cao, C. and Tsay, R. S. (1992). Nonlinear time series analysis of stock volatilities. *Journal of Applied Econometrics* **7**: s165–s185.

Ding, Z., Granger, C. W. J., and Engle, R. F. (1993). A long memory property of stock returns and a new model. *Journal of Empirical Finance* **1**: 83–106.

Engle, R. F. (1982). Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflations. *Econometrica* **50**: 987–1007.

Fernández, C. and Steel, M. F. J. (1998). On Bayesian modelling of fat tails and skewness. *Journal of the American Statistical Association* **93**: 359–371.

French, K. R., Schwert, G. W., and Stambaugh, R. F. (1987). Expected stock returns and volatility. *Journal of Financial Economics* **19**: 3–29.

Garman, M. B. and Klass, M. J. (1980). On the estimation of security price volatilities from historical data. *Journal of Business* **53**: 67–78.

Glosten, L. R., Jagannathan, R., and Runkle, D. E. (1993). On the relation between the expected value and the volatility of nominal excess return on stocks. *Journal of Finance* **48**: 1779–1801.

Harvey, A. C., Ruiz, E., and Shephard, N. (1994). Multivariate stochastic variance models. *Review of Economic Studies* **61**: 247–264.

Jacquier, E., Polson, N. G., and Rossi, P. (1994). Bayesian analysis of stochastic volatility models (with discussion). *Journal of Business & Economic Statistics* **12**: 371–417.

Lambert, P. and Laurent, S. (2001). Modelling financial time series using GARCH-type models and a skewed Student density. Working paper, Université de Liège.

McLeod, A. I. and Li, W. K. (1983). Diagnostic checking ARMA time series models using squared-residual autocorrelations. *Journal of Time Series Analysis* **4**: 269–273.

Melino, A. and Turnbull, S. M. (1990). Pricing foreign currency options with stochastic volatility. *Journal of Econometrics* **45**: 239–265.

Nelson, D. B. (1990). Stationarity and persistence in the GARCH(1,1) model. *Econometric Theory* **6**: 318–334.

Nelson, D. B. (1991). Conditional heteroskedasticity in asset returns: A new approach. *Econometrica* **59**: 347–370.

Nicholls, D. F. and Quinn, B. G. (1982). *Random Coefficient Autoregressive Models: An Introduction*, Lecture Notes in Statistics, 11. Springer, New York.

Parkinson, M. (1980). The extreme value method for estimating the variance of the rate of return. *Journal of Business* **53**: 61–65.

Ray, B. K. and Tsay, R. S. (2000). Long-range dependence in daily stock volatilities. *Journal of Business & Economic Statistics* **18**: 254–262.

Rogers, L. C. G. and Satchell, S. E. (1991). Estimating variance from high, low and closing prices. *Annals of Applied Probability* **1**: 504–512.

Taylor, S. J. (1994). Modeling stochastic volatility: A review and comparative study. *Mathematical Finance* **4**: 183–204.

Tong, H. (1978). On a threshold model. In C. H. Chen (Ed.). *Pattern Recognition and Signal Processing*. Sijhoff & Noordhoff, Amsterdam.

Tong, H. (1990). *Non-Linear Time Series: A Dynamical System Approach*. Oxford University Press, Oxford, UK.

Tsay, R. S. (1987). Conditional heteroscedastic time series models. *Journal of the American Statistical Association*, **82**: 590–604.

Yang, D. and Zhang, Q. (2000). Drift-independent volatility estimation based on high, low, open, and close prices. *Journal of Business* **73**: 477–491.

Zakoian, J. M. (1994). Threshold heteroscedastic models. *Journal of Economic Dynamics and Control* **18**: 931–955.