Conditional Heteroscedastic Models

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

Volatility is an important factor in options trading. Here volatility means the conditional variance 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, and we denote it by \( \ell \). If the holder can exercise her right any time on or before the expiration date, then the option is called an American call option. The well-known Black-Scholes option pricing formula states that the price of a European call option is

\[
    c_t = P_t \Phi(x) - K r^{-\ell} \Phi(x - \sigma_t \sqrt{\ell}),
\]

(36)

where \( x = \frac{\ln(P_t K r^{-\ell})}{\sigma_t \sqrt{\ell}} + \frac{1}{2}\sigma_t \sqrt{\ell} \) and \( P_t \) is the current price of the underlying stock, \( r \) is the risk-free interest rate, \( \sigma_t \) is the 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 $x$. The formula has several nice interpretations, but it suffices to say here that the conditional variance of the log return of the underlying stock plays an important role. This volatility evolves over time.

Volatility is also important in risk management. As discussed in financial engineering, volatility modeling provides a simple approach to calculating value at risk of a financial position. Finally, modeling the volatility of a time series can improve the efficiency in parameter estimation and the accuracy in interval forecast.

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 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), 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 later.

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 returns because there is only one observation in a trading day. If intraday data of the stock, such as 5-minute returns, are available, then one can estimate the daily volatility. The accuracy of such an estimate deserves a careful study, however. Furthermore, stock volatility consists of intraday volatility and variation between trading days. 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 (36) to deduce the conditional standard deviation $\sigma_t$. The resulting value of $\sigma_t^2$ is called the implied volatility of the un-
derlying stock. However, this implied volatility is derived under the log normal assumption for the return series. It might be very 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.

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. 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 $t$. 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 dependent. For illustration, Figure 3.1 shows the ACF and PACF of some functions of the monthly log stock returns of Intel Corporation from January 1973 to December 1997. The upper left panel shows the sample ACF of the return, which suggests no significant serial correlations except for a minor one at lag 7. The upper right panel shows the sample ACF of the absolute log returns (i.e., $|r_t|$), whereas the lower left panel shows the sample ACF of the squared returns $r_t^2$. These two plots clearly suggest that the monthly returns are not independent. Combining the three plots, it seems that the returns are indeed serially uncorrelated, but dependent. Volatility models attempt to capture such dependence in the return series.
Figure 3.1. Sample ACF and PACF of various functions of monthly log stock returns of Intel Corporation from January 1973 to December 1997: (a) ACF of the log returns, (b) ACF of the squared returns (lower left), (c) ACF of the absolute returns (upper right), and (d) PACF of the squared returns.

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

\[
\begin{align*}
\mu_t &= E(r_t | F_{t-1}), \\
\sigma_t^2 &= \text{Var}(r_t | F_{t-1}) = E[(r_t - \mu_t)^2 | F_{t-1}],
\end{align*}
\]

(37)

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 3.1, serial dependence of a stock return series $r_t$ is weak if it exists at all. Therefore, the equation for $\mu_t$ in (37) should be simple, and we assume that $r_t$ follows a simple time series model such as a stationary ARMA($p$, $q$) model. In other words, we entertain the model

$$r_t = \mu_t + a_t, \quad \mu_t = \phi_0 + \sum_{i=1}^{p} \phi_i r_{t-i} - \sum_{i=1}^{q} \theta_i a_{t-i},$$

(38)

for $r_t$, where $p$ and $q$ are non-negative integers.

Model (38) illustrates a possible financial application of the linear time series models 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. One may include some explanatory variables to the conditional mean equation and use a linear regression model with time series errors to capture the behavior of $\mu_t$. For example, a dummy variable can be used for the Mondays to study the effect of weekend on daily stock returns.

Combining (37) and (38), we have

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

(39)
The conditional heteroscedastic models of this chapter are concerned with the evolution of $\sigma_t^2$. The manner under which $\sigma_t^2$ 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 $\sigma_t^2$, whereas those in the second category use a stochastic equation to describe $\sigma_t^2$. The GARCH model belongs to the first category, and the stochastic volatility model is in the second category.

For simplicity in introducing volatility models, we assume that the model for the conditional mean is given. However, we estimate the conditional mean and variance equations jointly in empirical studies. Throughout the course, $a_t$ is referred to as the shock or mean-corrected return of an asset return at time $t$ and $\sigma_t$ is the positive square-root of $\sigma_t^2$. The model for $\mu_t$ in (38) is referred to as the mean equation for $r_t$ and the model for $\sigma_t^2$ is the volatility equation for $r_t$. Therefore, modeling conditional heteroscedasticity amounts to augmenting a dynamic equation to a time series model to govern the time evolution of the conditional variance of the shock.
3.3 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 mean corrected asset return $a_t$ 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

$$ a_t = \sigma_t \epsilon_t, \quad \sigma_t^2 = \alpha_0 + \alpha_1 a_{t-1}^2 + \cdots + \alpha_m a_{t-m}^2, \quad (40) $$

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 $\alpha_i$ must satisfy some regularity conditions to ensure that the unconditional variance of $a_t$ is finite. In practice, $\epsilon_t$ is often assumed to follow the standard normal or a standardized Student-t 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 $\sigma_t^2$ for the mean-corrected return $a_t$. 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 we use the word *tend* because a large variance does not necessarily produce a large variate. 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.

Figure 3.2. (a) Time plot of 10-minute returns of the exchange rate between Deutsche Mark and Dollar, and (b) the squared returns.

The ARCH effect also occurs in other financial time series. Figure 3.2 shows the time plots of (a) the percentage changes in DeutscheMark/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 changes occurred occasionally, but there exist certain stable periods. Figure 3.3(a) shows the sample ACF of
the percentage change series. Clearly, the series has no serial correlation. Figure 3.3(b) shows the sample PACF of the squared series of percentage changes. It is seen that there are some big spikes in the PACF. Such spikes suggest that the percentage changes are not independent and have some ARCH effects.

![Sample ACF of return series](image1)

![Sample PACF of squared returns](image2)

Figure 3.3. (a) Sample autocorrelation function of the return series of Mark/Dollar exchange rate, and (b) sample partial autocorrelation function of the squared returns.

**Remark:** Some authors use $h_t$ to denote the conditional variance in (40). In this case, the shock becomes $a_t = \sqrt{h_t}\epsilon_t$. 

94
3.3.1 Properties of ARCH Models

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

\[ a_t = \sigma_t \epsilon_t, \quad \sigma_t^2 = \alpha_0 + \alpha_1 a_{t-1}^2, \]

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

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

Because \( a_t \) is a stationary process with \( E(a_t) = 0 \), \( \text{Var}(a_t) = \text{Var}(a_{t-1}) = E(a_{t-1}^2) \). Therefore, we have \( \text{Var}(a_t) = \alpha_0 + \alpha_1 \text{Var}(a_t) \) and \( \text{Var}(a_t) = \frac{\alpha_0}{1-\alpha_1} \). Because the variance of \( a_t \) must be positive, we need \( 0 \leq \alpha_1 < 1 \). Third, in some applications, we need higher order moments of \( a_t \) to exist and, hence, \( \alpha_1 \) 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 $\epsilon_t$ in (40), 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,

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

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

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

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 $\alpha_1$ must also satisfy the condition $1 - 3\alpha_1^2 > 0$; that is, $0 \leq \alpha_1 < \frac{1}{3}$; and (b) the unconditional kurtosis of $a_t$ is

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

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 $\alpha_i \geq 0$ in (40) can be relaxed. It is a condition to ensure that the conditional variance $\sigma_t^2$ 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

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

(41)

where $A_{m,t-1} = (a_{t-1}, \ldots, a_{t-m})'$ and $\Omega$ is a $m \times m$ non-negative definite matrix. The ARCH($m$) model in (40) requires $\Omega$ to be diagonal. Thus, Engle’s model uses a parsimonious approach to approximate a quadratic function. A simple way to achieve (41) is to employ a random-coefficient model for $a_t$; see the CHARMA and RCA models discussed later.

3.3.2 Weaknesses of ARCH Models

The advantages of ARCH models include properties discussed in the previous subsection. 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 price of a financial asset responds differently to positive and negative shocks.

2. The ARCH model is rather restrictive. For instance, \( \alpha_1^2 \) of an ARCH(1) model must be in the interval \([0, \frac{1}{3}]\) if the series is to have a finite fourth moment. The constraint becomes complicated for higher order ARCH models.

3. The ARCH model does not provide any new insight for understanding the source of variations of a financial time series. They only provide 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.

3.3.3 Building an ARCH Model

A simple way to build an ARCH model consists of three steps: (1) build an econometric model (e.g., an ARMA model) for the return series to remove any linear dependence in the data, and
use the residual series of the model to test for ARCH effects; (2) specify the ARCH order and perform estimation; and (3) check the fitted ARCH model carefully and refine it if necessary. More details are given later.

Modeling the Mean Effect and Testing

An ARMA model is built for the observed time series to remove any serial correlations in the data. For most asset return series, this step 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. The squared series $a_t^2$ is used to check for conditional heteroscedasticity, where $a_t = r_t - \mu_t$ is the residual of the ARMA model. Two tests are available here. The first test is to check the usual Ljung-Box statistics of $a_t^2$. 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 $\alpha_i = 0 (i = 1, \cdots, m)$ in the linear regression

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

where $e_t$ denotes the error term, $m$ is a prespecified positive integer, and $T$ is the sample size. Let $SSR_0 = \sum_{t=m+1}^{T} (a_t^2 - \bar{\omega})^2$, where $\bar{\omega}$ is the sample mean of $a_t^2$, and $SSR_1 = \sum_{t=m+1}^{T} \hat{e}_t^2$,
where $\hat{e}_t$ is the least squares residual of the prior linear regression. Then we have

$$F = \frac{(SSR_0 - SSR_1)/m}{SSR_1/(T - 2m - 1)},$$

which is asymptotically distributed as a chi-squared distribution with $m$ degrees of freedom under the null hypothesis.

**Order Determination**

If the test statistic $F$ is significant, then conditional heteroscedasticity of $a_t$ is detected, and we 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 (40), 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 $\sigma_t^2$. Therefore, we expect that $a_t^2$ is linearly related to $a_{t-1}^2, \cdots, a_{t-m}^2$ 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 $\sigma_t^2$, 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**

Two likelihood functions are commonly used in ARCH estimation. Under the normality assumption, the likelihood function of an ARCH\((m)\) model is

\[
f(a_1, \cdots, a_T|\alpha) = f(a_T|F_{T-1}) f(a_{T-1}|F_{T-2}) \cdots \\
\times f(a_{m+1}|F_m) f(a_1, \cdots, a_m|\alpha) \\
= \prod_{t=m+1}^{T} \frac{1}{\sqrt{2\pi\sigma_t^2}} \exp \left( -\frac{a_t^2}{2\sigma_t^2} \right) f(a_1, \cdots, a_m|\alpha),
\]

where \(\alpha = (\alpha_0, \alpha_1, \cdots, \alpha_m)'\) and \(f(a_1, \cdots, a_m|\alpha)\) is the joint probability density function of \(a_1, \cdots, a_m\). Since the exact form of \(f(a_1, \cdots, a_m|\alpha)\) 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.

\[
f(a_{m+1}, \cdots, a_T|\alpha, a_1, \cdots, a_m) = \prod_{t=m+1}^{T} \frac{1}{\sqrt{2\pi\sigma_t^2}} \exp \left( -\frac{a_t^2}{2\sigma_t^2} \right),
\]
where $\sigma_t^2$ can be evaluated recursively. We refer to estimates obtained by maximizing the prior likelihood function as the conditional maximum likelihood estimates (MLE) under normality.

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

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

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

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

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 $\epsilon_t$ 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 $\text{Var}(x_v) = \frac{v}{v-2}$ for $v > 2$, and we use $\epsilon_t = \frac{x_v}{\sqrt{\frac{v}{v-2}}}$. The probability density function of $\epsilon_t$ is

$$f(\epsilon_t|v) = \frac{\Gamma\left(\frac{v+1}{2}\right)}{\Gamma\left(\frac{v}{2}\right)\sqrt{(v-2)\pi}} \left(1 + \frac{\epsilon_t^2}{v-2}\right)^{-\frac{v+1}{2}}, \quad v > 2, \quad (42)$$
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

\[
f(a_{m+1}, \ldots, a_T|\alpha, A_m)\]

\[
= \prod_{t=m+1}^T \frac{\Gamma(v+1)}{\Gamma(v/2) \sqrt{(v-2)\pi} \sigma_t} \frac{1}{(v-2)\sigma^2} \left[ 1 + \frac{a_t^2}{(v-2)\sigma^2} \right]^{-\frac{v+1}{2}},
\]

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 MLE 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 3 and 6 is often used if it is prespecifed.

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

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

(43)

If one wishes to estimate \( v \) jointly with other parameters, then the log likelihood function involving degrees of freedom

\[
\ell(a_{m+1}, \ldots, a_T|\alpha, v, A_m)
\]

\[
= (T - m) \left[ \ln(\Gamma(v+1/2)) - \ln(\Gamma(v/2)) - \frac{1}{2} \ln((v-2)\pi) \right]
\]

\[
+ \ell(a_{m+1}, \ldots, a_T|\alpha, A_m),
\]

where the second term is given in (43).
Model Checking

For an ARCH model, the standardized shocks

\[ \tilde{a}_t = \frac{a_t}{\sigma_t} \]

are iid random variates following either a standard normal or standardized Student-\(t\) distribution. 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.

Forecasting

Forecasts of the ARCH model in (40) 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 \(\sigma_{h+1}^2\) 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 $\ell$-step ahead forecast for $\sigma^2_{h+\ell}$ is

$$\sigma^2_{h}(\ell) = \alpha_0 + \sum_{i=1}^{m} \alpha_i \sigma^2_{h}(\ell - i),$$

where $\sigma^2_{h}(\ell - i) = a^2_{h+\ell-i}$ if $\ell - i \leq 0$.

### 3.3.4 Examples

In this subsection, 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 stock returns of Intel Corporation. The sample ACF and PACF of the squared returns in Figure 3.1 clearly show the existence of conditional heteroscedasticity. Thus, it is unnecessary to perform any statistical tests to confirm the need of ARCH modeling, and we proceed to identify the order of an ARCH model. The sample PACF in the lower right panel of Figure 3.1 indicates that an ARCH(3) model might be appropriate. Consequently, we specify the model

$$
\begin{align*}
    r_t &= \mu + a_t, \\
    a_t &= \sigma_t \epsilon_t, \\
    \sigma^2_t &= \alpha_0 + \alpha_1 a^2_{t-1} + \alpha_2 a^2_{t-2} + \alpha_3 a^2_{t-3}
\end{align*}
$$

for the monthly log returns of Intel stock. Assuming that $\epsilon_t$ are
iid standard normal, we obtain the fitted model
\[
\begin{align*}
    r_t &= 0.0196 + a_t, \\
    \sigma_t^2 &= 0.0090 + 0.2973a_{t-1}^2 + 0.0090a_{t-2}^2 + 0.0626a_{t-3}^2,
\end{align*}
\]
where the standard errors of the parameters are 0.0062, 0.0013, 0.0887, 0.0645, and 0.0777, respectively. While the estimates meet the general requirement of an ARCH(3) model, the estimates of \( \alpha_2 \) and \( \alpha_3 \) appear to be statistically nonsignificant at the 5\% level. Therefore, the model can be simplified.

Dropping the two nonsignificant parameters, we obtain the model
\[
\begin{align*}
    r_t &= 0.0213 + a_t, \\
    \sigma_t^2 &= 0.00998 + 0.4437a_{t-1}^2,
\end{align*}
\]
where the standard errors of the parameters are 0.0062, 0.00124, and 0.0938, respectively. All the estimates are highly significant. Figure 3.4 shows the standardized shocks and the sample ACF of some functions of the standardized shocks. The Ljung-Box statistics of the standardized shocks \( \{\tilde{a}_t\} \) give \( Q(10) = 12.53 \) with \( p \) value 0.25 and those of \( \{\tilde{a}_t^2\} \) give \( Q(10) = 17.23 \) with \( p \) value 0.07. Consequently, the ARCH(1) model in (44) is adequate for the data at the 5\% significance level.
Figure 3.4. Model checking statistics of the Gaussian ARCH(1) model in (44) for the monthly log stock returns of Intel from January 1973 to December 1997: parts (a), (b), and (c) show the sample ACF of the standardized shocks, their squared series, and absolute series, respectively, and (d) is the time plot of standardized shocks.

The ARCH(1) model in (44) has some interesting properties. First, the expected monthly log return for Intel stock is about 2.1%, which is remarkable. Second, $\hat{\alpha}_1^2 = 0.444^2 < \frac{1}{3}$ so that the unconditional fourth moment of the monthly log return of Intel stock exists. Third, the unconditional variance of $r_t$ is
\[
\frac{0.00998}{1-0.4437} = 0.0179. \text{ Finally, the ARCH(1) model can be used to predict the monthly volatility of Intel stock returns.}
\]

**t Innovation**

For comparison, we also fit an ARCH(1) model to the series, assuming that \( \epsilon_t \) follows a standardized Student-\( t \) distribution with 5 degrees of freedom. The resulting model is

\[
\begin{aligned}
& r_t = 0.0222 + a_t, \\
& \sigma_t^2 = 0.0121 + 0.3029a_{t-1}^2,
\end{aligned}
\]  

(45)

where the standard errors of the parameters are 0.0019, 0.1443, and 0.0061, respectively. All the estimates are significant at the 5\% level, but the \( t \) ratio of \( \hat{\alpha}_1 \) is only 2.10. The unconditional variance of \( a_t \) is \( \frac{0.0121}{1-0.3029} = 0.0174 \), which is close to that obtained under normality. The Ljung-Box statistics of the standardized shocks give \( Q(10) = 13.66 \) with \( p \)-value 0.19, confirming that the mean equation is adequate. However, the Ljung-Box statistics for the squared standardized shocks show \( Q(10) = 23.83 \) with \( p \) value 0.008. The volatility equation is inadequate at the 5\% level. We refine the model by considering an ARCH(2) model and obtain

\[
\begin{aligned}
& r_t = 0.0225 + a_t, \\
& \sigma_t^2 = 0.0113 + 0.226a_{t-1}^2 + 0.108a_{t-2}^2,
\end{aligned}
\]  

(46)
where the standard errors of the parameters are 0.006, 0.002, 0.135, and 0.094, respectively. The coefficient of $a_{t-1}^2$ is marginally significant at the 10% level, but that of $a_{t-2}^2$ is only slightly greater than its standard error. The Ljung-Box statistics for the squared standardized shocks give $Q(10) = 8.82$ with $p$ value 0.55. Consequently, the fitted ARCH(2) model appears to be adequate.

Comparing models (44), (45), and (46), we see that (a) using a heavy-tailed distribution for $\epsilon_t$ reduces the ARCH effect, and (b) the difference among the three models is small for this particular instance. Finally, a more appropriate conditional heteroscedastic model for this data set is a GARCH(1, 1) model, which is discussed in the next section.

**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 3.2(a). As shown in Figure 3.3(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 subsection, we specify an ARCH(3) model for the series. Using the conditional Gaussian
likelihood function, we obtain the fitted model
\[ \sigma_t^2 = 0.22 \times 10^{-6} + 0.328a_{t-1}^2 + 0.073a_{t-2}^2 + 0.103a_{t-3}^2, \]
where all the estimates are statistically significant at the 5% significant level, and the standard errors of the parameters are 0.46 \times 10^{-8}, 0.0162, 0.0160, and 0.0147, respectively. Model checking, using the standardized shock \( \tilde{a}_t \), indicates that the model is adequate.

**Remark:** The estimation of conditional heteroscedastic models of this chapter is carried out by the Regression Analysis of Time Series (RATS) package. There are other softwares available, including Eviews, Scientific Computing Associates (SCA), and S-Plus.

### 3.4 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. 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 \( r_t \), we assume that the mean equation of the process can be adequately
described by an ARMA model. Let $a_t = r_t - \mu_t$ be the mean-corrected log return. Then $a_t$ follows a GARCH($m$, $s$) model if

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

where again $\{\epsilon_t\}$ is a sequence of iid random variables with mean 0 and variance 1, $\alpha_0 > 0$, $\alpha_i \geq 0$, $\beta_j \geq 0$, 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 $\sigma_t^2$ evolves over time. As before, $\epsilon_t$ is often assumed to be a standard normal or standardized Student-$t$ distribution. Equation (47) reduces to a pure ARCH($m$) model if $s = 0$.

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 $\sigma_{t-i}^2 = a_{t-i}^2 - \eta_{t-i}(i = 0, \cdots, s)$ into (47), we can rewrite the GARCH model as

$$a_i^2 = \alpha_0 + \sum_{i=1}^{\max(m,s)} (\alpha_i + \beta_i) a_{t-i}^2 + \eta_t - \sum_{j=1}^{s} \beta_j \eta_{t-j}. \quad (48)$$

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 (48) 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 $a^2_i$. Using the unconditional mean of an ARMA model, we have

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

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

$$\begin{cases} 
\sigma_t^2 = \alpha_0 + \alpha_1 a^2_{t-1} + \beta_1 \sigma^2_{t-1}, \\
0 \leq \alpha_1, \beta_1 \leq 1, \\
\alpha_1 + \beta_1 < 1.
\end{cases} \tag{49}$$

First, a large $a^2_{t-1}$ or $\sigma^2_{t-1}$ gives rise to a large $\sigma^2_t$. This means that a large $a^2_{t-1}$ tends to be followed by another large $a^2_t$, 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

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

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 (49) 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 $\sigma_h^2$ are known at the time $h$. 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 (49) 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

$$\sigma_h^2(\ell) = \alpha_0 + (\alpha_1 + \beta_1) \sigma_h^2(\ell - 1), \quad \ell > 1. \quad (50)$$

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 (50), we obtain that the $\ell$-step ahead forecast can be written as

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

Therefore,

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

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 $\text{Var}(a_t)$ exists.

### 3.4.1 An Illustrative Example

The modeling procedure of ARCH models can also be used to build a GARCH model. However, specifying the order of an 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 $\sigma_1^2$ is treated as fixed, then $\sigma_t^2$ 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 $\sigma^2_1$. The fitted model can be checked by using the standardized residual. $\tilde{a}_t = \frac{a_t}{\sigma_t}$ and its squared process.

Figure 3.5. Time series plot of the monthly excess returns of S&P 500 index.
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 3.5. Denote the excess return series by \( r_t \). Figure 3.6 shows the sample ACF 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
\[
\begin{align*}
  r_t &= 0.0062 + a_t + 0.0944a_{t-1} - 0.1407a_{t-3}, \\
  \hat{\sigma}_a &= 0.0576
\end{align*}
\]
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
\[
\begin{align*}
  r_t &= 0.088r_{t-1} - 0.023r_{t-2} - 0.123r_{t-3} + 0.0066 + a_t, \\
  \hat{\sigma}_a^2 &= 0.00333.
\end{align*}
\]
(51)
For the GARCH effects, we use the GARCH(1, 1) model
\[
a_t = \sigma_t \epsilon_t, \quad \sigma_t^2 = \alpha_0 + \beta_1 \sigma_{t-1}^2 + \alpha_1 a_{t-1}^2.
\]
A joint estimation of the AR(3)-GARCH(1, 1) model gives
\[
\begin{align*}
  r_t &= 0.021r_{t-1} - 0.034r_{t-2} - 0.013r_{t-3} + 0.0085 + a_t, \\
  \sigma_t^2 &= 0.000099 + 0.08476\sigma_{t-1}^2 + 0.1219a_{t-1}^2.
\end{align*}
\]
From the volatility equation, the implied unconditional variance of \( a_t \) is
\[
\frac{0.000099}{1 - 0.8476 - 0.1219} = 0.00325,
\]
which is very close to that of (51). However, \( t \) ratios of the parameters in the mean equation suggest that all AR coefficients
are insignificant at the 5% level. Therefore, we refine the model by dropping all AR coefficients. The refined model is

\[
\begin{align*}
    r_t &= 0.0065 + a_t, \\
    \sigma_t^2 &= 0.00014 + 0.8220\sigma_{t-1}^2 + 0.1352a_{t-1}^2.
\end{align*}
\]

The standard error of the parameter in the mean equation is 0.0015, whereas those of the parameters in the volatility equation are 0.00002, 0.0208, and 0.0166, respectively. The unconditional variance of \( a_t \) is \( \frac{0.0001}{1-0.822-0.1352} = 0.00324 \). This is a simple stationary GARCH(1, 1) model. Figure 3.7 shows the estimated volatility process and the standardized shocks \( \tilde{a}_t = \frac{a_t}{\sigma_t} \) for the GARCH(1, 1) model in (52). The \( \tilde{a}_t \) series looks like a white noise process. Figure 3.8 provides the sample ACF of the standardized shocks \( \tilde{a}_t \) and the squared process \( \tilde{a}_t^2 \). These ACFs fail to suggest any significant serial correlations in the two processes. More specifically, we have \( Q(10) = 10.32(0.41) \) and \( Q(20) = 22.66(0.31) \) for \( \tilde{a}_t \), and \( Q(10) = 8.83(0.55) \) and \( Q(20) = 15.82(0.73) \) 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. Note that the fitted model shows \( \hat{\alpha}_1 + \hat{\beta}_1 = 0.9572 \), which is close to 1. This phenomenon is commonly observed in practice and it leads to imposing the constraint \( \hat{\alpha}_1 + \hat{\beta}_1 = 1 \) in a GARCH(1, 1) model, resulting in an integrated GARCH (or IGARCH) model; see Section 3.5.
Figure 3.7. (a) Time series plot of estimated volatility for the monthly excess returns of S&P 500 index, and (b) the standardized shocks of the monthly excess returns of S&P 500 index. Both plots are based on the GARCH(1, 1) model in (52).
Figure 3.8. Model checking of the GARCH(1, 1) model in (52) for monthly excess returns of S&P 500 index: (a) Sample ACF of standardized shocks, and (b) sample ACF of the squared standardized shocks.

Finally, to forecast the volatility of monthly excess returns of S&P500 index, we can use the volatility equation in (52). For instance, at the forecast origin \( h \), we have \( \sigma_{h+1}^2 = 0.00014 + 0.822\sigma_h^2 + 0.1352a_h^2 \). The 1-step ahead forecast is then
\[
\sigma_h^2(1) = 0.00014 + 0.822\sigma_h^2 + 0.1352a_h^2,
\]
where \( a_h \) is the residual of the mean equation at time \( h \) and \( \sigma_h^2 \) is obtained from the volatility equation. The starting value
$\sigma_0^2$ is fixed at either zero or the unconditional variance of $a_t$. For multistep ahead forecasts, we use the recursive formula in (50). Table 3.1 shows some mean and volatility forecasts for the monthly excess return of S&P500 index with forecast origin $h = 792$ based on the GARCH(1, 1) model in (52).

<table>
<thead>
<tr>
<th>Horizon</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>$\infty$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Return</td>
<td>0.0065</td>
<td>0.0065</td>
<td>0.0065</td>
<td>0.0065</td>
<td>0.0065</td>
<td>0.0065</td>
</tr>
<tr>
<td>Volatility</td>
<td>0.00311</td>
<td>0.00312</td>
<td>0.00312</td>
<td>0.00313</td>
<td>0.00314</td>
<td>0.00324</td>
</tr>
</tbody>
</table>


t Innovation

Assuming that $\epsilon_t$ follows a standardized Student-$t$ distribution with 5 degrees of freedom, we re-estimate the GARCH(1, 1) model and obtain

$$
\begin{align*}
    r_t &= 0.0085 + a_t, \\
    \sigma_t^2 &= 0.00018 + 0.1272 a_{t-1}^2 + 0.8217 \sigma_{t-1}^2,
\end{align*}
$$

(53)

where the standard errors of the parameters are 0.0014, $0.55 \times 10^{-4}$, 0.0349, and 0.0382, respectively. This model is essentially
an IGARCH(1, 1) model as $\hat{\alpha}_1 + \hat{\beta}_1 \approx 0.96$, which is close to 1. The Ljung-Box statistics of the standardized residuals give $Q(10) = 10.45$ with $p$ value 0.40 and those of the $\{\tilde{a}_t^2\}$ series give $Q(10) = 9.33$ with $p$ value 0.50. Thus, the fitted GARCH(1, 1) model with Student-t distribution is adequate.

**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

$$\begin{align*}
 r_t &= 0.0083 + a_t, \\
 \sigma_t^2 &= 0.00017 + 0.1227a_{t-1}^2 + 0.8193\sigma_{t-1}^2, \\
\end{align*}$$

where the estimated degrees of freedom is 6.51. Standard errors of the estimates in (54) are close to those in (53). The standard error of the estimated degrees of freedom is 1.49. Consequently, we cannot reject the hypothesis of using a standardized Student-t distribution with 5 degrees of freedom at the 5% significance level.

### 3.4.2 Forecasting Evaluation

Since the volatility of an asset return is not directly observable, comparing the forecasting performance of different volatil-
ity 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 \( \sigma_h^2(\ell) \). However, such a finding is not surprising because \( a_{h+\ell}^2 \) alone is not an adequate measure of the volatility at time \( h + \ell \). 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 \( \sigma_{h+1}^2 \). But it is not an accurate estimate of \( \sigma_{h+1}^2 \) 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.

### 3.5 The Integrated GARCH Model

If the AR polynomial of the GARCH representation in (48) 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

\[
a_t = \sigma_t \epsilon_t, \quad \sigma_t^2 = \alpha_0 + \beta_1 \sigma_{t-1}^2 + (1 - \beta_1) a_{t-1}^2,
\]
where $\{\epsilon_t\}$ is defined as before and $1 > \beta_1 > 0$. For the monthly excess returns of S&P 500 index, an estimated IGARCH(1, 1) model is

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

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 (50) give

$$
\sigma_h^2(\ell) = \sigma_h^2(1) + (\ell - 1)\alpha_0, \quad \ell \geq 1,
$$

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 $\alpha_0$. Nelson (1990) studies some probability properties of the volatility process $\sigma_t^2$ under an
IGARCH model. The process $\sigma_t^2$ 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. This special IGARCH$(1, 1)$ model is the volatility model used in RiskMetrics, which is an approach for calculating Value at Risk.

3.6 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 mean. A simple GARCH$(1, 1)$-M model can be written as

\[
\begin{align*}
    r_t &= \mu + c\sigma_t^2 + a_t, \\
    a_t &= \sigma_t \epsilon_t, \\
    \sigma_t^2 &= \alpha_0 + \alpha_1 a_{t-1}^2 + \beta \sigma_{t-1}^2, 
\end{align*}
\]

where $\mu$ and $c$ are constant. The parameter $c$ is called the risk premium parameter. A positive $c$ indicates that the return is positively related to its past volatility. Other specifications of risk premium have also been used in the literature, including
\[ r_t = \mu + c\sigma_t + a_t. \]

The formulation of the GARCH-M model in (55) 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 for the monthly excess returns of S&P 500 index from January 1926 to December 1991. The fitted model is

\[
\begin{align*}
    r_t &= 0.0028 + 1.99\sigma_t^2 + a_t, \\
    \sigma^2 &= 0.00016 + 0.1328a_{t-1}^2 + 0.8137\sigma_{t-1}^2,
\end{align*}
\]

where the standard errors for the two parameters in the mean equation are 0.0022 and 0.7425, respectively, and those for the parameters in the volatility equation are 0.00002, 0.0220, and 0.0192, respectively. The estimated risk premium for the index return is positive and significant at the 5% level. The idea of risk premium applies to other GARCH models.

### 3.7 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 considers the weighted innovation
\[ g(\epsilon_t) = \theta \epsilon_t + \gamma [ |\epsilon_t| - E(|\epsilon_t|) ] , \] (56)
where \( \theta \) and \( \gamma \) are real constants. Both \( \epsilon_t \) 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 \( \epsilon_t \), \( E(|\epsilon_t|) = \sqrt{\frac{2}{\pi}} \). For the standardized Student-\( t \) distribution in (42), we have
\[
E(|\epsilon_t|) = \frac{2\sqrt{v - 2\Gamma(\frac{v+1}{2})}}{(v - 1)\Gamma(\frac{v}{2})\sqrt{\pi}}.
\]

An EGARCH\((m, s)\) model can be written as
\[
\begin{cases}
 a_t = \sigma_t \epsilon_t, \\
 \ln(\sigma_t^2) = \alpha_0 + \frac{1 + \beta_1 B + \cdots + \beta_s B^s}{1 - \alpha_1 B - \cdots - \alpha_mB^m} g(\epsilon_{t-1}),
\end{cases}
\] (57)
where \( \alpha_0 \) is a constant, \( B \) is the back-shift (or lag) operator such that \( Bg(\epsilon_t) = g(\epsilon_{t-1}) \), and \( 1 + \beta_1 B + \cdots + \beta_s B^s \) and \( 1 - \alpha_1 B - \cdots - \alpha_mB^m \) 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, (57) 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 \( \ln(\sigma^2_t) \) is \( \alpha_0 \). 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 \( a_t \).

To better understand the EGARCH model, let us consider the
simple model with order \((1, 0)\)

\[
a_t = \sigma_t \epsilon_t, \quad (1 - \alpha B) \ln(\sigma^2_t) = (1 - \alpha)\alpha_0 + g(\epsilon_{t-1}), \quad (58)
\]

where \( \{\epsilon_t\} \) is a sequence of iid standard Gaussian random vari-
ables and the subscript of \( \alpha_1 \) is omitted. In this case, \( E(|\epsilon_t|) = \sqrt{\frac{2}{\pi}} \) and the model for \( \ln(\sigma^2_t) \) becomes

\[
(1 - \alpha B) \ln(\sigma^2_t) = \begin{cases} 
\alpha_* + (\theta + \gamma)\epsilon_{t-1} & \text{if } \epsilon_{t-1} \geq 0, \\
\alpha_* + (\theta - \gamma)\epsilon_{t-1} & \text{if } \epsilon_{t-1} < 0,
\end{cases} \quad (59)
\]

where \( \alpha_* = (1 - \alpha)\alpha_0 - \sqrt{\frac{2}{\pi}} \gamma \). This is a nonlinear function similar
to that of the threshold autoregressive model (TAR) 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

$$
\sigma^2_t = \sigma^2_{t-1} \exp \alpha_* \begin{cases}
\exp \left( (\theta + \gamma) \frac{a_{t-1}}{\sqrt{\sigma^2_{t-1}}} \right), & \text{if } a_{t-1} \geq 0, \\
\exp \left( (\theta - \gamma) \frac{a_{t-1}}{\sqrt{\sigma^2_{t-1}}} \right), & \text{if } a_{t-1} < 0.
\end{cases}
$$

The coefficients $(\theta + \gamma)$ and $(\theta - \gamma)$ show the asymmetry in response to positive and negative $a_{t-1}$. The model is, therefore, nonlinear if $\gamma \neq 0$. 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.

### 3.7.1 An 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:

$$
\begin{align*}
\begin{cases}
\begin{align*}
       r_t &= \phi_0 + \phi_1 r_{t-1} + c \sigma_t^2 + a_t, \\
\ln(\sigma_t^2) &= \alpha_0 + \ln(1 + w N_t) + \frac{1 + \beta B}{1 - \alpha_1 B - \alpha_2 B^2} g(\epsilon_{t-1}).
\end{align*}
\end{cases}
\end{align*}
$$

(60)
where $\sigma_t^2$ 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 $t$, $\alpha_0$ and $w$ are real parameters, $g(\epsilon_t)$ is defined in (56), and $\epsilon_t$ follows a generalized error distribution with probability density function

$$f(x) = \frac{v \exp\left(-\frac{1}{2} \left| \frac{x}{\lambda} \right|^v \right)}{\lambda 2^{(1+\frac{1}{v})} \Gamma\left(\frac{1}{v}\right)}, \quad -\infty < x < \infty, \quad 0 < v \leq \infty,$$

where again $\Gamma(\cdot)$ is the gamma function and

$$\lambda = \left[ 2^{(-\frac{3}{v})} \frac{\Gamma\left(\frac{1}{v}\right)}{\Gamma\left(\frac{3}{v}\right)} \right]^\frac{1}{2}.$$

<table>
<thead>
<tr>
<th>Par.</th>
<th>$\alpha_0$</th>
<th>$w$</th>
<th>$\nu$</th>
<th>$\alpha_1$</th>
<th>$\alpha_2$</th>
<th>$\beta$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Est.</td>
<td>-10.06</td>
<td>.183</td>
<td>.156</td>
<td>1.929</td>
<td>-9.29</td>
<td>-9.78</td>
</tr>
<tr>
<td>Err.</td>
<td>.346</td>
<td>.028</td>
<td>.013</td>
<td>.015</td>
<td>.015</td>
<td>.006</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Par.</th>
<th>$\theta$</th>
<th>$\phi_0$</th>
<th>$\phi_1$</th>
<th>$c$</th>
<th>$\nu$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Est.</td>
<td>-.118</td>
<td>3.5 $\cdot 10^{-4}$</td>
<td>.205</td>
<td>-3.361</td>
<td>1.576</td>
</tr>
<tr>
<td>Err.</td>
<td>.009</td>
<td>9.9 $\cdot 10^{-5}$</td>
<td>.012</td>
<td>2.026</td>
<td>.032</td>
</tr>
</tbody>
</table>


Similar to a GARCH-M model, the parameter $c$ in (60) is the risk premium parameter. Table 3.2 gives the parameter estimates and their standard errors of the model. The mean equation of model (60) 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 $\sigma_t^2$ as a regressor to account for risk premium. The estimated risk premium is negative, but statistically insignificant.

### 3.7.2 Another 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, 0) model is entertained and the fitted model is

$$r_t=0.0105 + 0.092r_{t-1} + a_t, \quad a_t = \sigma_t \epsilon_t, \quad (61)$$

$$\ln(\sigma^2_t)=-5.496 + \frac{g(\epsilon_{t-1})}{1 - 0.856B}, \quad (62)$$

$$g(\epsilon_{t-1})=-0.0795\epsilon_{t-1} + 0.2647\left|\epsilon_{t-1}\right| - \frac{2}{\sqrt{\pi}}, \quad (63)$$

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 $\tilde{a}_t = \frac{a_t}{\sigma_t}$ 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, 0) model is adequate.

From the estimated volatility equation in (63) and using $\sqrt{\frac{2}{\pi}} \approx 0.7979$, 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

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

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 two 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.
3.7.3 Forecasting Using an EGARCH Model

We use the EGARCH(1, 0) 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

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

Taking exponentials, the model becomes

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

(64)

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, (64) 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

$$\hat{\sigma}_h^2(2) = \hat{\sigma}_h^{2\alpha_1} \exp((1 - \alpha_1) \alpha_0) E_h[\exp(g(\epsilon_{h+1}))].$$
where $E_h$ denotes a conditional expectation taking at the time origin $h$. The prior expectation can be obtained as follows:

$$E[\exp(g(\epsilon))] = \int_{-\infty}^{\infty} e^{\theta \epsilon + \gamma (|\epsilon| - \sqrt{\frac{2}{\pi}})} f(\epsilon) d\epsilon$$

$$= \exp(-\gamma \sqrt{\frac{2}{\pi}}) \left[ \int_{-\infty}^{\infty} e^{(\theta + \gamma) \epsilon} \frac{1}{\sqrt{2\pi}} e^{-\frac{\epsilon^2}{2}} d\epsilon + \int_{-\infty}^{0} e^{(\theta - \gamma) \epsilon} \frac{1}{\sqrt{2\pi}} e^{-\frac{\epsilon^2}{2}} d\epsilon \right]$$

$$= e^{-\gamma \sqrt{\frac{2}{\pi}}} \left[ e^{\frac{(\theta + \gamma)^2}{2}} \Phi(\theta + \gamma) + e^{\frac{(\theta - \gamma)^2}{2}} \Phi(\theta - \gamma) \right],$$

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

$$\hat{\sigma}^2_h(2) = \hat{\sigma}_h^{2\alpha_1}(1) e^{(1-\alpha_1)\alpha_0 - \gamma \sqrt{\frac{2}{\pi}}} \times \left[ e^{\frac{(\theta + \gamma)^2}{2}} \Phi(\theta + \gamma) + e^{\frac{(\theta - \gamma)^2}{2}} \Phi(\gamma - \theta) \right].$$

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

$$\hat{\sigma}^2_h(j) = \hat{\sigma}_h^{2\alpha_1}(j - 1) \exp(\omega) \times \left[ e^{\frac{(\theta + \gamma)^2}{2}} \Phi(\theta + \gamma) + e^{\frac{(\theta - \gamma)^2}{2}} \Phi(\gamma - \theta) \right],$$

where $\omega = (1 - \alpha_1)\alpha_0 - \gamma \sqrt{\frac{2}{\pi}}$. The values of $\Phi(\theta + \gamma)$ and $\Phi(\theta - \gamma)$ can be obtained from most statistical packages.

For illustration, consider the AR(1)-EGARCH(1, 0) model of the previous subsection for the monthly log returns of IBM stock.
Using the fitted EGARCH\((1, 0)\) 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 \(\hat{\sigma}^2_{864}(4) = 4.94 \times 10^{-3}\). These forecasts converge gradually to the sample variance \(4.37 \times 10^{-3}\) of the shock process \(a_t\) of (61).

### 3.8 The Charma Model

Many other econometric models have been proposed in the literature to describe the evolution of the conditional variance \(\sigma^2_t\) in (37). We mention the conditional heteroscedastic ARMA (CHARMA) model that uses random coefficients to produce conditional heteroscedasticity. 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

\[
\begin{align*}
\{r_t\} &= \mu_t + a_t, \\
\{a_t\} &= \delta_1 a_{t-1} + \delta_2 a_{t-2} + \cdots + \delta_m a_{t-m} + \eta_t,
\end{align*}
\tag{65}
\]

where \(\{\eta_t\}\) is a Gaussian white noise series with mean zero and variance \(\sigma^2_\eta\), \(\{\delta_t\} = \{(\delta_{1t}, \cdots, \delta_{mt})'\}\) is a sequence of iid random vectors with mean zero and non-negative definite covariance matrix \(\Omega\), and \(\{\delta_t\}\) is independent of \(\{\eta_t\}\). In this section, we use some basic properties of vector and matrix operations to
simplify the presentation. For $m > 0$, the model can be written as

$$a_t = a'_{t-1} \delta_t + \eta_t,$$

where $a_t = (a_{t-1}, \ldots, a_{t-m})'$ 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 (65) is then

$$
\sigma^2_t = \sigma^2_\eta + a'_{t-1} \text{Cov}(\delta_t) a_{t-1} \\
= \sigma^2_\eta + (a_{t-1}, \ldots, a_{t-m}) \Omega (a_{t-1}, \ldots, a_{t-m})',
$$

(66)

Denote the $(i, j)$th element of $\Omega$ by $\omega_{ij}$. Because the matrix is symmetric, we have $\omega_{ij} = \omega_{ji}$. If $m = 1$, then (65) reduces to $\sigma^2_t = \sigma^2_\eta + \omega_{11} a^2_{t-1}$, which is an ARCH(1) model. If $m = 2$, then (65) reduces to

$$
\sigma^2_t = \sigma^2_\eta + \omega_{11} a^2_{t-1} + 2 \omega_{12} a_{t-1} a_{t-2} + \omega_{22} a^2_{t-2},
$$

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 $\Omega$ is a diagonal matrix. Because $\Omega$ is a covariance matrix, which is non-negative definite, and $\sigma^2_\eta$ is a variance, which is positive, we have $\sigma^2_t \geq \sigma^2_\eta > 0$ for all $t$. In other words, the positiveness of $\sigma^2_t$ 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 harder to handle random coefficients than constant coefficients.

For illustration, we employ the CHARMA model

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

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

$$\begin{align*}
\begin{cases}
  r_t = & 0.00635 + a_t, \\
  \sigma_t^2 = & 0.00179 + (a_{t-1}, a_{t-2})^\top \hat{\Omega}(a_{t-1}, a_{t-2})^\top,
\end{cases}
\end{align*}$$

where

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

and the numbers in parentheses are standard errors. The cross-product term of $\hat{\Omega}$ has a $t$ ratio of $-1.63$, which is marginally
significant at the 10% level. If we refine the model to
\[
\begin{align*}
  r_t &= \phi_0 + a_t, \\
  a_t &= \delta_1 a_{t-1} + \delta_2 a_{t-2} + \delta_3 a_{t-3} + \eta_t,
\end{align*}
\]
but assume that \( \delta_3 t \) is uncorrelated with \((\delta_1 t, \delta_2 t)\), then we obtain the fitted model
\[
\begin{align*}
  r_t &= 0.0068 + a_t, \\
  \sigma_t^2 &= 0.00136 + (a_{t-1}, a_{t-2}, a_{t-3})' \hat{\Omega}(a_{t-1}, a_{t-2}, a_{t-3})',
\end{align*}
\]
where the elements of \( \hat{\Omega} \) and their standard errors, shown in parentheses, are
\[
\hat{\Sigma} = \begin{bmatrix}
  0.1212(0.0355) & -0.0622(0.0283) & 0 \\
  -0.0622(0.0283) & 0.1913(0.0254) & 0 \\
  0 & 0 & 0.2988(0.0420)
\end{bmatrix}.
\]
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
\[
\sigma_t^2 = 0.00136 + 0.12a_{t-1}^2 - 0.12a_{t-2}a_{t-2} + 0.19a_{t-2}^2 + 0.30a_{t-3}^2.
\]
The conditional variance is slightly larger when \( a_{t-1}a_{t-2} \) is negative.
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

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

(67)

where $\delta_t = (\delta_{1t}, \cdots, \delta_{mt})'$ and $\eta_t$ are random vector and variable defined in (65). Then the conditional variance of $a_t$ is

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

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

3.9 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

\[ r_t = \phi_0 + \sum_{i=1}^{p} (\phi_i + \delta_{it}) r_{t-i} + a_t. \] (68)

where \( p \) is a positive integer, \( \{\delta_t\} = \{(\delta_{1t}, \cdots, \delta_{pt})'\} \) is a sequence of independent random vectors with mean zero and covariance matrix \( \Omega_{\delta} \), and \( \{\delta_t\} \) is independent of \( \{a_t\} \). The conditional mean and variance of the RCA model in (68) are

\[
\begin{align*}
\mu_t &= E(a_t | F_{t-1}) = \sum_{i=1}^{p} \phi_i a_{t-i}, \\
\sigma_t^2 &= \sigma_a^2 + (r_{t-1}, \cdots, r_{t-p}) \Omega_{\delta}(r_{t-1}, \cdots, r_{t-p})',
\end{align*}
\]

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.

### 3.10 The 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 \( a_t \). 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 \( \sigma_t^2 \). A SV model is defined as
\[
a_t = \sigma_t \epsilon_t, \quad (1 - \alpha_1 B - \cdots - \alpha_m B^m) \ln(\sigma_t^2) = \alpha_0 + v_t, \tag{69}
\]
where \( \{\epsilon_t\} \) are iid \( N(0, 1) \), \( \{v_t\} \) are iid \( N(0, \sigma_v^2) \), \( \{\epsilon_t\} \) and \( \{v_t\} \) are independent, \( \alpha_0 \) is a constant, and all zeros of the polynomial \( 1 - \sum_{i=1}^{m} \alpha_i B^i \) are greater than 1 in modulus. Introducing the innovation \( v_t \) substantially increases the flexibility of the model in describing the evolution of \( \sigma_t^2 \), but it also increases the difficulty in parameter estimation. To estimate a SV model, we need a quasi-likelihood method via Kalman filtering or a Monte Carlo method. The difficulty in estimating a SV model is understandable because for each shock at the model uses two innovations \( \epsilon_t \) and \( v_t \).

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
\[
\ln(\sigma_t^2) \sim \mathcal{N} \left( \frac{\alpha_0}{1 - \alpha_1}, \frac{\sigma_v^2}{1 - \alpha_1^2} \right) \equiv \mathcal{N}(\mu_h, \sigma_h^2),
\]
and \( E(a_t^2) = \exp(\mu_h + \frac{1}{2\sigma_h^2}) \), \( E(a_t^4) = 3 \exp(2\mu_h^2 + 2\sigma_h^2) \), and \( \text{Corr}(a_t^2, a_{t-i}^2) = \frac{\exp(\sigma_h^2\alpha_i) - 1}{3 \exp(\sigma_h^2) - 1} \). Limited experience shows that SV models often provided improvements in model fitting, but their contributions to out-of-sample volatility forecasts received mixed results.
3.11 The 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 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. Figure 3.9 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, 1997. These sample ACFs are positive and of moderate magnitude, but decay slowly.
A simple long-memory stochastic volatility (LMSV) model can be written as

\[ a_t = \sigma_t \epsilon_t, \quad \sigma_t = \sigma \exp \left( \frac{u_t}{2} \right), \quad (1 - B)^d u_t = \eta_t, \]  \hspace{1cm} (70)

where \( \sigma > 0 \), \( \{\epsilon_t\} \) are iid \( N(0, 1) \), \( \{\eta_t\} \) are iid \( N(0, \sigma_\eta^2) \) and independent of \( \epsilon_t \), 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 (70), we have
\[
\ln(a_t^2) = \ln(\sigma^2) + u_t + \ln(\epsilon_t^2)
\]
\[
= [\ln(\sigma^2) + E(\ln \epsilon_t^2)] + u_t + [\ln(\epsilon_t^2) - E(\ln \epsilon_t^2)]
\]
\[
\equiv \mu + u_t + e_t.
\]
Thus, the \(\ln(a_t^2)\) series is a Gaussian long-memory signal plus a non-Gaussian white noise. Estimation of the long-memory stochastic volatility model is complicated, but the fractional difference parameter \(d\) can be estimated by using either a quasi-maximum likelihood method or a regression method.

3.12 An Alternative Approach

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 some interest due in part to the availability of high-frequency financial data, especially in the foreign exchange markets (e.g., Andersen, Bollerslev, Diebold, and Labys, 1999).

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

\[
r^m_t = \sum_{i=1}^n r_{t,i}.
\]

Assuming that the conditional variance and covariance exist, we have

\[
\text{Var}(r^m_t|F_{t-1}) = \sum_{i=1}^n \text{Var}(r_{t,i}|F_{t-1}) + 2 \sum_{i<j} \text{Cov}[(r_{t,i}r_{t,j})|F_{t-1}],
\]

(71)

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

\[
\text{Var}(r^m_t|F_{t-1}) = n \text{Var}(r_{t,1}),
\]

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

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

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

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

(72)

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

\[
\text{Var}(r^m_t|F_{t-1}) = n \text{Var}(r_{t,1}) + 2(n - 1)\text{Cov}(r_{t,1}, r_{t,2}),
\]
which can be estimated by
\[ \hat{\sigma}_m^2 = \frac{n}{n - 1} \sum_{i=1}^{n} (r_{t,i} - \bar{r}_t)^2 + 2 \sum_{i=1}^{n-1} (r_{t,i} - \bar{r}_t)(r_{t,i+1} - \bar{r}_t). \] (73)

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 (71). 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 (71) 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 (72) and (73) may not even be consistent.

**Example 3.4.** Consider the monthly volatility of the log returns of 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 (72) (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 (73)]. The third method applies a GARCH(1, 1) model to the monthly returns from January 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

\[
\begin{align*}
    r_t^m &= 0.658 + a_t, \\
    a_t &= \sigma_t \epsilon_t, \\
    \sigma_t^2 &= 3.349 + 0.086 a_{t-1}^2 + 0.735 \sigma_{t-1}^2,
\end{align*}
\]

where \( \{\epsilon_t\} \) is a standard Gaussian white noise series. Figure 3.10 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.
Figure 3.10. Time plots of estimated monthly volatility for the log returns of S&P 500 index from January 1980 to December 1999: (a) assumes that the daily log returns form a white noise series, (b) assumes that the 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.

**Remark:** In (72), 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 can be used as an estimate of monthly volatility.

### 3.13 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 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.
Figure 3.11. Time plots of monthly log returns for IBM stock and S&P 500 index. The sample period is from January 1926 to December 1999. The returns are in percentages and include dividends.

**Example 3.5.** 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.
Denote the monthly log return series by $r_t$. If Gaussian GARCH models are entertained, we obtain the GARCH(1, 1) model

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

(74)

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 $\tilde{a}_t = \frac{a_t}{\sigma_t}$, we obtain $Q(10) = 7.82(0.553)$ and $Q(20) = 21.22(0.325)$, where $p$ 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

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

(75)
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 \(\alpha_{10}\) and \(\beta_{10}\) are statistically non-significance at the 10\% level. Therefore, we refine the equation and obtain the model
\[
\begin{align*}
    r_t &= 1.21 + 0.099 r_{t-1} + a_t, \\
    a_t &= \sigma_t \epsilon_t, \\
    \sigma_t^2 &= 4.539 + 0.113 a_{t-1}^2 + 0.816 \sigma_{t-1}^2 - 5.154 u_t.
\end{align*}
\]
(76)
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 \(\tilde{a}_t = \frac{a_t}{\sigma_t}\) 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 (74) and (76), we obtain the following conclusions. First, because the coefficient \(-5.51\) is
significantly different from zero with $p$ value 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 (76) as

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

The negative constant term $-0.615 = 4.539 - 5.514$ is counterintuitive. However, since the standard errors of 4.539 and 5.51 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

$$
\begin{cases} 
\begin{align*} 
    r_t &= 1.21 + 0.099r_{t-1} + a_t, \\
    a_t &= \sigma_t \epsilon_t, \\
    \sigma_t^2 &= 0.114a_{t-1}^2 + 0.811\sigma_{t-1}^2 - 4.552(1 - u_t). 
\end{align*}
\end{cases}
$$
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 (77) can readily be used to assess the Summer effect on the IBM stock volatility. For illustration, based on the model in (77), the medians of $a_t^2$ and $\sigma_t^2$ 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. Ratio of the two volatilities is $\frac{64.3}{68.8} \approx 93\%$. Thus, there is a 7% reduction in the volatility of the monthly log return of IBM stock in the Summer months.

**Example 3.6.** 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 S&P 500 index. Using the \( r_t \) series and Gaussian GARCH models, we obtain the following special GARCH(2, 1) model

\[
\begin{align*}
    r_t &= 0.609 + a_t, \\
    a_t &= \sigma_t \epsilon_t, \\
    \sigma_t^2 &= 0.717 + 0.147 a_{t-2}^2 + 0.839 \sigma_{t-1}^2.
\end{align*}
\] (78)

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 \( \tilde{a}_t = \frac{a_t}{\sigma_t} \), we have \( Q(10) = 11.51(0.32) \) and \( Q(20) = 23.71(0.26) \), where the number in parentheses denotes \( p \) 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 1.244 is
the sample mean of $x_t$. The fitted model for $r_t$ becomes

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

(79)

The standard error of the parameter in the mean equation is 0.139 and those of the parameters in the volatility equation are 0.271, 0.020, 0.01, 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 $\tilde{a}_t = \frac{a_t}{\sigma_t}$ 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.

Table 3.3. Fitted Volatilities for the Monthly Log Returns of the S&P 500 Index from July to December 1999 Using Models with and without the Past Log Return of IBM Stock.

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Model (3.43)</td>
<td>26.30</td>
<td>26.01</td>
<td>24.73</td>
<td>21.69</td>
<td>20.71</td>
<td>22.46</td>
</tr>
<tr>
<td>Model (3.44)</td>
<td>23.32</td>
<td>23.13</td>
<td>22.46</td>
<td>20.00</td>
<td>19.45</td>
<td>18.27</td>
</tr>
</tbody>
</table>

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. Table 3.3 gives the fitted volatility of S&P 500 index from July to December of 1999 using models (78) and (78). From the table, the past value of IBM log stock return indeed contributes to the modeling of the S&P 500 index volatility.

3.14 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

\[ a_t = \sigma_t \epsilon_t, \quad \sigma_t^2 = \alpha_0 + \alpha_1 a_{t-1}^2 + \beta_1 \sigma_{t-1}^2, \]

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

\[ E(\epsilon_t) = 0, \quad \text{Var}(\epsilon_t) = 1, \quad E(\epsilon_t^4) = K_\epsilon + 3. \]

where \( K_\epsilon \) is the excess kurtosis of the innovation \( \epsilon_t \). Based on the assumption, we have

- \( \text{Var}(a_t) = E(\sigma_t^2) = \frac{\alpha_0}{1-(\alpha_1+\beta_1)}. \)
- \( E(a_t^4) = (K_\epsilon + 3)E(\sigma_t^4) \) 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
\[
E(\sigma_{1}^4) = \frac{\alpha_0^2(1 + \alpha_1 + \beta_1)}{[1 - (\alpha_1 + \beta_1)][1 - \alpha_1^2(K_\epsilon + 2) - (\alpha_1 + \beta_1)^2]}
\]
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
\[
K_a = \frac{E(a_t^4)}{(E(a_t^2))^2} - 3 \\
= \frac{(K_\epsilon + 3)[1 - (\alpha_1 + \beta_1)^2]}{1 - 2\alpha_1^2 - (\alpha_1 + \beta_1)^2 - K_\epsilon \alpha_1^2} - 3.
\]

This excess kurtosis can be written in an informative expression. First, consider the case that \(\epsilon_t\) 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 $\epsilon_t$ is not Gaussian. Using the prior result, we have

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

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 $K_a^{(g)} = \frac{6\alpha_1^2}{1 - 3\alpha_1^2}$ provided that $3\alpha_1^2 < 1$ and the excess kurtosis of $a_t$ is

\[
K_a = \frac{(K_\epsilon + 3)(1 - \alpha_1^2)}{1 - (K_\epsilon + 3)\alpha_1^2} - 3 \\
= \frac{K_\epsilon + 2K_\epsilon\alpha_1^2 + 6\alpha_1^2}{1 - 3\alpha_1^2 - K_\epsilon\alpha_1^2} \\
= \frac{K_\epsilon(1 - 3\alpha_1^2) + 6\alpha_1^2 + 5K_\epsilon\alpha_1^2}{1 - 3\alpha_1^2 - K_\epsilon\alpha_1^2} \\
= \frac{K_\epsilon + K_a^{(g)} + \frac{5}{6}K_\epsilon K_a^{(g)}}{1 - \frac{1}{6}K_\epsilon K_a^{(g)}}.
\]

The prior result shows that for a GARCH(1, 1) model the coefficient $\alpha_1$ 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 $\epsilon_t$. 

158
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 $\nu$ degrees of freedom, we have $E(\epsilon_t^4) = \frac{6}{\nu-4} + 3$ if $\nu > 4$. Therefore, the excess kurtosis of $\epsilon_t$ is $K_\epsilon = \frac{6}{\nu-4}$ for $\nu > 4$. This is part of the reason that we used $t_5$ in the chapter when the degrees of freedom of $t$ distribution are prespecified. The excess kurtosis of $a_t$ becomes

$$K_a = \frac{6 + (\nu + 1)K_a^{(g)}}{\nu - 4 - K_a^{(g)}}$$

provided that $1 - 2\alpha_1^2\frac{\nu-1}{\nu-4} - (\alpha_1 + \beta_1)^2 > 0$. 

159