Analysis of Financial Time Series

Third Edition

RUEY S. TSAY

The University of Chicago
Booth School of Business
Chicago, IL

WILEY

A JOHN WILEY & SONS, INC., PUBLICATION

CHAPTER 2

Linear Time Series Analysis and Its Applications

In this chapter, we discuss basic theories of linear time series analysis, introduce some simple econometric models useful for analyzing financial data, and apply the models to financial time series such as asset returns. Discussions of the concepts are brief with emphasis on those relevant to financial applications. Understanding the simple time series models introduced here will go a long way to better appreciate the more sophisticated financial econometric models of the later chapters. There are many time series textbooks available. For basic concepts of linear time series analysis, see Box, Jenkins, and Reinsel (1994, Chapters 2 and 3) and Brockwell and Davis (1996, Chapters 1–3).

Treating an asset return (e.g., log return r_t of a stock) as a collection of random variables over time, we have a time series $\{r_t\}$. Linear time series analysis provides a natural framework to study the dynamic structure of such a series. The theories of linear time series discussed include stationarity, dynamic dependence, autocorrelation function, modeling, and forecasting. The econometric models introduced include (a) simple autoregressive (AR) models, (b) simple moving-average (MA) models, (b) mixed autoregressive moving-average (ARMA) models, (c) seasonal models, (d) unit-root nonstationarity, (e) regression models with time series errors, and (f) fractionally differenced models for long-range dependence. For an asset return r_t, simple models attempt to capture the linear relationship between r_t and information available prior to time t. The information may contain the historical values of r_t and the random vector $\boldsymbol{Y}$ in Eq. (1.14), which describes the economic environment under which the asset price is determined. As such, correlation plays an important role in understanding these models. In particular, correlations between the variable of interest and its past values become the focus of linear time series analysis. These correlations are referred to as *serial correlations* or *autocorrelations*. They are the basic tool for studying a stationary time series.

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

2.1 STATIONARITY

The foundation of time series analysis is stationarity. A time series $\{r_t\}$ is said to be *strictly stationary* if the joint distribution of $(r_{t_1}, \ldots, r_{t_k})$ is identical to that of $(r_{t_1+t}, \ldots, r_{t_k+t})$ for all t, where k is an arbitrary positive integer and $(t_1, \ldots, t_k)$ is a collection of k positive integers. In other words, strict stationarity requires that the joint distribution of $(r_{t_1}, \ldots, r_{t_k})$ is invariant under time shift. This is a very strong condition that is hard to verify empirically. A weaker version of stationarity is often assumed. A time series $\{r_t\}$ is *weakly stationary* if both the mean of r_t and the covariance between r_t and $r_{t-\ell}$ are time invariant, where ℓ is an arbitrary integer. More specifically, $\{r_t\}$ is weakly stationary if (a) $E(r_t) = \mu$, which is a constant, and (b) $\mathrm{Cov}(r_t, r_{t-\ell}) = \gamma_\ell$, which only depends on ℓ. In practice, suppose that we have observed T data points $\{r_t | t = 1, \ldots, T\}$. The weak stationarity implies that the time plot of the data would show that the T values fluctuate with constant variation around a fixed level. In applications, weak stationarity enables one to make inference concerning future observations (e.g., prediction).

Implicitly, in the condition of weak stationarity, we assume that the first two moments of r_t are finite. From the definitions, if r_t is strictly stationary and its first two moments are finite, then r_t is also weakly stationary. The converse is not true in general. However, if the time series r_t is normally distributed, then weak stationarity is equivalent to strict stationarity. In this book, we are mainly concerned with weakly stationary series.

The covariance $\gamma_\ell = \mathrm{Cov}(r_t, r_{t-\ell})$ is called the lag-ℓ autocovariance of r_t. It has two important properties: (a) $\gamma_0 = \mathrm{Var}(r_t)$ and (b) $\gamma_{-\ell} = \gamma_\ell$. The second property holds because $\mathrm{Cov}(r_t, r_{t-(-\ell)}) = \mathrm{Cov}(r_{t-(-\ell)}, r_t) = \mathrm{Cov}(r_{t+\ell}, r_t) = \mathrm{Cov}(r_{t_1}, r_{t_1-\ell})$, where $t_1 = t + \ell$.

In the finance literature, it is common to assume that an asset return series is weakly stationary. This assumption can be checked empirically provided that a sufficient number of historical returns are available. For example, one can divide the data into subsamples and check the consistency of the results obtained across the subsamples.

2.2 CORRELATION AND AUTOCORRELATION FUNCTION

The correlation coefficient between two random variables X and Y is defined as

$$\rho_{x,y} = \frac{\mathrm{Cov}(X, Y)}{\sqrt{\mathrm{Var}(X)\mathrm{Var}(Y)}} = \frac{E[(X - \mu_x)(Y - \mu_y)]}{\sqrt{E(X - \mu_x)^2 E(Y - \mu_y)^2}},$$

where μ_x and μ_y are the mean of X and Y, respectively, and it is assumed that the variances exist. This coefficient measures the strength of linear dependence between X and Y, and it can be shown that $-1 \le \rho_{x,y} \le 1$ and $\rho_{x,y} = \rho_{y,x}$. The two random variables are uncorrelated if $\rho_{x,y} = 0$. In addition, if both X and Y are normal random variables, then $\rho_{x,y} = 0$ if and only if X and Y are independent. When the sample $\{(x_t, y_t)\}_{t=1}^T$ is available, the correlation can be consistently estimated by its sample counterpart

$$\hat{\rho}_{x,y} = \frac{\sum_{t=1}^T (x_t - \bar{x})(y_t - \bar{y})}{\sqrt{\sum_{t=1}^T (x_t - \bar{x})^2 \sum_{t=1}^T (y_t - \bar{y})^2}},$$

where $\bar{x} = \sum_{t=1}^T x_t/T$ and $\bar{y} = \sum_{t=1}^T y_t/T$ are the sample mean of X and Y, respectively.

***Autocorrelation Function* (ACF)**

Consider a weakly stationary return series r_t. When the linear dependence between r_t and its past values r_{t-i} is of interest, the concept of correlation is generalized to autocorrelation. The correlation coefficient between r_t and $r_{t-\ell}$ is called the lag-ℓ *autocorrelation* of r_t and is commonly denoted by ρ_ℓ, which under the weak stationarity assumption is a function of ℓ only. Specifically, we define

$$\rho_\ell = \frac{\mathrm{Cov}(r_t, r_{t-\ell})}{\sqrt{\mathrm{Var}(r_t)\mathrm{Var}(r_{t-\ell})}} = \frac{\mathrm{Cov}(r_t, r_{t-\ell})}{\mathrm{Var}(r_t)} = \frac{\gamma_\ell}{\gamma_0}, \tag{2.1}$$

where the property $\mathrm{Var}(r_t) = \mathrm{Var}(r_{t-\ell})$ for a weakly stationary series is used. From the definition, we have $\rho_0 = 1$, $\rho_\ell = \rho_{-\ell}$, and $-1 \le \rho_\ell \le 1$. In addition, a weakly stationary series r_t is not serially correlated if and only if $\rho_\ell = 0$ for all $\ell > 0$.

For a given sample of returns $\{r_t\}_{t=1}^T$, let $\bar{r}$ be the sample mean (i.e., $\bar{r} = \sum_{t=1}^T r_t/T$). Then the lag-1 sample autocorrelation of r_t is

$$\hat{\rho}_1 = \frac{\sum_{t=2}^T (r_t - \bar{r})(r_{t-1} - \bar{r})}{\sum_{t=1}^T (r_t - \bar{r})^2}.$$

Under some general conditions, $\hat{\rho}_1$ is a consistent estimate of ρ_1. For example, if $\{r_t\}$ is an independent and identically distributed (iid) sequence and $E(r_t^2) < \infty$, then $\hat{\rho}_1$ is asymptotically normal with mean zero and variance $1/T$; see Brockwell and Davis (1991, Theorem 7.2.2). This result can be used in practice to test the null hypothesis $H_0 : \rho_1 = 0$ versus the alternative hypothesis $H_a : \rho_1 \neq 0$. The test statistic is the usual t ratio, which is $\sqrt{T}\hat{\rho}_1$ and follows asymptotically the standard normal distribution. The null hypothesis H_0 is rejected if the t ratio is large in magnitude or, equivalently, the p value of the t ratio is small, say less than 0.05. In general, the lag-ℓ sample autocorrelation of r_t is defined as

$$\hat{\rho}_\ell = \frac{\sum_{t=\ell+1}^T (r_t - \bar{r})(r_{t-\ell} - \bar{r})}{\sum_{t=1}^T (r_t - \bar{r})^2}, \qquad 0 \le \ell < T - 1. \tag{2.2}$$

If $\{r_t\}$ is an iid sequence satisfying $E(r_t^2) < \infty$, then $\hat{\rho}_\ell$ is asymptotically normal with mean zero and variance $1/T$ for any fixed positive integer ℓ. More generally, if r_t is a weakly stationary time series satisfying $r_t = \mu + \sum_{i=0}^q \psi_i a_{t-i}$, where

$\psi_0 = 1$ and $\{a_j\}$ is a sequence of iid random variables with mean zero, then $\hat{\rho}_\ell$ is asymptotically normal with mean zero and variance $(1 + 2\sum_{i=1}^{q} \rho_i^2)/T$ for $\ell > q$. This is referred to as Bartlett's formula in the time series literature; see Box, Jenkins, and Reinsel (1994). For more information about the asymptotic distribution of sample autocorrelations, see Fuller (1976, Chapter 6) and Brockwell and Davis (1991, Chapter 7).

Testing Individual ACF

For a given positive integer ℓ, the previous result can be used to test $H_0 : \rho_\ell = 0$ vs. $H_a : \rho_\ell \neq 0$. The test statistic is

$$t \text{ ratio} = \frac{\hat{\rho}_\ell}{\sqrt{(1 + 2\sum_{i=1}^{\ell-1} \hat{\rho}_i^2)/T}}.$$

If $\{r_t\}$ is a stationary Gaussian series satisfying $\rho_j = 0$ for $j > \ell$, the t ratio is asymptotically distributed as a standard normal random variable. Hence, the decision rule of the test is to reject H_0 if $|t \text{ ratio}| > Z_{\alpha/2}$, where $Z_{\alpha/2}$ is the $100(1 - \alpha/2)$th percentile of the standard normal distribution. For simplicity, many software packages use $1/T$ as the asymptotic variance of $\hat{\rho}_\ell$ for all $\ell \neq 0$. They essentially assume that the underlying time series is an iid sequence.

In finite samples, $\hat{\rho}_\ell$ is a biased estimator of ρ_ℓ. The bias is in the order of $1/T$, which can be substantial when the sample size T is small. In most financial applications, T is relatively large so that the bias is not serious.

Portmanteau Test

Financial applications often require to test jointly that several autocorrelations of r_t are zero. Box and Pierce (1970) propose the Portmanteau statistic

$$Q^*(m) = T \sum_{\ell=1}^{m} \hat{\rho}_\ell^2$$

as a test statistic for the null hypothesis $H_0 : \rho_1 = \cdots = \rho_m = 0$ against the alternative hypothesis $H_a : \rho_i \neq 0$ for some $i \in \{1, \ldots, m\}$. Under the assumption that $\{r_t\}$ is an iid sequence with certain moment conditions, $Q^*(m)$ is asymptotically a chi-squared random variable with m degrees of freedom.

Ljung and Box (1978) modify the $Q^*(m)$ statistic as below to increase the power of the test in finite samples,

$$Q(m) = T(T + 2) \sum_{\ell=1}^{m} \frac{\hat{\rho}_\ell^2}{T - \ell}. \tag{2.3}$$

The decision rule is to reject H_0 if $Q(m) > \chi_\alpha^2$, where χ_α^2 denotes the $100(1 - \alpha)$th percentile of a chi-squared distribution with m degrees of freedom. Most software packages will provide the p value of $Q(m)$. The decision rule is then to reject H_0 if the p value is less than or equal to α, the significance level.

In practice, the choice of m may affect the performance of the $Q(m)$ statistic. Several values of m are often used. Simulation studies suggest that the choice of $m \approx \ln(T)$ provides better power performance. This general rule needs modification in analysis of seasonal time series for which autocorrelations with lags at multiples of the seasonality are more important.

The statistics $\hat{\rho}_1, \hat{\rho}_2, \ldots$ defined in Eq. (2.2) is called the *sample autocorrelation function* (ACF) of r_t. It plays an important role in linear time series analysis. As a matter of fact, a linear time series model can be characterized by its ACF, and linear time series modeling makes use of the sample ACF to capture the linear dynamic of the data. Figure 2.1 shows the sample autocorrelation functions of monthly simple and log returns of IBM stock from January 1926 to December 2008. The two sample ACFs are very close to each other, and they suggest that the serial correlations of monthly IBM stock returns are very small, if any. The sample ACFs are all within their two standard error limits, indicating that they are not significantly different from zero at the 5% level. In addition, for the simple returns, the Ljung–Box statistics give $Q(5) = 3.37$ and $Q(10) = 13.99$, which correspond to p values of 0.64 and 0.17, respectively, based on chi-squared

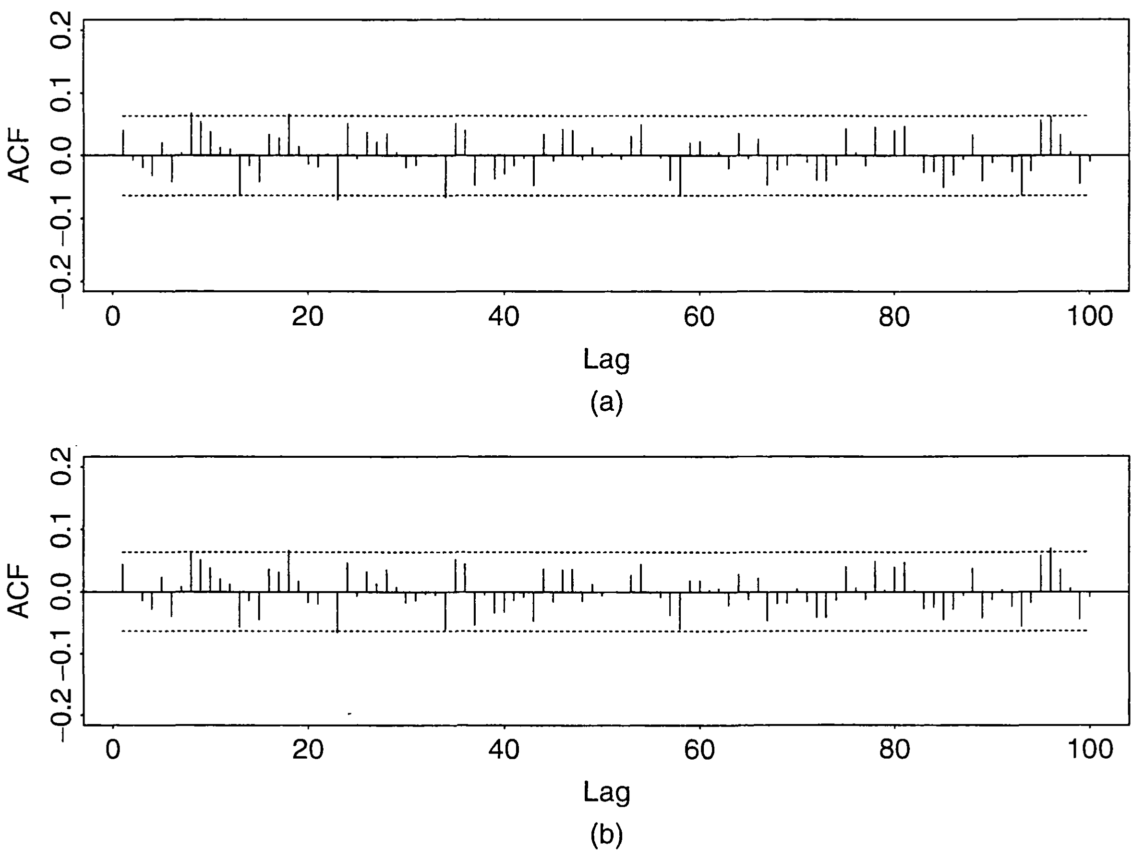

Figure 2.1 Sample autocorrelation functions of monthly (a) simple returns and (b) log returns of IBM stock from January 1926 to December 2008. In each plot, two horizontal dashed lines denote two standard error limits of sample ACF.

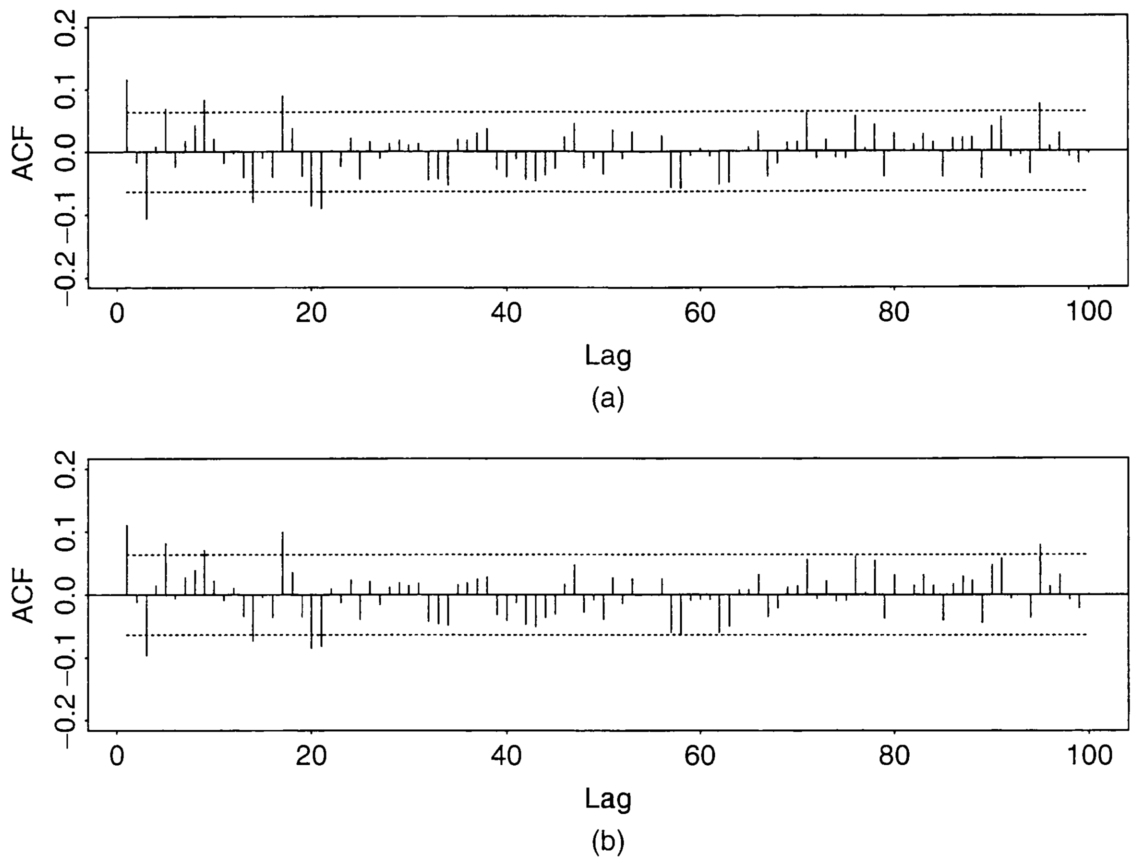

Figure 2.2 Sample autocorrelation functions of monthly (a) simple returns and (b) log returns of value-weighted index of U.S. markets from January 1926 to December 2008. In each plot, two horizontal dashed lines denote two standard error limits of sample ACF.

distributions with 5 and 10 degrees of freedom. For the log returns, we have $Q(5) = 3.52$ and $Q(10) = 13.39$ with p values 0.62 and 0.20, respectively. The joint tests confirm that monthly IBM stock returns have no significant serial correlations. Figure 2.2 shows the same for the monthly returns of the value-weighted index from the Center for Research in Security Prices (CRSP), at the University of Chicago. There are some significant serial correlations at the 5% level for both return series. The Ljung–Box statistics give $Q(5) = 29.71$ and $Q(10) = 39.55$ for the simple returns and $Q(5) = 28.38$ and $Q(10) = 36.16$ for the log returns. The p values of these four test statistics are all less than 0.0001, suggesting that monthly returns of the value-weighted index are serially correlated. Thus, the monthly market index return seems to have stronger serial dependence than individual stock returns.

In the finance literature, a version of the capital asset pricing model (CAPM) theory is that the return $\{r_t\}$ of an asset is not predictable and should have no autocorrelations. Testing for zero autocorrelations has been used as a tool to check the efficient market assumption. However, the way by which stock prices are determined and index returns are calculated might introduce autocorrelations in the observed return series. This is particularly so in analysis of high-frequency financial data. We discuss some of these issues, such as bid–ask bounce and nonsynchronous trading, in Chapter 5.

R Demonstration

The following output has been edited and % denotes explanation:

```
> da=read.table("m-ibm3dx2608.txt",header=T) % Load data
> da[1,]    % Check the 1st row of the data
      date       rtn     vwrtn    ewrtn    sprtn
1 19260130 -0.010381 0.000724 0.023174 0.022472
> sibm=da[,2] % Get the IBM simple returns
> Box.test(sibm,lag=5,type='Ljung') % Ljung-Box statistic Q(5)
      Box-Ljung test

data: sibm
X-squared = 3.3682, df = 5, p-value = 0.6434

> libm=log(sibm+1)  % Log IBM returns
> Box.test(libm,lag=5,type='Ljung')

      Box-Ljung test

data: libm
X-squared = 3.5236, df = 5, p-value = 0.6198
```

S-Plus Demonstration

Output edited.

```
> module(finmetrics)
> da=read.table("m-ibm3dx2608.txt",header=T) % Load data
> da[1,]    % Check the 1st row of the data
      date       rtn      vwrtn ewrtn    sprtn
1 19260130 -0.010381 0.000724 0.023174 0.022472
> sibm=da[,2] % Get IBM simple returns
> autocorTest(sibm,lag=5) % Ljung-Box Q(5) test

Test for Autocorrelation: Ljung-Box
Null Hypothesis: no autocorrelation
Test Statistics:
Test Stat 3.3682
  p.value 0.6434

Dist. under Null: chi-square with 5 degrees of freedom
   Total Observ.: 996
> libm=log(sibm+1) % IBM log returns
> autocorTest(libm,lag=5)

Test for Autocorrelation: Ljung-Box
Null Hypothesis: no autocorrelation
Test Statistics:
Test Stat 3.5236
  p.value 0.6198
```

2.3 WHITE NOISE AND LINEAR TIME SERIES

White Noise

A time series r_t is called a white noise if $\{r_t\}$ is a sequence of independent and identically distributed random variables with finite mean and variance. In particular, if r_t is normally distributed with mean zero and variance σ^2, the series is called a Gaussian white noise. For a white noise series, all the ACFs are zero. In practice, if all sample ACFs are close to zero, then the series is a white noise series. Based on Figures 2.1 and 2.2, the monthly returns of IBM stock are close to white noise, whereas those of the value-weighted index are not.

The behavior of sample autocorrelations of the value-weighted index returns indicates that for some asset returns it is necessary to model the serial dependence before further analysis can be made. In what follows, we discuss some simple time series models that are useful in modeling the dynamic structure of a time series. The concepts presented are also useful later in modeling volatility of asset returns.

Linear Time Series

A time series r_t is said to be linear if it can be written as

$$r_t = \mu + \sum_{i=0}^{\infty} \psi_i a_{t-i}, \tag{2.4}$$

where μ is the mean of r_t, $\psi_0 = 1$, and $\{a_t\}$ is a sequence of iid random variables with mean zero and a well-defined distribution (i.e., $\{a_t\}$ is a white noise series). It will be seen later that a_t denotes the new information at time t of the time series and is often referred to as the *innovation* or *shock* at time t. In this book, we are mainly concerned with the case where a_t is a continuous random variable. Not all financial time series are linear, however. We study nonlinearity and nonlinear models in Chapter 4.

For a linear time series in Eq. (2.4), the dynamic structure of r_t is governed by the coefficients ψ_i, which are called the ψ *weights* of r_t in the time series literature. If r_t is weakly stationary, we can obtain its mean and variance easily by using the independence of $\{a_t\}$ as

$$E(r_t) = \mu, \qquad \mathrm{Var}(r_t) = \sigma_a^2 \sum_{i=0}^{\infty} \psi_i^2, \tag{2.5}$$

where σ_a^2 is the variance of a_t. Because $\mathrm{Var}(r_t) < \infty$, $\{\psi_i^2\}$ must be a convergent sequence, that is, $\psi_i^2 \to 0$ as $i \to \infty$. Consequently, for a stationary series, impact of the remote shock a_{t-i} on the return r_t vanishes as i increases.

The lag-ℓ autocovariance of r_t is

$$\gamma_\ell = \mathrm{Cov}(r_t, r_{t-\ell}) = E\left[\left(\sum_{i=0}^{\infty} \psi_i a_{t-i}\right)\left(\sum_{j=0}^{\infty} \psi_j a_{t-\ell-j}\right)\right]$$

$$= E\left(\sum_{i,j=0}^{\infty} \psi_i \psi_j a_{t-i} a_{t-\ell-j}\right)$$

$$= \sum_{j=0}^{\infty} \psi_{j+\ell} \psi_j E(a_{t-\ell-j}^2) = \sigma_a^2 \sum_{j=0}^{\infty} \psi_j \psi_{j+\ell}. \tag{2.6}$$

Consequently, the ψ weights are related to the autocorrelations of r_t as follows:

$$\rho_\ell = \frac{\gamma_\ell}{\gamma_0} = \frac{\sum_{i=0}^{\infty} \psi_i \psi_{i+\ell}}{1 + \sum_{i=1}^{\infty} \psi_i^2}, \qquad \ell \geq 0, \tag{2.7}$$

where $\psi_0 = 1$. Linear time series models are econometric and statistical models used to describe the pattern of the ψ weights of r_t. For a weakly stationary time series, $\psi_i \to 0$ as $i \to \infty$ and, hence, ρ_ℓ converges to zero as ℓ increases. For asset returns, this means that, as expected, the linear dependence of current return r_t on the remote past return $r_{t-\ell}$ diminishes for large ℓ.

2.4 SIMPLE AR MODELS

The fact that the monthly return r_t of CRSP value-weighted index has a statistically significant lag-1 autocorrelation indicates that the lagged return r_{t-1} might be useful in predicting r_t. A simple model that makes use of such predictive power is

$$r_t = \phi_0 + \phi_1 r_{t-1} + a_t, \tag{2.8}$$

where $\{a_t\}$ is assumed to be a white noise series with mean zero and variance σ_a^2. This model is in the same form as the well-known simple linear regression model in which r_t is the dependent variable and r_{t-1} is the explanatory variable. In the time series literature, model (2.8) is referred to as an autoregressive (AR) model of order 1 or simply an AR(1) model. This simple model is also widely used in stochastic volatility modeling when r_t is replaced by its log volatility; see Chapters 3 and 12.

The AR(1) model in Eq. (2.8) has several properties similar to those of the simple linear regression model. However, there are some significant differences between the two models, which we discuss later. Here it suffices to note that an AR(1) model implies that, conditional on the past return r_{t-1}, we have

$$E(r_t | r_{t-1}) = \phi_0 + \phi_1 r_{t-1}, \qquad \text{Var}(r_t | r_{t-1}) = \text{Var}(a_t) = \sigma_a^2.$$

That is, given the past return r_{t-1}, the current return is centered around $\phi_0 + \phi_1 r_{t-1}$ with standard deviation σ_a. This is a Markov property such that conditional on r_{t-1}, the return r_t is not correlated with r_{t-i} for $i > 1$. Obviously, there are situations in which r_{t-1} alone cannot determine the conditional expectation of r_t and a more flexible model must be sought. A straightforward generalization of the AR(1) model is the AR(p) model:

$$r_t = \phi_0 + \phi_1 r_{t-1} + \cdots + \phi_p r_{t-p} + a_t, \tag{2.9}$$

where p is a nonnegative integer and $\{a_t\}$ is defined in Eq. (2.8). This model says that the past p variables r_{t-i} ($i = 1, \ldots, p$) jointly determine the conditional expectation of r_t given the past data. The AR(p) model is in the same form as a multiple linear regression model with lagged values serving as the explanatory variables.

2.4.1 Properties of AR Models

For effective use of AR models, it pays to study their basic properties. We discuss properties of AR(1) and AR(2) models in detail and give the results for the general AR(p) model.

AR(1) Model

We begin with the sufficient and necessary condition for weak stationarity of the AR(1) model in Eq. (2.8). Assuming that the series is weakly stationary, we have $E(r_t) = \mu$, $\text{Var}(r_t) = \gamma_0$, and $\text{Cov}(r_t, r_{t-j}) = \gamma_j$, where μ and γ_0 are constant and γ_j is a function of j, not t. We can easily obtain the mean, variance, and autocorrelations of the series as follows. Taking the expectation of Eq. (2.8) and because $E(a_t) = 0$, we obtain

$$E(r_t) = \phi_0 + \phi_1 E(r_{t-1}).$$

Under the stationarity condition, $E(r_t) = E(r_{t-1}) = \mu$ and hence

$$\mu = \phi_0 + \phi_1 \mu \qquad \text{or} \qquad E(r_t) = \mu = \frac{\phi_0}{1 - \phi_1}.$$

This result has two implications for r_t. First, the mean of r_t exists if $\phi_1 \neq 1$. Second, the mean of r_t is zero if and only if $\phi_0 = 0$. Thus, for a stationary AR(1) process, the constant term ϕ_0 is related to the mean of r_t via $\phi_0 = (1 - \phi_1)\mu$ and $\phi_0 = 0$ implies that $E(r_t) = 0$.

Next, using $\phi_0 = (1 - \phi_1)\mu$, the AR(1) model can be rewritten as

$$r_t - \mu = \phi_1 (r_{t-1} - \mu) + a_t. \tag{2.10}$$

By repeated substitutions, the prior equation implies that

$$\begin{aligned} r_t - \mu &= a_t + \phi_1 a_{t-1} + \phi_1^2 a_{t-2} + \cdots \\ &= \sum_{i=0}^{\infty} \phi_1^i a_{t-i}. \end{aligned} \tag{2.11}$$

This equation expresses an AR(1) model in the form of Eq. (2.4) with $\psi_i = \phi_1^i$. Thus, $r_t - \mu$ is a linear function of a_{t-i} for $i \geq 0$. Using this property and the independence of the series $\{a_t\}$, we obtain $E[(r_t - \mu)a_{t+1}] = 0$. By the stationarity assumption, we have $\text{Cov}(r_{t-1}, a_t) = E[(r_{t-1} - \mu)a_t] = 0$. This latter result can also be seen from the fact that r_{t-1} occurred before time t and a_t does not depend on any past information. Taking the square, then the expectation of Eq. (2.10), we obtain

$$\text{Var}(r_t) = \phi_1^2 \text{Var}(r_{t-1}) + \sigma_a^2,$$

where σ_a^2 is the variance of a_t, and we make use of the fact that the covariance between r_{t-1} and a_t is zero. Under the stationarity assumption, $\text{Var}(r_t) = \text{Var}(r_{t-1})$, so that

$$\text{Var}(r_t) = \frac{\sigma_a^2}{1 - \phi_1^2}$$

provided that $\phi_1^2 < 1$. The requirement of $\phi_1^2 < 1$ results from the fact that the variance of a random variable is bounded and nonnegative. Consequently, the weak stationarity of an AR(1) model implies that $-1 < \phi_1 < 1$, that is, $|\phi_1| < 1$. Yet if $|\phi_1| < 1$, then by Eq. (2.11) and the independence of the $\{a_t\}$ series, we can show that the mean and variance of r_t are finite and time invariant; see Eq. (2.5). In addition, by Eq. (2.6), all the autocovariances of r_t are finite. Therefore, the AR(1) model is weakly stationary. In summary, the necessary and sufficient condition for the AR(1) model in Eq. (2.8) to be weakly stationary is $|\phi_1| < 1$.

Using $\phi_0 = (1 - \phi_1)\mu$, one can rewrite a stationary AR(1) model as

$$r_t = (1 - \phi_1)\mu + \phi_1 r_{t-1} + a_t.$$

This model is often used in the finance literature with ϕ_1 measuring the persistence of the dynamic dependence of an AR(1) time series.

Autocorrelation Function of an AR(1) Model

Multiplying Eq. (2.10) by a_t, using the independence between a_t and r_{t-1}, and taking expectation, we obtain

$$E[a_t(r_t - \mu)] = \phi_1 E[a_t(r_{t-1} - \mu)] + E(a_t^2) = E(a_t^2) = \sigma_a^2,$$

where σ_a^2 is the variance of a_t. Multiplying Eq. (2.10) by $r_{t-\ell} - \mu$, taking expectation, and using the prior result, we have

$$\gamma_\ell = \begin{cases} \phi_1 \gamma_1 + \sigma_a^2 & \text{if } \ell = 0 \\ \phi_1 \gamma_{\ell-1} & \text{if } \ell > 0, \end{cases}$$

where we use $\gamma_\ell = \gamma_{-\ell}$. Consequently, for a weakly stationary AR(1) model in Eq. (2.8), we have

$$\mathrm{Var}(r_t) = \gamma_0 = \frac{\sigma^2}{1-\phi_1^2} \quad \text{and} \quad \gamma_\ell = \phi_1 \gamma_{\ell-1}, \quad \text{for} \quad \ell > 0.$$

From the latter equation, the ACF of r_t satisfies

$$\rho_\ell = \phi_1 \rho_{\ell-1}, \quad \text{for} \quad \ell > 0.$$

Because $\rho_0 = 1$, we have $\rho_\ell = \phi_1^\ell$. This result says that the ACF of a weakly stationary AR(1) series decays exponentially with rate ϕ_1 and starting value $\rho_0 = 1$. For a positive ϕ_1, the plot of ACF of an AR(1) model shows a nice exponential decay. For a negative ϕ_1, the plot consists of two alternating exponential decays with rate ϕ_1^2. Figure 2.3 shows the ACF of two AR(1) models with $\phi_1 = 0.8$ and $\phi_1 = -0.8$.

AR(2) Model

An AR(2) model assumes the form

$$r_t = \phi_0 + \phi_1 r_{t-1} + \phi_2 r_{t-2} + a_t. \tag{2.12}$$

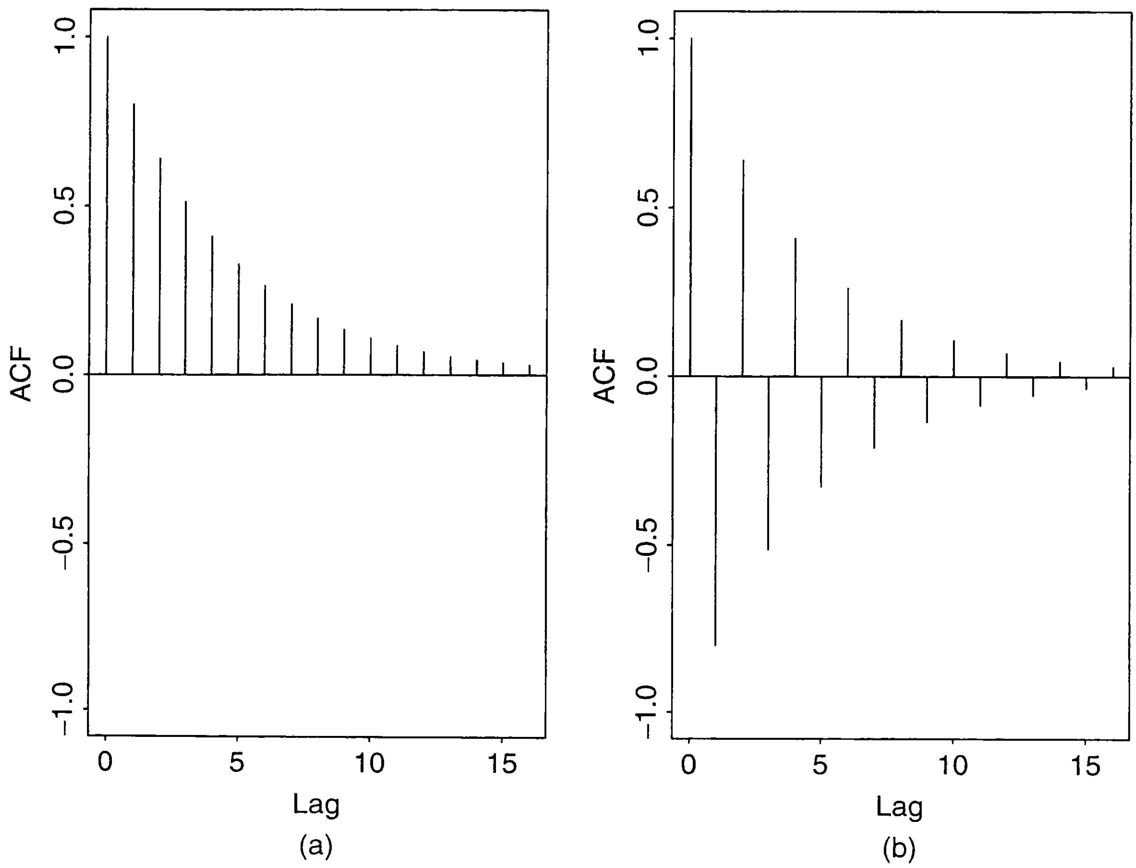

Figure 2.3 Autocorrelation function of an AR(1) model: (a) for $\phi_1 = 0.8$ and (b) for $\phi_1 = -0.8$.

Using the same technique as that of the AR(1) case, we obtain

$$E(r_t) = \mu = \frac{\phi_0}{1 - \phi_1 - \phi_2}$$

provided that $\phi_1 + \phi_2 \neq 1$. Using $\phi_0 = (1 - \phi_1 - \phi_2)\mu$, we can rewrite the AR(2) model as

$$(r_t - \mu) = \phi_1(r_{t-1} - \mu) + \phi_2(r_{t-2} - \mu) + a_t.$$

Multiplying the prior equation by $(r_{t-\ell} - \mu)$, we have

$$\begin{aligned}(r_{t-\ell} - \mu)(r_t - \mu) &= \phi_1(r_{t-\ell} - \mu)(r_{t-1} - \mu) \\ &\quad + \phi_2(r_{t-\ell} - \mu)(r_{t-2} - \mu) + (r_{t-\ell} - \mu)a_t.\end{aligned}$$

Taking expectation and using $E[(r_{t-\ell} - \mu)a_t] = 0$ for $\ell > 0$, we obtain

$$\gamma_\ell = \phi_1\gamma_{\ell-1} + \phi_2\gamma_{\ell-2}, \quad \text{for} \quad \ell > 0.$$

This result is referred to as the *moment equation* of a stationary AR(2) model. Dividing the above equation by γ_0, we have the property

$$\rho_\ell = \phi_1\rho_{\ell-1} + \phi_2\rho_{\ell-2}, \quad \text{for} \quad \ell > 0, \tag{2.13}$$

for the ACF of r_t. In particular, the lag-1 ACF satisfies

$$\rho_1 = \phi_1\rho_0 + \phi_2\rho_{-1} = \phi_1 + \phi_2\rho_1.$$

Therefore, for a stationary AR(2) series r_t, we have $\rho_0 = 1$,

$$\begin{aligned}\rho_1 &= \frac{\phi_1}{1 - \phi_2} \\ \rho_\ell &= \phi_1\rho_{\ell-1} + \phi_2\rho_{\ell-2}, \quad \ell \geq 2.\end{aligned}$$

The result of Eq. (2.13) says that the ACF of a stationary AR(2) series satisfies the second-order difference equation

$$(1 - \phi_1 B - \phi_2 B^2)\rho_\ell = 0,$$

where B is called the *back-shift* operator such that $B\rho_\ell = \rho_{\ell-1}$. This difference equation determines the properties of the ACF of a stationary AR(2) time series. It also determines the behavior of the forecasts of r_t. In the time series literature, some people use the notation L instead of B for the back-shift operator. Here L stands for *lag* operator. For instance, $Lr_t = r_{t-1}$ and $L\psi_k = \psi_{k-1}$.

Corresponding to the prior difference equation, there is a second-order polynomial equation:

$$1 - \phi_1 x - \phi_2 x^2 = 0. \tag{2.14}$$

Solutions of this equation are

$$x = \frac{\phi_1 \pm \sqrt{\phi_1^2 + 4\phi_2}}{-2\phi_2}.$$

In the time series literature, inverses of the two solutions are referred to as the *characteristic roots* of the AR(2) model. Denote the two characteristic roots by ω_1 and ω_2. If both ω_i are real valued, then the second-order difference equation of the model can be factored as $(1 - \omega_1 B)(1 - \omega_2 B)$ and the AR(2) model can be regarded as an AR(1) model operates on top of another AR(1) model. The ACF of r_t is then a mixture of two exponential decays. If $\phi_1^2 + 4\phi_2 < 0$, then ω_1 and ω_2 are complex numbers (called a *complex-conjugate pair*), and the plot of ACF of r_t would show a picture of damping sine and cosine waves. In business and economic applications, complex characteristic roots are important. They give rise to the behavior of business cycles. It is then common for economic time series models to have complex-valued characteristic roots. For an AR(2) model in Eq. (2.12) with a pair of complex characteristic roots, the *average* length of the stochastic cycles is

$$k = \frac{2\pi}{\cos^{-1}[\phi_1/(2\sqrt{-\phi_2})]},$$

where the cosine inverse is stated in radians. If one writes the complex solutions as $a \pm bi$, where $i = \sqrt{-1}$, then we have $\phi_1 = 2a$, $\phi_2 = -(a^2 + b^2)$, and

$$k = \frac{2\pi}{\cos^{-1}(a/\sqrt{a^2 + b^2})},$$

where $\sqrt{a^2 + b^2}$ is the absolute value of $a \pm bi$. See Example 2.1 for an illustration.

Figure 2.4 shows the ACF of four stationary AR(2) models. Part (b) is the ACF of the AR(2) model $(1 - 0.6B + 0.4B^2)r_t = a_t$. Because $\phi_1^2 + 4\phi_2 = 0.36 + 4 \times (-0.4) = -1.24 < 0$, this particular AR(2) model contains two complex characteristic roots, and hence its ACF exhibits damping sine and cosine waves. The other three AR(2) models have real-valued characteristic roots. Their ACFs decay exponentially.

Example 2.1. As an illustration, consider the quarterly growth rate of U.S. real gross national product (GNP), seasonally adjusted, from the second quarter of 1947 to the first quarter of 1991. This series shown in Figure 2.5 is also used in Chapter 4 as an example of nonlinear economic time series. Here we simply employ an AR(3) model for the data. Denoting the growth rate by r_t, we can use

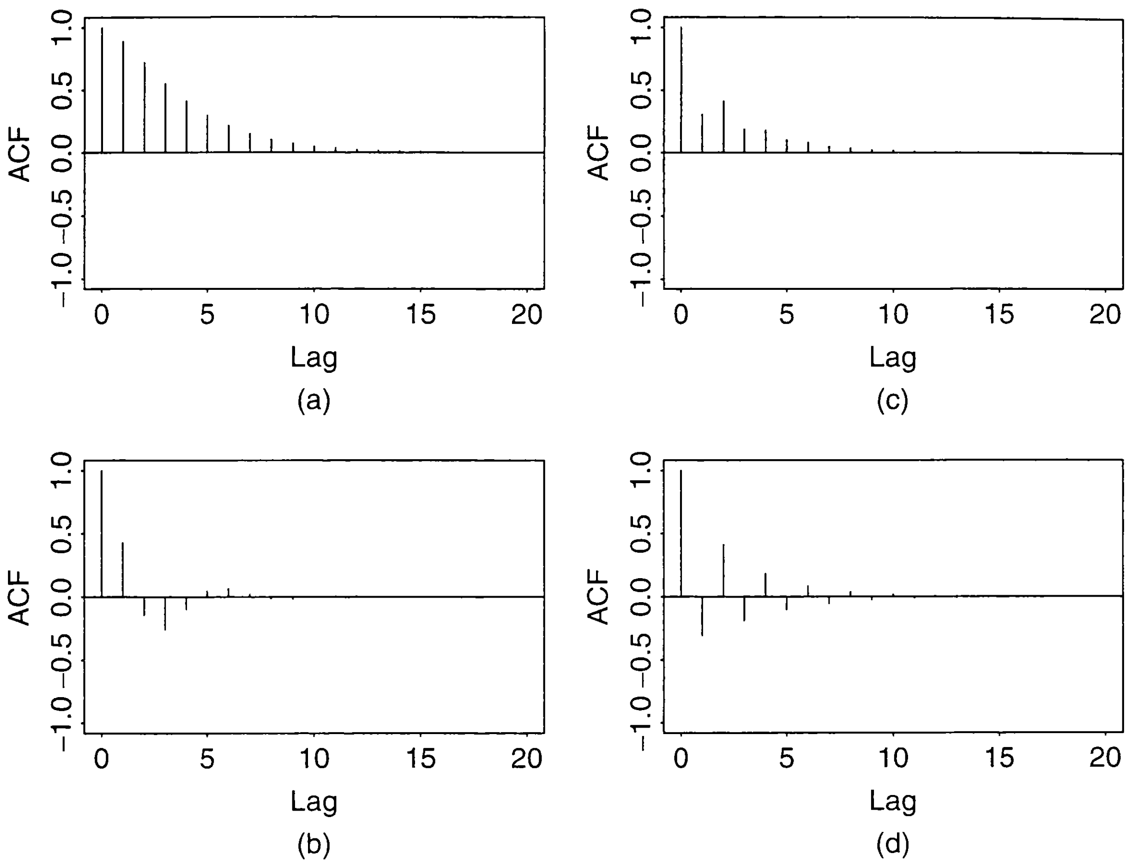

Figure 2.4 Autocorrelation function of an AR(2) model: (a) $\phi_1 = 1.2$ and $\phi_2 = -0.35$, (b) $\phi_1 = 0.6$ and $\phi_2 = -0.4$, (c) $\phi_1 = 0.2$ and $\phi_2 = 0.35$, and (d) $\phi_1 = -0.2$ and $\phi_2 = 0.35$.

the model building procedure of the next subsection to estimate the model. The fitted model is

$$r_t = 0.0047 + 0.348r_{t-1} + 0.179r_{t-2} - 0.142r_{t-3} + a_t, \qquad \hat{\sigma}_a = 0.0097. \quad (2.15)$$

Rewriting the model as

$$r_t - 0.348r_{t-1} - 0.179r_{t-2} + 0.142r_{t-3} = 0.0047 + a_t,$$

we obtain a corresponding third-order difference equation

$$1 - 0.348B - 0.179B^2 + 0.141B^3 = 0,$$

which can be factored approximately as

$$(1 + 0.521B)(1 - 0.869B + 0.274B^2) = 0.$$

The first factor $(1 + 0.521B)$ shows an exponentially decaying feature of the GNP growth rate. Focusing on the second-order factor $1 - 0.869B - (-0.274)B^2 = 0$, we have $\phi_1^2 + 4\phi_2 = 0.869^2 + 4(-0.274) = -0.341 < 0$. Therefore, the second factor of the AR(3) model confirms the existence of stochastic business cycles

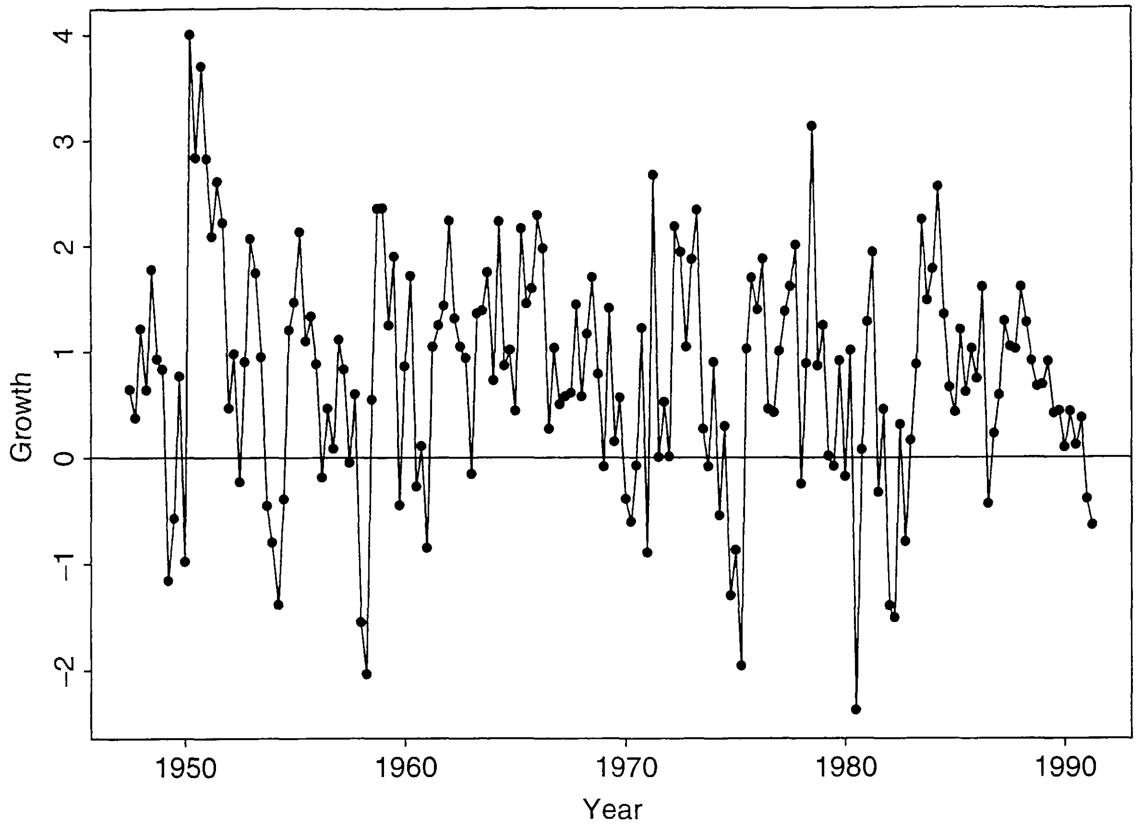

Figure 2.5 Time plot of growth rate of U.S. quarterly real GNP from 1947.II to 1991.I. Data are seasonally adjusted and in percentages.

in the quarterly growth rate of U.S. real GNP. This is reasonable as the U.S. economy went through expansion and contraction periods. The average length of the stochastic cycles is approximately

$$k = \frac{2(3.14159)}{\cos^{-1}[\phi_1/(2\sqrt{-\phi_2})]} = 10.62 \text{ quarters},$$

which is about 3 years. If one uses a nonlinear model to separate U.S. economy into "expansion" and "contraction" periods, the data show that the average duration of contraction periods is about three quarters and that of expansion periods is about 3 years; see the analysis in Chapter 4. The average duration of 10.62 quarters is a compromise between the two separate durations. The periodic feature obtained here is common among growth rates of national economies. For example, similar features can be found for many OECD (Organization for Economic Cooperation and Development) countries.

R Demonstration

The R demonstration for Example 2.1, where % denotes explanation, follows:

```
> gnp=scan(file='dgnp82.txt') % Load data
 % To create a time-series object
> gnp1=ts(gnp,frequency=4,start=c(1947,2))
> plot(gnp1)
```

```
> points(gnp1,pch='*')

> m1=ar(gnp,method=''mle'') % Find the AR order
> m1$order    % An AR(3) is selected based on AIC
[1] 3
> m2=arima(gnp,order=c(3,0,0)) % Estimation
> m2
Call:
arima(x = gnp, order = c(3, 0, 0))

Coefficients:
         ar1     ar2      ar3  intercept
      0.3480  0.1793  -0.1423     0.0077
s.e.  0.0745  0.0778   0.0745     0.0012

sigma^2 estimated as 9.427e-05: log likelihood=565.84,
   aic=-1121.68

% In R, ''intercept'' denotes the mean of the series.
% Therefore, the constant term is obtained below:
> (1-.348-.1793+.1423)*0.0077
[1] 0.0047355
> sqrt(m2$sigma2)  % Residual standard error
[1] 0.009709322

> p1=c(1,-m2$coef[1:3]) % Characteristic equation
> roots=polyroot(p1) % Find solutions
> roots
[1] 1.590253+1.063882i -1.920152+0.000000i 1.590253-1.063882i
> Mod(roots) % Compute the absolute values of the solutions
[1] 1.913308 1.920152 1.913308
 % To compute average length of business cycles:
> k=2*pi/acos(1.590253/1.913308)
> k
[1] 10.65638
```

Stationarity

The stationarity condition of an AR(2) time series is that the absolute values of its two characteristic roots are less than 1, that is, its two characteristic roots are less than 1 in modulus. Equivalently, the two solutions of the characteristic equation are greater than 1 in modulus. Under such a condition, the recursive equation in (2.13) ensures that the ACF of the model converges to 0 as the lag ℓ increases. This convergence property is a necessary condition for a stationary time series. In fact, the condition also applies to the AR(1) model where the polynomial equation is $1 - \phi_1 x = 0$. The characteristic root is $w = 1/x = \phi_1$, which must be less than 1 in modulus for r_t to be stationary. As shown before, $\rho_\ell = \phi_1^\ell$ for a stationary AR(1) model. The condition implies that $\rho_\ell \to 0$ as $\ell \to \infty$.

AR(p) Model

The results of the AR(1) and AR(2) models can readily be generalized to the general AR(p) model in Eq. (2.9). The mean of a stationary series is

$$E(r_t) = \frac{\phi_0}{1 - \phi_1 - \cdots - \phi_p}$$

provided that the denominator is not zero. The associated characteristic equation of the model is

$$1 - \phi_1 x - \phi_2 x^2 - \cdots - \phi_p x^p = 0.$$

If all the solutions of this equation are greater than 1 in modulus, then the series r_t is stationary. Again, inverses of the solutions are the *characteristic roots* of the model. Thus, stationarity requires that all characteristic roots are less than 1 in modulus. For a stationary AR(p) series, the ACF satisfies the difference equation

$$(1 - \phi_1 B - \phi_2 B^2 - \cdots - \phi_p B^p)\rho_\ell = 0, \qquad \text{for} \quad \ell > 0.$$

The plot of ACF of a stationary AR(p) model would then show a mixture of damping sine and cosine patterns and exponential decays depending on the nature of its characteristic roots.

2.4.2 Identifying AR Models in Practice

In application, the order p of an AR time series is unknown. It must be specified empirically. This is referred to as the *order determination* (or order specification) of AR models, and it has been extensively studied in the time series literature. Two general approaches are available for determining the value of p. The first approach is to use the partial autocorrelation function, and the second approach uses some information criteria.

***Partial Autocorrelation Function* (PACF)**

The PACF of a stationary time series is a function of its ACF and is a useful tool for determining the order p of an AR model. A simple, yet effective way to introduce PACF is to consider the following AR models in consecutive orders:

$$\begin{aligned}
r_t &= \phi_{0,1} + \phi_{1,1} r_{t-1} + e_{1t}, \\
r_t &= \phi_{0,2} + \phi_{1,2} r_{t-1} + \phi_{2,2} r_{t-2} + e_{2t}, \\
r_t &= \phi_{0,3} + \phi_{1,3} r_{t-1} + \phi_{2,3} r_{t-2} + \phi_{3,3} r_{t-3} + e_{3t}, \\
r_t &= \phi_{0,4} + \phi_{1,4} r_{t-1} + \phi_{2,4} r_{t-2} + \phi_{3,4} r_{t-3} + \phi_{4,4} r_{t-4} + e_{4t}, \\
&\;\;\vdots
\end{aligned}$$

where $\phi_{0,j}$, $\phi_{i,j}$, and $\{e_{jt}\}$ are, respectively, the constant term, the coefficient of r_{t-i}, and the error term of an AR(j) model. These models are in the form of a multiple linear regression and can be estimated by the least-squares method. As a matter of fact, they are arranged in a sequential order that enables us to apply the idea of partial F test in multiple linear regression analysis. The estimate $\hat{\phi}_{1,1}$ of the first equation is called the lag-1 sample PACF of r_t. The estimate $\hat{\phi}_{2,2}$ of the second equation is the lag-2 sample PACF of r_t. The estimate $\hat{\phi}_{3,3}$ of the third equation is the lag-3 sample PACF of r_t, and so on.

From the definition, the lag-2 PACF $\hat{\phi}_{2,2}$ shows the added contribution of r_{t-2} to r_t over the AR(1) model $r_t = \phi_0 + \phi_1 r_{t-1} + e_{1t}$. The lag-3 PACF shows the added contribution of r_{t-3} to r_t over an AR(2) model, and so on. Therefore, for an AR(p) model, the lag-p sample PACF should not be zero, but $\hat{\phi}_{j,j}$ should be close to zero for all $j > p$. We make use of this property to determine the order p. For a stationary Gaussian AR(p) model, it can be shown that the sample PACF has the following properties:

- $\hat{\phi}_{p,p}$ converges to ϕ_p as the sample size T goes to infinity.
- $\hat{\phi}_{\ell,\ell}$ converges to zero for all $\ell > p$.
- The asymptotic variance of $\hat{\phi}_{\ell,\ell}$ is $1/T$ for $\ell > p$.

These results say that, for an AR(p) series, the sample PACF cuts off at lag p.

As an example, consider the monthly simple returns of CRSP value-weighted index from January 1926 to December 2008. Table 2.1 gives the first 12 lags of sample PACF of the series. With $T = 996$, the asymptotic standard error of the sample PACF is approximately 0.032. Therefore, using the 5% significant level, we identify an AR(3) or AR(9) model for the data (i.e., $p = 3$ or 9). If the 1% significant level is used, we specify an AR(3) model.

As another example, Figure 2.6 shows the PACF of the GNP growth rate series of Example 2.1. The two dotted lines of the plot denote the approximate two standard error limits $\pm(2/\sqrt{176})$. The plot suggests an AR(3) model for the data because the first three lags of sample PACF appear to be large.

TABLE 2.1 Sample Partial Autocorrelation Function and Some Information Criteria for the Monthly Simple Returns of CRSP Value-Weighted Index from January 1926 to December 2008

p	1	2	3	4	5	6
PACF	0.115	−0.030	−0.102	0.033	0.062	−0.050
AIC	−5.838	−5.837	−5.846	−5.845	−5.847	−5.847
BIC	−5.833	−5.827	−5.831	−5.825	−5.822	−5.818
p	7	8	9	10	11	12
PACF	0.031	0.052	0.063	0.005	−0.005	0.011
AIC	−5.846	−5.847	−5.849	−5.847	−5.845	−5.843
BIC	−5.812	−5.807	−5.805	−5.798	−5.791	−5.784

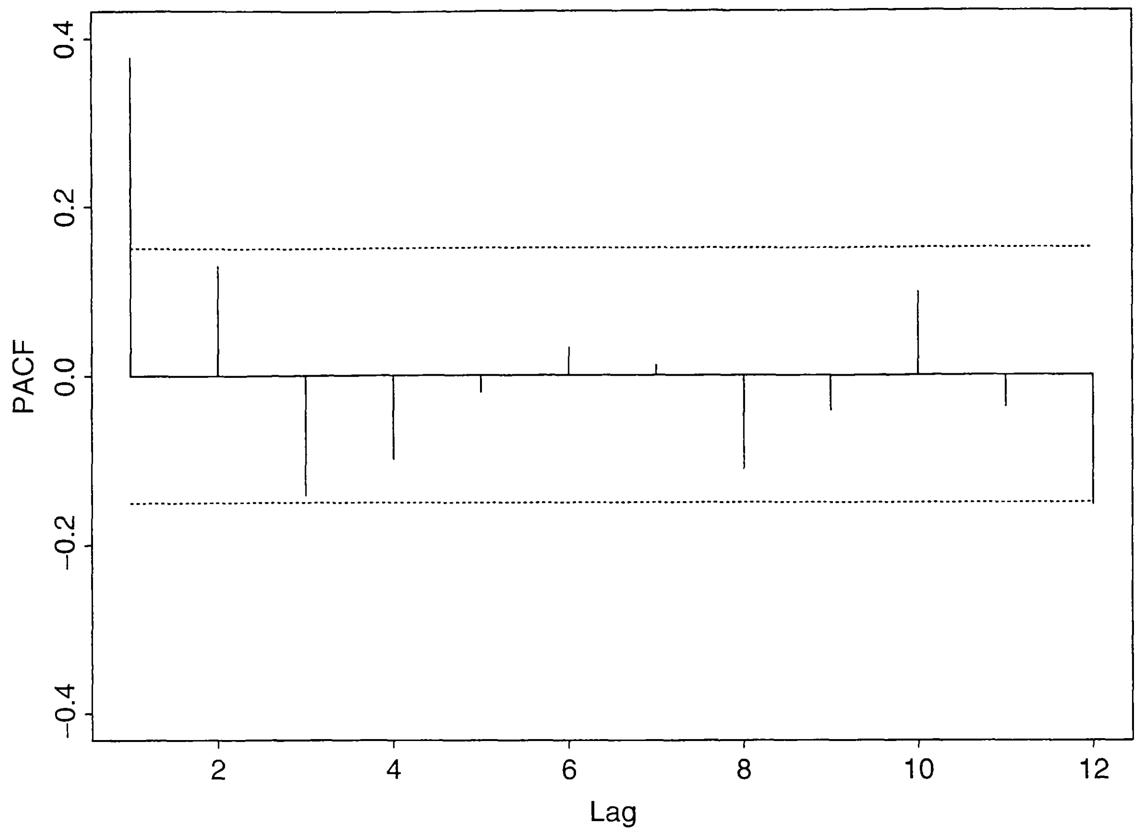

Figure 2.6 Sample partial autocorrelation function of U.S. quarterly real GNP growth rate from 1947.II to 1991.I. Dotted lines give approximate pointwise 95% confidence interval.

Information Criteria

There are several information criteria available to determine the order p of an AR process. All of them are likelihood based. For example, the well-known *Akaike information criterion* (AIC) (Akaike, 1973) is defined as

$$\mathrm{AIC} = \frac{-2}{T}\ln(\text{likelihood}) + \frac{2}{T} \times (\text{number of parameters}), \tag{2.16}$$

where the likelihood function is evaluated at the maximum-likelihood estimates and T is the sample size. For a Gaussian AR(ℓ) model, AIC reduces to

$$\mathrm{AIC}(\ell) = \ln(\tilde{\sigma}_\ell^2) + \frac{2\ell}{T}$$

where $\tilde{\sigma}_\ell^2$ is the maximum-likelihood estimate of σ_a^2, which is the variance of a_t, and T is the sample size; see Eq. (1.18). The first term of the AIC in Eq. (2.16) measures the goodness of fit of the AR(ℓ) model to the data, whereas the second term is called the *penalty function* of the criterion because it penalizes a candidate model by the number of parameters used. Different penalty functions result in different information criteria.

Another commonly used criterion function is the Schwarz–Bayesian information criterion (BIC). For a Gaussian AR(ℓ) model, the criterion is

$$\mathrm{BIC}(\ell) = \ln(\tilde{\sigma}_\ell^2) + \frac{\ell\ln(T)}{T}.$$

The penalty for each parameter used is 2 for AIC and $\ln(T)$ for BIC. Thus, compared with AIC, BIC tends to select a lower AR model when the sample size is moderate or large.

Selection Rule

To use AIC to select an AR model in practice, one computes AIC(ℓ) for $\ell = 0, \ldots, P$, where p is a prespecified positive integer and selects the order k that has the minimum AIC value. The same rule applies to BIC.

Table 2.1 also gives the AIC and BIC for $p = 1, \ldots, 12$. The AIC values are close to each other with minimum -5.849 occurring at $p = 9$, suggesting that an AR(9) model is preferred by the criterion. The BIC, on the other hand, attains its minimum value -5.833 at $p = 1$ with -5.831 as a close second at $p = 3$. Thus, the BIC selects an AR(1) model for the value-weighted return series. This example shows that different approaches or criteria to order determination may result in different choices of p. There is no evidence to suggest that one approach outperforms the other in a real application. Substantive information of the problem under study and simplicity are two factors that also play an important role in choosing an AR model for a given time series.

Again, consider the growth rate series of U.S. quarterly real GNP of Example 2.1. The AIC obtained from R also identifies an AR(3) model. Note that the AIC value of the `ar` command in R has been adjusted so that the minimum AIC is zero.

```
> gnp=scan(file='q-gnp4791.txt')
> ord=ar(gnp,method=''mle'')
> ord$aic
 [1] 27.847 2.742 1.603 0.000 0.323 2.243
 [7]  4.052 6.025 5.905 7.572 7.895 9.679
> ord$order
[1] 3
```

Parameter Estimation

For a specified AR(p) model in Eq. (2.9), the conditional least-squares method, which starts with the $(p + 1)$th observation, is often used to estimate the parameters. Specifically, conditioning on the first p observations, we have

$$r_t = \phi_0 + \phi_1 r_{t-1} + \cdots + \phi_p r_{t-p} + a_t, \qquad t = p + 1, \ldots, T,$$

which is in the form of a multiple linear regression and can be estimated by the least-squares method. Denote the estimate of ϕ_i by $\hat{\phi}_i$. The *fitted model* is

$$\hat{r}_t = \hat{\phi}_0 + \hat{\phi}_1 r_{t-1} + \cdots + \hat{\phi}_p r_{t-p},$$

and the associated residual is

$$\hat{a}_t = r_t - \hat{r}_t.$$

The series $\{\hat{a}_t\}$ is called the *residual series*, from which we obtain

$$\hat{\sigma}_a^2 = \frac{\sum_{t=p+1}^{T} \hat{a}_t^2}{T - 2p - 1}.$$

If the conditional-likelihood method is used, the estimates of ϕ_i remain unchanged, but the estimate of σ_a^2 becomes $\tilde{\sigma}_a^2 = \hat{\sigma}_a^2 \times (T - 2p - 1)/(T - p)$. In some packages, $\tilde{\sigma}_a^2$ is defined as $\hat{\sigma}_a^2 \times (T - 2p - 1)/T$. For illustration, consider an AR(3) model for the monthly simple returns of the value-weighted index in Table 2.1. The fitted model is

$$r_t = 0.0091 + 0.116 r_{t-1} - 0.019 r_{t-2} - 0.104 r_{t-3} + \hat{a}_t, \qquad \hat{\sigma}_a = 0.054.$$

The standard errors of the coefficients are 0.002, 0.032, 0.032, and 0.032, respectively. Except for the lag-2 coefficient, all parameters are statistically significant at the 1% level.

For this example, the AR coefficients of the fitted model are small, indicating that the serial dependence of the series is weak, even though it is statistically significant at the 1% level. The significance of $\hat{\phi}_0$ of the entertained model implies that the expected mean return of the series is positive. In fact, $\hat{\mu} = 0.0091/(1 - 0.116 + 0.019 + 0.104) = 0.009$, which is small but has an important long-term implication. It implies that the long-term return of the index can be substantial. Using the multiperiod simple return defined in Chapter 1, the average annual simple gross return is $[\prod_{t=1}^{996}(1 + R_t)]^{12/996} - 1 \approx 0.093$. In other words, the monthly simple returns of the CRSP value-weighted index grew about 9.3% per annum from 1926 to 2008, supporting the common belief that equity market performs well in the long term. A one-dollar investment at the beginning of 1926 would be worth about \$1593 at the end of 2008.

```
> vw=read.table('m-ibm3dx.txt',header=T)[,3]
> t1=prod(vw+1)
> t1
[1] 1592.953
> t1^(12/996)-1
[1] 0.0929
```

Model Checking

A fitted model must be examined carefully to check for possible model inadequacy. If the model is adequate, then the residual series should behave as a white noise. The ACF and the Ljung–Box statistics in Eq. (2.3) of the residuals can be used to check the closeness of $\hat{a}_t$ to a white noise. For an AR(p) model, the Ljung–Box statistic $Q(m)$ follows asymptotically a chi-squared distribution with $m - g$ degrees of freedom, where g denotes the number of AR coefficients used in the model. The adjustment in the degrees of freedom is made based on the number of constraints added to the residuals $\hat{a}_t$ from fitting the AR(p) to an AR(0) model. If a fitted model is found to be inadequate, it must be refined. For instance, if some of the estimated AR coefficients are not significantly different from zero, then the model should be simplified by trying to remove those insignificant parameters. If residual ACF shows additional serial correlations, then the model should be extended to take care of those correlations.

Remark. Most time series packages do not adjust the degrees of freedom when applying the Ljung–Box statistics $Q(m)$ to a residual series. This is understandable when $m \leq g$. □

Consider the residual series of the fitted AR(3) model for the monthly value-weighted simple returns. We have $Q(12) = 16.35$ with a p value 0.060 based on its asymptotic chi-squared distribution with 9 degrees of freedom. Thus, the null hypothesis of no residual serial correlation in the first 12 lags is barely not rejected at the 5% level. However, since the lag-2 AR coefficient is not significant at the 5% level, one can refine the model as

$$r_t = 0.0088 + 0.114r_{t-1} - 0.106r_{t-3} + a_t, \qquad \hat{\sigma}_a = 0.0536,$$

where all the estimates are now significant at the 1% level. The residual series gives $Q(12) = 16.83$ with a p value 0.078 (based on χ^2_{10}). The model is adequate in modeling the dynamic linear dependence of the data.

R Demonstration

In the following R demonstration, % denotes an explanation:

```
> vw=read.table('m-ibm3dx2608.txt',header=T)[,3]
> m3=arima(vw,order=c(3,0,0))
> m3
Call:
arima(x = vw, order = c(3, 0, 0))

Coefficients:
         ar1      ar2      ar3   intercept
      0.1158  -0.0187  -0.1042      0.0089
s.e.  0.0315   0.0317   0.0317      0.0017

sigma^2 estimated as 0.002875: log likelihood=1500.86,
   aic=-2991.73

> (1-.1158+.0187+.1042)*mean(vw) % Compute
   the intercept phi(0).
[1] 0.00896761
> sqrt(m3$sigma2)  % Compute standard error of residuals
[1] 0.0536189

> Box.test(m3$residuals,lag=12,type='Ljung')
```

```
      Box-Ljung test

data: m3$residuals   % R uses 12 degrees of freedom

X-squared = 16.3525, df = 12, p-value = 0.1756

> pv=1-pchisq(16.35,9) % Compute p-value using 9 degrees
   of freedom
> pv
[1] 0.05992276
 % To fix the AR(2) coef to zero:
> m3=arima(vw,order=c(3,0,0),fixed=c(NA,0,NA,NA))
 % The subcommand 'fixed' is used to fix parameter values,
 % where NA denotes estimation and 0 means fixing the
    parameter to 0.
 % The ordering of the parameters can be found using m3$coef.
> m3
Call:
arima(x = vw, order = c(3, 0, 0), fixed = c(NA, 0, NA, NA))

Coefficients:
          ar1  ar2      ar3  intercept
       0.1136    0  -0.1063     0.0089
s.e.   0.0313    0   0.0315     0.0017

sigma^2 estimated as 0.002876: log likelihood=1500.69,
   aic=-2993.38
> (1-.1136+.1063)*.0089 % Compute phi(0)
[1] 0.00883503
> sqrt(m3$sigma2)  % Compute residual standard error
[1] 0.05362832

> Box.test(m3$residuals,lag=12,type='Ljung')

      Box-Ljung test

data: m3$residuals
X-squared = 16.8276, df = 12, p-value = 0.1562

> pv=1-pchisq(16.83,10)
> pv
[1] 0.0782113
```

S-Plus Demonstration

The following S-Plus output has been edited:

```
> vw=read.table('m-ibm3dx2608.txt',header=T)[,3]
> ar3=OLS(vw ar(3))
```

```
> summary(ar3)
Call:
OLS(formula = vw ~ ar(3))

Residuals:
     Min       1Q  Median      3Q     Max
 -0.2863  -0.0263  0.0034  0.0297  0.3689

Coefficients:
               Value  Std. Error t value Pr(>|t|)
(Intercept)   0.0091  0.0018      5.1653  0.0000
       lag1   0.1148  0.0316      3.6333  0.0003
       lag2  -0.0188  0.0318     -0.5894  0.5557
       lag3  -0.1043  0.0318     -3.2763  0.0011

Regression Diagnostics:
         R-Squared 0.0246
Adjusted R-Squared 0.0216
Durbin-Watson Stat 1.9913

Residual Diagnostics:
                 Stat  P-Value
Jarque-Bera 1656.3928   0.0000
  Ljung-Box   50.1279   0.0087

Residual standard error: 0.05375 on 989 degrees of freedom

> autocorTest(ar3$residuals,lag=12)

Test for Autocorrelation: Ljung-Box
Null Hypothesis: no autocorrelation

Test Statistics:
Test Stat 16.5668
  p.value  0.1666    % S-Plus uses 12 degrees of freedom

Dist. under Null: chi-square with 12 degrees of freedom
  Total Observ.: 993
> 1-pchisq(16.57,9) % Compute p-value with 9 degrees
   of freedom
[1] 0.05589128
```

2.4.3 Goodness of Fit

A commonly used statistic to measure *goodness of fit* of a stationary model is the R square (R^2) defined as

$$R^2 = 1 - \frac{\text{residual sum of squares}}{\text{total sum of squares}}.$$

For a stationary AR(p) time series model with T observations $\{r_t | t = 1, \ldots, T\}$, the measure becomes

$$R^2 = 1 - \frac{\sum_{t=p+1}^{T} \hat{a}_t^2}{\sum_{t=p+1}^{T} (r_t - \bar{r})^2},$$

where $\bar{r} = \sum_{t=p+1}^{T} r_t/(T - p)$. It is easy to show that $0 \le R^2 \le 1$. Typically, a larger R^2 indicates that the model provides a closer fit to the data. However, this is only true for a stationary time series. For the unit-root nonstationary series discussed later in this chapter, R^2 of an AR(1) fit converges to one when the sample size increases to infinity, regardless of the true underlying model of r_t.

For a given data set, it is well known that R^2 is a nondecreasing function of the number of parameters used. To overcome this weakness, an *adjusted* R^2 is proposed, which is defined as

$$\begin{aligned} \text{Adj} - R^2 &= 1 - \frac{\text{variance of residuals}}{\text{variance of } r_t} \\ &= 1 - \frac{\hat{\sigma}_a^2}{\hat{\sigma}_r^2}, \end{aligned}$$

where $\hat{\sigma}_r^2$ is the sample variance of r_t. This new measure takes into account the number of parameters used in the fitted model. However, it is no longer between 0 and 1.

2.4.4 Forecasting

Forecasting is an important application of time series analysis. For the AR(p) model in Eq. (2.9), suppose that we are at the time index h and are interested in forecasting $r_{h+\ell}$, where $\ell \ge 1$. The time index h is called the *forecast origin* and the positive integer ℓ is the *forecast horizon*. Let $\hat{r}_h(\ell)$ be the forecast of $r_{h+\ell}$ using the minimum squared error loss function. In other words, the forecast $\hat{r}_k(\ell)$ is chosen such that

$$E\{[r_{h+\ell} - \hat{r}_h(\ell)]^2 | F_h\} \le \min_g E[(r_{h+\ell} - g)^2 | F_h],$$

where g is a function of the information available at time h (inclusive), that is, a function of F_h. We referred to $\hat{r}_h(\ell)$ as the ℓ-step ahead forecast of r_t at the forecast origin h. Let F_h be the collection of information available at the forecast origin h.

1-Step-Ahead Forecast

From the AR(p) model, we have

$$r_{h+1} = \phi_0 + \phi_1 r_h + \cdots + \phi_p r_{h+1-p} + a_{h+1}.$$

Under the minimum squared error loss function, the point forecast of r_{h+1} given F_h is the conditional expectation

$$\hat{r}_h(1) = E(r_{h+1}|F_h) = \phi_0 + \sum_{i=1}^{p} \phi_i r_{h+1-i},$$

and the associated forecast error is

$$e_h(1) = r_{h+1} - \hat{r}_h(1) = a_{h+1}.$$

Consequently, the variance of the 1-step-ahead forecast error is $\text{Var}[e_h(1)] = \text{Var}(a_{h+1}) = \sigma_a^2$. If a_t is normally distributed, then a 95% 1-step-ahead interval forecast of r_{h+1} is $\hat{r}_h(1) \pm 1.96 \times \sigma_a$. For the linear model in Eq. (2.4), a_{t+1} is also the 1-step-ahead forecast error at the forecast origin t. In the econometric literature, a_{t+1} is referred to as the *shock* to the series at time $t+1$.

In practice, estimated parameters are often used to compute point and interval forecasts. This results in a *conditional forecast* because such a forecast does not take into consideration the uncertainty in the parameter estimates. In theory, one can consider parameter uncertainty in forecasting, but it is much more involved. A natural way to consider parameter and model uncertainty in forecasting is Bayesian forecasting with Markov chan Monte Carlo (MCMC) methods. See Chapter 12 for further discussion. For simplicity, we assume that the model is given in this chapter. When the sample size used in estimation is sufficiently large, then the conditional forecast is close to the unconditional one.

2-Step-Ahead Forecast

Next consider the forecast of r_{h+2} at the forecast origin h. From the AR(p) model, we have

$$r_{h+2} = \phi_0 + \phi_1 r_{h+1} + \cdots + \phi_p r_{h+2-p} + a_{h+2}.$$

Taking conditional expectation, we have

$$\hat{r}_h(2) = E(r_{h+2}|F_h) = \phi_0 + \phi_1 \hat{r}_h(1) + \phi_2 r_h + \cdots + \phi_p r_{h+2-p}$$

and the associated forecast error

$$e_h(2) = r_{h+2} - \hat{r}_h(2) = \phi_1[r_{h+1} - \hat{r}_h(1)] + a_{h+2} = a_{h+2} + \phi_1 a_{h+1}.$$

The variance of the forecast error is $\text{Var}[e_h(2)] = (1+\phi_1^2)\sigma_a^2$. Interval forecasts of r_{h+2} can be computed in the same way as those for r_{h+1}. It is interesting to see that $\text{Var}[e_h(2)] \geq \text{Var}[e_h(1)]$, meaning that as the forecast horizon increases the uncertainty in forecast also increases. This is in agreement with common sense that we are more uncertain about r_{h+2} than r_{h+1} at the time index h for a linear time series.

Multistep-Ahead Forecast

In general, we have

$$r_{h+\ell} = \phi_0 + \phi_1 r_{h+\ell-1} + \cdots + \phi_p r_{h+\ell-p} + a_{h+\ell}.$$

The ℓ-step-ahead forecast based on the minimum squared error loss function is the conditional expectation of $r_{h+\ell}$ given F_h, which can be obtained as

$$\hat{r}_h(\ell) = \phi_0 + \sum_{i=1}^{p} \phi_i \hat{r}_h(\ell - i),$$

where it is understood that $\hat{r}_h(i) = r_{h+i}$ if $i \leq 0$. This forecast can be computed recursively using forecasts $\hat{r}_h(i)$ for $i = 1, \ldots, \ell - 1$. The ℓ-step-ahead forecast error is $e_h(\ell) = r_{h+\ell} - \hat{r}_h(\ell)$. It can be shown that for a stationary AR(p) model, $\hat{r}_h(\ell)$ converges to $E(r_t)$ as $\ell \to \infty$, meaning that for such a series long-term point forecast approaches its unconditional mean. This property is referred to as the *mean reversion* in the finance literature. For an AR(1) model, the speed of mean reversion is measured by the *half-life* defined as $\ell = \ln(0.5)/\ln(|\phi_1|)$. The variance of the forecast error then approaches the unconditional variance of r_t. Note that for an AR(1) model in (2.8), let $x_t = r_t - E(r_t)$ be the mean-adjusted series. It is easy to see that the ℓ-step-ahead forecast of $x_{h+\ell}$ at the forecast orign h is $\hat{x}_h(\ell) = \phi_1^\ell x_h$. The half-life is the forecast horizon such that $\hat{x}_h(\ell) = \frac{1}{2} x_h$. That is, $\phi_1^\ell = \frac{1}{2}$. Thus, $\ell = \ln(0.5)/\ln(|\phi_1|)$.

Table 2.2 contains the 1-step- to 12-step ahead forecasts and the standard errors of the associated forecast errors at the forecast origin 984 for the monthly simple return of the value-weighted index using an AR(3) model that was reestimated using the first 984 observations. The fitted model is

$$r_t = 0.0098 + 0.1024 r_{t-1} - 0.0201 r_{t-2} - 0.1090 r_{t-3} + a_t,$$

where $\hat{\sigma}_a = 0.054$. The actual returns of 2008 are also given in Table 2.2. Because of the weak serial dependence in the series, the forecasts and standard deviations of forecast errors converge to the sample mean and standard deviation of the data quickly. For the first 984 observations, the sample mean and standard error are 0.0095 and 0.0540, respectively.

Figure 2.7 shows the corresponding out-of-sample prediction plot for the monthly simple return series of the value-weighted index. The forecast origin $t = 984$ corresponds to December 2007. The prediction plot includes the two standard error limits of the forecasts and the actual observed returns for 2008. The forecasts and actual returns are marked by $\circ$ and $\bullet$, respectively. From the plot, except for the return of October 2008, all actual returns are within the 95% prediction intervals.

TABLE 2.2 Multistep Ahead Forecasts of an AR(3) Model for Monthly Simple Returns of CRSP Value-Weighted Index

Step	1	2	3	4	5	6
Forecast	0.0076	0.0161	0.0118	0.0099	0.0089	0.0093
Std. Error	0.0534	0.0537	0.0537	0.0540	0.0540	0.0540
Actual	−0.0623	−0.0220	−0.0105	0.0511	0.0238	−0.0786
Step	7	8	9	10	11	12
Forecast	0.0095	0.0097	0.0096	0.0096	0.0096	0.0096
Std. Error	0.0540	0.0540	0.0540	0.0540	0.0540	0.0540
Actual	−0.0132	0.0110	−0.0981	−0.1847	−0.0852	0.0215

[a]The forecast origin is $h = 984$.

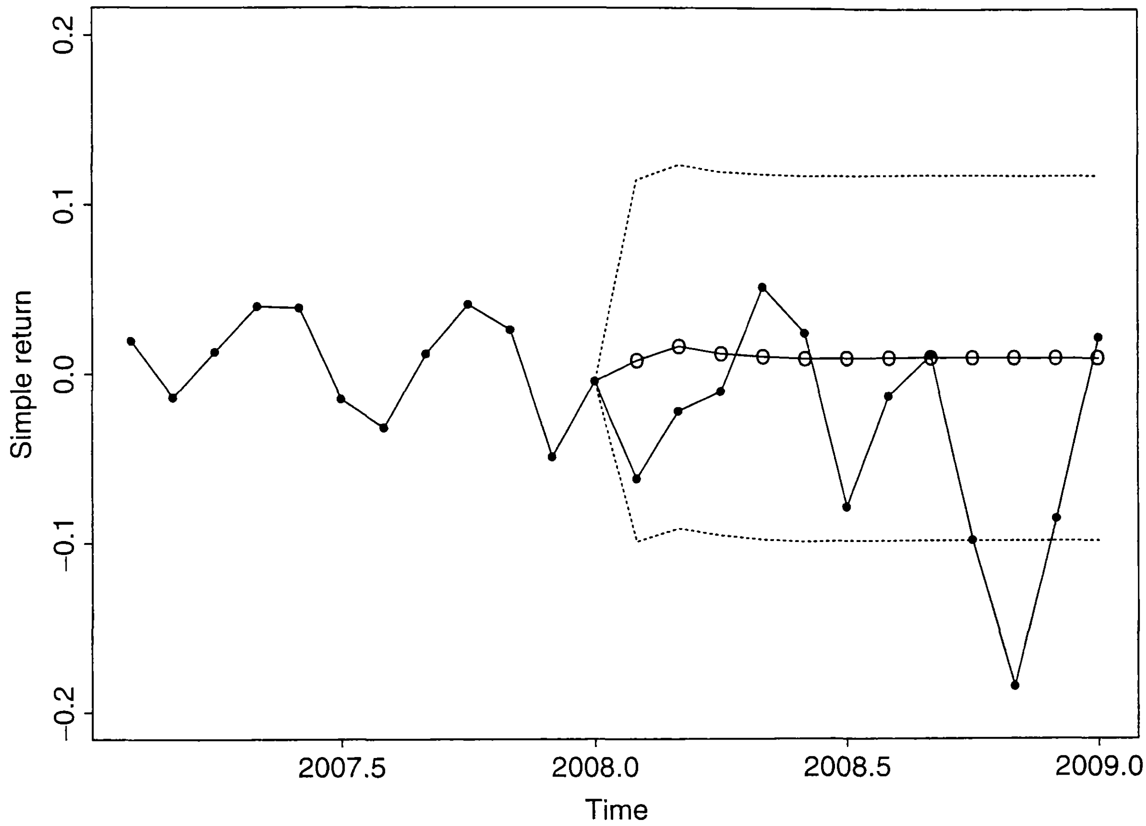

Figure 2.7 Plot of 1- to 12-step-ahead out-of-sample forecasts for monthly simple returns of CRSP value-weighted index. Forecast origin is $t = 984$, which is December 2007. Forecasts are denoted by "o" and actual observations by "•". Two dashed lines denote two standard error limits of the forecasts.

2.5 SIMPLE MA MODELS

We now turn to another class of simple models that are also useful in modeling return series in finance. These models are the moving-average (MA) models. As is shown in Chapter 5, the bid–ask bounce in stock trading may introduce an MA(1) structure in a return series. There are several ways to introduce MA models. One approach is to treat the model as a simple extension of white noise series. Another approach is to treat the model as an infinite-order AR model with some parameter constraints. We adopt the second approach.

There is no particular reason, but simplicity, to assume a priori that the order of an AR model is finite. We may entertain, at least in theory, an AR model with infinite order as

$$r_t = \phi_0 + \phi_1 r_{t-1} + \phi_2 r_{t-2} + \cdots + a_t.$$

However, such an AR model is not realistic because it has infinite many parameters. One way to make the model practical is to assume that the coefficients ϕ_i's satisfy some constraints so that they are determined by a finite number of parameters. A special case of this idea is

$$r_t = \phi_0 - \theta_1 r_{t-1} - \theta_1^2 r_{t-2} - \theta_1^3 r_{t-3} - \cdots + a_t, \tag{2.17}$$

where the coefficients depend on a single parameter θ_1 via $\phi_i = -\theta_1^i$ for $i \geq 1$. For the model in Eq. (2.17) to be stationary, θ_1 must be less than 1 in absolute value; otherwise, θ_1^i and the series will explode. Because $|\theta_1| < 1$, we have $\theta_1^i \to 0$ as $i \to \infty$. Thus, the contribution of r_{t-i} to r_t decays exponentially as i increases. This is reasonable as the dependence of a stationary series r_t on its lagged value r_{t-i}, if any, should decay over time.

The model in Eq. (2.17) can be rewritten in a rather compact form. To see this, rewrite the model as

$$r_t + \theta_1 r_{t-1} + \theta_1^2 r_{t-2} + \cdots = \phi_0 + a_t. \tag{2.18}$$

The model for r_{t-1} is then

$$r_{t-1} + \theta_1 r_{t-2} + \theta_1^2 r_{t-3} + \cdots = \phi_0 + a_{t-1}. \tag{2.19}$$

Multiplying Eq. (2.19) by θ_1 and subtracting the result from Eq. (2.18), we obtain

$$r_t = \phi_0(1 - \theta_1) + a_t - \theta_1 a_{t-1},$$

which says that except for the constant term r_t is a weighted average of shocks a_t and a_{t-1}. Therefore, the model is called an MA model of order 1 or MA(1) model for short. The general form of an MA(1) model is

$$r_t = c_0 + a_t - \theta_1 a_{t-1} \quad \text{or} \quad r_t = c_0 + (1 - \theta_1 B)a_t, \tag{2.20}$$

where c_0 is a constant and $\{a_t\}$ is a white noise series. Similarly, an MA(2) model is in the form

$$r_t = c_0 + a_t - \theta_1 a_{t-1} - \theta_2 a_{t-2}, \tag{2.21}$$

and an MA(q) model is

$$r_t = c_0 + a_t - \theta_1 a_{t-1} - \cdots - \theta_q a_{t-q}, \tag{2.22}$$

or $r_t = c_0 + (1 - \theta_1 B - \cdots - \theta_q B^q)a_t$, where $q > 0$.

2.5.1 Properties of MA Models

Again, we focus on the simple MA(1) and MA(2) models. The results of MA(q) models can easily be obtained by the same techniques.

Stationarity

Moving-average models are always weakly stationary because they are finite linear combinations of a white noise sequence for which the first two moments are time invariant. For example, consider the MA(1) model in Eq. (2.20). Taking expectation of the model, we have

$$E(r_t) = c_0,$$

which is time invariant. Taking the variance of Eq. (2.20), we have

$$\text{Var}(r_t) = \sigma_a^2 + \theta_1^2\sigma_a^2 = (1+\theta_1^2)\sigma_a^2,$$

where we use the fact that a_t and a_{t-1} are uncorrelated. Again, Var(r_t) is time invariant. The prior discussion applies to general MA(q) models, and we obtain two general properties. First, the constant term of an MA model is the mean of the series [i.e., $E(r_t) = c_0$]. Second, the variance of an MA(q) model is

$$\text{Var}(r_t) = (1+\theta_1^2+\theta_2^2+\cdots+\theta_q^2)\sigma_a^2.$$

Autocorrelation Function

Assume for simplicity that $c_0 = 0$ for an MA(1) model. Multiplying the model by $r_{t-\ell}$, we have

$$r_{t-\ell}r_t = r_{t-\ell}a_t - \theta_1 r_{t-\ell}a_{t-1}.$$

Taking expectation, we obtain

$$\gamma_1 = -\theta_1\sigma_a^2 \qquad \text{and} \qquad \gamma_\ell = 0, \qquad \text{for} \quad \ell > 1.$$

Using the prior result and the fact that $\text{Var}(r_t) = (1+\theta_1^2)\sigma_a^2$, we have

$$\rho_0 = 1, \qquad \rho_1 = \frac{-\theta_1}{1+\theta_1^2}, \qquad \rho_\ell = 0, \qquad \text{for} \qquad \ell > 1.$$

Thus, for an MA(1) model, the lag-1 ACF is not zero, but all higher order ACFs are zero. In other words, the ACF of an MA(1) model cuts off at lag 1. For the MA(2) model in Eq. (2.21), the autocorrelation coefficients are

$$\rho_1 = \frac{-\theta_1+\theta_1\theta_2}{1+\theta_1^2+\theta_2^2}, \qquad \rho_2 = \frac{-\theta_2}{1+\theta_1^2+\theta_2^2}, \qquad \rho_\ell = 0, \quad \text{for} \quad \ell > 2.$$

Here the ACF cuts off at lag 2. This property generalizes to other MA models. For an MA(q) model, the lag-q ACF is not zero, but $\rho_\ell = 0$ for $\ell > q$. Consequently, an MA(q) series is only linearly related to its first q-lagged values and hence is a "finite-memory" model.

Invertibility

Rewriting a zero-mean MA(1) model as $a_t = r_t + \theta_1 a_{t-1}$, one can use repeated substitutions to obtain

$$a_t = r_t + \theta_1 r_{t-1} + \theta_1^2 r_{t-2} + \theta_1^3 r_{t-3} + \cdots.$$

This equation expresses the current shock a_t as a linear combination of the present and past returns. Intuitively, θ_1^j should go to zero as j increases because the remote return r_{t-j} should have very little impact on the current shock, if any. Consequently, for an MA(1) model to be plausible, we require $|\theta_1| < 1$. Such an MA(1) model is said to be *invertible*. If $|\theta_1| = 1$, then the MA(1) model is noninvertible. See Section 2.6.5 for further discussion on invertibility.

2.5.2 Identifying MA Order

The ACF is useful in identifying the order of an MA model. For a time series r_t with ACF ρ_ℓ, if $\rho_q \neq 0$, but $\rho_\ell = 0$ for $\ell > q$, then r_t follows an MA(q) model.

Figure 2.8 shows the time plot of monthly simple returns of the CRSP equal-weighted index from January 1926 to December 2008 and the sample ACF of the series. The two dashed lines shown on the ACF plot denote the two standard error

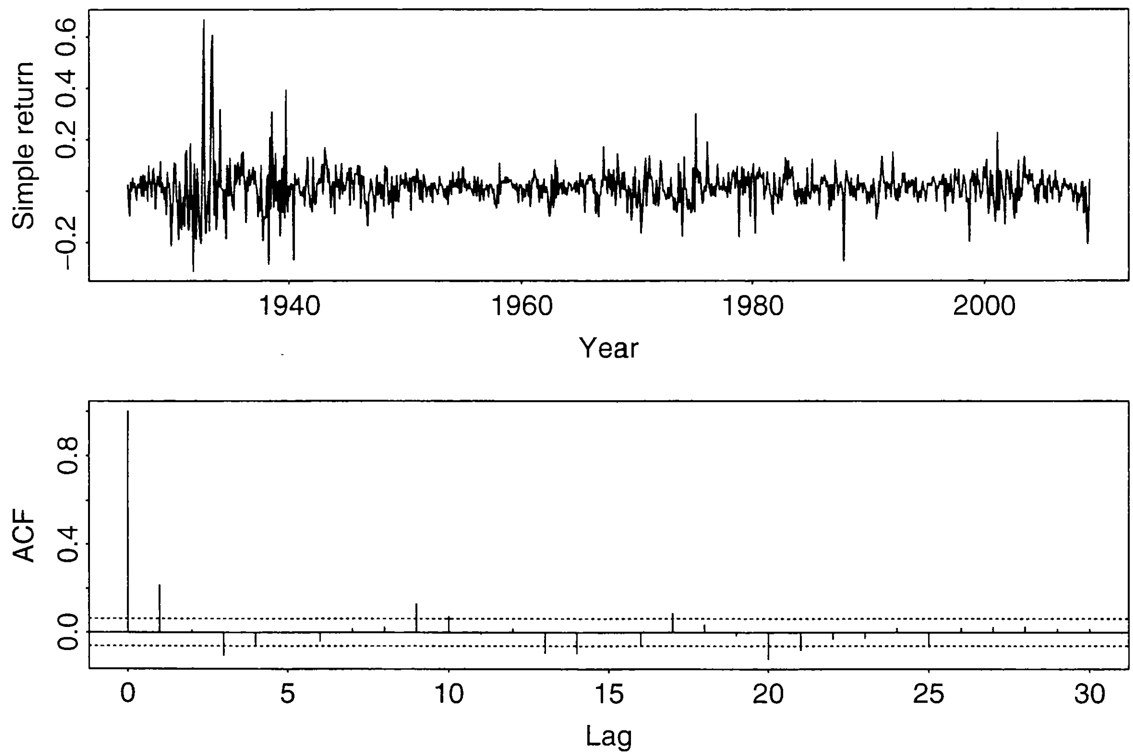

Figure 2.8 Time plot and sample autocorrelation function of monthly simple returns of CRSP equal-weighted index from January 1926 to December 2008.

limits. It is seen that the series has significant ACF at lags 1, 3, and 9. There are some marginally significant ACF at higher lags, but we do not consider them here. Based on the sample ACF, the following MA(9) model

$$r_t = c_0 + a_t - \theta_1 a_{t-1} - \theta_3 a_{t-3} - \theta_9 a_{t-9}$$

is identified for the series. Note that, unlike the sample PACF, sample ACF provides information on the nonzero MA lags of the model.

2.5.3 Estimation

Maximum-likelihood estimation is commonly used to estimate MA models. There are two approaches for evaluating the likelihood function of an MA model. The first approach assumes that the initial shocks (i.e., a_t for $t \le 0$) are zero. As such, the shocks needed in likelihood function calculation are obtained recursively from the model, starting with $a_1 = r_1 - c_0$ and $a_2 = r_2 - c_0 + \theta_1 a_1$. This approach is referred to as the *conditional-likelihood method* and the resulting estimates the conditional maximum-likelihood estimates. The second approach treats the initial shocks a_t, $t \le 0$, as additional parameters of the model and estimate them jointly with other parameters. This approach is referred to as the *exact-likelihood method*. The exact-likelihood estimates are preferred over the conditional ones, especially when the MA model is close to being noninvertible. The exact method, however, requires more intensive computation. If the sample size is large, then the two types of maximum-likelihood estimates are close to each other. For details of conditional- and exact-likelihood estimates of MA models, readers are referred to Box, Jenkins, and Reinsel (1994) or Chapter 8.

For illustration, consider the monthly simple return series of the CRSP equal-weighted index and the specified MA(9) model. The conditional maximum-likelihood method produces the fitted model

$$r_t = 0.012 + a_t + 0.189a_{t-1} - 0.121a_{t-3} + 0.122a_{t-9}, \quad \hat{\sigma}_a = 0.0714, \qquad (2.23)$$

where standard errors of the coefficient estimates are 0.003, 0.031, 0.031, and 0.031, respectively. The Ljung–Box statistics of the residuals give $Q(12) = 17.5$ with a p value 0.041, which is based on an asymptotic chi-squared distribution with 9 degrees of freedom. The model needs some refinements in modeling the linear dynamic dependence of the data. The p value would be 0.132 if 12 degrees of freedom are used. The exact maximum-likelihood method produces the fitted model

$$r_t = 0.012 + a_t + 0.191a_{t-1} - 0.120a_{t-3} + 0.123a_{t-9}, \qquad \hat{\sigma}_a = 0.0714, \quad (2.24)$$

where standard errors of the estimates are 0.003, 0.031, 0.031, and 0.031, respectively. The Ljung–Box statistics of the residuals give $Q(12) = 17.6$. The corresponding p values are 0.040 and 0.128, respectively, when the degrees of freedom are 9 and 12. Again, this fitted model is only marginally adequate. Comparing models (2.23) and (2.24), we see that for this particular instance, the difference between the conditional- and exact-likelihood methods is negligible.

2.5.4 Forecasting Using MA Models

Forecasts of an MA model can easily be obtained. Because the model has finite memory, its point forecasts go to the mean of the series quickly. To see this, assume that the forecast origin is h and let F_h denote the information available at time h. For the 1-step-ahead forecast of an MA(1) process, the model says

$$r_{h+1} = c_0 + a_{h+1} - \theta_1 a_h.$$

Taking the conditional expectation, we have

$$\hat{r}_h(1) = E(r_{h+1}|F_h) = c_0 - \theta_1 a_h,$$
$$e_h(1) = r_{h+1} - \hat{r}_h(1) = a_{h+1}.$$

The variance of the 1-step-ahead forecast error is $\text{Var}[e_h(1)] = \sigma_a^2$. In practice, the quantity a_h can be obtained in several ways. For instance, assume that $a_0 = 0$, then $a_1 = r_1 - c_0$, and we can compute a_t for $2 \le t \le h$ recursively by using $a_t = r_t - c_0 + \theta_1 a_{t-1}$. Alternatively, it can be computed by using the AR representation of the MA(1) model; see Section 2.6.5. Of course, a_t is the residual series of a fitted MA(1) model. Thus, a_h is readily available from the estimation.

For the 2-step-ahead forecast from the equation

$$r_{h+2} = c_0 + a_{h+2} - \theta_1 a_{h+1},$$

we have

$$\hat{r}_h(2) = E(r_{h+2}|F_h) = c_0,$$
$$e_h(2) = r_{h+2} - \hat{r}_h(2) = a_{h+2} - \theta_1 a_{h+1}.$$

The variance of the forecast error is $\text{Var}[e_h(2)] = (1 + \theta_1^2)\sigma_a^2$, which is the variance of the model and is greater than or equal to that of the 1-step-ahead forecast error. The prior result shows that for an MA(1) model the 2-step-ahead forecast of the series is simply the unconditional mean of the model. This is true for any forecast origin h. More generally, $\hat{r}_h(\ell) = c_0$ for $\ell \ge 2$. In summary, for an MA(1) model, the 1-step-ahead point forecast at the forecast origin h is $c_0 - \theta_1 a_h$ and the multistep ahead forecasts are c_0, which is the unconditional mean of the model. If we plot the forecasts $\hat{r}_h(\ell)$ versus ℓ, we see that the forecasts form a horizontal line after one step. Thus, for MA(1) models, mean reverting only takes one time period.

Similarly, for an MA(2) model, we have

$$r_{h+\ell} = c_0 + a_{h+\ell} - \theta_1 a_{h+\ell-1} - \theta_2 a_{h+\ell-2},$$

from which we obtain

$$\hat{r}_h(1) = c_0 - \theta_1 a_h - \theta_2 a_{h-1},$$
$$\hat{r}_h(2) = c_0 - \theta_2 a_h,$$
$$\hat{r}_h(\ell) = c_0, \quad \text{for} \quad \ell > 2.$$

Thus, the multistep-ahead forecasts of an MA(2) model go to the mean of the series after two steps. The variances of forecast errors go to the variance of the series after two steps. In general, for an MA(q) model, multistep-ahead forecasts go to the mean after the first q steps.

Table 2.3 gives some out-of-sample forecasts of an MA(9) model in the form of Eq. (2.24) for the monthly simple returns of the equal-weighted index at the forecast origin $h = 986$ (February 2008). The model parameters are reestimated using the first 986 observations. The sample mean and standard error of the estimation subsample are 0.0128 and 0.0736, respectively. As expected, the table shows that (a) the 10-step-ahead forecast is the sample mean, and (b) the standard deviations of the forecast errors converge to the standard deviation of the series as the forecast horizon increases. In this particular case, the point forecasts deviate substantially from the observed returns because of the worldwide financial crisis caused by the subprime mortgage problem and the collapse of Lehman Brothers.

Summary

A brief summary of AR and MA models is in order. We have discussed the following properties:

- For MA models, ACF is useful in specifying the order because ACF cuts off at lag q for an MA(q) series.
- For AR models, PACF is useful in order determination because PACF cuts off at lag p for an AR(p) process.
- An MA series is always stationary, but for an AR series to be stationary, all of its characteristic roots must be less than 1 in modulus.

TABLE 2.3 Out-of-Sample Forecast Performance of an MA(9) Model for Monthly Simple Returns of CRSP Equal-Weighted Index[a]

Step	1	2	3	4	5
Forecast	0.0043	0.0136	0.0150	0.0144	0.0120
Std. Error	0.0712	0.0724	0.0729	0.0729	0.0729
Actual	−0.0260	0.0312	0.0322	−0.0871	−0.0010
Step	6	7	8	9	10
Forecast	0.0019	0.0122	0.0056	0.0085	0.0128
Std. Error	0.0729	0.0729	0.0729	0.0729	0.0734
Actual	0.0141	−0.1209	−0.2060	−0.1366	0.0431

[a]The forecast origin is February 2008 With $h = 986$. The model is estimated by the exact maximum-likelihood method.

- For a stationary series, the multistep-ahead forecasts converge to the mean of the series, and the variances of forecast errors converge to the variance of the series as the forecast horizon increases.

2.6 SIMPLE ARMA MODELS

In some applications, the AR or MA models discussed in the previous sections become cumbersome because one may need a high-order model with many parameters to adequately describe the dynamic structure of the data. To overcome this difficulty, the autoregressive moving-average (ARMA) models are introduced; see Box, Jenkins, and Reinsel (1994). Basically, an ARMA model combines the ideas of AR and MA models into a compact form so that the number of parameters used is kept small, achieving parsimony in parameterization. For the return series in finance, the chance of using ARMA models is low. However, the concept of ARMA models is highly relevant in volatility modeling. As a matter of fact, the generalized autoregressive conditional heteroscedastic (GARCH) model can be regarded as an ARMA model, albeit nonstandard, for the a_t^2 series; see Chapter 3 for details. In this section, we study the simplest ARMA(1,1) model.

A time series r_t follows an ARMA(1,1) model if it satisfies

$$r_t - \phi_1 r_{t-1} = \phi_0 + a_t - \theta_1 a_{t-1}, \tag{2.25}$$

where $\{a_t\}$ is a white noise series. The left-hand side of the Eq. (2.25) is the AR component of the model and the right-hand side gives the MA component. The constant term is ϕ_0. For this model to be meaningful, we need $\phi_1 \neq \theta_1$; otherwise, there is a cancellation in the equation and the process reduces to a white noise series.

2.6.1 Properties of ARMA(1,1) Models

Properties of ARMA(1,1) models are generalizations of those of AR(1) models with some minor modifications to handle the impact of the MA(1) component. We start with the stationarity condition. Taking expectation of Eq. (2.25), we have

$$E(r_t) - \phi_1 E(r_{t-1}) = \phi_0 + E(a_t) - \theta_1 E(a_{t-1}).$$

Because $E(a_i) = 0$ for all i, the mean of r_t is

$$E(r_t) = \mu = \frac{\phi_0}{1 - \phi_1}$$

provided that the series is weakly stationary. This result is exactly the same as that of the AR(1) model in Eq. (2.8).

Next, assuming for simplicity that $\phi_0 = 0$, we consider the autocovariance function of r_t. First, multiplying the model by a_t and taking expectation, we have

$$E(r_t a_t) = E(a_t^2) - \theta_1 E(a_t a_{t-1}) = E(a_t^2) = \sigma_a^2. \tag{2.26}$$

Rewriting the model as

$$r_t = \phi_1 r_{t-1} + a_t - \theta_1 a_{t-1}$$

and taking the variance of the prior equation, we have

$$\text{Var}(r_t) = \phi_1^2 \text{Var}(r_{t-1}) + \sigma_a^2 + \theta_1^2 \sigma_a^2 - 2\phi_1\theta_1 E(r_{t-1}a_{t-1}).$$

Here we make use of the fact that r_{t-1} and a_t are uncorrelated. Using Eq. (2.26), we obtain

$$\text{Var}(r_t) - \phi_1^2 \text{Var}(r_{t-1}) = (1 - 2\phi_1\theta_1 + \theta_1^2)\sigma_a^2.$$

Therefore, if the series r_t is weakly stationary, then $\text{Var}(r_t) = \text{Var}(r_{t-1})$ and we have

$$\text{Var}(r_t) = \frac{(1 - 2\phi_1\theta_1 + \theta_1^2)\sigma_a^2}{1 - \phi_1^2}.$$

Because the variance is positive, we need $\phi_1^2 < 1$ (i.e., $|\phi_1| < 1$). Again, this is precisely the same stationarity condition as that of the AR(1) model.

To obtain the autocovariance function of r_t, we assume $\phi_0 = 0$ and multiply the model in Eq. (2.25) by $r_{t-\ell}$ to obtain

$$r_t r_{t-\ell} - \phi_1 r_{t-1} r_{t-\ell} = a_t r_{t-\ell} - \theta_1 a_{t-1} r_{t-\ell}.$$

For $\ell = 1$, taking expectation and using Eq. (2.26) for $t - 1$, we have

$$\gamma_1 - \phi_1 \gamma_0 = -\theta_1 \sigma_a^2,$$

where $\gamma_\ell = \text{Cov}(r_t, r_{t-\ell})$. This result is different from that of the AR(1) case for which $\gamma_1 - \phi_1\gamma_0 = 0$. However, for $\ell = 2$ and taking expectation, we have

$$\gamma_2 - \phi_1 \gamma_1 = 0,$$

which is identical to that of the AR(1) case. In fact, the same technique yields

$$\gamma_\ell - \phi_1 \gamma_{\ell-1} = 0, \quad \text{for} \quad \ell > 1. \tag{2.27}$$

In terms of ACF, the previous results show that for a stationary ARMA(1,1) model

$$\rho_1 = \phi_1 - \frac{\theta_1 \sigma_a^2}{\gamma_0}, \qquad \rho_\ell = \phi_1 \rho_{\ell-1}, \quad \text{for} \quad \ell > 1.$$

Thus, the ACF of an ARMA(1,1) model behaves very much like that of an AR(1) model except that the exponential decay starts with lag 2. Consequently, the ACF of an ARMA(1,1) model does not cut off at any finite lag.

Turning to PACF, one can show that the PACF of an ARMA(1,1) model does not cut off at any finite lag either. It behaves very much like that of an MA(1) model except that the exponential decay starts with lag 2 instead of lag 1.

In summary, the stationarity condition of an ARMA(1,1) model is the same as that of an AR(1) model, and the ACF of an ARMA(1,1) exhibits a similar pattern like that of an AR(1) model except that the pattern starts at lag 2.

2.6.2 General ARMA Models

A general ARMA(p, q) model is in the form

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

where $\{a_t\}$ is a white noise series and p and q are nonnegative integers. The AR and MA models are special cases of the ARMA(p, q) model. Using the back-shift operator, the model can be written as

$$(1 - \phi_1 B - \cdots - \phi_p B^p) r_t = \phi_0 + (1 - \theta_1 B - \cdots - \theta_q B^q) a_t. \tag{2.28}$$

The polynomial $1 - \phi_1 B - \cdots - \phi_p B^p$ is the AR polynomial of the model. Similarly, $1 - \theta_1 B - \cdots - \theta_q B^q$ is the MA polynomial. We require that there are no common factors between the AR and MA polynomials; otherwise the order (p, q) of the model can be reduced. Like a pure AR model, the AR polynomial introduces the characteristic equation of an ARMA model. If all of the solutions of the characteristic equation are less than 1 in absolute value, then the ARMA model is weakly stationary. In this case, the unconditional mean of the model is $E(r_t) = \phi_0/(1 - \phi_1 - \cdots - \phi_p)$.

2.6.3 Identifying ARMA Models

The ACF and PACF are not informative in determining the order of an ARMA model. Tsay and Tiao (1984) propose a new approach that uses the extended autocorrelation function (EACF) to specify the order of an ARMA process. The basic idea of EACF is relatively simple. If we can obtain a consistent estimate of the AR component of an ARMA model, then we can derive the MA component. From the derived MA series, we can use ACF to identify the order of the MA component.

The derivation of EACF is relatively involved; see Tsay and Tiao (1984) for details. Yet the function is easy to use. The output of EACF is a two-way table, where the rows correspond to AR order p and the columns to MA order q. The theoretical version of EACF for an ARMA(1,1) model is given in Table 2.4. The

TABLE 2.4 Theoretical EACF Table for an ARMA(1,1) Model, Where X Denotes Nonzero, O Denotes Zero, and * Denotes Either Zero or Nonzero[a]

	MA							
AR	0	1	2	3	4	5	6	7
0	X	X	X	X	X	X	X	X
1	X	O	O	O	O	O	O	O
2	*	X	O	O	O	O	O	O
3	*	*	X	O	O	O	O	O
4	*	*	*	X	O	O	O	O
5	*	*	*	*	X	O	O	O

[a]This latter category does not play any role in identifying the order (1,1).

key feature of the table is that it contains a triangle of O with the upper left vertex located at the order (1,1). This is the characteristic we use to identify the order of an ARMA process. In general, for an ARMA(p, q) model, the triangle of O will have its upper left vertex at the (p, q) position.

For illustration, consider the monthly log stock returns of the 3M Company from February 1946 to December 2008. There are 755 observations. The return series and its sample ACF are shown in Figure 2.9. The ACF indicates that there are no significant serial correlations in the data at the 1% level. Table 2.5 shows the sample EACF and a corresponding simplified table for the series. The simplified table is constructed by using the following notation:

1. X denotes that the absolute value of the corresponding EACF is greater than or equal to $2/\sqrt{T}$, which is twice of the asymptotic standard error of the EACF.
2. O denotes that the corresponding EACF is less than $2/\sqrt{T}$ in modulus.

The simplified table exhibits a triangular pattern of O with its upper left vertex at the order $(p, q) = (0, 0)$. A few exceptions of X appear when $q = 2$, 5, 9, and 11. However, the EACF table shows that the values of sample ACF corresponding to those X are around 0.08 or 0.09. These ACFs are only slightly greater than $2/\sqrt{755} = 0.073$. Indeed, if 1% critical value is used, those X would become O in the simplified EACF table. Consequently, the EACF suggests that the monthly log returns of 3M stock follow an ARMA(0,0) model (i.e., a white noise series). This is in agreement with the result suggested by the sample ACF in Figure 2.9.

The information criteria discussed earlier can also be used to select ARMA(p, q) models. Typically, for some prespecified positive integers P and Q, one computes AIC (or BIC) for ARMA(p, q) models, where $0 \leq p \leq P$ and $0 \leq q \leq Q$, and selects the model that gives the minimum AIC (or BIC). This approach requires maximum-likelihood estimation of many models and in some cases may encounter the difficulty of overfitting in estimation.

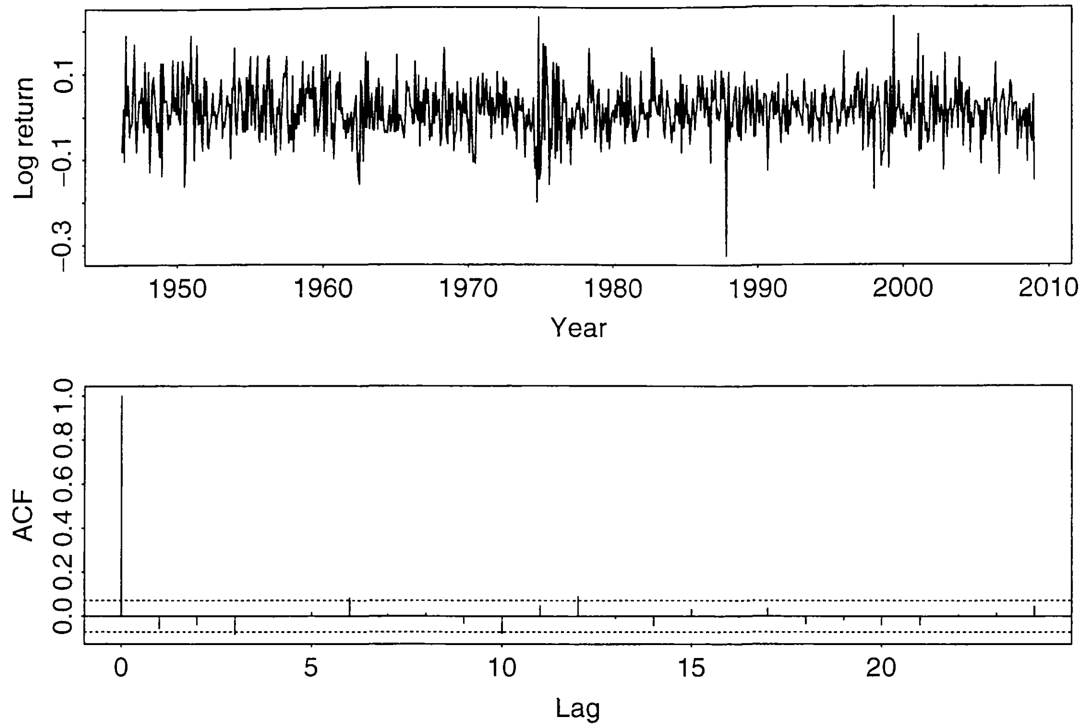

Figure 2.9 Time plot and sample autocorrelation function of monthly log stock returns of 3M Company from February 1946 to December 2008.

Once an ARMA(p, q) model is specified, its parameters can be estimated by either the conditional or exact-likelihood method. In addition, the Ljung–Box statistics of the residuals can be used to check the adequacy of a fitted model. If the model is correctly specified, then $Q(m)$ follows asymptotically a chi-squared distribution with $m - g$ degrees of freedom, where g denotes the number of AR or MA coefficients fitted in the model.

2.6.4 Forecasting Using an ARMA Model

Like the behavior of ACF, forecasts of an ARMA(p, q) model have similar characteristics as those of an AR(p) model after adjusting for the impacts of the MA component on the lower horizon forecasts. Denote the forecast origin by h and the available information by F_h. The 1-step-ahead forecast of r_{h+1} can be easily obtained from the model as

$$\hat{r}_h(1) = E(r_{h+1}|F_h) = \phi_0 + \sum_{i=1}^{p} \phi_i r_{h+1-i} - \sum_{i=1}^{q} \theta_i a_{h+1-i},$$

and the associated forecast error is $e_h(1) = r_{h+1} - \hat{r}_h(1) = a_{h+1}$. The variance of 1-step-ahead forecast error is $\text{Var}[e_h(1)] = \sigma_a^2$. For the ℓ-step-ahead forecast, we have

$$\hat{r}_h(\ell) = E(r_{h+\ell}|F_h) = \phi_0 + \sum_{i=1}^{p} \phi_i \hat{r}_h(\ell - i) - \sum_{i=1}^{q} \theta_i a_h(\ell - i),$$

TABLE 2.5 Sample Extended Autocorrelation Function and a Simplified Table for the Monthly Log Returns of 3M Stock from February 1946 to December 2008

Sample Extended Autocorrelation Function

	MA Order: q												
p	0	1	2	3	4	5	6	7	8	9	10	11	12
0	−0.06	−0.04	−0.08	−0.00	0.02	0.08	0.01	0.01	−0.03	−0.08	0.05	0.09	−0.01
1	−0.47	0.01	−0.07	−0.02	0.00	0.08	−0.03	0.00	−0.01	−0.07	0.04	0.09	−0.02
2	−0.38	−0.35	−0.07	0.02	−0.01	0.08	0.03	0.01	0.00	−0.03	0.02	0.04	0.04
3	−0.18	0.14	0.38	−0.02	0.00	0.04	−0.02	0.02	−0.00	−0.03	0.02	0.01	0.04
4	0.42	0.03	0.45	−0.01	0.00	0.00	−0.01	0.03	0.01	0.00	0.02	−0.00	0.01
5	−0.11	0.21	0.45	0.01	0.20	−0.01	−0.00	0.04	−0.01	−0.01	0.03	0.01	0.03
6	−0.21	−0.25	0.24	0.31	0.17	−0.04	−0.00	0.04	−0.01	−0.03	0.01	0.01	0.04

Simplified EACF Table

	MA Order: q												
p	0	1	2	3	4	5	6	7	8	9	10	11	12
0	O	O	X	O	O	X	O	O	O	X	O	X	O
1	X	O	O	O	O	X	O	O	O	O	O	X	O
2	X	X	O	O	O	X	O	O	O	O	O	O	O
3	X	X	X	O	O	O	O	O	O	O	O	O	O
4	X	O	X	O	O	O	O	O	O	O	O	O	O
5	X	X	X	O	X	O	O	O	O	O	O	O	O
6	X	X	X	X	X	O	O	O	O	O	O	O	O

where it is understood that $\hat{r}_h(\ell - i) = r_{h+\ell-i}$ if $\ell - i \le 0$ and $a_h(\ell - i) = 0$ if $\ell - i > 0$ and $a_h(\ell - i) = a_{h+\ell-i}$ if $\ell - i \le 0$. Thus, the multistep-ahead forecasts of an ARMA model can be computed recursively. The associated forecast error is

$$e_h(\ell) = r_{h+\ell} - \hat{r}_h(\ell),$$

which can be computed easily via a formula to be given in Eq. (2.34).

2.6.5 Three Model Representations for an ARMA Model

In this section, we briefly discuss three model representations for a stationary ARMA(p, q) model. The three representations serve three different purposes. Knowing these representations can lead to a better understanding of the model. The first representation is the ARMA(p, q) model in Eq. (2.28). This representation is compact and useful in parameter estimation. It is also useful in computing recursively multistep-ahead forecasts of r_t; see the discussion in the last section.

For the other two representations, we use long division of two polynomials. Given two polynomials $\phi(B) = 1 - \sum_{i=1}^{p} \phi_i B^i$ and $\theta(B) = 1 - \sum_{i=1}^{q} \theta_i B^i$, we can obtain, by long division, that

$$\frac{\theta(B)}{\phi(B)} = 1 + \psi_1 B + \psi_2 B^2 + \cdots \equiv \psi(B) \tag{2.29}$$

and

$$\frac{\phi(B)}{\theta(B)} = 1 - \pi_1 B - \pi_2 B^2 - \cdots \equiv \pi(B). \tag{2.30}$$

For instance, if $\phi(B) = 1 - \phi_1 B$ and $\theta(B) = 1 - \theta_1 B$, then

$$\psi(B) = \frac{1-\theta_1 B}{1-\phi_1 B} = 1 + (\phi_1 - \theta_1)B + \phi_1(\phi_1 - \theta_1)B^2 + \phi_1^2(\phi_1 - \theta_1)B^3 + \cdots,$$

$$\pi(B) = \frac{1-\phi_1 B}{1-\theta_1 B} = 1 - (\phi_1 - \theta_1)B - \theta_1(\phi_1 - \theta_1)B^2 - \theta_1^2(\phi_1 - \theta_1)B^3 - \cdots.$$

From the definition, $\psi(B)\pi(B) = 1$. Making use of the fact that $Bc = c$ for any constant (because the value of a constant is time invariant), we have

$$\frac{\phi_0}{\theta(1)} = \frac{\phi_0}{1-\theta_1-\cdots-\theta_q} \quad \text{and} \quad \frac{\phi_0}{\phi(1)} = \frac{\phi_0}{1-\phi_1-\cdots-\phi_p}.$$

AR Representation

Using the result of long division in Eq. (2.30), the ARMA(p, q) model can be written as

$$r_t = \frac{\phi_0}{1-\theta_1-\cdots-\theta_q} + \pi_1 r_{t-1} + \pi_2 r_{t-2} + \pi_3 r_{t-3} + \cdots + a_t. \tag{2.31}$$

This representation shows the dependence of the current return r_t on the past returns r_{t-i}, where $i > 0$. The coefficients $\{\pi_i\}$ are referred to as the π *weights* of an ARMA model. To show that the contribution of the lagged value r_{t-i} to r_t is diminishing as i increases, the π_i coefficient should decay to zero as i increases. An ARMA(p, q) model that has this property is said to be invertible. For a pure AR model, $\theta(B) = 1$ so that $\pi(B) = \phi(B)$, which is a finite-degree polynomial. Thus, $\pi_i = 0$ for $i > p$, and the model is invertible. For other ARMA models, a sufficient condition for invertibility is that all the zeros of the polynomial $\theta(B)$ are greater than unity in modulus. For example, consider the MA(1) model $r_t = (1 - \theta_1 B)a_t$. The zero of the first-order polynomial $1 - \theta_1 B$ is $B = 1/\theta_1$. Therefore, an MA(1) model is invertible if $|1/\theta_1| > 1$. This is equivalent to $|\theta_1| < 1$.

From the AR representation in Eq. (2.31), an invertible ARMA(p, q) series r_t is a linear combination of the current shock a_t and a weighted average of the past values. The weights decay exponentially for more remote past values.

MA Representation

Again, using the result of long division in Eq. (2.29), an ARMA(p, q) model can also be written as

$$r_t = \mu + a_t + \psi_1 a_{t-1} + \psi_2 a_{t-2} + \cdots = \mu + \psi(B)a_t, \tag{2.32}$$

where $\mu = E(r_t) = \phi_0/(1 - \phi_1 - \cdots - \phi_p)$. This representation shows explicitly the impact of the past shock a_{t-i} $(i > 0)$ on the current return r_t. The coefficients $\{\psi_i\}$ are referred to as the *impulse response function* of the ARMA model. For a weakly stationary series, the ψ_i coefficients decay exponentially as i increases. This is understandable as the effect of shock a_{t-i} on the return r_t should diminish over time. Thus, for a stationary ARMA model, the shock a_{t-i} does not have a permanent impact on the series. If $\phi_0 \neq 0$, then the MA representation has a constant term, which is the mean of r_t [i.e., $\phi_0/(1 - \phi_1 - \cdots - \phi_p)$].

The MA representation in Eq. (2.32) is also useful in computing the variance of a forecast error. At the forecast origin h, we have the shocks $a_h, a_{h-1}, \ldots$. Therefore, the ℓ-step-ahead point forecast is

$$\hat{r}_h(\ell) = \mu + \psi_\ell a_h + \psi_{\ell+1} a_{h-1} + \cdots, \tag{2.33}$$

and the associated forecast error is

$$e_h(\ell) = a_{h+\ell} + \psi_1 a_{h+\ell-1} + \cdots + \psi_{\ell-1} a_{h+1}.$$

Consequently, the variance of ℓ-step-ahead forecast error is

$$\text{Var}[e_h(\ell)] = (1 + \psi_1^2 + \cdots + \psi_{\ell-1}^2)\sigma_a^2, \tag{2.34}$$

which, as expected, is a nondecreasing function of the forecast horizon ℓ.

Finally, the MA representation in Eq. (2.32) provides a simple proof of mean reversion of a stationary time series. The stationarity implies that ψ_i approaches zero as $i \to \infty$. Therefore, by Eq. (2.33), we have $\hat{r}_h(\ell) \to \mu$ as $\ell \to \infty$. Because $\hat{r}_h(\ell)$ is the conditional expectation of $r_{h+\ell}$ at the forecast origin h, the result says that in the long term the return series is expected to approach its mean, that is, the series is mean reverting. Furthermore, using the MA representation in Eq. (2.32), we have $\text{Var}(r_t) = (1 + \sum_{i=1}^{\infty} \psi_i^2)\sigma_a^2$. Consequently, by Eq. (2.34), we have $\text{Var}[e_h(\ell)] \to \text{Var}(r_t)$ as $\ell \to \infty$. The speed by which $\hat{r}_h(\ell)$ approaches μ determines the speed of mean reverting.

2.7 UNIT-ROOT NONSTATIONARITY

So far we have focused on return series that are stationary. In some studies, interest rates, foreign exchange rates, or the price series of an asset are of interest. These series tend to be nonstationary. For a price series, the nonstationarity is mainly due to the fact that there is no fixed level for the price. In the time series literature, such a nonstationary series is called unit-root nonstationary time series. The best known example of unit-root nonstationary time series is the random-walk model.

2.7.1 Random Walk

A time series $\{p_t\}$ is a random walk if it satisfies

$$p_t = p_{t-1} + a_t, \tag{2.35}$$

where p_0 is a real number denoting the starting value of the process and $\{a_t\}$ is a white noise series. If p_t is the log price of a particular stock at date t, then p_0 could be the log price of the stock at its initial public offering (IPO) (i.e., the logged IPO price). If a_t has a symmetric distribution around zero, then conditional on p_{t-1}, p_t has a 50–50 chance to go up or down, implying that p_t would go up or down at random. If we treat the random-walk model as a special AR(1) model, then the coefficient of p_{t-1} is unity, which does not satisfy the weak stationarity condition of an AR(1) model. A random-walk series is, therefore, not weakly stationary, and we call it a unit-root nonstationary time series.

The random-walk model has widely been considered as a statistical model for the movement of logged stock prices. Under such a model, the stock price is not predictable or mean reverting. To see this, the 1-step-ahead forecast of model (2.35) at the forecast origin h is

$$\hat{p}_h(1) = E(p_{h+1}|p_h, p_{h-1}, \ldots) = p_h,$$

which is the log price of the stock at the forecast origin. Such a forecast has no practical value. The 2-step-ahead forecast is

$$\begin{aligned}\hat{p}_h(2) &= E(p_{h+2}|p_h, p_{h-1}, \ldots) = E(p_{h+1} + a_{h+2}|p_h, p_{h-1}, \ldots)\\ &= E(p_{h+1}|p_h, p_{h-1}, \ldots) = \hat{p}_h(1) = p_h,\end{aligned}$$

which again is the log price at the forecast origin. In fact, for any forecast horizon $\ell > 0$, we have

$$\hat{p}_h(\ell) = p_h.$$

Thus, for all forecast horizons, point forecasts of a random-walk model are simply the value of the series at the forecast origin. Therefore, the process is not mean reverting.

The MA representation of the random-walk model in Eq. (2.35) is

$$p_t = a_t + a_{t-1} + a_{t-2} + \cdots.$$

This representation has several important practical implications. First, the ℓ-step-ahead forecast error is

$$e_h(\ell) = a_{h+\ell} + \cdots + a_{h+1},$$

so that $\text{Var}[e_h(\ell)] = \ell\sigma_a^2$, which diverges to infinity as $\ell \to \infty$. The length of an interval forecast of $p_{h+\ell}$ will approach infinity as the forecast horizon-increases. This result says that the usefulness of point forecast $\hat{p}_h(\ell)$ diminishes as ℓ increases, which again implies that the model is not predictable. Second, the unconditional variance of p_t is unbounded because $\text{Var}[e_h(\ell)]$ approaches infinity as ℓ increases. Theoretically, this means that p_t can assume any real value for a sufficiently large t. For the log price p_t of an individual stock, this is plausible. Yet for market indexes, negative log price is very rare if it happens at all. In this sense, the adequacy of a random-walk model for market indexes is questionable. Third, from the representation, $\psi_i = 1$ for all i. Thus, the impact of any past shock a_{t-i} on p_t does not decay over time. Consequently, the series has a strong memory as it remembers all of the past shocks. In economics, the shocks are said to have a permanent effect on the series. The strong memory of a unit-root time series can be seen from the sample ACF of the observed series. The sample ACFs are all approaching 1 as the sample size increases.

2.7.2 Random Walk with Drift

As shown by empirical examples considered so far, the log return series of a market index tends to have a small and positive mean. This implies that the model for the log price is

$$p_t = \mu + p_{t-1} + a_t, \tag{2.36}$$

where $\mu = E(p_t - p_{t-1})$ and $\{a_t\}$ is a zero-mean white noise series. The constant term μ of model (2.36) is very important in financial study. It represents the time trend of the log price p_t and is often referred to as the *drift* of the model. To see this, assume that the initial log price is p_0. Then we have

$$\begin{aligned}
p_1 &= \mu + p_0 + a_1, \\
p_2 &= \mu + p_1 + a_2 = 2\mu + p_0 + a_2 + a_1, \\
&\vdots \\
p_t &= t\mu + p_0 + a_t + a_{t-1} + \cdots + a_1.
\end{aligned}$$

The last equation shows that the log price consists of a time trend $t\mu$ and a pure random-walk process $\sum_{i=1}^{t} a_i$. Because $\text{Var}(\sum_{i=1}^{t} a_i) = t\sigma_a^2$, where σ_a^2 is the variance of a_t, the conditional standard deviation of p_t is $\sqrt{t}\sigma_a$, which grows at a slower rate than the conditional expectation of p_t. Therefore, if we graph p_t against the time index t, we have a time trend with slope μ. A positive slope μ implies that the log price eventually goes to infinity. In contrast, a negative μ implies that the log price would converge to $-\infty$ as t increases. Based on the above discussion, it is then not surprising to see that the log return series of the CRSP value- and equal-weighted indexes have a small, but statistically significant, positive mean.

To illustrate the effect of the drift parameter on the price series, we consider the monthly log stock returns of the 3M Company from February 1946 to December 2008. As shown by the sample EACF in Table 2.5, the series has no significant serial correlation. The series thus follows the simple model

$$r_t = 0.0103 + a_t, \qquad \hat{\sigma}_a = 0.0637, \tag{2.37}$$

where 0.0103 is the sample mean of r_t and has a standard error 0.0023. The mean of the monthly log returns of 3M stock is, therefore, significantly different from zero at the 1% level. As a matter of fact, the one-sample test of zero mean shows a t ratio of 4.44 with a p value close to 0. We use the log return series to construct two log price series, namely

$$p_t = \sum_{i=1}^{t} r_i \qquad \text{and} \qquad p_t^* = \sum_{i=1}^{t} a_i,$$

where a_i is the mean-corrected log return in Eq. (2.37) (i.e., $a_t = r_t - 0.0103$). The p_t is the log price of 3M stock, assuming that the initial log price is zero (i.e., the log price of January 1946 was zero). The p_t^* is the corresponding log price if the mean of log returns was zero. Figure 2.10 shows the time plots of p_t and p_t^* as well as a straight line $y_t = 0.0103 \times t + 1946$, where t is the time sequence of the returns and 1946 is the starting year of the stock. From the plots, the importance of the constant 0.0103 in Eq. (2.37) is evident. In addition, as expected, it represents the slope of the upward trend of p_t.

Interpretation of the Constant Term

From the previous discussions, it is important to understand the meaning of a constant term in a time series model. First, for an MA(q) model in Eq. (2.22), the constant term is simply the mean of the series. Second, for a stationary AR(p) model in Eq. (2.9) or ARMA(p, q) model in Eq. (2.28), the constant term is related to the mean via $\mu = \phi_0/(1 - \phi_1 - \cdots - \phi_p)$. Third, for a random walk with drift, the constant term becomes the time slope of the series. These different interpretations for the constant term in a time series model clearly highlight the difference between dynamic and usual linear regression models.

Another important difference between dynamic and regression models is shown by an AR(1) model and a simple linear regression model,

$$r_t = \phi_0 + \phi_1 r_{t-1} + a_t \qquad \text{and} \qquad y_t = \beta_0 + \beta_1 x_t + a_t.$$

For the AR(1) model to be meaningful, the coefficient ϕ_1 must satisfy $|\phi_1| \leq 1$. However, the coefficient β_1 can assume any fixed real number.

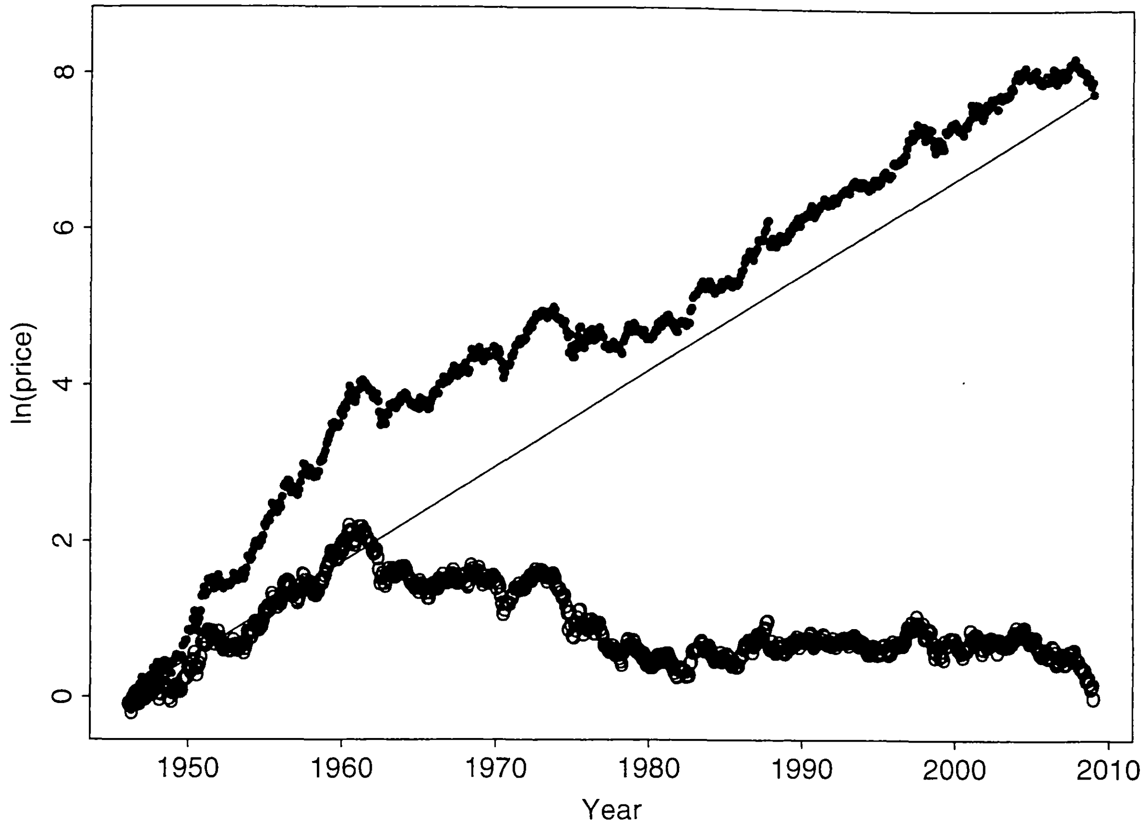

Figure 2.10 Time plots of log prices for 3M stock from February 1946 to December 2008, assuming that log price of January 1946 was zero. The "o" line is for log price without time trend. Straight line is $y_t = 0.0103 \times t + 1946$.

2.7.3 Trend-Stationary Time Series

A closely related model that exhibits linear trend is the trend-stationary time series model,

$$p_t = \beta_0 + \beta_1 t + r_t,$$

where r_t is a stationary time series, for example, a stationary AR(p) series. Here p_t grows linearly in time with rate β_1 and hence can exhibit behavior similar to that of a random-walk model with drift. However, there is a major difference between the two models. To see this, suppose that p_0 is fixed. The random-walk model with drift assumes the mean $E(p_t) = p_0 + \mu t$ and variance $\text{Var}(p_t) = t\sigma_a^2$, both of them are time dependent. On the other hand, the trend-stationary model assumes the mean $E(p_t) = \beta_0 + \beta_1 t$, which depends on time, and variance $\text{Var}(p_t) = \text{Var}(r_t)$, which is finite and time invariant. The trend-stationary series can be transformed into a stationary one by removing the time trend via a simple linear regression analysis. For analysis of trend-stationary time series, see the method of Section 2.9.

2.7.4 General Unit-Root Nonstationary Models

Consider an ARMA model. If one extends the model by allowing the AR polynomial to have 1 as a characteristic root, then the model becomes the well-known autoregressive integrated moving-average (ARIMA) model. An ARIMA model is said to be unit-root nonstationary because its AR polynomial has a unit root. Like a random-walk model, an ARIMA model has strong memory because the ψ_i coefficients in its MA representation do not decay over time to zero, implying that the past shock a_{t-i} of the model has a permanent effect on the series. A conventional approach for handling unit-root nonstationarity is to use *differencing*.

Differencing

A time series y_t is said to be an ARIMA$(p, 1, q)$ process if the change series $c_t = y_t - y_{t-1} = (1 - B)y_t$ follows a stationary and invertible ARMA(p, q) model. In finance, price series are commonly believed to be nonstationary, but the log return series, $r_t = \ln(P_t) - \ln(P_{t-1})$, is stationary. In this case, the log price series is unit-root nonstationary and hence can be treated as an ARIMA process. The idea of transforming a nonstationary series into a stationary one by considering its change series is called *differencing* in the time series literature. More formally, $c_t = y_t - y_{t-1}$ is referred to as the first differenced series of y_t. In some scientific fields, a time series y_t may contain multiple unit roots and needs to be differenced multiple times to become stationary. For example, if both y_t and its first differenced series $c_t = y_t - y_{t-1}$ are unit-root nonstationary, but $s_t = c_t - c_{t-1} = y_t - 2y_{t-1} + y_{t-2}$ is weakly stationary, then y_t has double unit roots, and s_t is the second differenced series of y_t. In addition, if s_t follows an ARMA(p, q) model, then y_t is an ARIMA$(p, 2, q)$ process. For such a time series, if s_t has a nonzero mean, then y_t has a quadratic time function and the quadratic time coefficient is related to the mean of s_t. The seasonally adjusted series of U.S. quarterly gross domestic product implicit price deflator might have double unit roots. However, the mean of the second differenced series is not significantly different from zero; see the Exercises of this chapter. Box, Jenkins, and Reinsel (1994) discuss many properties of general ARIMA models.

2.7.5 Unit-Root Test

To test whether the log price p_t of an asset follows a random walk or a random walk with drift, we employ the models

$$p_t = \phi_1 p_{t-1} + e_t \tag{2.38}$$

$$p_t = \phi_0 + \phi_1 p_{t-1} + e_t, \tag{2.39}$$

where e_t denotes the error term, and consider the null hypothesis $H_0 : \phi_1 = 1$ versus the alternative hypothesis $H_a : \phi_1 < 1$. This is the well-known unit-root testing problem; see Dickey and Fuller (1979). A convenient test statistic is the t ratio of the least-squares (LS) estimate of ϕ_1 under the null hypothesis. For Eq. (2.38), the LS method gives

$$\hat{\phi}_1 = \frac{\sum_{t=1}^{T} p_{t-1} p_t}{\sum_{t=1}^{T} p_{t-1}^2}, \qquad \hat{\sigma}_e^2 = \frac{\sum_{t=1}^{T} (p_t - \hat{\phi}_1 p_{t-1})^2}{T - 1},$$

where $p_0 = 0$ and T is the sample size. The t ratio is

$$\text{DF} \equiv t \text{ ratio} = \frac{\hat{\phi}_1 - 1}{\text{std}(\hat{\phi}_1)} = \frac{\sum_{t=1}^{T} p_{t-1} e_t}{\hat{\sigma}_e \sqrt{\sum_{t=1}^{T} p_{t-1}^2}},$$

which is commonly referred to as the Dickey–Fuller (DF) test. If $\{e_t\}$ is a white noise series with finite moments of order slightly greater than 2, then the DF statistic converges to a function of the standard Brownian motion as $T \to \infty$; see Chan and Wei (1988) and Phillips (1987) for more information. If ϕ_0 is zero but Eq. (2.39) is employed anyway, then the resulting t ratio for testing $\phi_1 = 1$ will converge to another nonstandard asymptotic distribution. In either case, simulation is used to obtain critical values of the test statistics; see Fuller (1976, Chapter 8) for selected critical values. Yet if $\phi_0 \neq 0$ and Eq. (2.39) is used, then the t ratio for testing $\phi_1 = 1$ is asymptotically normal. However, large sample sizes are needed for the asymptotic normal distribution to hold. Standard Brownian motion is introduced in Chapter 6.

For many economic time series, ARIMA(p, d, q) models might be more appropriate than the simple model in Eq. (2.39). In the econometric literature, AR(p) models are often used. Denote the series by x_t. To verify the existence of a unit root in an AR(p) process, one may perform the test $H_0 : \beta = 1$ vs. $H_a : \beta < 1$ using the regression

$$x_t = c_t + \beta x_{t-1} + \sum_{i=1}^{p-1} \phi_i \, \Delta x_{t-i} + e_t, \tag{2.40}$$

where c_t is a deterministic function of the time index t and $\Delta x_j = x_j - x_{j-1}$ is the differenced series of x_t. In practice, c_t can be zero or a constant or $c_t = \omega_0 + \omega_1 t$. The t ratio of $\hat{\beta} - 1$,

$$\text{ADF-test} = \frac{\hat{\beta} - 1}{\text{std}(\hat{\beta})},$$

where $\hat{\beta}$ denotes the least-squares estimate of β, is the well-known *augmented Dickey–Fuller* (ADF) unit-root test. Note that because of the first differencing, Eq. (2.40) is equivalent to an AR(p) model with deterministic function c_t. Equation (2.40) can also be rewritten as

$$\Delta x_t = c_t + \beta_c x_{t-1} + \sum_{i=1}^{p-1} \phi_i \Delta x_{t-i} + e_t,$$

where $\beta_c = \beta - 1$. One can then test the equivalent hypothesis $H_0 : \beta_c = 0$ vs. $H_a : \beta_c < 0$.

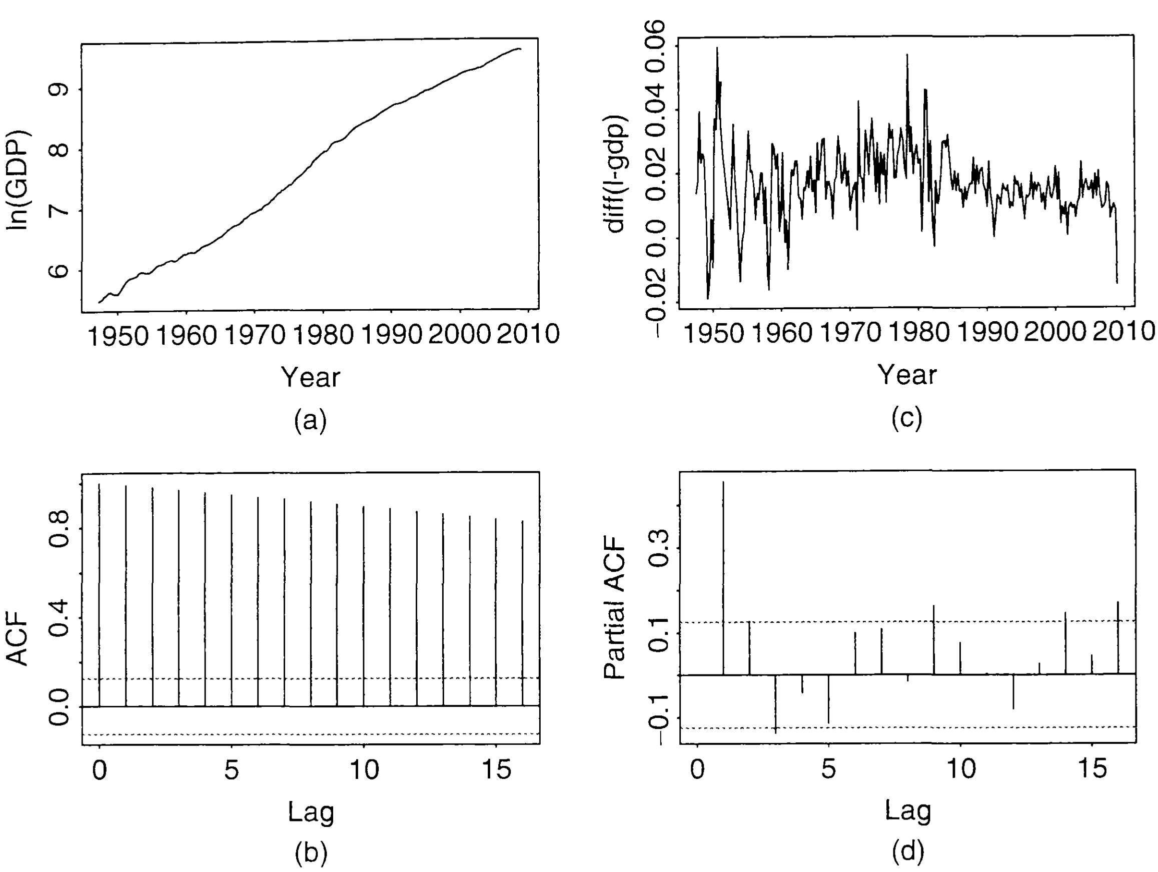

Figure 2.11 Log series of U.S. quarterly GDP from 1947.I to 2008.IV: (a) time plot of logged GDP series, (b) sample ACF of log GDP data, (c) time plot of first differenced series, and (d) sample PACF of differenced series.

Example 2.2. Consider the log series of U.S. quarterly GDP from 1947.I to 2008.IV. The series exhibits an upward trend, showing the growth of the U.S. economy, and has high sample serial correlations; see the lower left panel of Figure 2.11. The first differenced series, representing the growth rate of U.S. GDP and also shown in Figure 2.11, seems to vary around a fixed mean level, even though the variability appears to be smaller in recent years. To confirm the observed phenomenon, we apply the ADF unit-root test to the log series. Based on the sample PACF of the differenced series shown in Figure 2.11, we choose $p = 10$. Other values of p are also used, but they do not alter the conclusion of the test. With $p = 10$, the ADF test statistic is -1.701 with a p value 0.4297, indicating that the unit-root hypothesis cannot be rejected. From the attached S-Plus output, $\hat{\beta} = 1 + \hat{\beta}_c = 1 - 0.0008 = 0.9992$.

R Demonstration

```
> library(fUnitRoots)
> da=read.table("q-gdp4708.txt",header=T)
> gdp=log(da[,4])
> m1=ar(diff(gdp),method='mle')
> m1$order
[1] 10
> adfTest(gdp,lags=10,type=c("c"))
```

```
Title:
 Augmented Dickey-Fuller Test

Test Results:
 PARAMETER:
  Lag Order: 10
 STATISTIC:
  Dickey-Fuller: -1.6109
 P VALUE:  0.4569
```

S-Plus Demonstration

The following output has been edited:

```
> adft=unitroot(gdp,trend='c',method='adf',lags=10)
> summary(adft)

Test for Unit Root: Augmented DF Test
Null Hypothesis: there is a unit root
   Type of Test: t-test
 Test Statistic: -1.701
        P-value: 0.4297

Coefficients:
          Value  Std. Error t value Pr(>|t|)
   lag1 -0.0008  0.0005    -1.7006  0.0904
   lag2  0.3799  0.0659     5.7637  0.0000
   lag3  0.1883  0.0696     2.7047  0.0074
   ...
   lag10 0.1784  0.0637     2.8023  0.0055
constant 0.0134  0.0045     2.9636  0.0034

Regression Diagnostics:
         R-Squared 0.2877
Adjusted R-Squared 0.2564
Durbin-Watson Stat 1.9940

Residual standard error: 0.009318 on 234 degrees of freedom
```

As another example, consider the log series of the S&P 500 index from January 3, 1950, to April 16, 2008, for 14,462 observations. The series is shown in Figure 2.12. Testing for a unit root in the index is relevant if one wishes to verify empirically that the Index follows a random walk with drift. To this end, we use $c_t = \omega_0 + \omega_1 t$ in applying the ADF test. Furthermore, we choose $p = 15$ based on the sample PACF of the first differenced series. The resulting test statistic is -1.998 with a p value 0.602. Thus, the unit-root hypothesis cannot be rejected at any reasonable significance level. The constant term is statistically significant, whereas the estimate of the time trend is not at the usual 5% level. The latter is

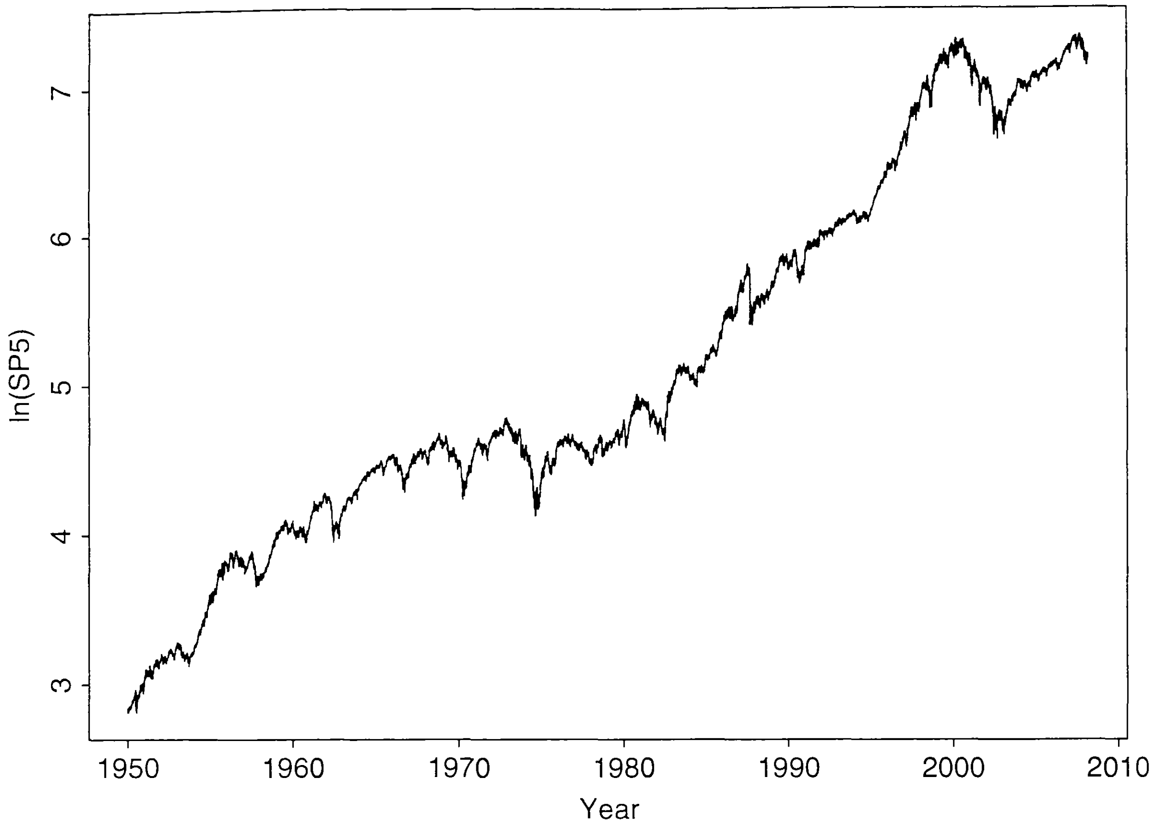

Figure 2.12 Time plot of logarithm of daily S&P 500 index from January 3, 1950, to April 16, 2008.

significant at the 10% level, however. In summary, for the period from January 1950 to April 2008, the log series of the S&P 500 index contains a unit root and a positive drift, but there is no strong evidence of a time trend.

R Demonstration

```
> library(fUnitRoots)
> da=read.table("d-sp55008.txt",header=T)
> sp5=log(da[,7])
> m2=ar(diff(sp5),method='mle')
> m2$order
[1] 2
> adfTest(sp5,lags=2,type=("ct"))

Title:
 Augmented Dickey-Fuller Test

Test Results:
 PARAMETER:
  Lag Order: 2
 STATISTIC:
  Dickey-Fuller: -2.0179
 P VALUE: 0.5708

> adfTest(sp5,lags=15,type=("ct"))
```

```
Title:
 Augmented Dickey-Fuller Test

Test Results:
 PARAMETER:
  Lag Order: 15
 STATISTIC:
  Dickey-Fuller: -1.9946
 P VALUE:  0.5807
```

S-Plus Demonstration

The following output has been edited:

```
> adft=unitroot(sp5,method='adf',trend='ct',lags=15)
> summary(adft)

Test for Unit Root: Augmented DF Test

 Null Hypothesis: there is a unit root
    Type of Test: t-test
  Test Statistic: -1.998
         P-value: 0.602

Coefficients:
           Value Std. Error t value Pr(>|t|)
    lag1 -0.0005  0.0003    -1.9977  0.0458
    lag2  0.0722  0.0083     8.7374  0.0000
    lag3 -0.0386  0.0083    -4.6532  0.0000
    lag4 -0.0071  0.0083    -0.8548  0.3927
    ...
   lag15  0.0133  0.0083     1.6122  0.1069
constant  0.0019  0.0008     2.3907  0.0168
    time  0.0020  0.0011     1.8507  0.0642

Regression Diagnostics:
         R-Squared 0.0081
Adjusted R-Squared 0.0070
Durbin-Watson Stat 1.9995

Residual standard error: 0.008981 on 14643 degrees of freedom
```

2.8 SEASONAL MODELS

Some financial time series such as quarterly earnings per share of a company exhibits certain cyclical or periodic behavior. Such a time series is called a *seasonal time series*. Figure 2.13(a) shows the time plot of quarterly earnings per share of Johnson & Johnson from the first quarter of 1960 to the last quarter of 1980.

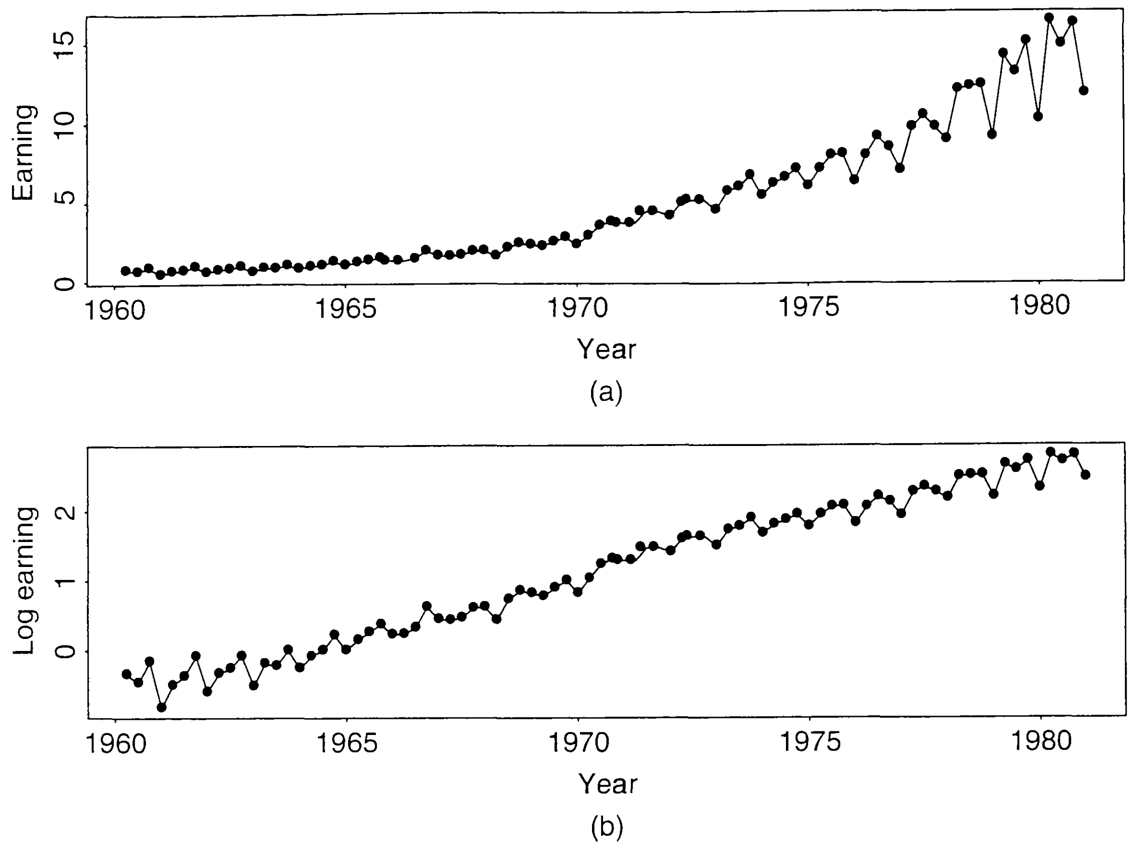

Figure 2.13 Time plots of quarterly earnings per share of Johnson & Johnson from 1960 to 1980: (a) observed earnings and (b) log earnings.

The data obtained from Shumway and Stoffer (2000) possess some special characteristics. In particular, the earnings grew exponentially during the sample period and had a strong seasonality. Furthermore, the variability of earnings increased over time. The cyclical pattern repeats itself every year so that the periodicity of the series is 4. If monthly data are considered (e.g., monthly sales of Wal-Mart stores), then the periodicity is 12. Seasonal time series models are also useful in pricing weather-related derivatives and energy futures because most environmental time series exhibit strong seasonal behavior.

Analysis of seasonal time series has a long history. In some applications, seasonality is of secondary importance and is removed from the data, resulting in a seasonally adjusted time series that is then used to make inference. The procedure to remove seasonality from a time series is referred to as *seasonal adjustment*. Most economic data published by the U.S. government are seasonally adjusted (e.g., the growth rate of gross domestic product and the unemployment rate). In other applications such as forecasting, seasonality is as important as other characteristics of the data and must be handled accordingly. Because forecasting is a major objective of financial time series analysis, we focus on the latter approach and discuss some econometric models that are useful in modeling seasonal time series.

2.8.1 Seasonal Differencing

Figure 2.13(b) shows the time plot of log earnings per share of Johnson & Johnson. We took the log transformation for two reasons. First, it is used to handle the exponential growth of the series. Indeed, the plot confirms that the growth is linear in the log scale. Second, the transformation is used to stablize the variability of the series. Again, the increasing pattern in variability of Figure 2.13(a) disappears in the new plot. Log transformation is commonly used in analysis of financial and economic time series. In this particular instance, all earnings are positive so that no adjustment is needed before taking the transformation. In some cases, one may need to add a positive constant to every data point before taking the transformation.

Denote the log earnings by x_t. The upper left panel of Figure 2.14 shows the sample ACF of x_t, which indicates that the quarterly log earnings per share has strong serial correlations. A conventional method to handle such strong serial correlations is to consider the first differenced series of x_t [i.e., $\Delta x_t = x_t - x_{t-1} = (1 - B)x_t$]. The lower left plot of Figure 2.14 gives the sample ACF of Δx_t. The ACF is strong when the lag is a multiple of periodicity 4. This is a well-documented behavior of sample ACF of a seasonal time series. Following the procedure of Box, Jenkins, and Reinsel (1994, Chapter 9), we take another difference of the data, that is,

$$\Delta_4(\Delta x_t) = (1 - B^4)\Delta x_t = \Delta x_t - \Delta x_{t-4} = x_t - x_{t-1} - x_{t-4} + x_{t-5}.$$

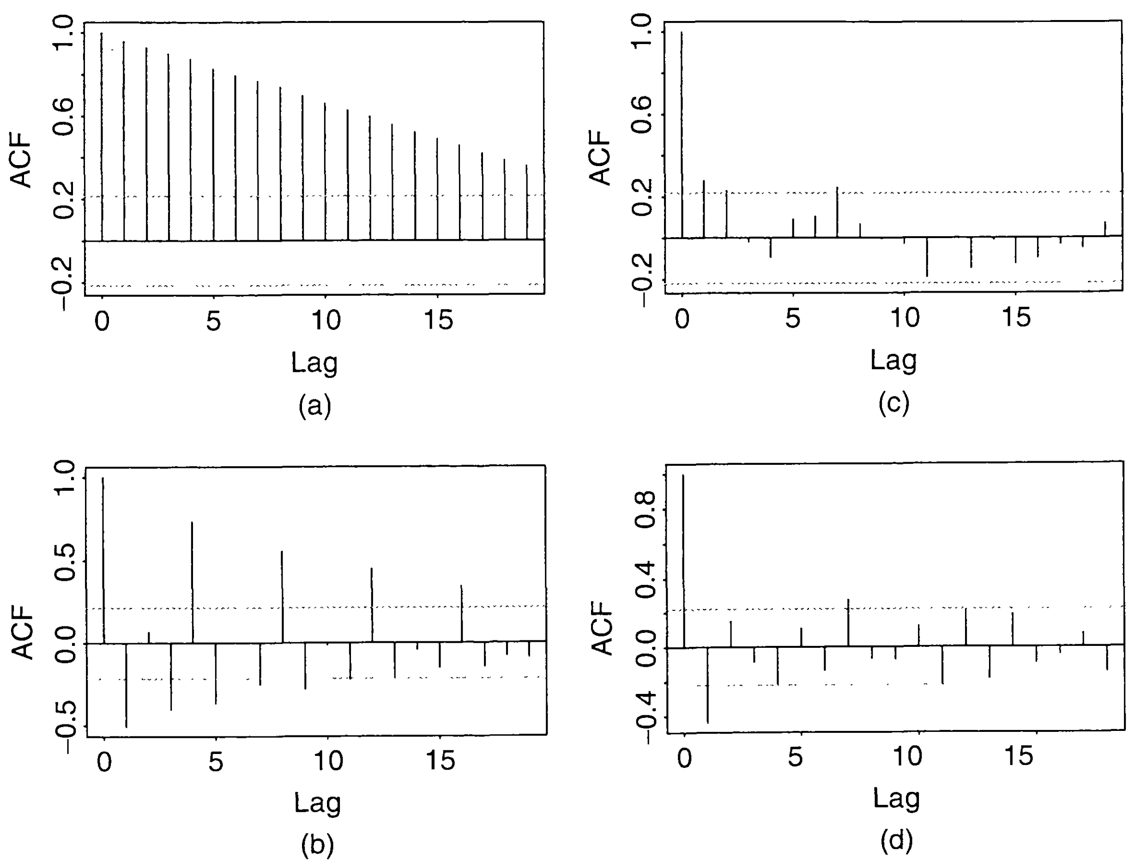

Figure 2.14 Sample ACF of log series of quarterly earnings per share of Johnson & Johnson from 1960 to 1980. (a) log earnings, (b) first differenced series, (c) seasonally differenced series, and (d) series with regular and seasonal differencing.

The operation $\Delta_4 = (1 - B^4)$ is called a *seasonal differencing*. In general, for a seasonal time series y_t with periodicity s, seasonal differencing means

$$\Delta_s y_t = y_t - y_{t-s} = (1 - B^s)y_t.$$

The conventional difference $\Delta y_t = y_t - y_{t-1} = (1 - B)y_t$ is referred to as the *regular differencing*. The lower right plot of Figure 2.14 shows the sample ACF of $\Delta_4 \Delta x_t$, which has a significant negative ACF at lag 1 and a marginal negative correlation at lag 4. For completeness, Figure 2.14 also gives the sample ACF of the seasonally differenced series $\Delta_4 x_t$.

2.8.2 Multiplicative Seasonal Models

The behavior of the sample ACF of $(1 - B^4)(1 - B)x_t$ in Figure 2.14 is common among seasonal time series. It led to the development of the following special seasonal time series model:

$$(1 - B^s)(1 - B)x_t = (1 - \theta B)(1 - \Theta B^s)a_t, \tag{2.41}$$

where s is the periodicity of the series, a_t is a white noise series, $|\theta| < 1$, and $|\Theta| < 1$. This model is referred to as the *airline model* in the literature; see Box, Jenkins, and Reinsel (1994, Chapter 9). It has been found to be widely applicable in modeling seasonal time series. The AR part of the model simply consists of the regular and seasonal differences, whereas the MA part involves two parameters. Focusing on the MA part (i.e., on the model),

$$w_t = (1 - \theta B)(1 - \Theta B^s)a_t = a_t - \theta a_{t-1} - \Theta a_{t-s} + \theta\Theta a_{t-s-1},$$

where $w_t = (1 - B^s)(1 - B)x_t$ and $s > 1$. It is easy to obtain that $E(w_t) = 0$ and

$$\begin{aligned}
\mathrm{Var}(w_t) &= (1 + \theta^2)(1 + \Theta^2)\sigma_a^2, \\
\mathrm{Cov}(w_t, w_{t-1}) &= -\theta(1 + \Theta^2)\sigma_a^2, \\
\mathrm{Cov}(w_t, w_{t-s+1}) &= \theta\Theta\sigma_a^2, \\
\mathrm{Cov}(w_t, w_{t-s}) &= -\Theta(1 + \theta^2)\sigma_a^2, \\
\mathrm{Cov}(w_t, w_{t-s-1}) &= \theta\Theta\sigma_a^2, \\
\mathrm{Cov}(w_t, w_{t-\ell}) &= 0, \quad \text{for} \quad \ell \neq 0, 1, s-1, s, s+1.
\end{aligned}$$

Consequently, the ACF of the w_t series is given by

$$\rho_1 = \frac{-\theta}{1 + \theta^2}, \qquad \rho_s = \frac{-\Theta}{1 + \Theta^2}, \qquad \rho_{s-1} = \rho_{s+1} = \rho_1\rho_s = \frac{\theta\Theta}{(1 + \theta^2)(1 + \Theta^2)},$$

and $\rho_\ell = 0$ for $\ell > 0$ and $\ell \neq 1, s-1, s, s+1$. For example, if w_t is a quarterly time series, then $s = 4$ and for $\ell > 0$, the ACF ρ_ℓ is nonzero at lags 1, 3, 4, and 5 only.

It is interesting to compare the prior ACF with those of the MA(1) model $y_t = (1 - \theta B)a_t$ and the MA(s) model $z_t = (1 - \Theta B^s)a_t$. The ACF of y_t and z_t series are

$$\rho_1(y) = \frac{-\theta}{1+\theta^2} \quad \text{and} \quad \rho_\ell(y) = 0, \quad \ell > 1,$$

$$\rho_s(z) = \frac{-\Theta}{1+\Theta^2} \quad \text{and} \quad \rho_\ell(z) = 0, \quad \ell > 0, \neq s.$$

We see that (i) $\rho_1 = \rho_1(y)$, (ii) $\rho_s = \rho_s(z)$, and (iii) $\rho_{s-1} = \rho_{s+1} = \rho_1(y) \times \rho_s(z)$. Therefore, the ACF of w_t at lags $(s-1)$ and $(s+1)$ can be regarded as the *interaction* between lag-1 and lag-s serial dependence, and the model of w_t is called a *multiplicative* seasonal MA model. In practice, a multiplicative seasonal model says that the dynamics of the regular and seasonal components of the series are approximately orthogonal.

The model

$$w_t = (1 - \theta B - \Theta B^s)a_t, \tag{2.42}$$

where $|\theta| < 1$ and $|\Theta| < 1$, is a nonmultiplicative seasonal MA model. It is easy to see that for the model in Eq. (2.42), $\rho_{s+1} = 0$. A multiplicative model is more parsimonious than the corresponding nonmultiplicative model because both models use the same number of parameters, but the multiplicative model has more nonzero ACFs.

Example 2.3. In this example we apply the airline model to the log series of quarterly earnings per share of Johnson & Johnson from 1960 to 1980. Based on the exact-likelihood method, the fitted model is

$$(1 - B)(1 - B^4)x_t = (1 - 0.678B)(1 - 0.314B^4)a_t, \qquad \hat{\sigma}_a = 0.089,$$

where standard errors of the two MA parameters are 0.080 and 0.101, respectively. The Ljung–Box statistics of the residuals show $Q(12) = 10.0$ with a p value of 0.44. The model appears to be adequate.

To illustrate the forecasting performance of the prior seasonal model, we reestimate the model using the first 76 observations and reserve the last 8 data points for forecasting evaluation. We compute 1-step- to 8-step-ahead forecasts and their standard errors of the fitted model at the forecast origin $h = 76$. An antilog transformation is taken to obtain forecasts of earnings per share using the relationship between normal and lognormal distributions given in Chapter 1. Figure 2.15 shows the forecast performance of the model, where the observed data are in solid line, point forecasts are shown by dots, and the dashed lines show 95% interval forecasts. The forecasts show a strong seasonal pattern and are close to the observed

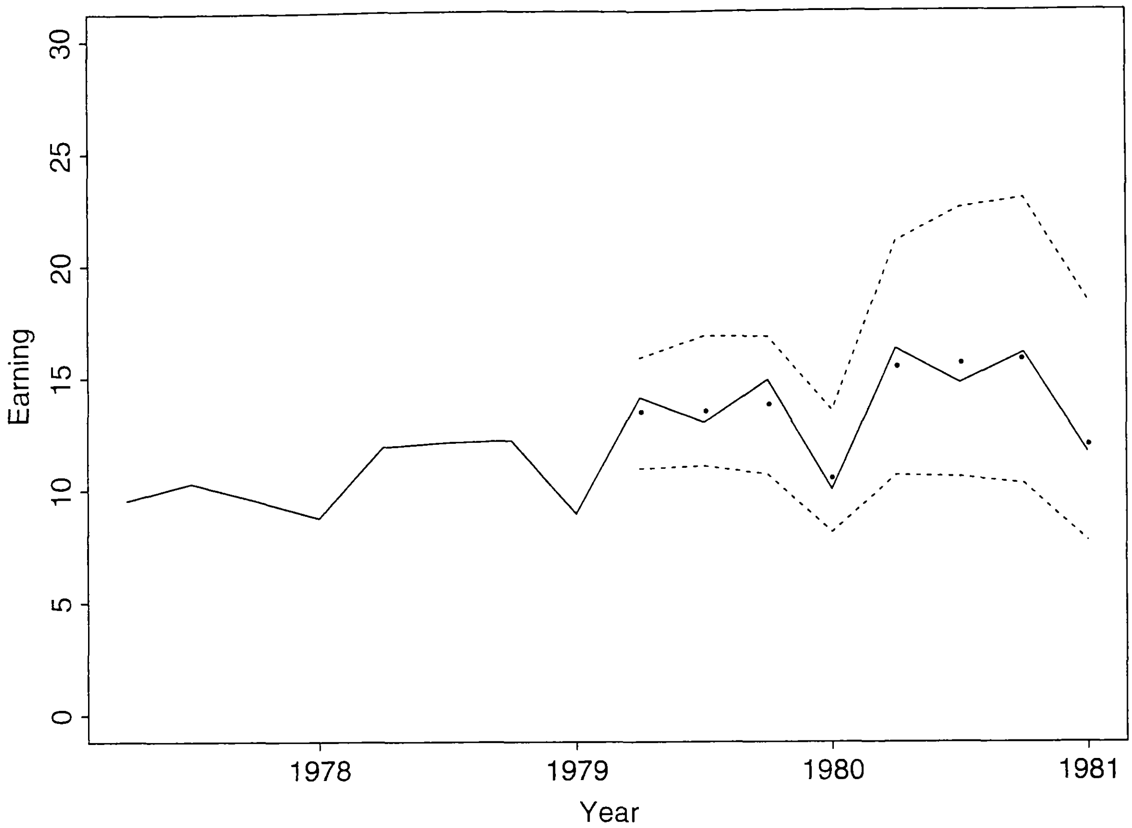

Figure 2.15 Out-of-sample point and interval forecasts for quarterly earnings of Johnson & Johnson. Forecast origin is fourth quarter of 1978. In plot, solid line shows actual observations, dots represent point forecasts, and dashed lines show 95% interval forecasts.

data. Finally, for an alternative approach to modeling the quarterly earnings data, see Example 11.3.

When the seasonal pattern of a time series is stable over time (e.g., close to a deterministic function), dummy variables may be used to handle the seasonality. This approach is taken by some analysts. However, deterministic seasonality is a special case of the multiplicative seasonal model discussed before. Specifically, if $\Theta = 1$, then model (2.41) contains a deterministic seasonal component. Consequently, the same forecasts are obtained by using either dummy variables or a multiplicative seasonal model when the seasonal pattern is deterministic. Yet use of dummy variables can lead to inferior forecasts if the seasonal pattern is not deterministic. In practice, we recommend that the exact-likelihood method should be used to estimate a multiplicative seasonal model, especially when the sample size is small or when there is the possibility of having a deterministic seasonal component.

Example 2.4. To demonstrate deterministic seasonal behavior, consider the monthly simple returns of the CRSP Decile 1 Index from January 1970 to December 2008 for 468 observations. The series is shown in Figure 2.16(a), and the time plot does not show any clear pattern of seasonality. However, the sample ACF of the return series shown in Figure 2.16(b) contains significant lags at 12, 24, and 36 as well as lag 1. If seasonal ARMA models are entertained, a model in the form

$$(1 - \phi_1 B)(1 - \phi_{12} B^{12})R_t = (1 - \theta_{12} B^{12})a_t$$

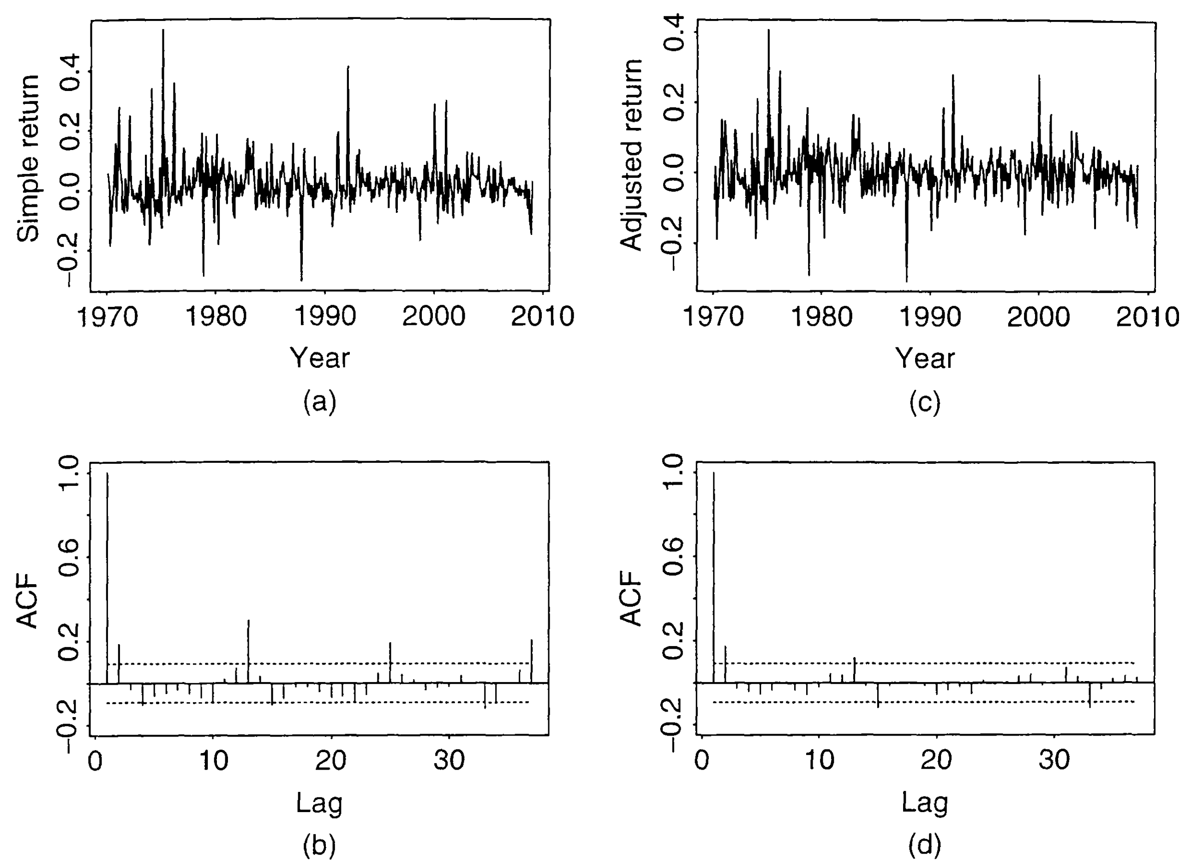

Figure 2.16 Monthly simple returns of CRSP Decile 1 index from January 1970 to December 2008: (a) time plot of the simple returns, (b) sample ACF of simple returns, (c) time plot of simple returns after adjusting for January effect, and (d) sample ACF of adjusted simple returns.

is identified, where R_t denotes the monthly simple return. Using the conditional-likelihood method, the fitted model is

$$(1 - 0.18B)(1 - 0.87B^{12})R_t = (1 - 0.74B^{12})a_t, \qquad \tilde{\sigma}_a = 0.069.$$

See the attached SCA (Scientific Computing Associates) output below. The estimates of the seasonal AR and MA coefficients are of similar magnitude. If the exact-likelihood method is used, we have

$$(1 - 0.188B)(1 - 0.951B^{12})R_t = (1 - 0.997B^{12})a_t, \qquad \tilde{\sigma}_a = 0.063.$$

The cancellation between seasonal AR and MA factors is clearly seen. This highlights the usefulness of using the exact-likelihood method and, the estimation result suggests that the seasonal behavior might be deterministic. To further confirm this assertion, we define the dummy variable for January, that is,

$$\mathrm{Jan}_t = \begin{cases} 1 & \text{if } t \text{ is January,} \\ 0 & \text{otherwise,} \end{cases}$$

and employ the simple linear regression

$$R_t = \beta_0 + \beta_1 \mathrm{Jan}_t + e_t.$$

The fitted model is $R_t = 0.0029 + 0.1253\text{Jan}_t + e_t$, where the standard errors of the estimates are 0.0033 and 0.0115, respectively. The right panel of Figure 2.16 shows the time plot and sample ACF of the residual series of the prior simple linear regression. From the sample ACF, serial correlations at lags 12, 24, and 36 largely disappear, suggesting that the seasonal pattern of the Decile 1 returns has been successfully removed by the January dummy variable. Consequently, the seasonal behavior in the monthly simple return of Decile 1 is mainly due to the *January effect*.

R Demonstration

The following output has been edited and % denotes explanation:

```
> da=read.table("m-deciles08.txt",header=T)
> d1=da[,2]
> jan=rep(c(1,rep(0,11)),39) % Create January dummy.
> m1=lm(d1 jan)
> summary(m1)
Call:
lm(formula = d1 ~ jan)

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) 0.002864  0.003333    0.859   0.391
jan         0.125251  0.011546   10.848  <2e-16 ***
---

Residual standard error: 0.06904 on 466 degrees of freedom
Multiple R-squared: 0.2016,     Adjusted R-squared: 0.1999

> m2=arima(d1,order=c(1,0,0),seasonal=list(order=c(1,0,1),
+ period=12))
> m2
Coefficients:
        ar1    sar1     sma1  intercept
     0.1769  0.9882  -0.9144     0.0118
s.e. 0.0456  0.0093   0.0335     0.0129

sigma^2 estimated as 0.004717: log likelihood=584.07,
   aic=-1158.14
> tsdiag(m2,gof=36)   % plot not shown.

> m2=arima(d1,order=c(1,0,0),seasonal=list(order=c(1,0,1),
+ period=12),include.mean=F)
> m2
Call:
arima(x=d1,order=c(1,0,0),seasonal=list(order=c(1,0,1),
   period=12),include.mean = F)
```

```
Coefficients:
         ar1    sar1     sma1
      0.1787 0.9886 -0.9127 % Slightly differ from those of SCA.
s.e.  0.0456 0.0089  0.0335

sigma^2 estimated as 0.00472: log likelihood=583.68,
   aic=-1159.36
```

SCA Demonstration

The following output has been edited:

```
input date,dec1,d2,d9,d10. file 'm-deciles08.txt'.
 --
tsm m1. model (1)(12)dec1=(12)noise.
 --
estim m1. hold resi(r1).    % Conditional MLE estimation

 SUMMARY FOR UNIVARIATE TIME SERIES MODEL --   M1
--------------------------------------------------------------
 VAR  TYPE OF   ORIGINAL    DIFFERENCING
      VARIABLE OR CENTERED

 DEC1 RANDOM    ORIGINAL    NONE
--------------------------------------------------------------
PAR.   VAR.  NUM./  FACTOR  ORDER CONS-    VALUE  STD    T
LABEL  NAME  DENOM.               TRAINT          ERROR  VALUE

1      D1      MA      1      12    NONE     .7388  .0488  15.14
2      D1      AR      1       1    NONE     .1765  .0447   3.95
3      D1      AR      2      12    NONE     .8698  .0295  29.49

 EFFECTIVE NUMBER OF OBSERVATIONS . .            455
 R-SQUARE . . . . . . . . . . . . . .          0.199
 RESIDUAL STANDARD ERROR. . . . . . .  0.689906E-01
 RESIDUAL STANDARD ERROR. . . . . . .  0.705662E-01
 --
estim m1. method exact. hold resi(r1)  % Exact MLE estimation

SUMMARY FOR UNIVARIATE TIME SERIES MODEL --   M1
--------------------------------------------------------------
VAR.   TYPE OF   ORIGINAL    DIFFERENCING
       VAR.     OR CENTERED

DEC1   RANDOM    ORIGINAL     NONE
--------------------------------------------------------------
PAR.   VARI.  NUM./  FACTOR ORDER CONS-  VALUE  STD     T
LABEL  NAME   DENOM.              TRAINT        ERROR  VALUE

1      D1       MA      1     12    NONE   .9968  .0150   66.31
```

```
2       D1       AR      1      1     NONE    .1884   .0448     4.21
3       D1       AR      2     12     NONE    .9505   .0070   135.46

 EFFECTIVE NUMBER OF OBSERVATIONS . .               455
 R-SQUARE . . . . . . . . . . . . . .             0.328
 RESIDUAL STANDARD ERROR. . . . . . . .     0.631807E-01
```

2.9 REGRESSION MODELS WITH TIME SERIES ERRORS

In many applications, the relationship between two time series is of major interest. An obvious example is the *market model* in finance that relates the excess return of an individual stock to that of a market index. The term structure of interest rates is another example in which the time evolution of the relationship between interest rates with different maturities is investigated. These examples lead naturally to the consideration of a linear regression in the form

$$y_t = \alpha + \beta x_t + e_t, \tag{2.43}$$

where y_t and x_t are two time series and e_t denotes the error term. The least-squares (LS) method is often used to estimate model (2.43). If $\{e_t\}$ is a white noise series, then the LS method produces consistent estimates. In practice, however, it is common to see that the error term e_t is serially correlated. In this case, we have a regression model with time series errors, and the LS estimates of α and β may not be consistent.

A regression model with time series errors is widely applicable in economics and finance, but it is one of the most commonly misused econometric models because the serial dependence in e_t is often overlooked. It pays to study the model carefully.

We introduce the model by considering the relationship between two U.S. weekly interest rate series:

1. r_{1t}: the 1-year Treasury constant maturity rate
2. r_{3t}: the 3-year Treasury constant maturity rate

Both series have 2467 observations from January 5, 1962, to April 10, 2009, and are measured in percentages. The series are obtained from the Federal Reserve Bank of St Louis. Strictly speaking, we should model the two interest series jointly using multivariate time series analysis in Chapter 8. However, for simplicity, we focus here on regression-type analysis and ignore the issue of simultaneity.

Figure 2.17 shows the time plots of the two interest rates with a solid line denoting the 1-year rate and a dashed line the 3-year rate. Figure 2.18(a) plots r_{1t} versus r_{3t}, indicating that, as expected, the two interest rates are highly correlated. A naive way to describe the relationship between the two interest rates is to use

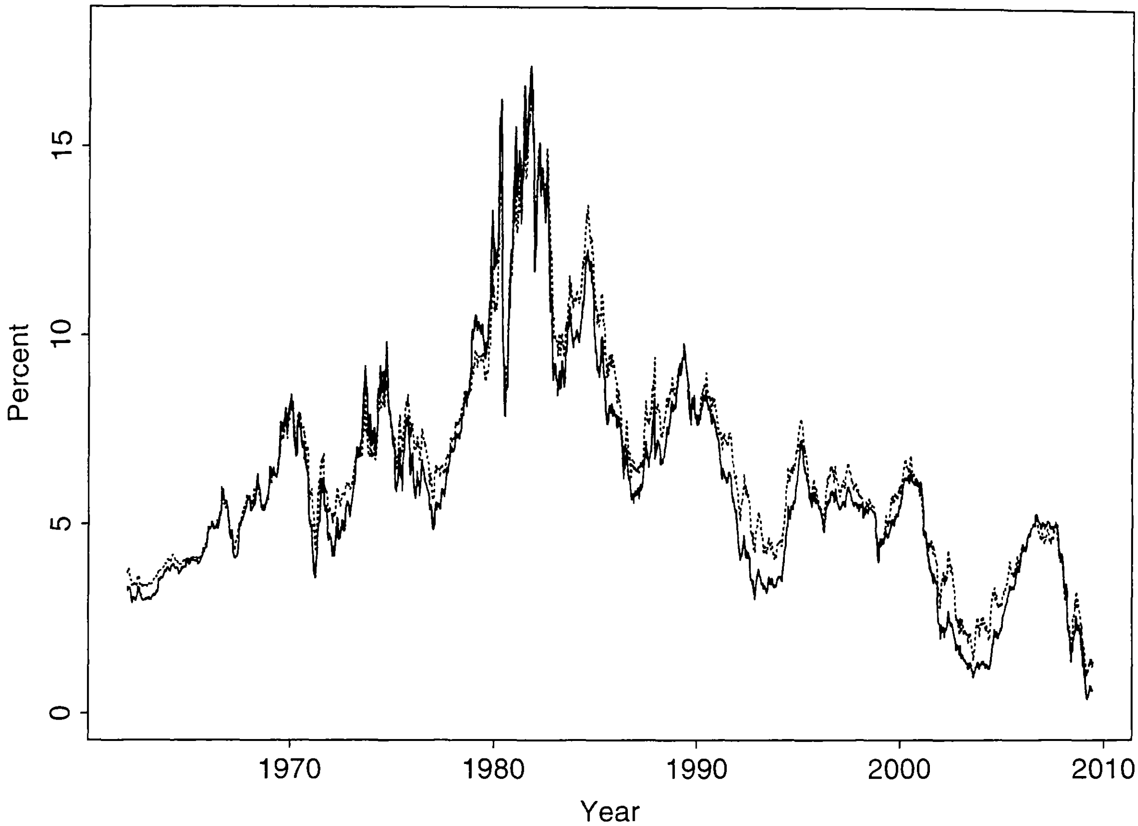

Figure 2.17 Time plots of U.S. weekly interest rates (in percentages) from January 5, 1962, to April 10, 2009. Solid line is Treasury 1-year constant maturity rate and dashed line Treasury 3-year constant maturity rate.

the simple model $r_{3t} = \alpha + \beta r_{1t} + e_t$. This results in a fitted model

$$r_{3t} = 0.832 + 0.930 r_{1t} + e_t, \qquad \hat{\sigma}_e = 0.523 \tag{2.44}$$

with $R^2 = 96.5\%$, where the standard errors of the two coefficients are 0.024 and 0.004, respectively. Model (2.44) confirms the high correlation between the two interest rates. However, the model is seriously inadequate, as shown by Figure 2.19, which gives the time plot and ACF of its residuals. In particular, the sample ACF of the residuals is highly significant and decays slowly, showing the pattern of a unit-root nonstationary time series. The behavior of the residuals suggests that marked differences exist between the two interest rates. Using the modern econometric terminology, if one assumes that the two interest rate series are unit-root nonstationary, then the behavior of the residuals of Eq. (2.44) indicates that the two interest rates are not *cointegrated*; see Chapter 8 for discussion of cointegration. In other words, the data fail to support the hypothesis that there exists a long-term equilibrium between the two interest rates. In some sense, this is not surprising because the pattern of "inverted yield curve" did occur during the data span. By inverted yield curve we mean the situation under which interest rates are inversely related to their time to maturities.

The unit-root behavior of both interest rates and the residuals of Eq. (2.44) leads to the consideration of the change series of interest rates. Let

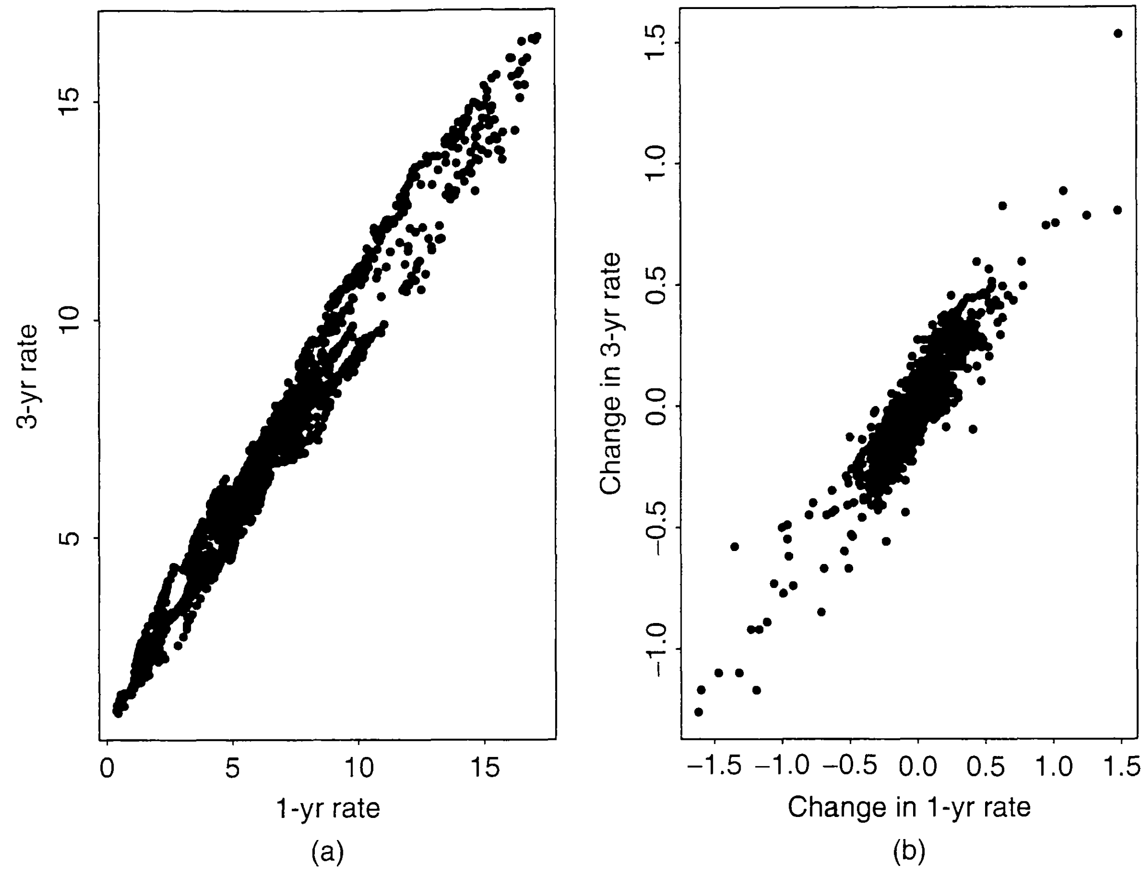

Figure 2.18 Scatterplots of U.S. weekly interest rates from January 5, 1962, to April 10, 2009: (a) 3-year rate vs. 1-year rate and (b) changes in 3-year rate vs. changes in 1-year rate.

1. $c_{1t} = r_{1t} - r_{1,t-1} = (1 - B)r_{1t}$ for $t \geq 2$: changes in the 1-year interest rate
2. $c_{3t} = r_{3t} - r_{3,t-1} = (1 - B)r_{3t}$ for $t \geq 2$: changes in the 3-year interest rate

and consider the linear regression $c_{3t} = \beta c_{1t} + e_t$. Figure 2.20 shows time plots of the two change series, whereas Figure 2.18(b) provides a scatterplot between them. The change series remain highly correlated with a fitted linear regression model given by

$$c_{3t} = 0.792c_{1t} + e_t, \qquad \hat{\sigma}_e = 0.0690, \tag{2.45}$$

with $R^2 = 82.5\%$. The standard error of the coefficient is 0.0073. This model further confirms the strong linear dependence between interest rates. Figure 2.21 shows the time plot and sample ACF of the residuals of Eq. (2.45). Once again, the ACF shows some significant serial correlations in the residuals, but magnitudes of the correlations are much smaller. This weak serial dependence in the residuals can be modeled by using the simple time series models discussed in the previous sections, and we have a linear regression with time series errors.

The main objective of this section is to discuss a simple approach for building a linear regression model with time series errors. The approach is straightforward. We employ a simple time series model discussed in this chapter for the residual series

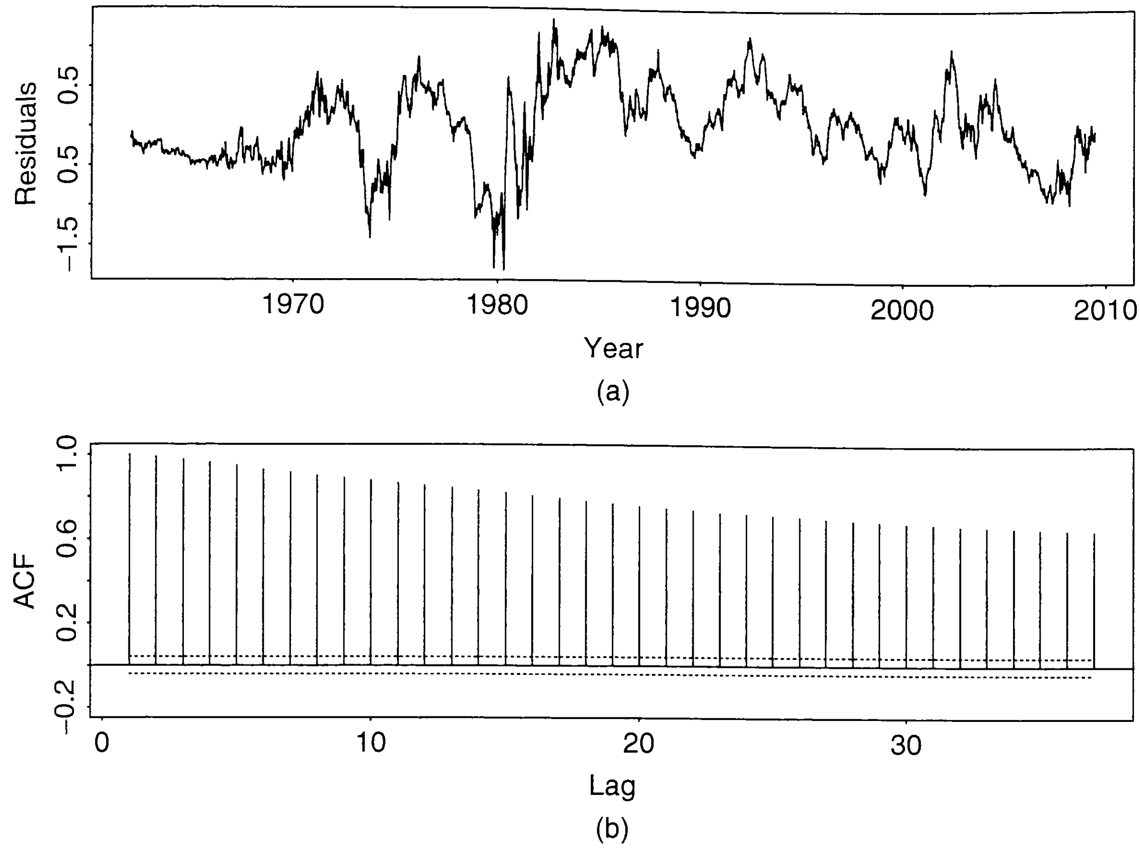

Figure 2.19 Residual series of linear regression (2.44) for two U.S. weekly interest rates: (a) time plot and (b) sample ACF.

and estimate the whole model jointly. For illustration, consider the simple linear regression in Eq. (2.45). Because residuals of the model are serially correlated, we shall identify a simple ARMA model for the residuals. From the sample ACF of the residuals shown in Figure 2.21, we specify an MA(1) model for the residuals and modify the linear regression model to

$$c_{3t} = \beta c_{1t} + e_t, \qquad e_t = a_t - \theta_1 a_{t-1}, \tag{2.46}$$

where $\{a_t\}$ is assumed to be a white noise series. In other words, we simply use an MA(1) model, without the constant term, to capture the serial dependence in the error term of Eq. (2.45). The resulting model is a simple example of linear regression with time series errors. In practice, more elaborated time series models can be added to a linear regression equation to form a general regression model with time series errors.

Estimating a regression model with time series errors was not easy before the advent of modern computers. Special methods such as the Cochrane–Orcutt estimator have been proposed to handle the serial dependence in the residuals; see Greene (2003, p. 273). By now, the estimation is as easy as that of other time series models. If the time series model used is stationary and invertible, then one can estimate the model jointly via the maximum-likelihood method. This is the approach we take by using either the SCA or R package. R and S-Plus demonstrations are given

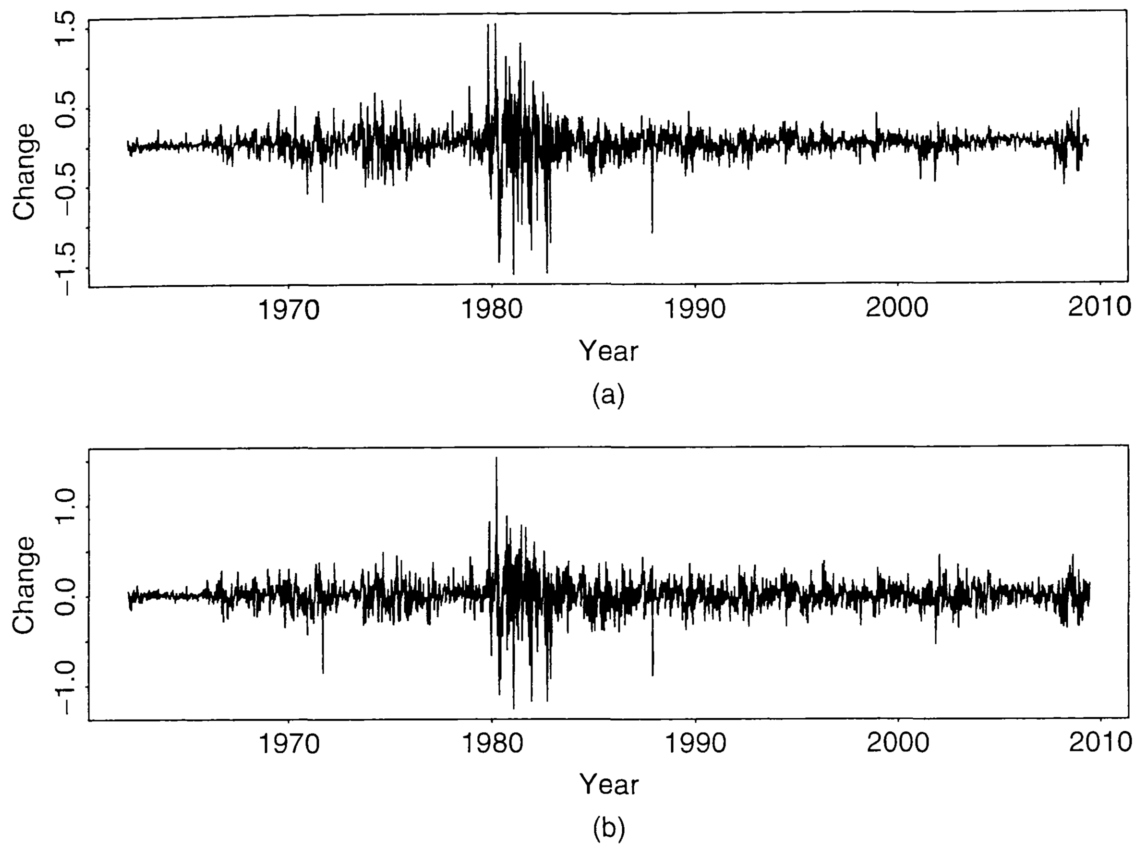

Figure 2.20 Time plots of change series of U.S. weekly interest rates from January 12, 1962, to April 10, 2009: (a) changes in Treasury 1-year constant maturity rate and (b) changes in Treasury 3-year constant maturity rate.

later. For the U.S. weekly interest rate data, the fitted version of model (2.46) is

$$c_{3t} = 0.794c_{1t} + e_t, \qquad e_t = a_t + 0.1823a_{t-1}, \qquad \hat{\sigma}_a = 0.0678, \tag{2.47}$$

with $R^2 = 83.1\%$. The standard errors of the parameters are 0.0075 and 0.0196, respectively. The model no longer has a significant lag-1 residual ACF, even though some minor residual serial correlations remain at lags 4, 6, and 7. The incremental improvement of adding additional MA parameters at lags 4, 6, and 7 to the residual equation is small and the result is not reported here.

Comparing the models in Eqs. (2.44), (2.45), and (2.47), we make the following observations. First, the high R^2 96.5% and coefficient 0.930 of model (2.44) are misleading because the residuals of the model show strong serial correlations. Second, for the change series, R^2 and the coefficient of c_{1t} of models (2.45) and (2.47) are close. In this particular instance, adding the MA(1) model to the change series only provides a marginal improvement. This is not surprising because the estimated MA coefficient is small numerically, even though it is statistically highly significant. Third, the analysis demonstrates that it is important to check residual serial dependence in linear regression analysis.

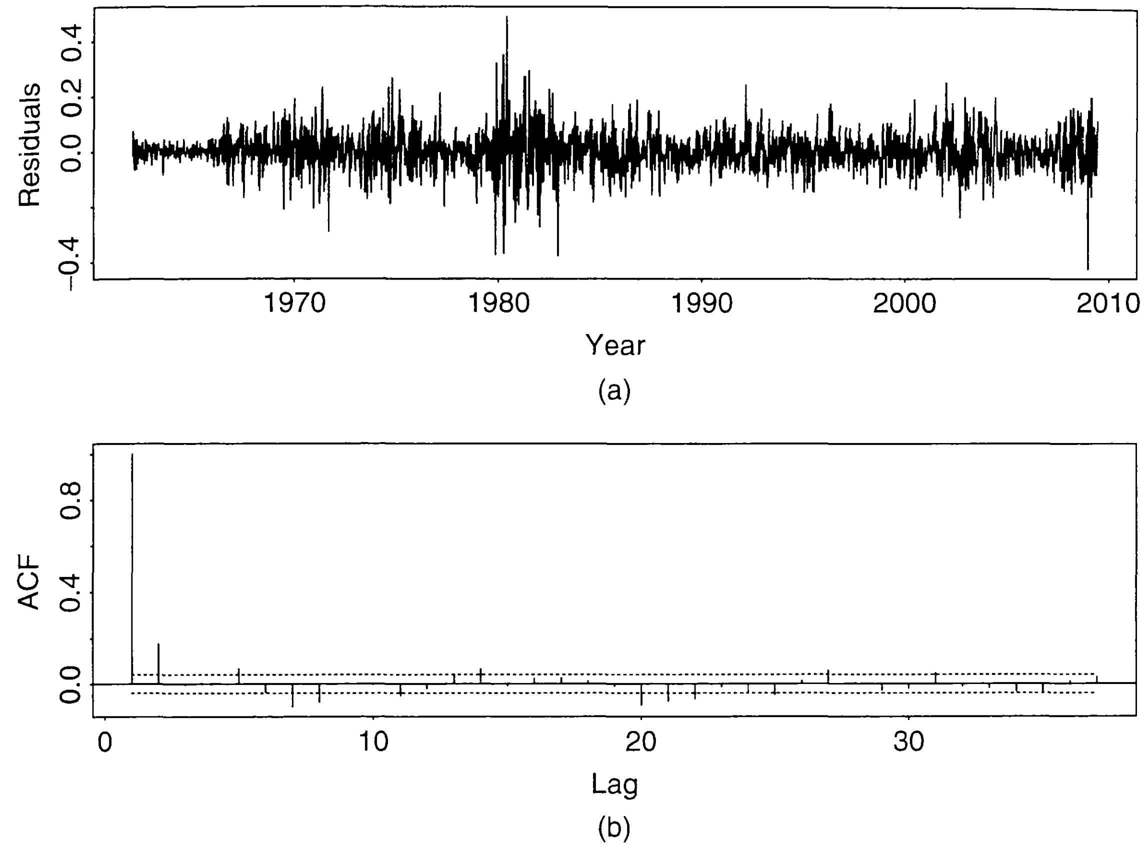

Figure 2.21 Residual series of linear regression (2.45) for two change series of U.S. weekly interest rates: (a) time plot and (b) sample ACF.

From Eq. (2.47), the model shows that the two weekly interest rate series are related as

$$r_{3t} = r_{3,t-1} + 0.794(r_{1t} - r_{1,t-1}) + a_t + 0.182a_{t-1}.$$

The interest rates are concurrently and serially correlated.

R Demonstration

The following output has been edited.

```
> r1=read.table("w-gs1yr.txt",header=T)[,4]
> r3=read.table("w-gs3yr.txt",header=T)[,4]
> m1=lm(r3 r1)
> summary(m1)
Call:
lm(formula = r3 ~ r1)
Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept ) 0.83214    0.02417   34.43  <2e-16 ***
r1           0.92955    0.00357  260.40  <2e-16 ***
```

```
---
Residual standard error: 0.5228 on 2465 degrees of freedom
Multiple R-squared: 0.9649,    Adjusted R-squared: 0.9649

> plot(m1$residuals,type='l')
> acf(m1$residuals,lag=36)
> c1=diff(r1)
> c3=diff(r3)
> m2=lm(c3 -1+c1)
> summary(m2)
Call:
lm(formula = c3 ~ -1 + c1)
Coefficients:
   Estimate Std. Error t value  Pr(>|t|)
c1  0.791935   0.007337   107.9   <2e-16 ***
---
Residual standard error: 0.06896 on 2465 degrees of freedom
Multiple R-squared: 0.8253,   Adjusted R-squared: 0.8253

> acf(m2$residuals,lag=36)

> m3=arima(c3,order=c(0,0,1),xreg=c1,include.mean=F)
> m3
Call:
arima(x = c3, order = c(0, 0, 1), xreg = c1, include.mean = F)
Coefficients:
         ma1       c1
      0.1823   0.7936
s.e.  0.0196   0.0075

sigma^2 estimated as 0.0046: log likelihood=3136.62,
   aic=-6267.23
>
> rsq=(sum(c3^2)-sum(m3$residuals^2))/sum(c3^2)
> rsq
[1] 0.8310077
```

Summary

We outline a general procedure for analyzing linear regression models with time series errors:

1. Fit the linear regression model and check serial correlations of the residuals.
2. If the residual series is unit-root nonstationary, take the first difference of both the dependent and explanatory variables. Go to step 1. If the residual series appears to be stationary, identify an ARMA model for the residuals and modify the linear regression model accordingly.

3. Perform a joint estimation via the maximum-likelihood method and check the fitted model for further improvement.

To check the serial correlations of residuals, we recommend that the Ljung–Box statistics be used instead of the Durbin–Watson (DW) statistic because the latter only considers the lag-1 serial correlation. There are cases in which serial dependence in residuals appears at higher order lags. This is particularly so when the time series involved exhibits some seasonal behavior.

Remark. For a residual series e_t with T observations, the Durbin–Watson statistic is

$$\mathrm{DW} = \frac{\sum_{t=2}^{T}(e_t - e_{t-1})^2}{\sum_{t=1}^{T} e_t^2}.$$

Straightforward calculation shows that $\mathrm{DW} \approx 2(1 - \hat{\rho}_1)$, where $\hat{\rho}_1$ is the lag-1 ACF of $\{e_t\}$. □

In S-Plus, regression models with time series errors can be analyzed by the command `OLS` (ordinary least squares) if the residuals assume an AR model. Also, to identify a lagged variable, the command is `tslag`, for example, `y = tslag(r,1)`. For the interest rate series, the relevant commands follow, where % denotes explanation of the command:

```
> r1t=read.table("w-gs1yr.txt",header=T)[,4] %load data
> r3t=read.table("w-gs3yr.txt",header=T)[,4]
> fit=OLS(r3t~r1t)      % fit the first regression
> summary(fit)
> c3t=diff(r3t)         % take difference
> c1t=diff(r1t)
> fit1=OLS(c3t~c1t)     % fit second regression
> summary(fit1)
> fit2=OLS(c3t~c1t+tslag(c3t,1)+tslag(c1t,1), na.rm=T)
> summary(fit2)
```

See the output in the next section for more information.

2.10 CONSISTENT COVARIANCE MATRIX ESTIMATION

Consider again the regression model in Eq. (2.43). There may exist situations in which the error term e_t has serial correlations and/or conditional heteroscedasticity, but the main objective of the analysis is to make inference concerning the regression coefficients α and β. See Chapter 3 for discussion of conditional heteroscedasticity. In situations under which the OLS estimates of the coefficients remain consistent, methods are available to provide consistent estimate of the covariance matrix of the coefficient estimates. Two such methods are widely used. The first method is called the heteroscedasticity consistent (HC) estimator; see Eicker (1967) and White (1980). The second method is called the heteroscedasticity and autocorrelation consistent (HAC) estimator; see Newey and West (1987).

For ease in discussion, we shall rewrite the regression model as

$$y_t = \boldsymbol{x}_t'\boldsymbol{\beta} + e_t, \qquad t = 1, \ldots, T, \tag{2.48}$$

where y_t is the dependent variable, $\boldsymbol{x}_t = (x_{1t}, \ldots, x_{kt})'$ is a k-dimensional vector of explanatory variables including constant, and $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_k)'$ is the parameter vector. Here $\boldsymbol{c}'$ denotes the transpose of the vector $\boldsymbol{c}$. The LS estimate of $\boldsymbol{\beta}$ and the associate covariance matrix are

$$\hat{\boldsymbol{\beta}} = \left[\sum_{t=1}^{T} \boldsymbol{x}_t\boldsymbol{x}_t'\right]^{-1} \sum_{t=1}^{T} \boldsymbol{x}_t y_t, \qquad \mathrm{Cov}(\hat{\boldsymbol{\beta}}) = \sigma_e^2 \left[\sum_{t=1}^{T} \boldsymbol{x}_t\boldsymbol{x}_t'\right]^{-1},$$

where σ_e^2 is the variance of e_t and is estimated by the variance of the residuals of the regression. In the presence of serial correlations or conditional heteroscedasticity, the prior covariance matrix estimator is inconsistent, often resulting in inflating the t ratios of $\hat{\boldsymbol{\beta}}$.

The estimator of White (1980) is

$$\mathrm{Cov}(\hat{\boldsymbol{\beta}})_{\mathrm{HC}} = \left[\sum_{t=1}^{T} \boldsymbol{x}_t\boldsymbol{x}_t'\right]^{-1} \left[\sum_{t=1}^{T} \hat{e}_t^2 \boldsymbol{x}_t\boldsymbol{x}_t'\right] \left[\sum_{t=1}^{T} \boldsymbol{x}_t\boldsymbol{x}_t'\right]^{-1}, \tag{2.49}$$

where $\hat{e}_t = y_t - \boldsymbol{x}_t'\hat{\boldsymbol{\beta}}$ is the residual at time t. The estimator of Newey and West (1987) is

$$\mathrm{Cov}(\hat{\boldsymbol{\beta}})_{\mathrm{HAC}} = \left[\sum_{t=1}^{T} \boldsymbol{x}_t\boldsymbol{x}_t'\right]^{-1} \widehat{\boldsymbol{C}}_{\mathrm{HAC}} \left[\sum_{t=1}^{T} \boldsymbol{x}_t\boldsymbol{x}_t'\right]^{-1}, \tag{2.50}$$

where

$$\widehat{\boldsymbol{C}}_{\mathrm{HAC}} = \sum_{t=1}^{T} \hat{e}_t^2 \boldsymbol{x}_t\boldsymbol{x}_t' + \sum_{j=1}^{\ell} w_j \sum_{t=j+1}^{T} (\boldsymbol{x}_t \hat{e}_t e_{t-j} \boldsymbol{x}_{t-j}' + \boldsymbol{x}_{t-j} \hat{e}_{t-j} \hat{e}_t \boldsymbol{x}_t'),$$

where ℓ is a truncation parameter and w_j is a weight function such as the Bartlett weight function defined by

$$w_j = 1 - \frac{j}{\ell + 1}.$$

Other weight functions can also be used. Newey and West (1987) suggest choosing ℓ to be the integer part of $4(T/100)^{2/9}$. This estimator essentially uses a nonparametric method to estimate the covariance matrix of $\{\sum_{t=1}^{T} \hat{e}_t \boldsymbol{x}_t\}$.

For illustration, we employ the first differenced interest rate series in Eq. (2.45). The t ratio of the coefficient of c_{1t} is 107.91 if both serial correlation and heteroscedasticity in the residuals are ignored, it becomes 48.44 when the HC estimator is used, and it reduces to 39.92 when the HAC estimator is used. The S-Plus demonstration below also uses a regression that includes lagged values $c_{1,t-1}$ and $c_{3,t-1}$ as regressors to take care of serial correlations in the residuals. One can also apply the HC or HAC estimator to the fitted model to refine the t ratios of the coefficient estimates.

S-Plus Demonstration

The following output has been edited and % denotes explanation:

```
> module(finmetrics)
> r1=read.table("w-gs1yr.txt",header=T)[,4] % Load data
> r3=read.table("w-gs3yr.txt",header=T)[,4]
> c1=diff(r1) % Take 1st difference
> c3=diff(r3)

> reg.fit=OLS(c3~c1) % Fit a simple linear regression.
> summary(reg.fit)
Call:
OLS(formula = c3 ~ c1)

Residuals:
     Min        1Q    Median       3Q      Max
 -0.4246   -0.0358   -0.0012   0.0347   0.4892

Coefficients:
                Value Std. Error   t value  Pr(>|t|)
(Intercept)   -0.0001     0.0014   -0.0757   0.9397
         c1    0.7919     0.0073  107.9063   0.0000

Regression Diagnostics:
         R-Squared 0.8253
Adjusted R-Squared 0.8253
Durbin-Watson Stat 1.6456

Residual Diagnostics:
                   Stat   P-Value
Jarque-Bera  1644.6146    0.0000
  Ljung-Box   230.0477    0.0000

Residual standard error: 0.06897 on 2464 degrees of freedom

> summary(reg.fit,correction="white") % Use HC the estimator
```

```
Coefficients:
                 Value Std. Error  t value  Pr(>|t|)
(Intercept)  -0.0001    0.0014   -0.0757   0.9396
         c1   0.7919    0.0163   48.4405   0.0000

> summary(reg.fit,correction="nw") % Use the HAC estimator

Coefficients:
                 Value Std. Error  t value  Pr(>|t|)
(Intercept)  -0.0001   0.0016    -0.0678   0.9459
         c1   0.7919   0.0198    39.9223   0.0000

 % Below, fit a regression model with time series errors.
> reg.ts=OLS(c3~c1+tslag(c3,1)+tslag(c1,1),na.rm=T)
> summary(reg.ts)
Call:
OLS(formula = c3 ~ c1 + tslag(c3, 1)+tslag(c1, 1), na.rm = T)

Residuals:
     Min        1Q    Median       3Q      Max
 -0.4481   -0.0355   -0.0008   0.0341   0.4582

Coefficients:
                  Value Std. Error   t value  Pr(>|t|)
 (Intercept)   -0.0001    0.0014    -0.0636   0.9493
          c1    0.7971    0.0077   103.6320   0.0000
tslag(c3, 1)    0.1766    0.0198     8.9057   0.0000
tslag(c1, 1)   -0.1580    0.0174    -9.0583   0.0000

Regression Diagnostics:
         R-Squared 0.8312
Adjusted R-Squared 0.8310
Durbin-Watson Stat 1.9865

Residual Diagnostics:
                 Stat   P-Value
Jarque-Bera 1620.5090    0.0000
  Ljung-Box  131.6048    0.0000

Residual standard error: 0.06785 on 2461 degrees of freedom
```

Let $\hat{\beta}_j$ be the jth element of $\hat{\boldsymbol{\beta}}$. When $k > 1$, the HC variance of $\hat{\beta}_j$ in Eq. (2.49) can be obtained by using an auxiliary regression. Let $\boldsymbol{x}_{-j,t}$ be the $(k-1)$-dimensional vector obtained by removing the element x_{jt} from $\boldsymbol{x}_t$. Consider the auxiliary regression

$$x_{jt} = \boldsymbol{x}'_{-j,t}\boldsymbol{\gamma} + v_t, \qquad t = 1, \ldots, T. \tag{2.51}$$

Let $\hat{v}_t$ be the least-squares residual of this auxiliary regression. It can be shown that

$$\text{Var}(\hat{\beta}_j)_{\text{HC}} = \frac{\sum_{t=1}^{T} \hat{e}_t^2 \hat{v}_t^2}{\left(\sum_{t=1}^{T} \hat{v}_t^2\right)^2},$$

where $\hat{e}_t$ is the residual of original regression in Eq. (2.48). The auxiliary regression is simply a step taken to achieve orthogonality between $\hat{v}_t$ and the rest of the regressors so that the formula in Eq. (2.49) can be simplified.

2.11 LONG-MEMORY MODELS

We have discussed that for a stationary time series the ACF decays exponentially to zero as lag increases. Yet for a unit-root nonstationary time series, it can be shown that the sample ACF converges to 1 for all fixed lags as the sample size increases; see Chan and Wei (1988) and Tiao and Tsay (1983). There exist some time series whose ACF decays slowly to zero at a polynomial rate as the lag increases. These processes are referred to as long-memory time series. One such example is the fractionally differenced process defined by

$$(1 - B)^d x_t = a_t, \qquad -0.5 < d < 0.5, \tag{2.52}$$

where $\{a_t\}$ is a white noise series. Properties of model (2.52) have been widely studied in the literature (e.g., Hosking, 1981). We now summarize some of these properties:

1. If $d < 0.5$, then x_t is a weakly stationary process and has the infinite MA representation

$$\begin{aligned} x_t &= a_t + \sum_{i=1}^{\infty} \psi_i a_{t-i} \quad \text{with} \quad \psi_k \\ &= \frac{d(1+d)\cdots(k-1+d)}{k!} = \frac{(k+d-1)!}{k!(d-1)!}. \end{aligned}$$

2. If $d > -0.5$, then x_t is invertible and has the infinite AR representation

$$\begin{aligned} x_t &= \sum_{i=1}^{\infty} \pi_i x_{t-i} + a_t \quad \text{with} \quad \pi_k \\ &= \frac{-d(1-d)\cdots(k-1-d)}{k!} = \frac{(k-d-1)!}{k!(-d-1)!}. \end{aligned}$$

3. For $-0.5 < d < 0.5$, the ACF of x_t is

$$\rho_k = \frac{d(1+d)\cdots(k-1+d)}{(1-d)(2-d)\cdots(k-d)}, \qquad k = 1, 2, \ldots.$$

   In particular, $\rho_1 = d/(1-d)$ and

$$\rho_k \approx \frac{(-d)!}{(d-1)!} k^{2d-1} \quad \text{as} \quad k \to \infty.$$

4. For $-0.5 < d < 0.5$, the PACF of x_t is $\phi_{k,k} = d/(k-d)$ for $k = 1, 2, \ldots$.
5. For $-0.5 < d < 0.5$, the spectral density function $f(\omega)$ of x_t, which is the Fourier transform of the ACF of x_t, satisfies

$$f(\omega) \sim \omega^{-2d}, \quad \text{as} \quad \omega \to 0, \tag{2.53}$$

   where $\omega \in [0, 2\pi]$ denotes the frequency.

Of particular interest here is the behavior of ACF of x_t when $d < 0.5$. The property says that $\rho_k \sim k^{2d-1}$, which decays at a polynomial, instead of exponential, rate. For this reason, such an x_t process is called a long-memory time series. A special characteristic of the spectral density function in Eq. (2.53) is that the spectrum diverges to infinity as $\omega \to 0$. However, the spectral density function of a stationary ARMA process is bounded for all $\omega \in [0, 2\pi]$.

Earlier we used the binomial theorem for noninteger powers

$$(1-B)^d = \sum_{k=0}^{\infty} (-1)^k \begin{pmatrix} d \\ k \end{pmatrix} B^k, \qquad \begin{pmatrix} d \\ k \end{pmatrix} = \frac{d(d-1)\cdots(d-k+1)}{k!}.$$

If the fractionally differenced series $(1-B)^d x_t$ follows an ARMA(p, q) model, then x_t is called an ARFIMA(p, d, q) process, which is a generalized ARIMA model by allowing for noninteger d.

In practice, if the sample ACF of a time series is not large in magnitude, but decays slowly, then the series may have long memory. As an illustration, Figure 2.22 shows the sample ACFs of the absolute series of daily simple returns for the CRSP value- and equal-weighted indexes from January 2, 1970, to December 31, 2008. The ACFs are relatively small in magnitude but decay very slowly; they appear to be significant at the 5% level even after 300 lags. For more information about the behavior of sample ACF of absolute return series, see Ding, Granger, and Engle (1993). For the pure fractionally differenced model in Eq. (2.52), one can estimate d using either a maximum-likelihood method or a regression method with logged periodogram at the lower frequencies. Finally, long-memory models have attracted some attention in the finance literature in part because of the work on fractional Brownian motion in the continuous-time models.

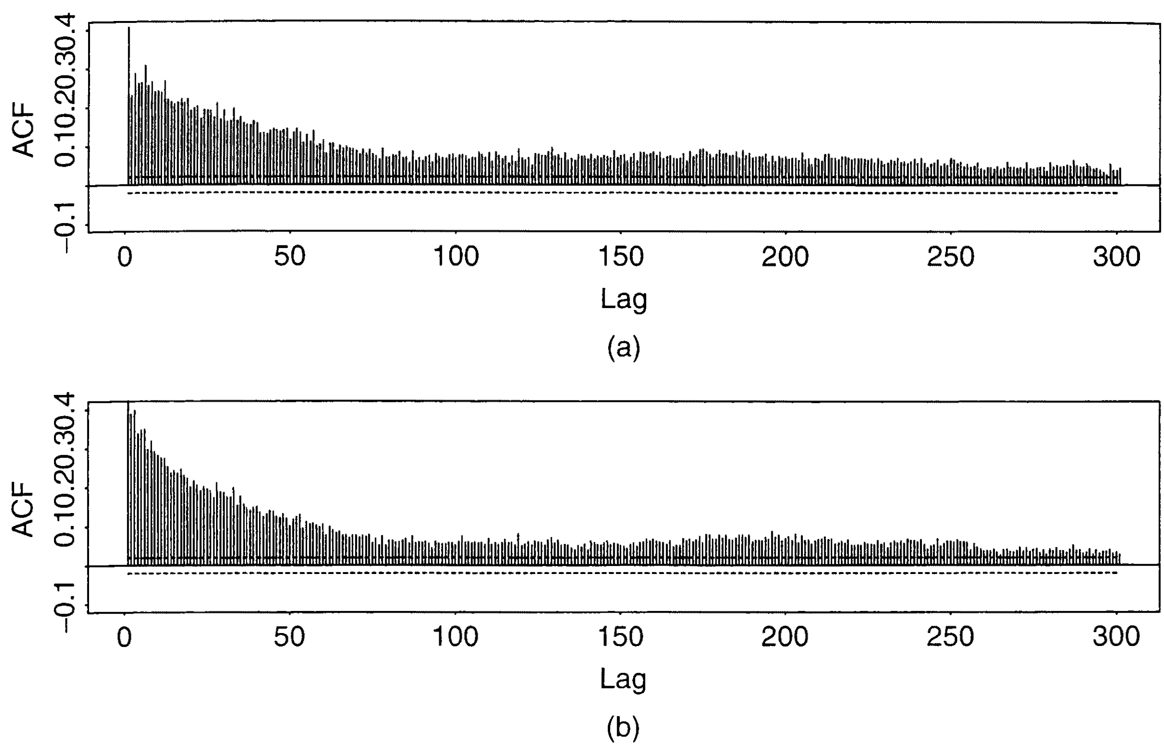

Figure 2.22 Sample autocorrelation function of absolute series of daily simple returns for CRSP value- and equal-weighted indexes: (a) value-weighted index return and (b) equal-weighted index return. Sample period is from January 2, 1970, to December 31, 2008.

APPENDIX: SOME SCA COMMANDS

In this appendix, we give the SCA commands used in Section 2.9. The 1-year maturity interest rates are in the file `w-gs1yr.txt` and the 3-year rates are in the file `w-gs3yr.txt`.

```
-- load the data into SCA, denote the data by rate1 and rate3.
input year,mom,day,rate1. file 'w-gs1yr.txt'
--
input year,mon,day,rate3. file 'w-gs3yr.txt'
-- specify a simple linear regression model.
tsm m1. model rate3=b0+(b1)rate1+noise.
-- estimate the specified model and store residual in r1.
estim m1. hold resi(r1).
-- compute 10 lags of residual acf.
acf r1. maxl 10.
-- difference the two series, denote the new series by c1t
   and c3t
diff old rate1,rate3. new c1t, c3t. compress.
-- specify a linear regression model for the differenced data
tsm m2. model c3t=h0+(h1)c1t+noise.
-- estimation
estim m2. hold resi(r2).
-- compute residual acf.
```

```
acf r2. maxl 10.
-- specify a regression model with time series errors.
tsm m3. model c3t=g0+(g1)c1t+(1)noise.
-- estimate the model using the exact likelihood method.
estim m3. method exact. hold resi(r3).
-- compute residual acf.
acf r3. maxl 10.
-- refine the model to include more MA lags.
tsm m4. model c3t=g0+(g1)c1t+(1,4,6,7)noise.
-- estimation
estim m4. method exact. hold resi(r4).
-- compute residual acf.
acf r4. maxl 10.
-- exit SCA
stop
```

EXERCISES

If not specifically specified, use 5% significance level to draw conclusions in the exercises.

2.1. Suppose that the simple return of a monthly bond index follows the MA(1) model

$$R_t = a_t + 0.2a_{t-1}, \qquad \sigma_a = 0.025.$$

Assume that $a_{100} = 0.01$. Compute the 1-step- and 2-step-ahead forecasts of the return at the forecast origin $t = 100$. What are the standard deviations of the associated forecast errors? Also compute the lag-1 and lag-2 autocorrelations of the return series.

2.2. Suppose that the daily log return of a security follows the model

$$r_t = 0.01 + 0.2r_{t-2} + a_t,$$

where $\{a_t\}$ is a Gaussian white noise series with mean zero and variance 0.02. What are the mean and variance of the return series r_t? Compute the lag-1 and lag-2 autocorrelations of r_t. Assume that $r_{100} = -0.01$, and $r_{99} = 0.02$. Compute the 1- and 2-step-ahead forecasts of the return series at the forecast origin $t = 100$. What are the associated standard deviations of the forecast errors?

2.3. Consider the monthly U.S. unemployment rate from January 1948 to March 2009 in the file `m-unrate.txt`. The data are seasonally adjusted and obtained from the Federal Reserve Bank of St Louis. Build a time series model for the series and use the model to forecast the unemployment rate for the April, May, June, and July of 2009. In addition, does the fitted model imply the existence of business cycles? Why? (Note that there are more than one model fits the data well. You only need an adequate model.)

2.4. Consider the monthly simple returns of the Decile 1, Decile 2, Decile 9, and Decile 10 of NYSE/AMEX/NASDAQ based on market capitalization. The data span is from January 1970 to December 2008, and the data are obtained from CRSP.

(a) For the return series of Decile 2 and Decile 10, test the null hypothesis that the first 12 lags of autocorrelations are zero at the 5% level. Draw your conclusion.

(b) Build an ARMA model for the return series of Decile 2. Perform model checking and write down the fitted model.

(c) Use the fitted ARMA model to produce 1- to 12-step-ahead forecasts of the series and the associated standard errors of forecasts.

2.5. Consider the daily simple returns of IBM stock from 1970 to 2008 in the file `d-ibm3dx7008.txt`. Compute the first 100 lags of ACF of the absolute series of daily simple returns of IBM stock. Is there evidence of long-range dependence? Why?

2.6. Consider the demand of electricity of a manufacturing sector in the United States. The data are logged, denote the demand of a fixed day of each month, and are in `power6.txt`. Build a time series model for the series and use the fitted model to produce 1- to 24-step-ahead forecasts.

2.7. Consider the daily simple returns of IBM stock, CRSP value-weighted index, CRSP equal-weighted index, and the S&P composite index from January 1980 to December 2008. The index returns include dividend distributions. The data file is `d-ibm3dxwkdays8008.txt`, which has 12 columns. The columns are (year, month, day, IBM, VW, EW, SP, M, T, W, H, F), where M, T, W, R, and F denotes indicator variables for Monday to Friday, respectively. Use a regression model to study the effects of trading days on the equal-weighted index returns. What is the fitted model? Are the weekday effects significant in the returns at the 5% level? Use the HAC estimator of the covariance matrix to obtain the t ratio of regression estimates. Does the HAC estimator change the conclusion of weekday effects? Are there serial correlations in the regression residuals? If yes, build a regression model with time series error to study weekday effects.

2.8. Consider the data set of the previous question, but focus on the daily simple returns of the S&P composite index. Perform the necessary data analysis and statistical tests using the 5% significance level to answer the following questions:

(a) Is there any weekday effect on the daily simple returns of the S&P composite index? You may employ a linear regression model to answer this question. Estimate the model, check its validity, and test the hypothesis that there is no Friday effect. Draw your conclusion.

(b) Check the residual serial correlations using Q(12) statistic. Are there any significant serial correlations in the residuals? If yes, build a regression model with time series errors for the data.

2.9. Now consider similar questions of the previous exercise for the IBM stock returns.

(a) Is there any weekday effect on the daily simple returns of IBM stock? Estimate your model and test the hypothesis that there is no Friday effect. Draw your conclusion.

(b) Are there serial correlations in the residuals? Use Q(12) to perform the test. Draw your conclusion.

(c) Refine the above model by using the technique of regression model with time series errors. In there a significant weekday effect based on the refined model?

2.10. Consider the weekly yields of Moody's Aaa and Baa seasoned bonds from January 5, 1962, to April 10, 2009. The data are obtained from the Federal Reserve Bank of St Louis. Weekly yields are averages of daily yields. Obtain the summary statistics (sample mean, standard deviation, skewness, excess kurtosis, minimum, and maximum) of the two yield series. Are the bond yields skewed? Do they have heavy tails? Answer the questions using 5% significance level.

2.11. Consider the monthly Aaa bond yields of the prior problem. Build a time series model for the series.

2.12. Again, consider the two bond yield series, that is, Aaa and Baa. What is the relationship between the two series? To answer this question, build a time series model using yields of Aaa bonds as the dependent variable and yields of Baa bonds as independent variable.

2.13. Consider the monthly log returns of CRSP equal-weighted index from January 1962 to December 1999 for 456 observations. You may obtain the data from CRSP directly or from the file `m-ew6299.txt` on the Web.

(a) Build an AR model for the series and check the fitted model.

(b) Build an MA model for the series and check the fitted model.

(c) Compute 1- and 2-step-ahead forecasts of the AR and MA models built in the previous two questions.

(d) Compare the fitted AR and MA models.

2.14. This problem is concerned with the dynamic relationship between the spot and futures prices of the S&P 500 index. The data file `sp5may.dat` has three columns: log(futures price), log(spot price), and cost-of-carry ($\times 100$). The data were obtained from the Chicago Mercantile Exchange for the S&P 500 stock index in May 1993 and its June futures contract. The time interval is 1 minute (intraday). Several authors used the data to study index futures arbitrage. Here we focus on the first two columns. Let f_t and s_t be the log prices of futures and spot, respectively. Consider $y_t = f_t - f_{t-1}$ and $x_t = s_t - s_{t-1}$. Build a regression model with time series errors between $\{y_t\}$ and $\{x_t\}$, with y_t being the dependent variable.

2.15. The quarterly gross domestic product implicit price deflator is often used as a measure of inflation. The file `q-gdpdef.txt` contains the data for the United States from the first quarter of 1947 to the last quarter of 2008. Data format is year, month, day, and deflator. The data are seasonally adjusted and equal to 100 for year 2000. Build an ARIMA model for the series and check the validity of the fitted model. Use the fitted model to predict the inflation for each quarter of 2009. The data are obtained from the Federal Reserve Bank of St Louis.

REFERENCES

Akaike, H. (1973). Information theory and an extension of the maximum likelihood principle. In B. N. Petrov and F. Csaki, (eds.). *2nd International Symposium on Information Theory*, pp. 267–281. Akademia Kiado, Budapest.

Box, G. E. P. and Pierce, D. (1970). Distribution of residual autocorrelations in autoregressive-integrated moving average time series models. *Journal of the American Statistical Association* **65**: 1509–1526.

Box, G. E. P., Jenkins, G. M., and Reinsel, G. C. (1994). *Time Series Analysis: Forecasting and Control*, 3rd ed. Prentice Hall, Englewood Cliffs, NJ.

Brockwell, P. J. and Davis, R. A. (1991). *Time Series: Theory and Methods*, 2nd ed. Springer, New York.

Brockwell, P. J. and Davis, R. A. (1996). *Introduction to Time Series and Forecasting*. Springer, New York.

Chan, N. H. and Wei, C. Z. (1988). Limiting distributions of least squares estimates of unstable autoregressive processes. *Annals of Statistics* **16**: 367–401.

Dickey, D. A. and Fuller, W. A. (1979). Distribution of the estimates for autoregressive time series with a unit root. *Journal of the American Statistical Association* **74**: 427–431.

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

Eicker, F. (1967). Limit theorems for regression with unequal and dependent Errors. In L. LeCam and J. Neyman (eds.). *Proceedings of the 5th Berkeley Symposium on Mathematical Statistics and Probability*. University of California Press, Berkeley.

Fuller, W. A. (1976). *Introduction to Statistical Time Series*, Wiley, New York.

Greene, W. H. (2003). *Econometric Analysis*, 5th ed. Prentice-Hall, Upper Saddle River, NJ.

Hosking, J. R. M. (1981). Fractional differencing. *Biometrika* **68**: 165–176.

Ljung, G. and Box, G. E. P. (1978). On a measure of lack of fit in time series models. *Biometrika* **66**: 67–72.

Newey, W. and West, K. (1987). A simple positive semidefinite, heteroscedasticity and autocorrelation consistent covariance matrix. *Econometrica* **55**: 863–898.

Phillips, P. C. B. (1987). Time series regression with a unit root. *Econometrica* **55**: 277–301.

Shumway, R. H. and Stoffer, D. S. (2000). *Time Series Analysis and Its Applications*. Springer, New York.

Tiao, G. C. and Tsay, R. S. (1983). Consistency properties of least squares estimates of autoregressive parameters in ARMA models. *Annals of Statistics* **11**: 856–871.

Tsay, R. S. and Tiao, G. C. (1984). Consistent estimates of autoregressive parameters and extended sample autocorrelation function for stationary and nonstationary ARMA models. *Journal of the American Statistical Association* **79**: 84–96.

White, H. (1980). A heteroscedasticity consistent covariance matrix estimator and a direct test for heteroscedasticity. *Econometrica* **48**: 827–838.