CHAPTER 1

Financial Time Series and Their Characteristics

Financial time series analysis is concerned with the theory and practice of asset valuation over time. It is a highly empirical discipline, but like other scientific fields theory forms the foundation for making inference. There is, however, a key feature that distinguishes financial time series analysis from other time series analysis. Both financial theory and its empirical time series contain an element of uncertainty. For example, there are various definitions of asset volatility, and for a stock return series, the volatility is not directly observable. As a result of the added uncertainty, statistical theory and methods play an important role in financial time series analysis.

The objective of this book is to provide some knowledge of financial time series, introduce some statistical tools useful for analyzing these series, and gain experience in financial applications of various econometric methods. We begin with the basic concepts of asset returns and a brief introduction to the processes to be discussed throughout the book. Chapter 2 reviews basic concepts of linear time series analysis such as stationarity and autocorrelation function, introduces simple linear models for handling serial dependence of the series, and discusses regression models with time series errors, seasonality, unit-root nonstationarity, and long-memory processes. The chapter also provides methods for consistent estimation of the covariance matrix in the presence of conditional heteroscedasticity and serial correlations. Chapter 3 focuses on modeling conditional heteroscedasticity (i.e., the conditional variance of an asset return). It discusses various econometric models developed recently to describe the evolution of volatility of an asset return over time. The chapter also discusses alternative methods to volatility modeling, including use of high-frequency transactions data and daily high and low prices of an asset. In Chapter 4, we address nonlinearity in financial time series, introduce test statistics that can discriminate nonlinear series from linear ones, and discuss several nonlinear models. The chapter also introduces nonparametric estimation

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

methods and neural networks and shows various applications of nonlinear models in finance. Chapter 5 is concerned with analysis of high-frequency financial data, the effects of market microstructure, and some applications of high-frequency finance. It shows that nonsynchronous trading and bid–ask bounce can introduce serial correlations in a stock return. It also studies the dynamic of time duration between trades and some econometric models for analyzing transactions data. In Chapter 6, we introduce continuous-time diffusion models and Ito's lemma. Black–Scholes option pricing formulas are derived, and a simple jump diffusion model is used to capture some characteristics commonly observed in options markets. Chapter 7 discusses extreme value theory, heavy-tailed distributions, and their application to financial risk management. In particular, it discusses various methods for calculating value at risk and expected shortfall of a financial position. Chapter 8 focuses on multivariate time series analysis and simple multivariate models with emphasis on the lead–lag relationship between time series. The chapter also introduces cointegration, some cointegration tests, and threshold cointegration and applies the concept of cointegration to investigate arbitrage opportunity in financial markets, including pairs trading. Chapter 9 discusses ways to simplify the dynamic structure of a multivariate series and methods to reduce the dimension. It introduces and demonstrates three types of factor model to analyze returns of multiple assets. In Chapter 10, we introduce multivariate volatility models, including those with time-varying correlations, and discuss methods that can be used to reparameterize a conditional covariance matrix to satisfy the positiveness constraint and reduce the complexity in volatility modeling. Chapter 11 introduces state-space models and the Kalman filter and discusses the relationship between state-space models and other econometric models discussed in the book. It also gives several examples of financial applications. Finally, in Chapter 12, we introduce some Markov chain Monte Carlo (MCMC) methods developed in the statistical literature and apply these methods to various financial research problems, such as the estimation of stochastic volatility and Markov switching models.

The book places great emphasis on application and empirical data analysis. Every chapter contains real examples and, in many occasions, empirical characteristics of financial time series are used to motivate the development of econometric models. Computer programs and commands used in data analysis are provided when needed. In some cases, the programs are given in an appendix. Many real data sets are also used in the exercises of each chapter.

1.1 ASSET RETURNS

Most financial studies involve returns, instead of prices, of assets. Campbell, Lo, and MacKinlay (1997) give two main reasons for using returns. First, for average investors, return of an asset is a complete and scale-free summary of the investment opportunity. Second, return series are easier to handle than price series because the former have more attractive statistical properties. There are, however, several definitions of an asset return.

Let P_t be the price of an asset at time index t. We discuss some definitions of returns that are used throughout the book. Assume for the moment that the asset pays no dividends.

One-Period Simple Return

Holding the asset for one period from date $t-1$ to date t would result in a *simple gross return*:

$$1 + R_t = \frac{P_t}{P_{t-1}} \qquad \text{or} \qquad P_t = P_{t-1}(1 + R_t). \tag{1.1}$$

The corresponding one-period *simple net return* or *simple return* is

$$R_t = \frac{P_t}{P_{t-1}} - 1 = \frac{P_t - P_{t-1}}{P_{t-1}}. \tag{1.2}$$

Multiperiod Simple Return

Holding the asset for k periods between dates $t-k$ and t gives a k-period simple gross return:

$$\begin{aligned} 1 + R_t[k] = \frac{P_t}{P_{t-k}} &= \frac{P_t}{P_{t-1}} \times \frac{P_{t-1}}{P_{t-2}} \times \cdots \times \frac{P_{t-k+1}}{P_{t-k}} \\ &= (1 + R_t)(1 + R_{t-1}) \cdots (1 + R_{t-k+1}) \\ &= \prod_{j=0}^{k-1} (1 + R_{t-j}). \end{aligned}$$

Thus, the k-period simple gross return is just the product of the k one-period simple gross returns involved. This is called a compound return. The k-period simple net return is $R_t[k] = (P_t - P_{t-k})/P_{t-k}$.

In practice, the actual time interval is important in discussing and comparing returns (e.g., monthly return or annual return). If the time interval is not given, then it is implicitly assumed to be one year. If the asset was held for k years, then the annualized (average) return is defined as

$$\text{Annualized}\{R_t[k]\} = \left[\prod_{j=0}^{k-1} (1 + R_{t-j})\right]^{1/k} - 1.$$

This is a geometric mean of the k one-period simple gross returns involved and can be computed by

$$\text{Annualized}\{R_t[k]\} = \exp\left[\frac{1}{k}\sum_{j=0}^{k-1} \ln(1 + R_{t-j})\right] - 1,$$

where $\exp(x)$ denotes the exponential function and $\ln(x)$ is the natural logarithm of the positive number x. Because it is easier to compute arithmetic average than geometric mean and the one-period returns tend to be small, one can use a first-order Taylor expansion to approximate the annualized return and obtain

$$\text{Annualized}\,\{R_t[k]\} \approx \frac{1}{k}\sum_{j=0}^{k-1} R_{t-j}. \tag{1.3}$$

Accuracy of the approximation in Eq. (1.3) may not be sufficient in some applications, however.

Continuous Compounding

Before introducing continuously compounded return, we discuss the effect of compounding. Assume that the interest rate of a bank deposit is 10% per annum and the initial deposit is \$1.00. If the bank pays interest once a year, then the net value of the deposit becomes $\$1(1 + 0.1) = \$1.1$ one year later. If the bank pays interest semiannually, the 6-month interest rate is $10\%/2 = 5\%$ and the net value is $\$1(1 + 0.1/2)^2 = \$1.1025$ after the first year. In general, if the bank pays interest m times a year, then the interest rate for each payment is $10\%/m$ and the net value of the deposit becomes $\$1(1 + 0.1/m)^m$ one year later. Table 1.1 gives the results for some commonly used time intervals on a deposit of \$1.00 with interest rate of 10% per annum. In particular, the net value approaches \$1.1052, which is obtained by exp(0.1) and referred to as the result of continuous compounding. The effect of compounding is clearly seen.

In general, the net asset value A of continuous compounding is

$$A = C\,\exp(r \times n), \tag{1.4}$$

where r is the interest rate per annum, C is the initial capital, and n is the number of years. From Eq. (1.4), we have

$$C = A\,\exp(-r \times n), \tag{1.5}$$

TABLE 1.1 Illustration of Effects of Compounding: Time Interval Is 1 Year and Interest Rate Is 10% per Annum

Type	Number of Payments	Interest Rate per Period	Net Value
Annual	1	0.1	\$1.10000
Semiannual	2	0.05	\$1.10250
Quarterly	4	0.025	\$1.10381
Monthly	12	0.0083	\$1.10471
Weekly	52	0.1/52	\$1.10506
Daily	365	0.1/365	\$1.10516
Continuously	∞		\$1.10517

which is referred to as the *present value* of an asset that is worth A dollars n years from now, assuming that the continuously compounded interest rate is r per annum.

Continuously Compounded Return

The natural logarithm of the simple gross return of an asset is called the continuously compounded return or *log return*:

$$r_t = \ln(1 + R_t) = \ln \frac{P_t}{P_{t-1}} = p_t - p_{t-1}, \tag{1.6}$$

where $p_t = \ln(P_t)$. Continuously compounded returns r_t enjoy some advantages over the simple net returns R_t. First, consider multiperiod returns. We have

$$\begin{aligned} r_t[k] &= \ln(1 + R_t[k]) = \ln[(1 + R_t)(1 + R_{t-1}) \cdots (1 + R_{t-k+1})] \\ &= \ln(1 + R_t) + \ln(1 + R_{t-1}) + \cdots + \ln(1 + R_{t-k+1}) \\ &= r_t + r_{t-1} + \cdots + r_{t-k+1}. \end{aligned}$$

Thus, the continuously compounded multiperiod return is simply the sum of continuously compounded one-period returns involved. Second, statistical properties of log returns are more tractable.

Portfolio Return

The simple net return of a portfolio consisting of N assets is a weighted average of the simple net returns of the assets involved, where the weight on each asset is the percentage of the portfolio's value invested in that asset. Let p be a portfolio that places weight w_i on asset i. Then the simple return of p at time t is $R_{p,t} = \sum_{i=1}^{N} w_i R_{it}$, where R_{it} is the simple return of asset i.

The continuously compounded returns of a portfolio, however, do not have the above convenient property. If the simple returns R_{it} are all small in magnitude, then we have $r_{p,t} \approx \sum_{i=1}^{N} w_i r_{it}$, where $r_{p,t}$ is the continuously compounded return of the portfolio at time t. This approximation is often used to study portfolio returns.

Dividend Payment

If an asset pays dividends periodically, we must modify the definitions of asset returns. Let D_t be the dividend payment of an asset between dates $t-1$ and t and P_t be the price of the asset at the end of period t. Thus, dividend is not included in P_t. Then the simple net return and continuously compounded return at time t become

$$R_t = \frac{P_t + D_t}{P_{t-1}} - 1, \qquad r_t = \ln(P_t + D_t) - \ln(P_{t-1}).$$

Excess Return

Excess return of an asset at time t is the difference between the asset's return and the return on some reference asset. The reference asset is often taken to be riskless such as a short-term U.S. Treasury bill return. The simple excess return and log excess return of an asset are then defined as

$$Z_t = R_t - R_{0t}, \qquad z_t = r_t - r_{0t}, \tag{1.7}$$

where R_{0t} and r_{0t} are the simple and log returns of the reference asset, respectively. In the finance literature, the excess return is thought of as the payoff on an arbitrage portfolio that goes long in an asset and short in the reference asset with no net initial investment.

Remark. A long financial position means owning the asset. A short position involves selling an asset one does not own. This is accomplished by borrowing the asset from an investor who has purchased it. At some subsequent date, the short seller is obligated to buy exactly the same number of shares borrowed to pay back the lender. Because the repayment requires equal shares rather than equal dollars, the short seller benefits from a decline in the price of the asset. If cash dividends are paid on the asset while a short position is maintained, these are paid to the buyer of the short sale. The short seller must also compensate the lender by matching the cash dividends from his own resources. In other words, the short seller is also obligated to pay cash dividends on the borrowed asset to the lender. □

Summary of Relationship

The relationships between simple return R_t and continuously compounded (or log) return r_t are

$$r_t = \ln(1 + R_t), \qquad R_t = e^{r_t} - 1.$$

If the returns R_t and r_t are in percentages, then

$$r_t = 100 \ln\left(1 + \frac{R_t}{100}\right), \qquad R_t = 100\left(e^{r_t/100} - 1\right).$$

Temporal aggregation of the returns produces

$$1 + R_t[k] = (1 + R_t)(1 + R_{t-1}) \cdots (1 + R_{t-k+1}),$$
$$r_t[k] = r_t + r_{t-1} + \cdots + r_{t-k+1}.$$

If the continuously compounded interest rate is r per annum, then the relationship between present and future values of an asset is

$$A = C \exp(r \times n), \qquad C = A \exp(-r \times n).$$

Example 1.1. If the monthly log return of an asset is 4.46%, then the corresponding monthly simple return is 100[exp(4.46/100) − 1] = 4.56%. Also, if the monthly log returns of the asset within a quarter are 4.46%, −7.34%, and 10.77%, respectively, then the quarterly log return of the asset is (4.46 − 7.34 + 10.77)% = 7.89%.

1.2 DISTRIBUTIONAL PROPERTIES OF RETURNS

To study asset returns, it is best to begin with their distributional properties. The objective here is to understand the behavior of the returns across assets and over time. Consider a collection of N assets held for T time periods, say, $t = 1, \ldots, T$. For each asset i, let r_{it} be its log return at time t. The log returns under study are $\{r_{it}; i = 1, \ldots, N; t = 1, \ldots, T\}$. One can also consider the simple returns $\{R_{it}; i = 1, \ldots, N; t = 1, \ldots, T\}$ and the log excess returns $\{z_{it}; i = 1, \ldots, N; t = 1, \ldots, T\}$.

1.2.1 Review of Statistical Distributions and Their Moments

We briefly review some basic properties of statistical distributions and the moment equations of a random variable. Let R^k be the k-dimensional Euclidean space. A point in R^k is denoted by $\boldsymbol{x} \in R^k$. Consider two random vectors $\boldsymbol{X} = (X_1, \ldots, X_k)'$ and $\boldsymbol{Y} = (Y_1, \ldots, Y_q)'$. Let $P(\boldsymbol{X} \in A, \boldsymbol{Y} \in B)$ be the probability that $\boldsymbol{X}$ is in the subspace $A \subset R^k$ and $\boldsymbol{Y}$ is in the subspace $B \subset R^q$. For most of the cases considered in this book, both random vectors are assumed to be continuous.

Joint Distribution

The function

$$F_{X,Y}(\boldsymbol{x}, \boldsymbol{y}; \boldsymbol{\theta}) = P(\boldsymbol{X} \leq \boldsymbol{x}, \boldsymbol{Y} \leq \boldsymbol{y}; \boldsymbol{\theta}),$$

where $\boldsymbol{x} \in R^p$, $\boldsymbol{y} \in R^q$, and the inequality $\leq$ is a component-by-component operation, is a joint distribution function of $\boldsymbol{X}$ and $\boldsymbol{Y}$ with parameter $\boldsymbol{\theta}$. Behavior of $\boldsymbol{X}$ and $\boldsymbol{Y}$ is characterized by $F_{X,Y}(\boldsymbol{x}, \boldsymbol{y}; \boldsymbol{\theta})$. If the joint probability density function $f_{x,y}(\boldsymbol{x}, \boldsymbol{y}; \boldsymbol{\theta})$ of $\boldsymbol{X}$ and $\boldsymbol{Y}$ exists, then

$$F_{X,Y}(\boldsymbol{x}, \boldsymbol{y}; \boldsymbol{\theta}) = \int_{-\infty}^{\boldsymbol{x}} \int_{-\infty}^{\boldsymbol{y}} f_{x,y}(\boldsymbol{w}, \boldsymbol{z}; \boldsymbol{\theta}) \, d\boldsymbol{z} \, d\boldsymbol{w}.$$

In this case, $\boldsymbol{X}$ and $\boldsymbol{Y}$ are continuous random vectors.

Marginal Distribution

The marginal distribution of $\boldsymbol{X}$ is given by

$$F_X(\boldsymbol{x}; \boldsymbol{\theta}) = F_{X,Y}(\boldsymbol{x}, \infty, \cdots, \infty; \boldsymbol{\theta}).$$

Thus, the marginal distribution of $\boldsymbol{X}$ is obtained by integrating out $\boldsymbol{Y}$. A similar definition applies to the marginal distribution of $\boldsymbol{Y}$.

If $k = 1$, X is a scalar random variable and the distribution function becomes

$$F_X(x) = P(X \leq x; \boldsymbol{\theta}),$$

which is known as the cumulative distribution function (CDF) of X. The CDF of a random variable is nondecreasing [i.e., $F_X(x_1) \leq F_X(x_2)$ if $x_1 \leq x_2$] and satisfies $F_X(-\infty) = 0$ and $F_X(\infty) = 1$. For a given probability p, the smallest real number x_p such that $p \leq F_X(x_p)$ is called the 100pth quantile of the random variable X. More specifically,

$$x_p = \inf_x \{x | p \leq F_X(x)\}.$$

We use the CDF to compute the p value of a test statistic in the book.

Conditional Distribution

The conditional distribution of $\boldsymbol{X}$ given $\boldsymbol{Y} \leq \boldsymbol{y}$ is given by

$$F_{X|Y\leq y}(\boldsymbol{x}; \boldsymbol{\theta}) = \frac{P(\boldsymbol{X} \leq \boldsymbol{x}, \boldsymbol{Y} \leq \boldsymbol{y}; \boldsymbol{\theta})}{P(\boldsymbol{Y} \leq \boldsymbol{y}; \boldsymbol{\theta})}.$$

If the probability density functions involved exist, then the conditional density of $\boldsymbol{X}$ given $\boldsymbol{Y} = \boldsymbol{y}$ is

$$f_{x|y}(\boldsymbol{x}; \boldsymbol{\theta}) = \frac{f_{x,y}(\boldsymbol{x}, \boldsymbol{y}; \boldsymbol{\theta})}{f_y(\boldsymbol{y}; \boldsymbol{\theta})}, \tag{1.8}$$

where the marginal density function $f_y(\boldsymbol{y}; \boldsymbol{\theta})$ is obtained by

$$f_y(\boldsymbol{y}; \boldsymbol{\theta}) = \int_{-\infty}^{\infty} f_{x,y}(\boldsymbol{x}, \boldsymbol{y}; \boldsymbol{\theta})\, d\boldsymbol{x}.$$

From Eq. (1.8), the relation among joint, marginal, and conditional distributions is

$$f_{x,y}(\boldsymbol{x}, \boldsymbol{y}; \boldsymbol{\theta}) = f_{x|y}(\boldsymbol{x}; \boldsymbol{\theta}) \times f_y(\boldsymbol{y}; \boldsymbol{\theta}). \tag{1.9}$$

This identity is used extensively in time series analysis (e.g., in maximum-likelihood estimation). Finally, $\boldsymbol{X}$ and $\boldsymbol{Y}$ are independent random vectors if and only if $f_{x|y}(\boldsymbol{x}; \boldsymbol{\theta}) = f_x(\boldsymbol{x}; \boldsymbol{\theta})$. In this case, $f_{x,y}(\boldsymbol{x}, \boldsymbol{y}; \boldsymbol{\theta}) = f_x(\boldsymbol{x}; \boldsymbol{\theta}) f_y(\boldsymbol{y}; \boldsymbol{\theta})$.

Moments of a Random Variable

The ℓth moment of a continuous random variable X is defined as

$$m'_\ell = E(X^\ell) = \int_{-\infty}^{\infty} x^\ell f(x)\, dx,$$

where E stands for expectation and $f(x)$ is the probability density function of X. The first moment is called the *mean* or *expectation* of X. It measures the central location of the distribution. We denote the mean of X by μ_x. The ℓth central moment of X is defined as

$$m_\ell = E[(X - \mu_x)^\ell] = \int_{-\infty}^{\infty} (x - \mu_x)^\ell f(x)\, dx$$

provided that the integral exists. The second central moment, denoted by σ_x^2, measures the variability of X and is called the *variance* of X. The positive square root, σ_x, of variance is the *standard deviation* of X. The first two moments of a random variable uniquely determine a normal distribution. For other distributions, higher order moments are also of interest.

The third central moment measures the symmetry of X with respect to its mean, whereas the fourth central moment measures the tail behavior of X. In statistics, *skewness* and *kurtosis*, which are normalized third and fourth central moments of X, are often used to summarize the extent of asymmetry and tail thickness. Specifically, the skewness and kurtosis of X are defined as

$$S(x) = E\left[\frac{(X-\mu_x)^3}{\sigma_x^3}\right], \qquad K(x) = E\left[\frac{(X-\mu_x)^4}{\sigma_x^4}\right].$$

The quantity $K(x) - 3$ is called the *excess kurtosis* because $K(x) = 3$ for a normal distribution. Thus, the excess kurtosis of a normal random variable is zero. A distribution with positive excess kurtosis is said to have heavy tails, implying that the distribution puts more mass on the tails of its support than a normal distribution does. In practice, this means that a random sample from such a distribution tends to contain more extreme values. Such a distribution is said to be *leptokurtic*. On the other hand, a distribution with negative excess kurtosis has short tails (e.g., a uniform distribution over a finite interval). Such a distribution is said to be *platykurtic*.

In application, skewness and kurtosis can be estimated by their sample counterparts. Let $\{x_1, \ldots, x_T\}$ be a random sample of X with T observations. The sample mean is

$$\hat{\mu}_x = \frac{1}{T}\sum_{t=1}^{T} x_t, \tag{1.10}$$

the sample variance is

$$\hat{\sigma}_x^2 = \frac{1}{T-1}\sum_{t=1}^{T}(x_t - \hat{\mu}_x)^2, \tag{1.11}$$

the sample skewness is

$$\hat{S}(x) = \frac{1}{(T-1)\hat{\sigma}_x^3}\sum_{t=1}^{T}(x_t - \hat{\mu}_x)^3, \tag{1.12}$$

and the sample kurtosis is

$$\hat{K}(x) = \frac{1}{(T-1)\hat{\sigma}_x^4}\sum_{t=1}^{T}(x_t - \hat{\mu}_x)^4. \tag{1.13}$$

Under the normality assumption, $\hat{S}(x)$ and $\hat{K}(x) - 3$ are distributed asymptotically as normal with zero mean and variances $6/T$ and $24/T$, respectively; see Snedecor and Cochran (1980, p. 78). These asymptotic properties can be used to test the normality of asset returns. Given an asset return series $\{r_1, \ldots, r_T\}$, to test the skewness of the returns, we consider the null hypothesis $H_0 : S(r) = 0$ versus the alternative hypothesis $H_a : S(r) \neq 0$. The t-ratio statistic of the sample skewness in Eq. (1.12) is

$$t = \frac{\hat{S}(r)}{\sqrt{6/T}}.$$

The decision rule is as follows. Reject the null hypothesis at the α significance level, if $|t| > Z_{\alpha/2}$, where $Z_{\alpha/2}$ is the upper $100(\alpha/2)$th quantile of the standard normal distribution. Alternatively, one can compute the p value of the test statistic t and reject H_0 if and only if the p value is less than α.

Similarly, one can test the excess kurtosis of the return series using the hypotheses $H_0 : K(r) - 3 = 0$ versus $H_a : K(r) - 3 \neq 0$. The test statistic is

$$t = \frac{\hat{K}(r) - 3}{\sqrt{24/T}},$$

which is asymptotically a standard normal random variable. The decision rule is to reject H_0 if and only if the p value of the test statistic is less than the significance level α. Jarque and Bera (1987) (JB) combine the two prior tests and use the test statistic

$$\text{JB} = \frac{\hat{S}^2(r)}{6/T} + \frac{[\hat{K}(r) - 3]^2}{24/T},$$

which is asymptotically distributed as a chi-squared random variable with 2 degrees of freedom, to test for the normality of r_t. One rejects H_0 of normality if the p value of the JB statistic is less than the significance level.

Example 1.2. Consider the daily simple returns of the International Business Machines (IBM) stock used in Table 1.2. The sample skewness and kurtosis of the returns are parts of the descriptive (or summary) statistics that can be obtained easily using various statistical software packages. Both R and S-Plus are used in the demonstration, where `d-ibm3dx7008.txt` is the data file name. Note that in R the *kurtosis* denotes excess kurtosis. From the output, the excess kurtosis is high, indicating that the daily simple returns of IBM stock have heavy tails. To test the symmetry of return distribution, we use the test statistic

$$t = \frac{0.0614}{\sqrt{6/9845}} = \frac{0.0614}{0.0247} = 2.49,$$

which gives a p value of about 0.013, indicating that the daily simple returns of IBM stock are significantly skewed to the right at the 5% level.

TABLE 1.2 Descriptive Statistics for Daily and Monthly Simple and Log Returns of Selected Indexes and Stocks[a]

Security	Start	Size	Mean	Standard Deviation	Skewness	Excess Kurtosis	Minimum	Maximum
				Daily Simple Returns (%)				
SP	70/01/02	9845	0.029	1.056	−0.73	22.81	−20.47	11.58
VW	70/01/02	9845	0.040	1.004	−0.62	18.02	−17.13	11.52
EW	70/01/02	9845	0.076	0.814	−0.77	17.08	−10.39	10.74
IBM	70/01/02	9845	0.040	1.693	0.06	9.92	−22.96	13.16
Intel	72/12/15	9096	0.108	2.891	−0.15	6.13	−29.57	26.38
3M	670/01/02	9845	0.045	1.482	−0.36	13.34	−25.98	11.54
Microsoft	86/03/14	5752	0.123	2.359	−0.13	9.92	−30.12	19.57
Citi-Grp	86/10/30	5592	0.067	2.602	1.80	55.25	−26.41	57.82
				Daily Log Returns (%)				
SP	70/01/02	9845	0.023	1.062	−1.17	30.20	−22.90	10.96
VW	70/01/02	9845	0.035	1.008	−0.94	21.56	−18.80	10.90
EW	70/01/02	9845	0.072	0.816	−1.00	17.76	−10.97	10.20
IBM	70/01/02	9845	0.026	1.694	−0.27	12.17	−26.09	12.37
Intel	72/12/15	9096	0.066	2.905	−0.54	7.81	−35.06	23.41
3M	70/01/02	9845	0.034	1.488	−0.78	20.57	−30.08	10.92
Microsoft	86/03/14	5752	0.095	2.369	−0.63	14.23	−35.83	17.87
Citi-Grp	86/10/30	5592	0.033	2.575	0.22	33.19	−30.66	45.63
				Monthly Simple Returns (%)				
SP	26/01	996	0.58	5.53	0.32	9.47	−29.94	42.22
VW	26/01	996	0.89	5.43	0.15	7.69	−29.01	38.37
EW	26/01	996	1.22	7.40	1.52	14.94	−31.28	66.59
IBM	26/01	996	1.35	7.15	0.44	3.43	−26.19	47.06
Intel	73/01	432	2.21	12.85	0.32	2.70	−44.87	62.50
3M	46/02	755	1.24	6.45	0.22	0.98	−27.83	25.80
Microsoft	86/04	273	2.62	11.08	0.66	1.96	−34.35	51.55
Citi-Grp	86/11	266	1.17	9.75	−0.47	1.77	−39.27	26.08
				Monthly Log Returns (%)				
SP	26/01	996	0.43	5.54	−0.52	7.93	−35.58	35.22
VW	26/01	996	0.74	5.43	−0.58	6.85	−34.22	32.47
EW	26/01	996	0.96	7.14	0.25	8.55	−37.51	51.04
IBM	26/01	996	1.09	7.03	−0.07	2.62	−30.37	38.57
Intel	73/01	432	1.39	12.80	−0.55	3.06	−59.54	48.55
3M	46/02	755	1.03	6.37	−0.08	1.25	−32.61	22.95
Microsoft	86/04	273	2.01	10.66	0.10	1.59	−42.09	41.58
Citi-Grp	86/11	266	0.68	10.09	−1.09	3.76	−49.87	23.18

[a]Returns are in percentages and the sample period ends on December 31, 2008. The statistics are defined in eqs. (1.10)–(1.13), and VW, EW and SP denote value-weighted, equal-weighted, and S&P composite index.

R Demonstration

In the following program code > is the prompt character and % denotes explanation:

```
> library(fBasics)  % Load the package fBasics.
> da=read.table("d-ibm3dx7008.txt",header=T)  % Load the data.
% header=T means 1st row of the data file contains
% variable names. The default is header=F, i.e., no names.

> dim(da)   % Find size of the data: 9845 rows and 5 columns.
[1] 9845    5
> da[1,]   % See the first row of the data
      Date      rtn   vwretd  ewretd   sprtrn % column names
1 19700102 0.000686 0.012137 0.03345 0.010211

> ibm=da[,2]   % Obtain IBM simple returns
> sibm=ibm*100   % Percentage simple returns

> basicStats(sibm)   % Compute the summary statistics
                   sibm
nobs        9845.000000 % Sample size
NAs            0.000000 % Number of missing values
Minimum      -22.963000
Maximum       13.163600
1. Quartile   -0.857100 % 25th percentile
3. Quartile    0.883300 % 75th percentile
Mean           0.040161 % Sample mean
Median         0.000000 % Sample median
Sum          395.387600 % Sum of the percentage simple returns
SE Mean        0.017058 % Standard error of the sample mean
LCL Mean       0.006724 % Lower bound of 95% conf.
                        % interval for mean
UCL Mean       0.073599 % Upper bound of 95% conf.
                        % interval for mean
Variance       2.864705 % Sample variance
Stdev          1.692544 % Sample standard error
Skewness       0.061399 % Sample skewness
Kurtosis       9.916359 % Sample excess kurtosis.

% Alternatively, one can use individual commands as follows:
> mean(sibm)
[1] 0.04016126
> var(sibm)
[1] 2.864705
> sqrt(var(sibm)) % Standard deviation
[1] 1.692544
> skewness(sibm)
[1] 0.06139878
attr(,"method")
```

```
[1] "moment"
> kurtosis(sibm)
[1] 9.91636
attr(,"method")
[1] "excess"

% Simple tests
> s1=skewness(sibm)
> t1=s1/sqrt(6/9845) % Compute test statistic
> t1
[1] 2.487093
> pv=2*(1-pnorm(t1)) % Compute p-value.
> pv
[1] 0.01287919

% Turn to log returns in percentages
> libm=log(ibm+1)*100
> t.test(libm)  % Test mean being zero.
        One Sample t-test
data:  libm
t = 1.5126, df = 9844, p-value = 0.1304
alternative hypothesis: true mean is not equal to 0
95 percent confidence interval:
 -0.007641473  0.059290531
% The result shows that the hypothesis of zero expected return
% cannot be rejected at the 5% or 10% level.

> normalTest(libm,method='jb') % Normality test
Title:
 Jarque - Bera Normality Test

Test Results:
  STATISTIC:
    X-squared: 60921.9343
  P VALUE:
    Asymptotic p Value: < 2.2e-16
% The result shows the normality for log-return is rejected.
```

S-Plus Demonstration

In the following program code > is the prompt character and % marks explanation:

```
> module(finmetrics) % Load the Finmetrics module.
> da=read.table("d-ibm3dx7008.txt",header=T) % Load data.
> dim(da)      % Obtain the size of the data set.
[1] 9845    5
> da[1,]     % See the first row of the data
      Date      rtn   vwretd  ewretd   sprtrn
1 19700102 0.000686 0.012137 0.03345 0.010211
```

```
> sibm=da[,2]*100  % Obtain percentage simple returns of
                   % IBM stock.
> summaryStats(sibm) % Obtain summary statistics

Sample Quantiles:
    min      1Q median     3Q   max
 -22.96 -0.8571      0 0.8833 13.16
Sample Moments:
    mean   std skewness kurtosis
 0.04016 1.693  0.06141    12.92
Number of Observations:  9845
% simple tests
> s1=skewness(sibm) % Compute skewness
> t=s1/sqrt(6/9845) % Perform test of skewness
> t
[1] 2.487851
> pv=2*(1-pnorm(t)) % Calculate p-value.
> pv
[1] 0.01285177

> libm=log(da[,2]+1)*100 % Turn to log-return
> t.test(libm)  % Test expected return being zero.
        One-sample t-Test
data:  libm
t = 1.5126, df = 9844, p-value = 0.1304
alternative hypothesis:  mean is not equal to 0
95 percent confidence interval:
 -0.007641473  0.059290531

> normalTest(libm,method='jb') % Normality test
Test for Normality: Jarque-Bera
Null Hypothesis: data is normally distributed

Test Stat 60921.93
  p.value     0.00
Dist. under Null: chi-square with 2 degrees of freedom
   Total Observ.: 9845
```

Remark. In S-Plus, kurtosis is the regular kurtosis, not excess kurtosis. That is, S-Plus does not subtract 3 from the sample kurtosis. Also, in many cases R and S-Plus use the same commands. □

1.2.2 Distributions of Returns

The most general model for the log returns $\{r_{it}; i = 1, \ldots, N; t = 1, \ldots, T\}$ is its joint distribution function:

$$F_r(r_{11}, \ldots, r_{N1}; r_{12}, \ldots, r_{N2}; \ldots; r_{1T}, \ldots, r_{NT}; \boldsymbol{Y}; \boldsymbol{\theta}), \tag{1.14}$$

where $\boldsymbol{Y}$ is a state vector consisting of variables that summarize the environment in which asset returns are determined and $\boldsymbol{\theta}$ is a vector of parameters that uniquely determines the distribution function $F_r(\cdot)$. The probability distribution $F_r(\cdot)$ governs the stochastic behavior of the returns r_{it} and $\boldsymbol{Y}$. In many financial studies, the state vector $\boldsymbol{Y}$ is treated as given and the main concern is the conditional distribution of $\{r_{it}\}$ given $\boldsymbol{Y}$. Empirical analysis of asset returns is then to estimate the unknown parameter $\boldsymbol{\theta}$ and to draw statistical inference about the behavior of $\{r_{it}\}$ given some past log returns.

The model in Eq. (1.14) is too general to be of practical value. However, it provides a general framework with respect to which an econometric model for asset returns r_{it} can be put in a proper perspective.

Some financial theories such as the capital asset pricing model (CAPM) of Sharpe (1964) focus on the joint distribution of N returns at a single time index t (i.e., the distribution of $\{r_{1t}, \ldots, r_{Nt}\}$). Other theories emphasize the dynamic structure of individual asset returns (i.e., the distribution of $\{r_{i1}, \ldots, r_{iT}\}$ for a given asset i). In this book, we focus on both. In the univariate analysis of Chapters 2–7, our main concern is the joint distribution of $\{r_{it}\}_{t=1}^{T}$ for asset i. To this end, it is useful to partition the joint distribution as

$$\begin{aligned} F(r_{i1}, \ldots, r_{iT}; \boldsymbol{\theta}) &= F(r_{i1})F(r_{i2}|r_{i1}) \cdots F(r_{iT}|r_{i,T-1}, \ldots, r_{i1}) \\ &= F(r_{i1}) \prod_{t=2}^{T} F(r_{it}|r_{i,t-1}, \ldots, r_{i1}), \end{aligned} \tag{1.15}$$

where, for simplicity, the parameter $\boldsymbol{\theta}$ is omitted. This partition highlights the temporal dependencies of the log return r_{it}. The main issue then is the specification of the conditional distribution $F(r_{it}|r_{i,t-1}, \cdot)$, in particular, how the conditional distribution evolves over time. In finance, different distributional specifications lead to different theories. For instance, one version of the random-walk hypothesis is that the conditional distribution $F(r_{it}|r_{i,t-1}, \ldots, r_{i1})$ is equal to the marginal distribution $F(r_{it})$. In this case, returns are temporally independent and, hence, not predictable.

It is customary to treat asset returns as continuous random variables, especially for index returns or stock returns calculated at a low frequency, and use their probability density functions. In this case, using the identity in Eq. (1.9), we can write the partition in Eq. (1.15) as

$$f(r_{i1}, \ldots, r_{iT}; \boldsymbol{\theta}) = f(r_{i1}; \boldsymbol{\theta}) \prod_{t=2}^{T} f(r_{it}|r_{i,t-1}, \ldots, r_{i1}; \boldsymbol{\theta}). \tag{1.16}$$

For high-frequency asset returns, discreteness becomes an issue. For example, stock prices change in multiples of a tick size on the New York Stock Exchange (NYSE). The tick size was $\frac{1}{8}$ of a dollar before July 1997 and was $\frac{1}{16}$ of a dollar from July 1997 to January 2001. Therefore, the tick-by-tick return of an individual stock listed on the NYSE is not continuous. We discuss high-frequency stock price changes and time durations between price changes later in Chapter 5.

Remark. On August 28, 2000, the NYSE began a pilot program with 7 stocks priced in decimals and the American Stock Exchange (AMEX) began a pilot program with 6 stocks and two options classes. The NYSE added 57 stocks and 94 stocks to the program on September 25 and December 4, 2000, respectively. All NYSE and AMEX stocks started trading in decimals on January 29, 2001. □

Equation (1.16) suggests that conditional distributions are more relevant than marginal distributions in studying asset returns. However, the marginal distributions may still be of some interest. In particular, it is easier to estimate marginal distributions than conditional distributions using past returns. In addition, in some cases, asset returns have weak empirical serial correlations, and, hence, their marginal distributions are close to their conditional distributions.

Several statistical distributions have been proposed in the literature for the marginal distributions of asset returns, including normal distribution, lognormal distribution, stable distribution, and scale mixture of normal distributions. We briefly discuss these distributions.

Normal Distribution

A traditional assumption made in financial study is that the simple returns $\{R_{it} | t = 1, \ldots, T\}$ are independently and identically distributed as normal with fixed mean and variance. This assumption makes statistical properties of asset returns tractable. But it encounters several difficulties. First, the lower bound of a simple return is -1. Yet the normal distribution may assume any value in the real line and, hence, has no lower bound. Second, if R_{it} is normally distributed, then the multiperiod simple return $R_{it}[k]$ is not normally distributed because it is a product of one-period returns. Third, the normality assumption is not supported by many empirical asset returns, which tend to have a positive excess kurtosis.

Lognormal Distribution

Another commonly used assumption is that the log returns r_t of an asset are independent and identically distributed (iid) as normal with mean μ and variance σ^2. The simple returns are then iid lognormal random variables with mean and variance given by

$$E(R_t) = \exp\left(\mu + \frac{\sigma^2}{2}\right) - 1, \quad \mathrm{Var}(R_t) = \exp(2\mu + \sigma^2)[\exp(\sigma^2) - 1]. \quad (1.17)$$

These two equations are useful in studying asset returns (e.g., in forecasting using models built for log returns). Alternatively, let m_1 and m_2 be the mean and variance of the simple return R_t, which is lognormally distributed. Then the mean and variance of the corresponding log return r_t are

$$E(r_t) = \ln\left[\frac{m_1 + 1}{\sqrt{1 + m_2/(1 + m_1)^2}}\right], \qquad \mathrm{Var}(r_t) = \ln\left[1 + \frac{m_2}{(1 + m_1)^2}\right].$$

Because the sum of a finite number of iid normal random variables is normal, $r_t[k]$ is also normally distributed under the normal assumption for $\{r_t\}$. In addition, there is no lower bound for r_t, and the lower bound for R_t is satisfied using $1 + R_t = \exp(r_t)$. However, the lognormal assumption is not consistent with all the properties of historical stock returns. In particular, many stock returns exhibit a positive excess kurtosis.

Stable Distribution

The stable distributions are a natural generalization of normal in that they are stable under addition, which meets the need of continuously compounded returns r_t. Furthermore, stable distributions are capable of capturing excess kurtosis shown by historical stock returns. However, nonnormal stable distributions do not have a finite variance, which is in conflict with most finance theories. In addition, statistical modeling using nonnormal stable distributions is difficult. An example of nonnormal stable distributions is the Cauchy distribution, which is symmetric with respect to its median but has infinite variance.

Scale Mixture of Normal Distributions

Recent studies of stock returns tend to use scale mixture or finite mixture of normal distributions. Under the assumption of scale mixture of normal distributions, the log return r_t is normally distributed with mean μ and variance σ^2 [i.e., $r_t \sim N(\mu, \sigma^2)$]. However, σ^2 is a random variable that follows a positive distribution (e.g., σ^{-2} follows a gamma distribution). An example of finite mixture of normal distributions is

$$r_t \sim (1 - X)N(\mu, \sigma_1^2) + XN(\mu, \sigma_2^2),$$

where X is a Bernoulli random variable such that $P(X = 1) = \alpha$ and $P(X = 0) = 1 - \alpha$ with $0 < \alpha < 1$, σ_1^2 is small, and σ_2^2 is relatively large. For instance, with $\alpha = 0.05$, the finite mixture says that 95% of the returns follow $N(\mu, \sigma_1^2)$ and 5% follow $N(\mu, \sigma_2^2)$. The large value of σ_2^2 enables the mixture to put more mass at the tails of its distribution. The low percentage of returns that are from $N(\mu, \sigma_2^2)$ says that the majority of the returns follow a simple normal distribution. Advantages of mixtures of normal include that they maintain the tractability of normal, have finite higher order moments, and can capture the excess kurtosis. Yet it is hard to estimate the mixture parameters (e.g., the α in the finite-mixture case).

Figure 1.1 shows the probability density functions of a finite mixture of normal, Cauchy, and standard normal random variable. The finite mixture of normal is $(1 - X)N(0, 1) + X \times N(0, 16)$ with X being Bernoulli such that $P(X = 1) =$

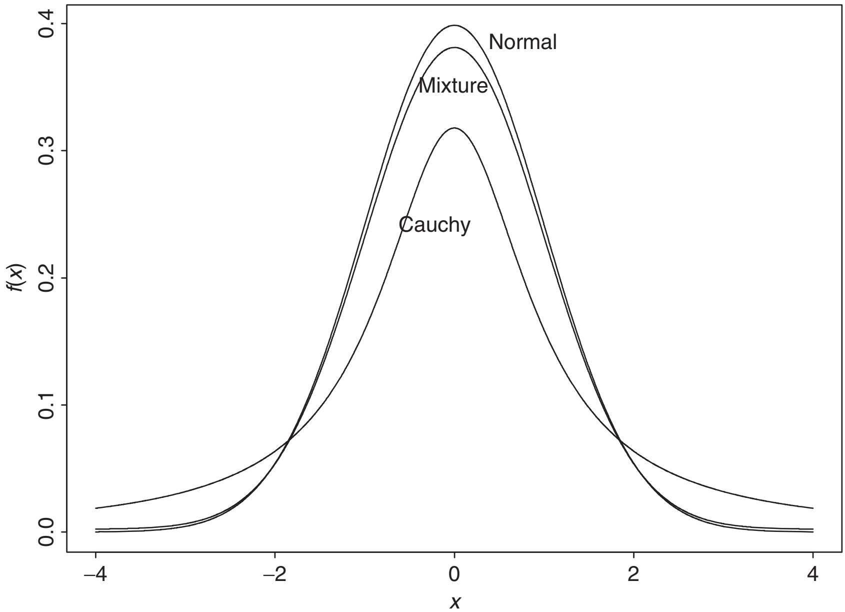

Figure 1.1 Comparison of finite mixture, stable, and standard normal density functions.

0.05, and the density function of Cauchy is

$$f(x) = \frac{1}{\pi(1 + x^2)}, \qquad -\infty < x < \infty.$$

It is seen that the Cauchy distribution has fatter tails than the finite mixture of normal, which, in turn, has fatter tails than the standard normal.

1.2.3 Multivariate Returns

Let $\boldsymbol{r}_t = (r_{1t}, \ldots, r_{Nt})'$ be the log returns of N assets at time t. The multivariate analyses of Chapters 8 and 10 are concerned with the joint distribution of $\{\boldsymbol{r}_t\}_{t=1}^T$. This joint distribution can be partitioned in the same way as that of Eq. (1.15). The analysis is then focused on the specification of the conditional distribution function $F(\boldsymbol{r}_t|\boldsymbol{r}_{t-1}, \ldots, \boldsymbol{r}_1, \boldsymbol{\theta})$. In particular, how the conditional expectation and conditional covariance matrix of $\boldsymbol{r}_t$ evolve over time constitute the main subjects of Chapters 8 and 10.

The mean vector and covariance matrix of a random vector $\boldsymbol{X} = (X_1, \ldots, X_p)$ are defined as

$$E(\boldsymbol{X}) = \boldsymbol{\mu}_x = [E(X_1), \ldots, E(X_p)]',$$
$$\mathrm{Cov}(\boldsymbol{X}) = \boldsymbol{\Sigma}_x = E[(\boldsymbol{X} - \boldsymbol{\mu}_x)(\boldsymbol{X} - \boldsymbol{\mu}_x)'],$$

provided that the expectations involved exist. When the data $\{\boldsymbol{x}_1, \ldots, \boldsymbol{x}_T\}$ of $\boldsymbol{X}$ are available, the sample mean and covariance matrix are defined as

$$\widehat{\boldsymbol{\mu}}_x = \frac{1}{T}\sum_{t=1}^{T} \boldsymbol{x}_t, \qquad \widehat{\boldsymbol{\Sigma}}_x = \frac{1}{T-1}\sum_{t=1}^{T}(\boldsymbol{x}_t - \widehat{\boldsymbol{\mu}}_x)(\boldsymbol{x}_t - \widehat{\boldsymbol{\mu}}_x)'.$$

These sample statistics are consistent estimates of their theoretical counterparts provided that the covariance matrix of $\boldsymbol{X}$ exists. In the finance literature, multivariate normal distribution is often used for the log return $\boldsymbol{r}_t$.

1.2.4 Likelihood Function of Returns

The partition of Eq. (1.15) can be used to obtain the likelihood function of the log returns $\{r_1, \ldots, r_T\}$ of an asset, where for ease in notation the subscript i is omitted from the log return. If the conditional distribution $f(r_t|r_{t-1}, \ldots, r_1, \boldsymbol{\theta})$ is normal with mean μ_t and variance σ_t^2, then $\boldsymbol{\theta}$ consists of the parameters in μ_t and σ_t^2, and the likelihood function of the data is

$$f(r_1, \ldots, r_T; \boldsymbol{\theta}) = f(r_1; \boldsymbol{\theta})\prod_{t=2}^{T} \frac{1}{\sqrt{2\pi}\sigma_t}\exp\left[\frac{-(r_t-\mu_t)^2}{2\sigma_t^2}\right], \tag{1.18}$$

where $f(r_1; \boldsymbol{\theta})$ is the marginal density function of the first observation r_1. The value of $\boldsymbol{\theta}$ that maximizes this likelihood function is the maximum-likelihood estimate (MLE) of $\boldsymbol{\theta}$. Since the log function is monotone, the MLE can be obtained by maximizing the log-likelihood function,

$$\ln\, f(r_1, \ldots, r_T; \boldsymbol{\theta}) = \ln\, f(r_1; \boldsymbol{\theta}) - \frac{1}{2}\sum_{t=2}^{T}\left[\ln(2\pi) + \ln(\sigma_t^2) + \frac{(r_t-\mu_t)^2}{\sigma_t^2}\right],$$

which is easier to handle in practice. The log-likelihood function of the data can be obtained in a similar manner if the conditional distribution $f(r_t|r_{t-1}, \ldots, r_1; \boldsymbol{\theta})$ is not normal.

1.2.5 Empirical Properties of Returns

The data used in this section are obtained from the Center for Research in Security Prices (CRSP) of the University of Chicago. Dividend payments, if any, are included in the returns. Figure 1.2 shows the time plots of monthly simple returns and log returns of IBM stock from January 1926 to December 2008. A *time plot* shows the data against the time index. The upper plot is for the simple returns. Figure 1.3 shows the same plots for the monthly returns of value-weighted market index. As expected, the plots show that the basic patterns of simple and log returns are similar.

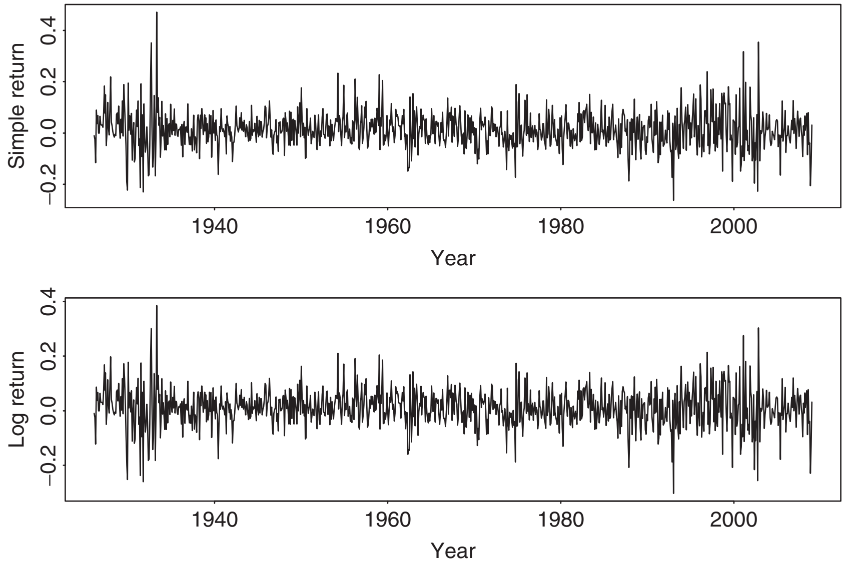

Figure 1.2 Time plots of monthly returns of IBM stock from January 1926 to December 2008. Upper panel is for simple returns, and lower panel is for log returns.

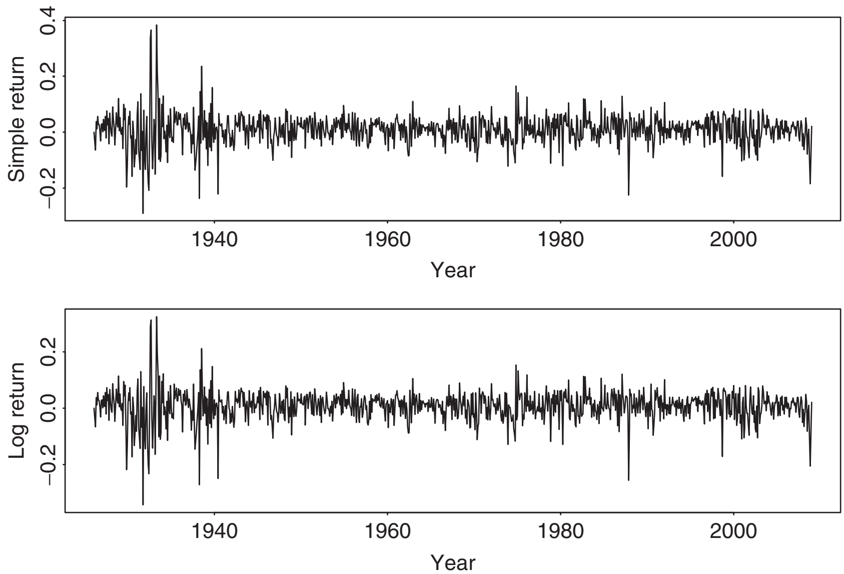

Figure 1.3 Time plots of monthly returns of value-weighted index from January 1926 to December 2008. Upper panel is for simple returns, and lower panel is for log returns.

Table 1.2 provides some descriptive statistics of simple and log returns for selected U.S. market indexes and individual stocks. The returns are for daily and monthly sample intervals and are in percentages. The data spans and sample sizes are also given in Table 1.2. From the table, we make the following observations. (a) Daily returns of the market indexes and individual stocks tend to have high excess kurtoses. For monthly series, the returns of market indexes have higher excess kurtoses than individual stocks. (b) The mean of a daily return series is close to zero, whereas that of a monthly return series is slightly larger. (c) Monthly returns have higher standard deviations than daily returns. (d) Among the daily returns, market indexes have smaller standard deviations than individual stocks. This is in agreement with common sense. (e) The skewness is not a serious problem for both daily and monthly returns. (f) The descriptive statistics show that the difference between simple and log returns is not substantial.

Figure 1.4 shows the empirical density functions of monthly simple and log returns of IBM stock from 1926 to 2008. Also shown, by a dashed line, in each graph is the normal probability density function evaluated by using the sample mean and standard deviation of IBM returns given in Table 1.2. The plots indicate that the normality assumption is questionable for monthly IBM stock returns. The empirical density function has a higher peak around its mean, but fatter tails than that of the corresponding normal distribution. In other words, the empirical density

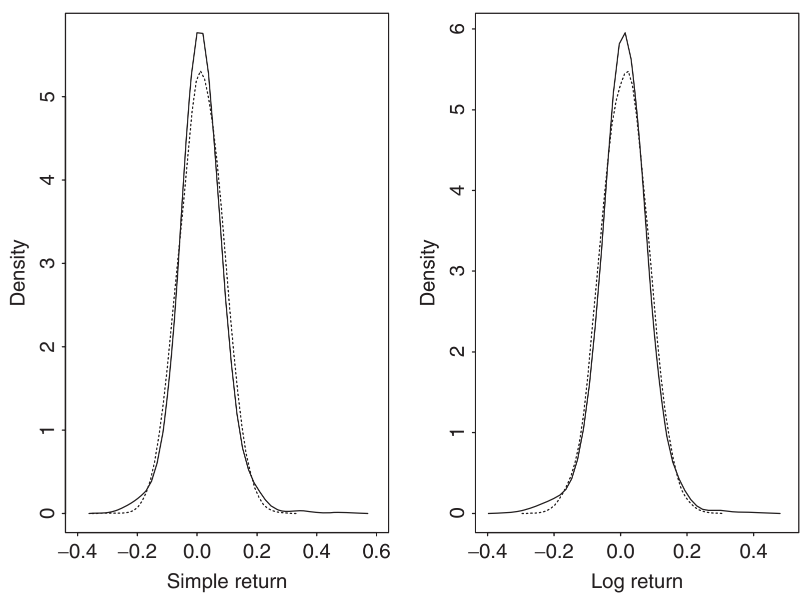

Figure 1.4 Comparison of empirical and normal densities for monthly simple and log returns of IBM stock. Sample period is from January 1926 to December 2008. Left plot is for simple returns and right plot for log returns. Normal density, shown by the dashed line, uses sample mean and standard deviation given in Table 1.2.

function is taller and skinnier, but with a wider support than the corresponding normal density.

1.3 PROCESSES CONSIDERED

Besides the return series, we also consider the volatility process and the behavior of extreme returns of an asset. The volatility process is concerned with the evolution of conditional variance of the return over time. This is a topic of interest because, as shown in Figures 1.2 and 1.3, the variabilities of returns vary over time and appear in clusters. In application, volatility plays an important role in pricing options and risk management. By extremes of a return series, we mean the large positive or negative returns. Table 1.2 shows that the minimum and maximum of a return series can be substantial. The negative extreme returns are important in risk management, whereas positive extreme returns are critical to holding a short position. We study properties and applications of extreme returns, such as the frequency of occurrence, the size of an extreme, and the impacts of economic variables on the extremes, in Chapter 7.

Other financial time series considered in the book include interest rates, exchange rates, bond yields, and quarterly earning per share of a company. Figure 1.5 shows the time plots of two U.S. monthly interest rates. They are the 10-year and 1-year

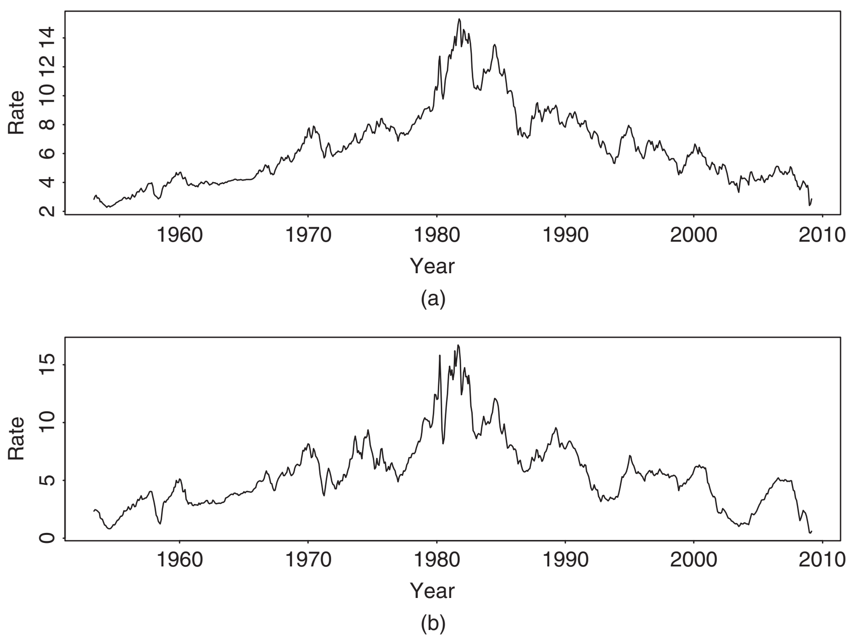

Figure 1.5 Time plots of monthly U.S. interest rates from April 1953 to February 2009: (a) 10-year Treasury constant maturity rate and (b) 1-year maturity rate.

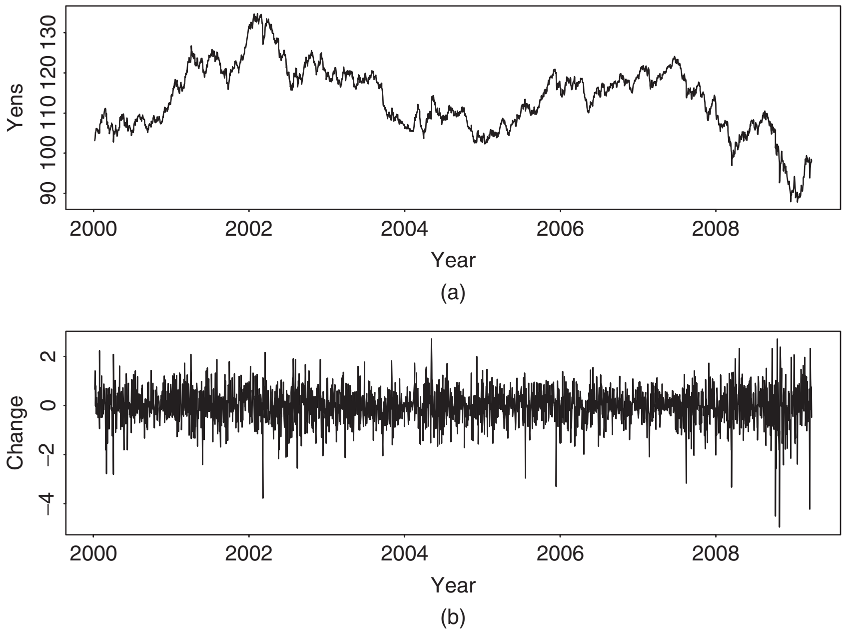

Figure 1.6 Time plot of daily exchange rate between U.S. dollar and Japanese yen from January 4, 2000, to March 27, 2009: (a) exchange rate and (b) changes in exchange rate.

Treasury constant maturity rates from April 1954 to February 2009. As expected, the two interest rates moved in unison, but the 1-year rates appear to be more volatile. Figure 1.6 shows the daily exchange rate between the U.S. dollar and the Japanese yen from January 4, 2000, to March 27, 2009. From the plot, the exchange rate encountered occasional big changes in the sampling period. Table 1.3 provides some descriptive statistics for selected U.S. financial time series. The monthly bond returns obtained from CRSP are Fama bond portfolio returns from January 1952 to December 2008. The interest rates are obtained from the Federal Reserve Bank of St. Louis. The weekly 3-month Treasury bill rate started on January 8, 1954, and the 6-month rate started on December 12, 1958. Both series ended on March 27, 2009. For the interest rate series, the sample means are proportional to the time to maturity, but the sample standard deviations are inversely proportional to the time to maturity. For the bond returns, the sample standard deviations are positively related to the time to maturity, whereas the sample means remain stable for all maturities. Most of the series considered have positive excess kurtoses.

With respect to the empirical characteristics of returns shown in Table 1.2, Chapters 2–4 focus on the first four moments of a return series and Chapter 7 on the behavior of minimum and maximum returns. Chapters 8 and 10 are concerned with moments of and the relationships between multiple asset returns, and Chapter 5 addresses properties of asset returns when the time interval is small. An introduction to mathematical finance is given in Chapter 6.

TABLE 1.3 Descriptive Statistics of Selected U.S. Financial Time Series[a]

Maturity	Mean	Standard Deviation	Skewness	Excess Kurtosis	Minimum	Maximum
Monthly Bond Returns: Jan. 1952 to Dec. 2008, $T = 684$						
1–12 months	0.45	0.35	2.47	13.14	−0.40	3.52
12–24 months	0.49	0.67	1.88	15.44	−2.94	6.85
24–36 months	0.52	0.98	1.37	12.92	−4.90	9.33
48–60 months	0.53	1.40	0.60	4.83	−5.78	10.06
61–120 months	0.55	1.69	0.65	4.79	−7.35	10.92
Monthly Treasury Rates: April 1953 to February 2009, $T = 671$						
1 year	5.59	2.98	1.02	1.32	0.44	16.72
3 years	5.98	2.85	0.95	0.95	1.07	16.22
5 years	6.19	2.77	0.97	0.82	1.52	15.93
10 years	6.40	2.69	0.95	0.61	2.29	15.32
Weekly Treasury Bill Rates: End on March 27, 2009.						
3 months	5.07	2.82	1.08	1.80	0.02	16.76
6 months	5.52	2.73	0.99	1.53	0.20	15.76

[a] The data are in percentages. The weekly 3-month Treasury bill rate started from January 8, 1954, and the 6-month rate started from December 12, 1958. The sample sizes for Treasury bill rates are 2882 and 2625, respectively. Data sources are given in the text.

APPENDIX: R PACKAGES

R is a free software available from http://www.r-project.org. One can click *CRAN* on its Web page to select a nearby *CRAN Mirror* to download and install the software and selected packages. For financial time series analysis, the `Rmetrics` of Diethelm Wuertz and his associates have produced many useful packages, including `fBasics`, `timeSeries`, `fGarch`, etc. We use many functions of these packages in this book. Further information concerning installing R and the commands used can be found either on the Web page of this book or on the author's teaching Web page.

R and S-Plus are objective-oriented software. They enable users to create many objects. For instance, one can use the command `ts` to create a time series object. Treating time series data as a time series object in R has some advantages, but it requires some learning to get used to it. It is, however, not necessary to create a time series object in R to perform the analyses discussed in this book. As an illustration, consider the monthly simple returns to the General Motors stock from January 1975 to December 2008; see Exercise 1.2. The data have 408 observations. The following R commands are used to illustrate the points:

```
> da=read.table("m-gm3dx7508.txt",header=T)  % Load data
> gm=da[,2]     % Column 2 contains GM stock returns
> gm1=ts(gm,frequency=12,start=c(1975,1))
% Creates a ts object.
> par(mfcol=c(2,1))   % Put two plots on a page.
```

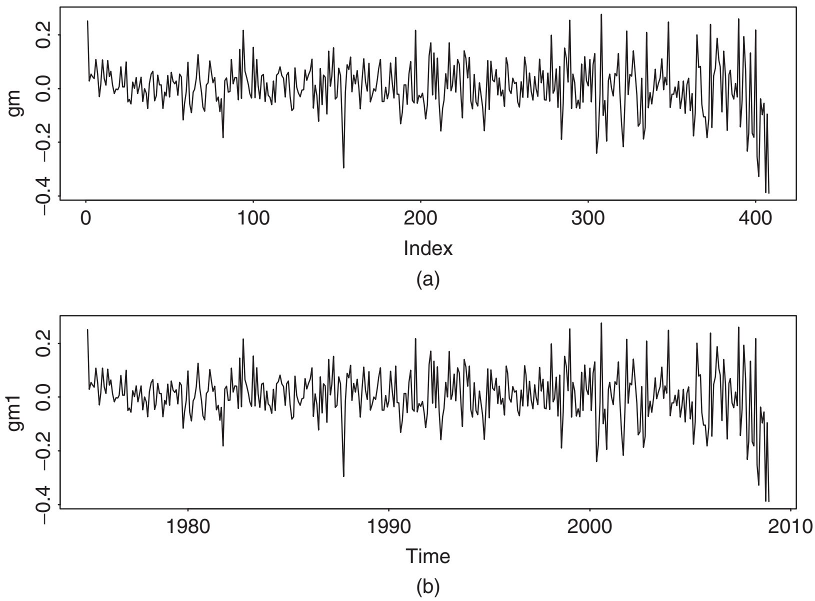

Figure 1.7 Time plots of monthly simple returns to General Motors stock from January 1975 to December 2008: (a) and (b) are without and with time series object, respectively.

```
> plot(gm,type='l')
> plot(gm1,type='l')
> acf(gm,lag=24)
> acf(gm1,lag=24)
```

In the `ts` command, `frequency = 12` says that the time unit is *year* and there are 12 equally spaced observations in each time unit, and start = c(1975,1) means the starting time is January 1975. Frequency and start are the two basic arguments needed in R to create a time series object. For further details, please use `help(ts)` in R to obtain details of the command. Here `gm1` is a time series object in R, but `gm` is not. Figures 1.7 and 1.8 show, respectively, the time plot and autocorrelation function (ACF) of the returns of GM stock. In each figure, the upper plot is produced without using time series object, whereas the lower plot is produced by a time series object. The upper and lower plots are identical except for the horizontal label. For the time plot, the time series object uses calendar time to label the x axis, which is preferred. On the other hand, for the ACF plot, the time series object uses fractions of time unit in the label, not the commonly used time lags.

EXERCISES

1.1. Consider the daily stock returns of American Express (AXP), Caterpillar (CAT), and Starbucks (SBUX) from January 1999 to December 2008. The

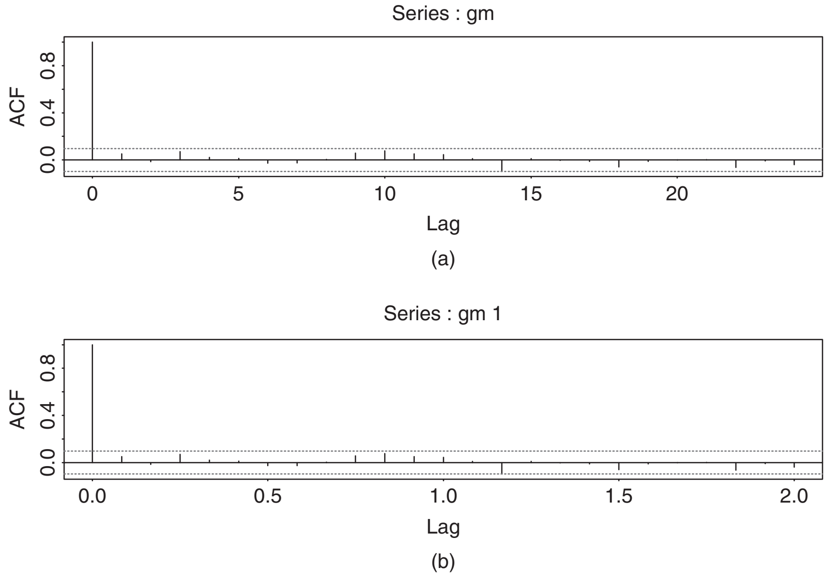

Figure 1.8 Sample ACFs of the monthly simple returns to General Motors stock from January 1975 to December 2008: (a) and (b) are without and with time series object, respectively.

data are simple returns given in the file `d-3stocks9908.txt` (date, axp, cat, sbux).

(a) Express the simple returns in percentages. Compute the sample mean, standard deviation, skewness, excess kurtosis, minimum, and maximum of the percentage simple returns.
(b) Transform the simple returns to log returns.
(c) Express the log returns in percentages. Compute the sample mean, standard deviation, skewness, excess kurtosis, minimum, and maximum of the percentage log returns.
(d) Test the null hypothesis that the mean of the log returns of each stock is zero. That is, perform three separate tests. Use 5% significance level to draw your conclusion.

1.2. Answer the same questions as in Exercise 1.1 but using monthly stock returns for General Motors (GM), CRSP value-weighted index (VW), CRSP equal-weighted index (EW), and S&P composite index from January 1975 to December 2008. The returns of the indexes include dividend distributions. Data file is `m-gm3dx7508.txt` (date, gm, vw, ew, sp).

1.3. Consider the monthly stock returns of S&P composite index from January 1975 to December 2008 in Exercise 1.2. Answer the following questions:

(a) What is the average annual log return over the data span?

(b) Assume that there were no transaction costs. If one invested \$1.00 on the S&P composite index at the beginning of 1975, what was the value of the investment at the end of 2008?

1.4. Consider the daily log returns of American Express stock from January 1999 to December 2008 as in Exercise 1.1. Use the 5% significance level to perform the following tests: (a) Test the null hypothesis that the skewness measure of the returns is zero. (b) Test the null hypothesis that the excess kurtosis of the returns is zero.

1.5. Daily foreign exchange rates (spot rates) can be obtained from the Federal Reserve Bank in Chicago. The data are the noon buying rates in New York City certified by the Federal Reserve Bank of New York. Consider the exchange rates between the U.S. dollar and the Canadian dollar, euro, U.K. pound, and the Japanese yen from January 4, 2000, to March 27, 2009. The data are also on the Web. (a) Compute the daily log return of each exchange rate. (b) Compute the sample mean, standard deviation, skewness, excess kurtosis, minimum, and maximum of the log returns of each exchange rate. (c) Discuss the empirical characteristics of the log returns of exchange rates. (d) Obtain a density plot of the daily long returns of dollar–euro exchange rate.

REFERENCES

Campbell, J. Y., Lo, A. W., and MacKinlay, A. C. (1997). *The Econometrics of Financial Markets*. Princeton University Press, Princeton, NJ.

Jarque, C. M. and Bera, A. K. (1987). A test of normality of observations and regression residuals. *International Statistical Review* **55**: 163–172.

Sharpe, W. (1964). Capital asset prices: A theory of market equilibrium under conditions of risk. *Journal of Finance* 19: 425–442.

Snedecor, G. W. and Cochran, W. G. (1980). *Statistical Methods*, 7th ed. Iowa State University Press, Ames, IA.

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

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

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

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

where the property $\text{Var}(r_t) = \text{Var}(r_{t-\ell})$ 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 $H_a : \rho_1 \ne 0$. 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

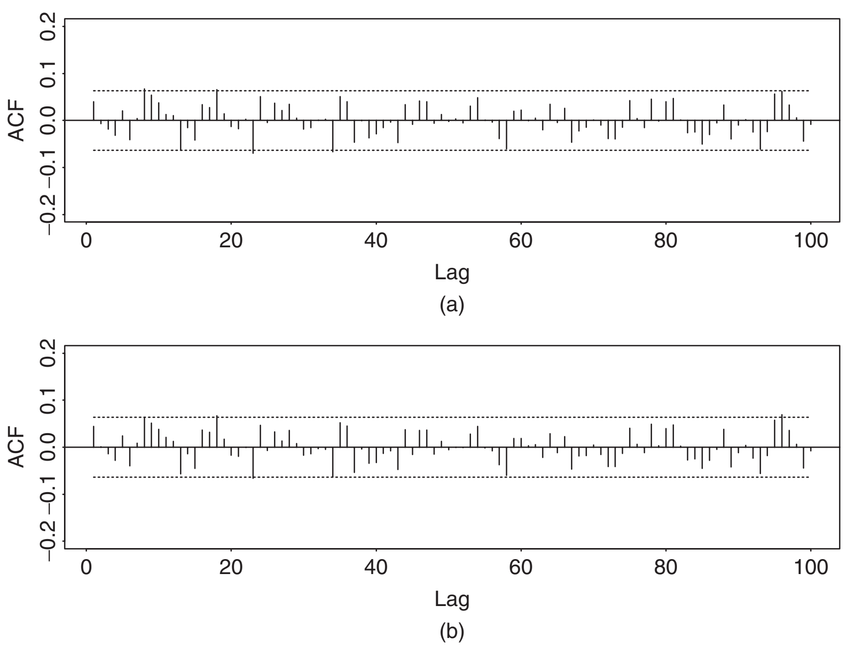

Figure 2.1 Sample autocorrelation functions of monthly (a) simple returns and (b) log returns of IBM stock from January 1926 to December 2008. In each plot, two horizontal dashed lines denote two standard error limits of sample ACF.

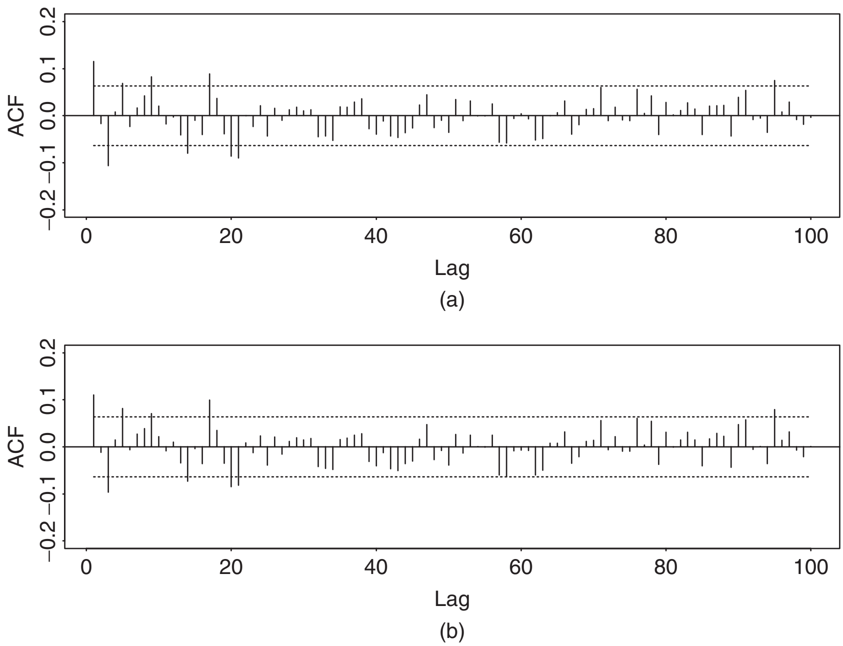

Figure 2.2 Sample autocorrelation functions of monthly (a) simple returns and (b) log returns of value-weighted index of U.S. markets from January 1926 to December 2008. In each plot, two horizontal dashed lines denote two standard error limits of sample ACF.

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
```