2

Statistical Properties of Financial Market Data

In this chapter, we describe the empirical characteristics of some statistical measures of asset returns. More specifically, in Section 2.1, we provide some definitions that are useful for the rest of the book. Although most of the statistical measures used are well-known, emphasis is placed on critical issues concerning the estimation and the applications of these statistical measures. Then, we consider the various aspects of the return distribution. In Section 2.2, we show that the standard assumption of normality is rejected for stock market returns mainly because of asymmetry and tail thickness of the return distribution. In Section 2.3, we discuss time dependence and demonstrate the strong volatility clustering in financial markets. Section 2.4 discusses correlation across asset returns, its time-varying nature, and the implication of changing dependence on asset allocation. However, correlation is not a valid measure for dependence when returns are non-normal. Other measures for dependence are provided in Chapter 6. Finally, when the distribution appears non-normal, it is very useful and relevant to consider multivariate moments such as co-skewness and co-kurtosis. This is covered in Section 2.5.

Financial modeling deals with capturing the main characteristics of return distributions. The statistical properties of financial market data identified in this chapter will serve as the backbone for the later chapters. For instance, Chapter 3 describes some theoretical models explaining some of the observed statistical properties, and Chapters 4 and 5 will develop the tools necessary for modeling time-dependency and non-normality, in practice.

2.1 Definitions of returns

Although prices are what we observe in financial markets, most empirical studies focus on returns. The reason is that, in general, prices are non-stationary whereas returns are stationary. There are several return definitions, each of which produces a different set of properties for returns. We more specifically focus on simple returns and log-returns.

2.1.1 Simple returns

The one-period simple return for holding an asset is

$$R_t = \frac{P_t - P_{t-1}}{P_{t-1}} \qquad \text{or} \qquad P_t = P_{t-1}\left(1 + R_t\right),$$

where P_t is the price (including dividends) of the asset at time t, and R_t is the one-period simple return from time $t-1$ to t.

If we hold the asset for k periods from $t-k$ to t, we have the k-period simple return

$$R_t\left[k\right] = \frac{P_t - P_{t-k}}{P_{t-k}},$$

or

$$P_t = P_{t-k}\left(1 + R_t\left[k\right]\right) = P_{t-k}\left(1 + R_{t-k+1}\right) \times \cdots \times \left(1 + R_t\right),$$

where $R_t\left[k\right]$ is the k-period simple return from $t-k$ to t. Therefore, the relation between the simple one-period return and k-period return is non-linear

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

In some cases, if returns are small, we may use the approximation

$$R_t\left[k\right] \approx \sum_{j=0}^{k-1} R_{t-j},$$

but it is too crude in many applications.

On the other hand, the simple return of an N-asset portfolio is simply the weighted average of the individual simple returns of the N assets, denoted $R_{i,t}$, for $i = 1, \cdots, N$. Let p denote the portfolio with investment weight ω_i on asset i and $\sum_{i=1}^{N} \omega_i = 1$. Then, the portfolio return is

$$R_{p,t} = \sum_{i=0}^{N} \omega_i R_{i,t}.$$

2.1.2 Log-returns

If a bank pays an annual interest of $R_t^{(m)}$, m times a year, the interest rate for each unit of investment is by definition $R_t^{(m)}/m$, and after one year the total value of the deposit is $\left(1 + \frac{1}{m} R_t^{(m)}\right)^m$. In the limiting case where the interest is cumulated continuously ($m \to \infty$), we have

$$\lim_{m \to \infty}\left(1 + \frac{1}{m} R_t^{(m)}\right)^m = e^{r_t}.$$

We deduce that r_t, the continuously compounded return (or log-return), is

$$r_t = \log(P_t) - \log(P_{t-1}) = p_t - p_{t-1},$$

where $p_t = \log(P_t)$ is the log-price.

A key advantage of the log-return is that the multiple-period return is simply the sum of one-period returns, so that

$$r_t[k] = \log(1 + R_t[k]) = \sum_{j=0}^{k-1} \log(1 + R_{t-j}) = \sum_{j=0}^{k-1} r_{t-j}.$$

Now, the log-return of a portfolio does not have the same convenient property as the case with simple return. Indeed, the portfolio log-return $r_{p,t}$ is

$$r_{p,t} = \log\left(\sum_{i=1}^{N} \omega_i e^{r_{i,t}}\right) \neq \sum_{i=1}^{N} \omega_i r_{i,t}.$$

But this problem is usually considered as minor in empirical applications.

2.1.3 Stylized facts

At first sight, there is no reason why commodity price, stock price, or exchange rate should behave in a particular fashion. However, many empirical studies[1] have identified a set of common features among financial data that are known as *stylized facts*. Cont (2001), in particular, provides a comprehensive survey of these stylized facts, which include:

1. *Fat tails*: The unconditional distribution of returns has fatter tails than that expected from a normal distribution. This means that, if we use the normal distribution to model financial returns, we will underestimate the number and magnitude of crashes and booms.
2. *Asymmetry*: The unconditional distribution is negatively skewed, suggesting that extreme negative returns are more frequent than extreme positive returns. The asymmetry and fat-tail phenomena persist even after adjusting for conditional heteroskedasticity (or changing volatility), meaning that the conditional distribution is also non-normal.
3. *Aggregated normality*: As the frequency of the returns lengthens, the return distribution gets closer to the normal distribution.
4. *Absence of serial correlation*: Returns generally do not display significant serial correlation, except at high frequency.

[1] See, for example, Bollerslev, Chou, and Kroner (1992), Pagan (1996), Shephard (1996), Campbell, Lo, and MacKinlay (1997), Cont (2001), and Gouriéroux and Jasiak (2001a).

5. *Volatility clustering*: Volatility of returns is serially correlated, suggesting that a large (positive or negative) return tends to be followed by another large (positive or negative) return. Among the proxies for volatility, absolute returns appear to be the most strongly serially correlated.
6. *Time-varying cross-correlation*: Correlation between asset returns tends to increase during high-volatility periods, in particular during crashes.

In the following sections, we will use financial market data to illustrate these stylized facts and to see how aggregation (or lengthening of data frequency) affects these stylized facts. The data consists of log-returns on four stock market indices; namely the Standard and Poor's 500 (SP500) from the US, the DAX 30 from Germany, the FTSE All Shares from the UK, and the Nikkei 225 from Japan. The sample period covers 2 January 1980 to 31 August 2004.

2.2 Distribution of returns

Figure 2.1 displays the log-returns on four stock market indices. Extreme negative returns appear to be more pronounced than extreme positive returns. Figure 2.2 displays the histogram of the log-returns, which show the empirical distributions as asymmetric and have tails that do not vanish as fast as those of the normal distribution. This is confirmed by the minimum and maximum returns reported in Table 2.1.

Table 2.1. *Mimimum and maximum of daily log-returns*

	SP500	DAX	FT-SE	Nikkei
Minimum	−22.833	−13.71	−11.914	−16.135
Maximum	8.709	7.553	5.698	12.430

2.2.1 Moments of a random variable

In this book, we will typically use non-Gaussian distributions, which are characterized by moments that are higher than the second order.[2] Consider a continuous random variable X (say the log-return series) with *cumulative distribution function* (*cdf*)

[2] In this book, we use the words *Gaussian* and *normal* interchangeably. The normal distribution was studied as early as 1733 by de Moivre. Later on, Laplace and Gauss used it to model least-square errors. Hence, the Gaussian distribution is sometimes referred to as the Gauss-Laplace distribution.

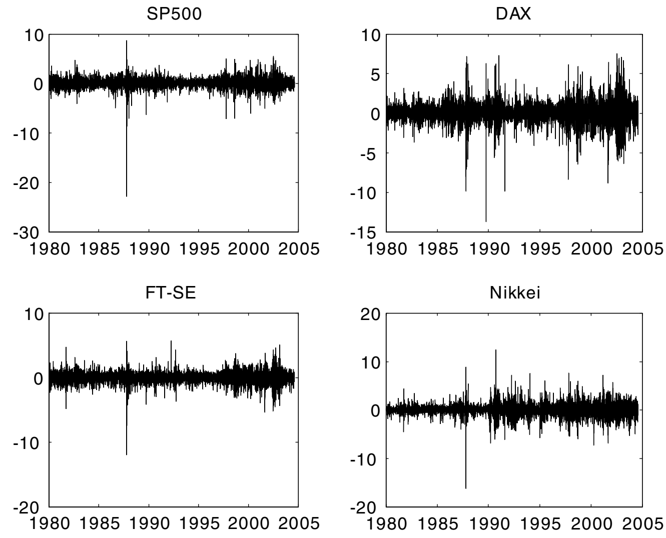

Fig. 2.1. *Evolution of daily log-returns.*

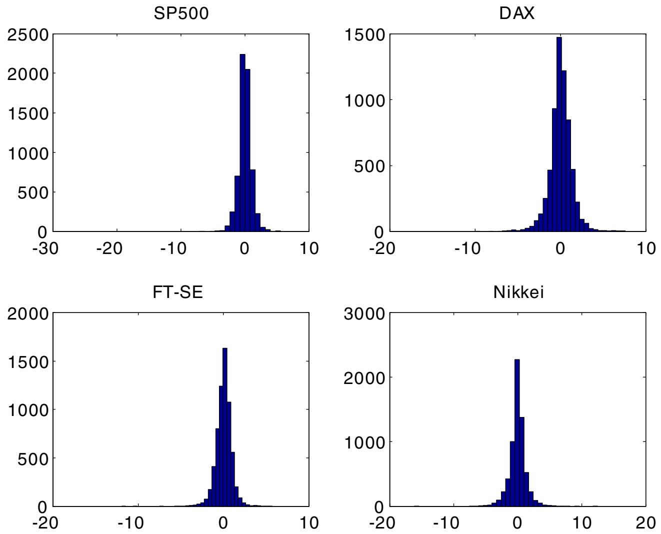

Fig. 2.2. *Histogram of daily log-returns.*

$$F_X(x) = \Pr[X \leq x] = \int_{-\infty}^{x} f_X(u)\,du,$$

where f_X is the *probability density function* (*pdf*). The uncentered moments of X are defined as

$$m_k = E\left[X^k\right] = \int_{-\infty}^{\infty} x^k f_X(x)\,dx, \quad \text{for } k = 1, 2, \cdots.$$

The first non-central moment $m_1 = E[X] = \mu$ is the mean of X.

Given the mean m_1, the centered moments of X are defined as

$$\mu_k = E\left[(X - m_1)^k\right] = \int_{-\infty}^{\infty} (x - m_1)^k f_X(x)\,dx, \quad \text{for } k = 1, 2, \cdots.$$

By construction, $\mu_1 = 0$. The second centered moment is the variance of X: $V[X] = \mu_2 = m_2 - m_1^2 = \sigma^2$. The third and fourth centered moments μ_3 and μ_4 are, respectively, the non-standardized skewness and kurtosis. The standardized skewness and kurtosis, or simply skewness and kurtosis, are defined as

$$s = Sk[X] = E\left[\left(\frac{X-\mu}{\sigma}\right)^3\right] = \frac{\mu_3}{\sigma^3},$$

$$\kappa = Ku[X] = E\left[\left(\frac{X-\mu}{\sigma}\right)^4\right] = \frac{\mu_4}{\sigma^4}.$$

Skewness, $Sk[X]$, depicts the asymmetry of the distribution. When it is negative, large negative realizations of X are more likely to appear than large positive realizations. *Kurtosis*, $Ku[X]$, captures the tail thickness of the distribution. A large kurtosis means that large realizations (either positive or negative) are more likely to be obtained than expected from a normal distribution.

Figures 2.3 and 2.4 illustrate how introducing asymmetry or fat tails would affect the shape of a distribution vis-à-vis the normal distribution. In Figure 2.3, the distribution with negative skewness leans to the right, since the long negative tail on the left has to be compensated by a greater probability mass on the shorter tail on the right. Note also that distributions with thicker tails have a thinner and higher peak in the center when compared with the normal distribution [3]

[3] In Figure 2.3, the asymmetric distributions are obtained from the so-called skewed Student t distribution with 10 degrees of freedom and skewness parameters -0.75 and 0.75, respectively, for the long-dash line and the solid bold line (see Section 5.2.3). In Figure 2.4, the fat-tailed distributions are Student t distributions with 5 and 2 degrees of freedom, respectively, for the long-dash line and the solid bold line.

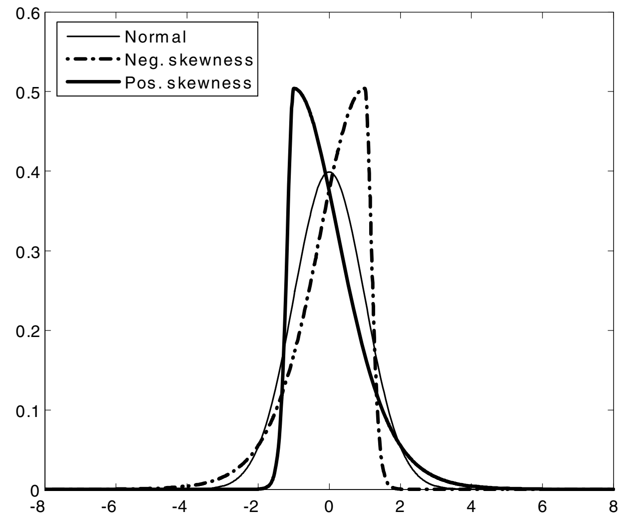

Fig. 2.3. *Examples of asymmetric distributions.*

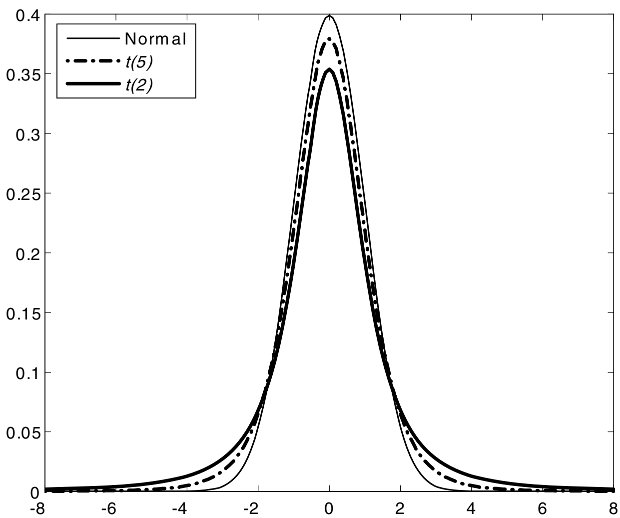

Fig. 2.4. *Examples of distributions with fat tails.*

The normal distribution has skewness always equal to zero and kurtosis always equal to 3. All further moments are either 0 (odd moments) or functions of μ and σ (even moments). Hence, a normal distribution is fully identified by its mean and variance. Also, as distributions are often compared with the normal distribution, the *excess kurtosis* (i.e., $\kappa - 3$) is often reported instead of the kurtosis itself. Most financial asset returns have kurtosis values greater than 3, and stock market return, in particular, is strongly featured by a negative skewness. Other financial series, such as interest rates, returns on commodities or on hedge funds, may have positive skewness. Finally, foreign exchange rates may have positive as well as negative skewness, depending on the way they are computed.

2.2.2 Empirical moments

Measures of location

Now, consider a time series of realized returns $\{r_t\}_{t=1}^T$. The most commonly used measures of location are *sample mean* $\bar{r}$ and *median* m. The sample mean is calculated as[4]

$$\bar{r} = \hat{\mu} = \frac{1}{T}\sum_{t=1}^{T} r_t.$$

If $\{r_t\}_{t=1}^T$ has a symmetric distribution, then $\bar{r}$ is the optimal measure of location. Median is defined as the 50th percentile of the sample. In other words, 50% of the sample has a value lower than $m = med[r]$, i.e.,

$$\Pr\left[r_t \leq m\right] = \Pr\left[r_t \geq m\right] = \frac{1}{2}.$$

Mean, as a measure for location, is sensitive to outliers. One erroneously recorded value could potentially move the mean away from the central part of the distribution. In contrast, the median is more robust against outliers, because it does not rely on the precise value of the realizations other than the median itself.

Measures of dispersion

Variance is a popular measure of dispersion, and it is the optimal measure for dispersion if returns have a normal distribution. The unbiased *sample variance*, $\hat{\sigma}^2$, can be calculated from a set of returns as[5]

[4] In this book, we denote estimates with a hat ^.

[5] Note that $\hat{\sigma}^2$ is not the Maximum-Likelihood (ML) estimator of the variance under normality. For a normal distribution, the ML estimator for variance is $\tilde{\sigma}^2 = \frac{1}{T}\sum_{t=1}^{T}(r_t - \bar{r})^2$.

$$\hat{\sigma}^2 = \frac{1}{T-1} \sum_{t=1}^{T} (r_t - \bar{r})^2 . \tag{2.1}$$

In finance, the standard deviation $\hat{\sigma}$, the square root of variance, is often used to mean volatility.[6] Since the statistical properties of sample mean make it a very inaccurate estimate of the true mean, especially for small samples, taking deviations around zero instead of the sample mean as in (2.1) typically increases volatility forecast accuracy. There are methods for estimating volatility that are designed to exploit or reduce the influence of extremes (see Ball and Torous, 1983, Parkinson, 1980, or Garman and Klass, 1980). But these methods generally rely on the assumption of normality of returns, which is not likely in the case of financial market series.

Another interesting measure of dispersion is the *mean absolute deviation* (MAD). It is computed as

$$MAD = \frac{1}{T} \sum_{t=1}^{T} |r_t - \bar{r}| .$$

The MAD is more robust to outliers than the variance, because it involves lower powers of returns.

The *inter-quartile range* (IQR) is the difference between the 75th and 25th percentiles of the data. Since only the middle 50% of the data affects this measure, it is robust against outliers.

Higher moments

The sample (standardized) skewness and kurtosis are computed, respectively, as

$$\hat{s} = \frac{1}{T} \sum_{t=1}^{T} \left(\frac{r_t - \bar{r}}{\hat{\sigma}} \right)^3 ,$$

and

$$\hat{\kappa} = \frac{1}{T} \sum_{t=1}^{T} \left(\frac{r_t - \bar{r}}{\hat{\sigma}} \right)^4 .$$

Under the assumption of normality, we have the following asymptotic distributions (see Kendall and Stuart, 1977)

$$\sqrt{T}\,(\hat{\mu} - \mu) \Rightarrow \mathcal{N}\left(0, \sigma^2\right) \qquad \sqrt{T}\left(\hat{\sigma}^2 - \sigma^2\right) \Rightarrow \mathcal{N}\left(0, 2\sigma^4\right),$$

$$\sqrt{T}\hat{s} \Rightarrow \mathcal{N}(0, 6) \qquad \sqrt{T}\,(\hat{\kappa} - 3) \Rightarrow \mathcal{N}(0, 24) ,$$

[6] While $\hat{\sigma}^2$ in equation (2.1) is an unbiased estimate of σ^2, the square root of $\hat{\sigma}^2$ is a biased estimate of σ due to Jensen inequality. Cox and Rubinstein (1985) explain how this bias can be corrected assuming a normal distribution for r_t. However, in most cases, the impact of this adjustment is small.

where "⇒" denotes asymptotic convergence. Due to the large value of their asymptotic variances, sample skewness and sample kurtosis are informative only for large samples. These asymptotic distributions can be used to perform statistical tests on the distributional assumption of returns. This is the basis of the Jarque-Bera test for normality described in the next subsection.

All the statistics described above are often classified as summary statistics. Table 2.2 presents the summary statistics of the four stock market returns measured at different data frequencies. If we rank the returns based on the summary statistics, we note that the ranks produced by the mean are similar to those produced by the median. Similarly, using standard deviation or MAD to rank returns does not produce vastly different results. As expected, median and MAD are less dispersed because they are more robust against outliers. Annualized variance, derived from daily, weekly, and monthly data, is similar to the variance of annual data. Although the same is not true for skewness and kurtosis, there is no systematic pattern between, e.g., skewness of daily and annual returns. For instance, the daily skewness of SP500 is larger than the annual skewness, but the opposite is true for skewness of FT-SE.

2.2.3 Testing for normality

It has been noted for a long time that most financial asset returns are non-normal (Mandelbrot, 1963, and Fama, 1963). This non-normality is strongly featured in two statistical phenomena: (i) Extreme events occur more often than predicted by a normal distribution (Mandelbrot, 1963, Fama, 1963, Blattberg and Gonedes, 1974, Kon, 1984), and (ii) crashes occur more often than booms (Fama, 1965, Arditti, 1971, Simkowitz and Beedles, 1978, Singleton and Wingender, 1986). Phenomenon (i) corresponds to excess kurtosis or *fat tails*, whereas phenomenon (ii) is associated with negative skewness or *asymmetry.*

Following Mandelbrot (1963), two related issues have been addressed in the finance literature; does the normality assumption hold for asset returns? If not, to what extent are returns non-normal? In this chapter, we focus on unconditional normality. That is, for the moment we assume that returns distribution does not change through time. In Chapter 5, we will relax this assumption and allow distribution to vary conditionally to some distributional variable. Various tests for normality have been developed based on the moments or on the density function of the distribution, or properties of the ranked series. We will describe here tests that are based on properties of the original distribution.

Tests based on moments

The most widely used test in this category is due to Jarque and Bera (1980) and Bera and Jarque (1981) who rely on the fact that skewness and excess kurtosis are both equal to zero for the normal distribution. Jarque and Bera

Table 2.2. *Summary statistics of log-returns*

	SP500	DAX	FT-SE	Nikkei
	Daily frequency (6,437 observations)			
Mean	0.036	0.032	0.035	0.008
Median	0.011	0.019	0.039	0.000
Std deviation	1.047	1.329	0.926	1.266
MAD	0.712	0.917	0.657	0.851
IQR	1.017	1.322	0.999	1.140
Ann. mean	9.105	7.943	8.870	2.047
Ann. std dev.	16.624	21.096	14.693	20.099
Skewness	−1.751	−0.472	−0.803	−0.158
Kurtosis	42.513	9.847	13.295	12.066
	Weekly frequency (1,288 observations)			
Mean	0.180	0.159	0.176	0.041
Median	0.266	0.244	0.307	0.194
Std deviation	2.443	3.014	2.234	2.899
MAD	1.717	2.202	1.602	2.123
IQR	2.638	3.279	2.387	3.144
Ann. mean	9.377	8.253	9.176	2.150
Ann. std dev.	17.614	21.736	16.110	20.906
Skewness	−1.862	−0.588	−1.188	−0.214
Kurtosis	27.146	6.423	12.436	6.010
	Monthly frequency (296 observations)			
Mean	0.769	0.683	0.737	0.167
Median	1.043	1.133	1.176	0.613
Std deviation	4.486	6.338	4.859	5.794
MAD	3.349	4.644	3.518	4.419
IQR	5.417	7.114	5.269	7.156
Ann. mean	9.231	8.202	8.845	2.006
Ann. std dev.	15.540	21.956	16.831	20.071
Skewness	−0.875	−0.849	−1.398	−0.418
Kurtosis	6.394	5.766	9.136	3.957
	Annual frequency (25 observations)			
Mean	9.556	8.692	9.009	1.941
Median	10.923	11.980	13.039	7.594
Std deviation	14.598	27.419	15.819	22.914
MAD	10.931	21.854	11.555	18.914
IQR	18.821	39.918	15.979	30.030
Skewness	−0.831	−0.695	−1.412	−0.355
Kurtosis	3.299	3.047	4.420	2.374

(1980) show, for *iid* and normally distributed random variables, that the standardized skewness and excess kurtosis are asymptotically independent and have the following asymptotic distributions

$$\frac{\hat{s}}{\sqrt{6/T}} \underset{H_0}{\Rightarrow} \mathcal{N}(0,1) \qquad \text{and} \qquad \frac{\hat{\kappa}-3}{\sqrt{24/T}} \underset{H_0}{\Rightarrow} \mathcal{N}(0,1)\,.$$

Bera and Jarque's idea for testing normality is to use the JB test statistic defined as

$$JB = T\left[\frac{\hat{s}^2}{6} + \frac{(\hat{\kappa}-3)^2}{24}\right],$$

which is asymptotically distributed as $\chi^2(2)$.

The most severe limitation of the JB test is that the asymptotic distribution holds only for very large samples. To take into account the finite sample bias, Doornik and Hansen (1994) propose an "omnibus" test for normality. First, they produce approximations for the finite sample distributions of skewness and kurtosis under the assumption of normality and with the additional assumptions that the kurtosis has a Gamma distribution and that $\hat{\kappa} > 1+\hat{s}^2$.[7] Using z_1 and z_2 to denote the finite-sample skewness and kurtosis, Doornik and Hansen (1994) show, with the normality assumption, that

$$\begin{aligned}
\tilde{W} &= z_1^2 + z_2^2 \underset{app}{\Rightarrow} \chi^2(2),\\
z_1 &= \frac{1}{\sqrt{\log(\omega)}}\log\left(g+\sqrt{1+g^2}\right),\\
z_2 &= \left[\left(\frac{\chi}{2\alpha}\right)^{1/3} - 1 + \frac{1}{9\alpha}\right]\sqrt{9\alpha},
\end{aligned}$$

where $\underset{app}{\Rightarrow}$ denotes approximate distribution and

$$\begin{aligned}
g &= \hat{s}^2\sqrt{\frac{\omega^2-1}{2}\frac{(T+1)(T+3)}{6(T-2)}},\\
\omega^2 &= -1+\sqrt{2(b_0-1)},\\
\chi &= 2b_1\left(\hat{\kappa}-1-\hat{s}^2\right),\\
\alpha &= b_2 + b_3\hat{s}^2,
\end{aligned}$$

and with the following correction factors for finite sample

[7] An approximation of finite sample distribution of skewness can also be found in D'Agostino (1970).

$$b_0 = \frac{3\left(T^2+27T-70\right)(T+1)(T+3)}{(T-2)(T+5)(T+7)(T+9)},$$
$$b_1 = \frac{(T+5)(T+7)\left(T^3+37T^2+11T-313\right)}{12\tau},$$
$$b_2 = \frac{(T-2)(T+5)(T+7)\left(T^2+27T-70\right)}{6\tau},$$
$$b_3 = \frac{(T-7)(T+5)(T+7)\left(T^2+2T-5\right)}{6\tau},$$
$$\tau = (T-3)(T+1)\left(T^2+15T-4\right).$$

A second drawback of the JB test is that the empirical skewness and kurtosis are computed for given values of mean and variance, both of which are subject to sampling errors. Richardson and Smith (1993) and Harvey (1995) independently propose joint estimation of the moments based on the GMM (Generalized Method of Moments) orthogonality conditions[8]

$$e_{1,t} = r_t - \bar{r},$$
$$e_{2,t} = (r_t - \bar{r})^2 - \sigma^2,$$
$$e_{3,t} = (r_t - \bar{r})^3 / \sigma^3 - s,$$
$$e_{4,t} = (r_t - \bar{r})^4 / \sigma^4 - \kappa.$$

Solving the orthogonality conditions by the method of moments yields a joint estimate $\hat{\theta}$ of the first four moments $\theta = \left(\mu, \sigma^2, s, \kappa\right)'$ and an associated covariance matrix Σ_T, which helps to remove the dependence on the asymptotic distribution of these moments. Under the null hypothesis of normality, we have $s = \kappa - 3 = 0$. The test is performed by a Wald test using the statistic

$$W = T\left(G\hat{\theta}\right)' \times (G\Sigma_T G)^{-1} \times \left(G\hat{\theta}\right),$$

where

$$G = \begin{pmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}.$$

Under the null hypothesis of normality, W is distributed as a $\chi^2(2)$.

For the case of weakly dependent data, Bai and Ng (2001) provide the sampling distributions of the skewness and kurtosis estimators and propose a joint test for normality that takes this sampling distribution into account.

---

[8] Richardson and Smith (1993) estimate all four moments and test whether skewness and excess kurtosis are jointly zero. On the other hand, Harvey (1995) performs the estimation under the null and uses the last two orthogonality conditions as over-identifying conditions.

Tests based on densities

The key test here is the Kolmogorov-Smirnov (KS) test that compares the empirical *cdf* $F_r(.)$ of the returns with the *cdf* of the normal distribution (or any other assumed distribution) $F^*(.)$ with parameter vector θ. Since the true distribution $F_r(.)$ is unknown, it is approximated by the empirical *cdf* $G(.)$ defined as

$$G(x) = \frac{1}{T}\sum_{t=1}^{T} 1_{\{r_t \leq x\}}.$$

The null hypothesis is $H_0 : G(.) = F^*(x;\theta)$ for all x versus the alternative hypothesis $H_a : G(.) \neq F^*(x;\theta)$ for at least one x. The KS test assumes that θ is known. If θ is unknown, the Lilliefors test, explained later on, should be used.

One of the simplest measures is the largest distance between the two functions $G(x)$ and $F^*(x;\theta)$. This is known as the KS test as suggested initially by Kolmogorov (1933). The test statistic is defined as

$$KS = \sup_{\{x\}} \quad |F^*(x;\theta) - G(x)|.$$

In practice, this test is very easy to implement:

1. Sort the sample data by increasing order and denote the new sample $\{\tilde{r}_t\}_{t=1}^{T}$, with $\tilde{r}_1 \leq \cdots \leq \tilde{r}_T$. Then, by construction, the empirical *cdf* is given by $G(\tilde{r}_t) = t/T$.
2. Evaluate the assumed *cdf* $F^*(\tilde{r}_t;\theta)$ for all values $\{\tilde{r}_t\}_{t=1}^{T}$. In the case of the normal distribution, we assumed that the mean μ and the variance σ^2 are known.
3. Compute the KS test statistic

$$KS = \sup_{\{t\}} \quad \left| F^*(\tilde{r}_t;\theta) - \frac{t}{T} \right|.$$

The critical values of this statistic have been tabulated.

A drawback of the KS test is that the mean μ and the variance σ^2 are unknown and are subject to sampling errors when they are estimated empirically. In this case, we can apply the Lilliefors modified KS_L test, which involves, first of all, the estimation of $\hat{\mu}$ and $\hat{\sigma}^2$ from the data. It is then followed by steps 1 to 3 above, replacing $\left(\mu, \sigma^2\right)$ with $\left(\hat{\mu}, \hat{\sigma}^2\right)$ in step 2 to produce $F^*(\tilde{r}_t;\hat{\theta})$ in step 3, and a new set of critical values in the final stage. At the 95% confidence level, the new critical value for KS_L is $0.805/\sqrt{T}$.

Top panel of Figure 2.5 displays the empirical *cdf* $F^*(\tilde{r}_t;\hat{\theta})$ of the SP500 return, and the bottom panel presents the quantile plot (or QQ-plot) of the empirical and the normal distributions. The Lilliefors statistic KS_L corresponds to the maximum difference between the two curves (here, we have

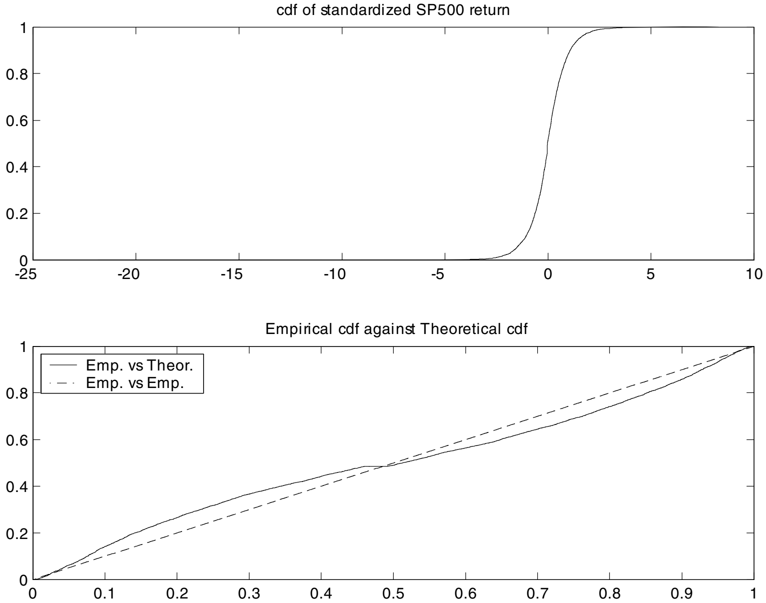

Fig. 2.5. *Comparison of the empirical distribution of the SP500 with the normal distribution as the theoretical distribution.*

$KS_L = 0.0692$). Given that the critical value is $0.805/\sqrt{6437} = 0.01$, the null hypothesis of normality is strongly rejected.

Table 2.3 presents various statistics to test normality as described above. The diagnosis on normality of returns is quite clear-cut. From daily to monthly frequencies, the different tests systematically reject the null hypothesis of normality. The rejection is due to both asymmetry and tail thickness of the distribution. In the case of annual returns, normality is not rejected. But in this instance, it is possible that the sample size is so small that the tests lack power and are simply unreliable.

2.3 Time dependency

Time dependency in asset returns can occur at several levels. We will discuss dependency in returns, volatility, and higher moments in turns in the following subsections. Time dependency is a crucial issue for several reasons: First, if return distribution is time dependent, then statistical tests using unconditional statistics and inferences derived thereof could be misleading. Second, if the time dependency can be fully exploited, it will help to produce better forecasts of level, volatility, and higher moments of returns. Such an improve-

Table 2.3. *Normality tests for log-returns*

	SP500	DAX	FT-SE	Nikkei
	Daily frequency (6,437 observations)			
Skewness	−1.751	−0.472	−0.803	−0.158
$t(s)$	−57.332	−15.468	−26.302	−5.186
Kurtosis	42.513	9.847	13.295	12.066
$t(\kappa)$	647.054	112.127	168.593	148.461
JB	421965.45	12811.81	29115.35	22067.65
p-value	0.000	0.000	0.000	0.000
KS_L	0.069	0.070	0.052	0.083
Critical value	0.010	0.010	0.010	0.010
	Weekly frequency (1,288 observations)			
Skewness	−1.862	−0.588	−1.188	−0.214
$t(s)$	−27.264	−8.606	−17.404	−3.138
Kurtosis	27.146	6.423	12.436	6.010
$t(\kappa)$	176.818	25.068	69.099	22.042
JB	32007.99	702.47	5077.56	495.69
p-value	0.000	0.000	0.000	0.000
KS_L	0.057	0.067	0.060	0.053
Critical value	0.022	0.022	0.022	0.022
	Monthly frequency (296 observations)			
Skewness	−0.875	−0.849	−1.398	−0.418
$t(s)$	−6.135	−5.956	−9.799	−2.931
Kurtosis	6.394	5.766	9.136	3.957
$t(\kappa)$	11.900	9.696	21.512	3.355
JB	179.246	129.487	558.767	19.844
p-value	0.000	0.000	0.000	0.000
KS_L	0.051	0.062	0.079	0.060
Critical value	0.047	0.047	0.047	0.047
	Annual frequency (25 observations)			
Skewness	−0.831	−0.695	−1.412	−0.355
$t(s)$	−1.662	−1.389	−2.825	−0.710
Kurtosis	3.299	3.047	4.420	2.374
$t(\kappa)$	0.299	0.047	1.420	−0.626
JB	2.850	1.932	9.996	0.895
p-value	0.085	0.155	0.001	0.327
KS_L	0.099	0.093	0.189	0.086
Critical value	0.161	0.161	0.161	0.161

ment is obviously very important for many finance applications such as risk management and asset allocation.

2.3.1 Serial correlation in returns

Here, the null hypothesis is that the first p serial correlations of returns are equal to 0, $H_0 : \rho_1 = \cdots = \rho_p = 0$, where the correlation of order j is estimated by

$$\hat{\rho}_j = \frac{\sum_{t=j+1}^{T} (r_t - \bar{r})(r_{t-j} - \bar{r})}{\sum_{t=1}^{T} (r_t - \bar{r})^2} \qquad \text{for} \qquad 0 \leq j < T - 1.$$

A simple test of H_0 could be based on the Ljung-Box Q statistic

$$Q_p = T(T+2) \sum_{j=1}^{p} \frac{1}{T-j} \hat{\rho}_j^2.$$

Under the null hypothesis of no serial correlation, the Q_p statistic is asymptotically distributed as $\chi^2(p)$. A common practice is to test H_0 repeatedly using several choices of p.

2.3.2 Serial correlation in volatility

To test for time dependency in volatility, we need a time-varying measure of volatility. There are at least two possible ways to approach this; first using mean-adjusted squared returns and secondly using absolute returns. Assume that returns have the following dynamics

$$r_t = \mu + \varepsilon_t, \qquad \text{with} \qquad \varepsilon_t = \sigma_t z_t,$$

where μ is the constant mean, ε_t the mean adjusted returns, σ_t the time-varying volatility, and z_t is an $\mathcal{N}(0,1)$ innovation. Then with information set $\mathcal{F}_{t-1}$ at time $t-1$,

$$E\left[\varepsilon_t^2 | \mathcal{F}_{t-1}\right] = \sigma_t^2 E\left[z_t^2 | \mathcal{F}_{t-1}\right] = \sigma_t^2$$

because z_t^2 is distributed as $\chi^2(1)$. Therefore, ε_t^2 can be viewed as a proxy for the volatility at time t. Alternatively, omitting μ for the moment, we have $r_t \sim \mathcal{N}\left(0, \sigma_t^2\right)$ and

$$E\left[|r_t|\right] = \sigma_t \sqrt{2/\pi}.$$

Consequently, $|r_t| / \sqrt{2/\pi}$ is a proxy for σ_t. It should be noted however that these two measures are noisy estimates of conditional volatility. See Chapter 4 for more detail and more sophisticated measures of conditional volatility.

Given these crude measures of volatility, the test for serial dependence in volatility can be executed in the same manner as the test for returns in the previous subsection using Ljung-Box Q statistic. Non-zero serial correlation in squared or absolute returns is evidence of volatility time dependence. Table 2.4 reports the Ljung-Box statistic for returns, squared returns, and absolute returns for $p = 10$ (and $p = 5$ in the case of annual data). The corresponding critical values are 18.307 for $p = 10$ (and 11.071 for $p = 5$). The results indicate that returns are serially correlated when they are sampled at the daily frequency but not at other frequencies. Squared and absolute returns are more strongly correlated, at least at the daily and weekly frequencies.

Table 2.4. *Ljung-Box test statistics of log-returns, squared returns, and absolute returns*

Frequency	SP500	DAX	FT-SE	Nikkei
		Returns		
Daily	21.205	29.882	69.534	29.347
Weekly	15.178	15.655	23.265	12.306
Monthly	9.537	5.412	6.660	7.418
Annual	11.727	4.946	2.838	2.280
		Squared returns		
Daily	438.200	2327.700	2923.500	776.500
Weekly	30.947	350.231	138.645	131.836
Monthly	5.653	21.260	3.592	52.751
Annual	1.850	1.530	8.154	8.345
		Absolute returns		
Daily	1784.700	4411.200	2780.600	2630.100
Weekly	194.380	537.175	190.139	238.459
Monthly	17.955	28.815	10.764	42.842
Annual	2.319	4.217	10.063	9.327

When sampled at monthly frequency, squared and absolute returns of DAX and Nikkei are still serially correlated reflecting the strong time dependence of volatility.

Indeed, the strong time dependence in volatility is an important feature in financial market modeling and gives rise to the birth of a huge class of conditional heteroskedasticity models (see Chapter 4). In fact, the autocorrelation coefficient of volatility proxies is significantly greater than zero even for very long horizons. To illustrate this point, we plot in Figure 2.6 the autocorrelogram of squared returns for up to 100 lags.[9] The figure shows that the serial correlation of DAX squared returns is significant up to lag 100. This phenomenon is now referred to as the long memory of volatility.

There has also been some debate as to whether volatility has a unit root (Perry, 1982, and Pagan and Schwert, 1990). That is, if

$$\sigma_t = \varphi \sigma_{t-1} + \varepsilon_t,$$

the issue concerns whether or not φ is indistinguishable from 1. Much of these discussions took place before the new concept of realized volatility developed. If we apply unit root test to the volatility series computed using intraday data, we categorically reject the null hypothesis of a unit root even at the 1% level (see Section 4.8).

[9] The 90% confidence interval reported in the figure is computed assuming that the estimator $\hat{\rho}_j$ is asymptotically normal $N(0, V)$, with V estimated by $\hat{V} = 1 + 2\sum_{i=1}^{j-1} \hat{\rho}_i^2$.

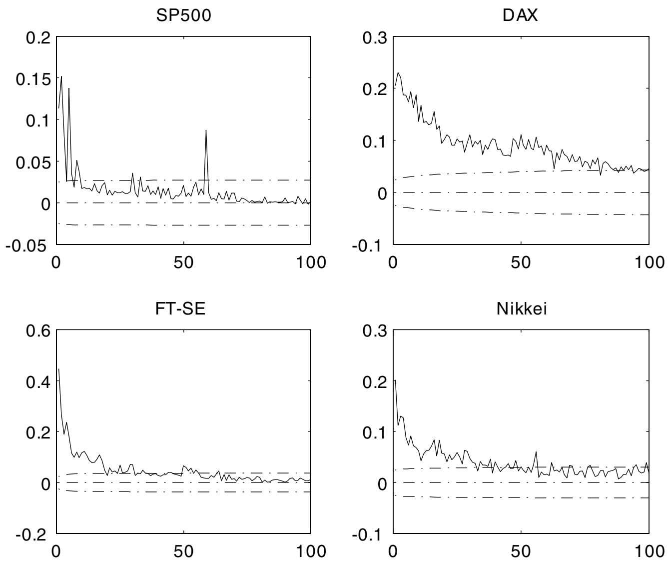

Fig. 2.6. *Autocorrelogram for daily squared returns.*

2.3.3 Volatility asymmetry

Another important feature of financial market volatility is that it is more affected by negative returns than by positive returns. To illustrate the point, we perform the following regressions

$$\varepsilon_t^2 = \omega + \alpha\varepsilon_{t-1}^2 + \beta\varepsilon_{t-1}^2 \times 1_{\{\varepsilon_{t-1}<0\}}, \tag{2.2}$$

or

$$|\varepsilon_t| = \omega + \alpha\,|\varepsilon_{t-1}| + \beta\,|\varepsilon_{t-1}| \times 1_{\{\varepsilon_{t-1}<0\}}, \tag{2.3}$$

where α measures the direct effect of past returns, and β captures the additional impact of negative return shocks. Table 2.5 reports the regression coefficients with their standard errors reported in parentheses. First note that α and β estimates are of similar magnitude for the case of absolute returns but not for the case of squared returns. For absolute returns, α ranges between 0.12 and 0.18, whereas β has a range of similar magnitude between 0.11 and 0.14. This means that, on average, a negative return shock has twice as much impact on volatility as a positive return shock. Black (1976) and Christie (1986) call this leverage effect, relating to the fact that when equity value decreases, leverage and, hence, risk and volatility increase. Campbell and Hentschel (1992) interpret the different responses as a stronger impact of bad news than good news. This asymmetric (or leverage) effect in volatility will be investigated further in Chapter 4.

Table 2.5. *Parameter estimates of volatlility asymmetry regressions (2.2) and (2.3)*

	SP500	DAX	FT-SE	Nikkei
		Squared returns		
ω	1.434	0.540	1.310	2.955
	(0.068)	(0.035)	(0.068)	(0.300)
α	0.134	0.237	0.079	0.061
	(0.021)	(0.024)	(0.019)	(0.048)
β	0.105	0.256	0.205	−0.003
	(0.025)	(0.027)	(0.024)	(0.050)
		Absolute returns		
ω	0.694	0.502	0.669	0.931
	(0.016)	(0.011)	(0.015)	(0.022)
α	0.178	0.180	0.154	0.127
	(0.015)	(0.015)	(0.015)	(0.017)
β	0.135	0.120	0.123	0.114
	(0.017)	(0.017)	(0.018)	(0.019)

2.3.4 Time-varying higher moments

One of the main features of asset returns is the volatility clustering, suggesting that the volatility of returns may vary over time depending on the arrival of new information. We may then ask if other moments of the distribution are also time-varying. To illustrate this issue, we plot in Figure 2.7 the evolution of the first four moments of the daily SP500 return, computed using a rolling window of 126 working days, which is about 6 months of data. The figure indicates that, beyond the October 1987 crash, there has been a recent upward trend in the evolution of the volatility (measured as the standard deviation of the series). Skewness and kurtosis are clearly affected by large jumps. The median value for the skewness is −0.11, and the median value for the kurtosis is 3.62. These values suggest that, over shorter sample periods, the normality assumption may hold. But, over a longer sample period, return is prone to large negative shocks that have a severe impact on the skewness and kurtosis estimates.

2.4 Linear dependence across returns

In this section, we turn to some stylized facts concerning the joint distribution of returns, which can be analyzed by separating the marginal distributions and the dependence structure. The marginal distributions are the univariate distributions that characterize the individual series. The dependence structure, on the other hand, describes how the individual series are related to each other, usually without the influence of the univariate distribution. Here, we focus on the dependence structure.

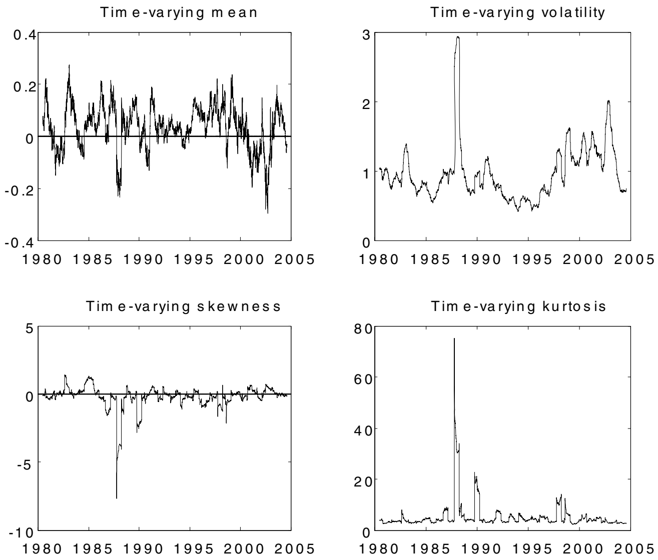

Fig. 2.7. *Time-varying moments for the daily SP500 return, computed over rolling windows of 6 months.*

2.4.1 Pearson's correlation coefficient

The most frequently used measure of dependence is the *linear correlation*, or Pearson's correlation

$$\rho[X,Y] = \frac{Cov[X,Y]}{\sqrt{V[X]\,V[Y]}},$$

where $Cov[X,Y]$ is the covariance between X and Y. The Pearson's correlation satisfies the constraint $-1 \leq \rho[X,Y] \leq 1$, and has as special cases: $\rho[X,X] = 1$ and $\rho[X,Y] = \rho[Y,X]$. Since $\rho[\alpha X+\beta, \gamma Y+\delta] = sign(\alpha\gamma)\,\rho[X,Y]$, correlation is invariant under strictly increasing linear transformations. When $|\rho[X,Y]| = 1$, we have perfect correlation (or perfect linear dependence), because it means that Y can be written as $Y = \alpha X + \beta$. When $\rho[X,Y] = 0$, the two series are uncorrelated. It is worth emphasizing that "no correlation" does not necessarily mean "independence". Correlation is a good measure for dependence if returns have elliptical distributions, which include normal and Student t distributions.[10] Correlation may not have any relationship with dependence for non-elliptical distributions.

A consistent estimator of the Pearson's correlation between two series $\{x_t\}_{t=1}^{T}$ and $\{y_t\}_{t=1}^{T}$ is

[10] Elliptical distributions are defined and analyzed in Section 6.2.

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

where $\bar{x}$ and $\bar{y}$ are the sample mean of x and y, respectively. The asymptotic distribution of $\hat{\rho}$ is

$$\sqrt{T}(\hat{\rho} - \rho) \Rightarrow \mathcal{N}(0,1).$$

Table 2.6 reports the correlation matrix for returns on the four stock market indices measured at various frequencies. All correlation coefficients reported in the table are positive. Moreover, the lower the data frequency, the higher the correlation coefficient estimate. Figure 2.8 presents the evolution of the correlation coefficients computed using a rolling period of 6 months. Except for the pair SP500 vs. Nikkei, all correlation coefficients exhibit an upward trend.

2.4.2 Test for equality of two correlation coefficients

There have been a lot of discussions in the literature on how the correlation changes when the financial markets are "agitated". This is an important issue in portfolio diversification. If correlation increases during turbulent periods and stock market crashes, then the benefits of diversification disappear when they are most needed. Previous studies on changing correlation estimated it

Table 2.6. *Correlation matrix of log-returns*

	SP500	DAX	FT-SE	Nikkei
		Daily frequency		
SP500	1	0.358	0.369	0.109
DAX	0.358	1	0.564	0.236
FT-SE	0.369	0.564	1	0.268
Nikkei	0.109	0.236	0.268	1
		Weekly frequency		
SP500	1	0.535	0.548	0.301
DAX	0.535	1	0.592	0.339
FT-SE	0.548	0.592	1	0.357
Nikkei	0.301	0.339	0.357	1
		Monthly frequency		
SP500	1	0.598	0.714	0.413
DAX	0.598	1	0.623	0.373
FT-SE	0.714	0.623	1	0.412
Nikkei	0.413	0.373	0.412	1
		Annual frequency		
SP500	1	0.711	0.798	0.405
DAX	0.711	1	0.753	0.474
FT-SE	0.798	0.753	1	0.629
Nikkei	0.405	0.474	0.629	1

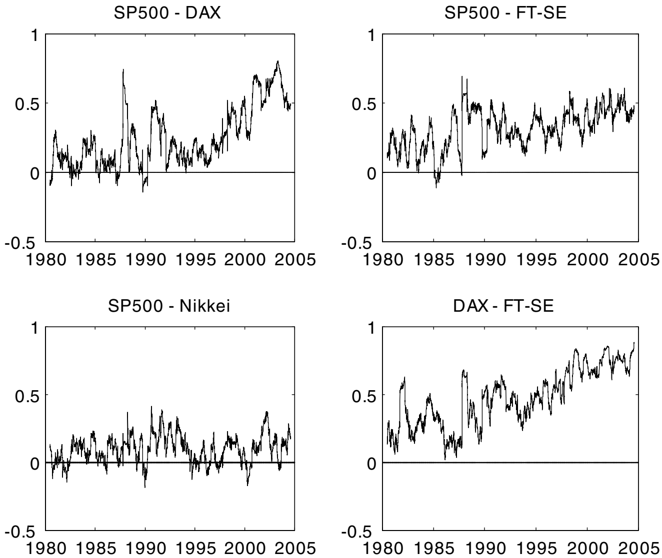

Fig. 2.8. *Time-varying correlations between daily stock market return computed using a rolling window of 6 months.*

either from different sample periods or conditional on one of the return series exceeding a threshold. The first type of studies includes Kaplanis (1988) and Ratner (1992) who find the correlation matrix to remain constant before and after the 1987 stock market crash. On the contrary, Koch and Koch (1991) find that correlation increases through time, but King, Sentana, and Wadhwani (1994) argue that the 1987 stock market crash is the main cause of the increase in correlation. However, Boyer, Gibson, and Loretan (1997), Loretan and English (1999), and Forbes and Rigobon (2002) show that the correlation coefficient between two series is biased, when it is computed conditionally on one of the series exceeding a threshold. Therefore, even when the breaking date is assumed to be known (corresponding to a well-established crash, for instance), unconditional correlation estimates have to be corrected before any testing procedure. As a consequence, a higher correlation coefficient could be a spurious outcome caused by higher volatility. To test for a change in the correlation relationship, it is necessary (i) to identify a data generating process that allows for the possibility of structural changes, (ii) to estimate the model parameters, and (iii) to test for changing correlations and possibly other structural breaks.

The above discussions highlight the complexities involved in estimating and testing correlation coefficients, the conditional nature of the distributional variables, and the complications caused by structural breaks. We will address these issues in a greater detail in Chapter 6. Here, we present the simplest case when there is no structural break and changing distributional characteristics other than possibly the correlation relationship. Before executing the test on correlation, it is also a good practice to first filter out the time dependence of the individual series. In some cases, the removal of common factor dependence is helpful in identifying the true dependence structure.

The test of the equality of the correlation coefficients of two non-overlapping periods T_1 and T_2 relies on the asymptotic distribution of the correlation coefficient. Since the distribution becomes unstable as $\rho \to 1$, Fisher (1915) introduces the z-transformation and recommends that the test be conducted on $z(\hat{\rho})$ instead of $\hat{\rho}$, where $\hat{\rho}$ is the estimated correlation coefficient. Formally,

$$z(\hat{\rho}) = \frac{1}{2}\log\left(\frac{1+\hat{\rho}}{1-\hat{\rho}}\right).$$

Under the assumption that two samples are drawn from two independent bivariate normal distributions with the same correlation coefficient, the difference between the estimated $z(\hat{\rho})$ for the two samples converges to the normal distribution $\mathcal{N}(0, 1/(T_1 - 3) + 1/(T_2 - 3))$.

2.4.3 Test for equality of two correlation matrices

Jennrich (1970) develops a test based on the normalized difference between two correlation matrices. Let R_1 and R_2 denote the correlation matrices for two independent subsamples of equal size $T_1 = T_2 = T$. Then the normalized difference is

$$Z = \sqrt{\frac{T}{2}}R^{-1}(R_1 - R_2),$$

where $R = \frac{1}{2}(R_1 + R_2)$ is the average correlation matrix over the two subsamples and for $R = (\hat{\rho}_{ij})$, the inverse is $R^{-1} = (\hat{\rho}^{ij})$, where $\hat{\rho}^{ij}$ denotes the components of the inverse of the correlation matrix. Then the test statistic χ^2 is defined as

$$\chi^2 = \frac{1}{2}\mathrm{tr}\left(Z^2\right) - \mathrm{diag}(Z)' S^{-1}\mathrm{diag}(Z), \tag{2.4}$$

where $S = \left(\delta_{ij} + \hat{\rho}^{ij}\hat{\rho}_{ij}\right)$ with

$$\delta_{ij} = \begin{cases} 1 \text{ if } i = j, \\ 0 \text{ otherwise,} \end{cases}$$

diag(A) denotes the diagonal of the square matrix A in a column form and tr(B) denotes the trace of B. The Jennrich test statistic χ^2 has an asymptotic chi-square distribution with $p(p-1)/2$ degrees of freedom, where p is the number of variables or the dimension of the correlation matrix.

2.5 Multivariate higher moments

The ability to compute *multivariate higher moments* is important in some financial applications such as asset allocation. In this section, we will first define multivariate higher moments (or co-moments of higher order). Then, we will describe how the moments of a portfolio return can be expressed in a very convenient way.

2.5.1 Multivariate co-skewness and co-kurtosis

If there are n assets in the portfolio, then the $\left(n, n^2\right)$ co-skewness matrix is defined as

$$\begin{aligned} M_3 &= E\left[(r-\mu)(r-\mu)' \otimes (r-\mu)'\right] = \{s_{ijk}\}, \\ s_{ijk} &= E\left[(r_i-\mu_i)\left(r_j-\mu_j\right)(r_k-\mu_k)\right] \quad \text{for } i,j,k = 1,\cdots,n, \end{aligned}$$

where r_i denotes the individual asset return i, μ is the mean, and $\otimes$ is the Kronecker product.[11] For instance, in the case of $n = 3$ assets, the resulting $(3,9)$ co-skewness matrix is

$$\begin{aligned} M_3 &= \left[\begin{array}{ccc|ccc|ccc} s_{111} & s_{112} & s_{113} & s_{211} & s_{212} & s_{213} & s_{311} & s_{312} & s_{313} \\ s_{121} & s_{122} & s_{123} & s_{221} & s_{222} & s_{223} & s_{321} & s_{322} & s_{323} \\ s_{131} & s_{132} & s_{133} & s_{231} & s_{232} & s_{233} & s_{331} & s_{332} & s_{333} \end{array}\right] \\ &= \left[\, S_{1jk} \; S_{2jk} \; S_{3jk} \,\right], \end{aligned}$$

where S_{1jk} denotes the (n,n) matrix with elements $\{s_{1jk}\}_{j,k=1,2,3}$.

The $\left(n, n^3\right)$ co-kurtosis matrix is defined as

$$\begin{aligned} M_4 &= E\left[(r-\mu)(r-\mu)' \otimes (r-\mu)' \otimes (r-\mu)'\right] = \{\kappa_{ijkl}\}, \\ \kappa_{ijkl} &= E\left[(r_i-\mu_i)\left(r_j-\mu_j\right)(r_k-\mu_k)(r_l-\mu_l)\right] \quad \text{for } i,j,k,l = 1,\cdots,n. \end{aligned}$$

Again, with the $n = 3$ example, the $(3,27)$ co-kurtosis matrix is

$$M_4 = \left[\, K_{11kl} \; K_{12kl} \; K_{13kl} \,\middle|\, \cdots \,\middle|\, K_{31kl} \; K_{32kl} \; K_{33kl} \,\right],$$

where K_{11kl} denotes the (n,n) matrix with elements $\{\kappa_{11kl}\}_{k,l=1,2,3}$.

These notations are extensions of the covariance matrix M_2. They have been used by Harvey et al. (2002), Prakash, Chang, and Pactwa (2003), and Athayde and Flôres (2004). It should be noted that, because of certain symmetries, not all the elements of these matrices have to be computed. For example, the dimension of the covariance matrix is (n,n), but only $n(n+1)/2$

[11] If $A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$ and $B = \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix}$, then $A \otimes B = \begin{bmatrix} a & 2a & b & 2b \\ 3a & 4a & 3b & 4b \\ c & 2c & d & 2d \\ 3c & 4c & 3d & 4d \end{bmatrix}$. In general, if A has dimension $(m \times n)$ and B has dimension $(p \times q)$, $A \otimes B$ has dimension $(mp \times nq)$.

elements have to be computed. Similarly, the co-skewness and co-kurtosis matrices have dimensions (n, n^2) and (n, n^3), but only involve, respectively, $n(n+1)(n+2)/6$ and $n(n+1)(n+2)(n+3)/24$ different elements.[12,13]

2.5.2 Computing moments of portfolio returns

Now, using these notations from the previous subsection, moments of the portfolio return can be computed in a very tractable way. For a given portfolio weight vector α, moments of the portfolio return are

$$\mu_p = \alpha' \mu \qquad \sigma_p^2 = \alpha' M_2 \alpha$$
$$s_p = \alpha' M_3 (\alpha \otimes \alpha) \qquad \kappa_p = \alpha' M_4 (\alpha \otimes \alpha \otimes \alpha).$$

The second- to fourth-order moments of the portfolio return can also be written in a slightly different form

$$\sigma_p^2 = E\left[\sum_{i=1}^{n} \alpha_i (r_i - \mu_i)\left(r_p - \mu_p\right)\right] = \alpha' \Sigma_p,$$
$$s_p = E\left[\sum_{i=1}^{n} \alpha_i (r_i - \mu_i)\left(r_p - \mu_p\right)^2\right] = \alpha' S_p,$$
$$\kappa_p = E\left[\sum_{i=1}^{n} \alpha_i (r_i - \mu_i)\left(r_p - \mu_p\right)^3\right] = \alpha' K_p,$$

where

$$\Sigma_p = E\left[(r_i - \mu_i)\left(r_p - \mu_p\right)\right] = M_2 \alpha,$$
$$S_p = E\left[(r_i - \mu_i)\left(r_p - \mu_p\right)^2\right] = M_3 (\alpha \otimes \alpha),$$
$$K_p = E\left[(r_i - \mu_i)\left(r_p - \mu_p\right)^3\right] = M_4 (\alpha \otimes \alpha \otimes \alpha),$$

are, respectively, the $(n, 1)$ vectors of covariances, co-skewness, and co-kurtosis between the asset returns and the portfolio return.[14]

[12] For $n = 3$, we have to compute 6 out of 9 elements for the covariance matrix, 10 out of 27 elements for the co-skewness matrix, and 15 out of 81 elements for the co-kurtosis matrix.

[13] There are other ways of calculating co-moments. In particular, it may be possible to improve the efficiency of co-moment measures by specifying the joint distribution. For instance, Simaan (1993) assumes all returns depend on a random factor with a non-spherical distribution. When no specific structure for the conditional joint distribution of returns is assumed, the measures of co-moment may suffer from inefficiency. See also Chapter 5.

[14] The notations Σ_p, S_p, and K_p are directly related to the notions of systematic risk and are widely used in the literature on higher-moment CAPM, see Kraus and Litzenberger (1976), Ingersoll (1987), Hwang and Satchell (1999), Jurczenko and Maillet (2001).