LECTURE 2

3 Time Series and their characteristics

Time series analysis deals with analysis of data recorded over time.

In time series, observations $X_1, X_2, \cdots, X_n$ are generally dependent. Dependent data series (X_t) are modelled as random sequences or random (stochastic) processes.

Definition 1 *A sequence of random variables* $\{X_t\}$ *indexed by discrete time* $t = ..., -1, 0, 1, 2, ...,$ *is called a* **time series**.
(If time t is continuous, $\{X_t\}$ *is called a random process.)*

3.1 Financial time series. Asset returns

Financial time series are concerned with the theory and practice of asset valuation over time (prices/returs).

Uncertainty. Empirical financial time series contain an element of *uncertainty.*
→ Therefore statistical theory and methods play an important role in time series analysis.
Comment. Most financial studies deal with returns, instead of prices, of assets. Returns for an investor is a complete scale free summary of the investment opportunity. Returns are easier to handle than prices; they 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 t. We present two definitions of an asset return.

One-period Simple return. Holding the asset for one period from time $t-1$ to t yields a *simple gross return*

$$1 + R_t = \frac{P_t}{P_{t-1}}, \qquad or \quad P_t = P_{t-1}(1 + R_t).$$

Definition: the simple net return or *simple return* is

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

Multiperiod Simple return. Holding the asset for k periods from time $t-k$ to t gives a k-period *simple gross return:*

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

Hence, k-period *simple gross return* is just the product of k one-period simple gross returns. It is also called a **compound return.**

Definition: the k-period net return, which is commonly called *simple return* is

$$R_t[k] = \frac{P_t}{P_{t-k}} - 1 = \frac{P_t - P_{t-k}}{P_{t-k}}.$$

Continuous compounding

The effect of compounding: assume that the interest rate of a bank is 10% per annum and the initial deposit is \$1. [Use $P_t = P_{t-k}(1+R_t[k])$]

- If the bank pays interest rate once a year then the net value of deposit becomes $1(1+0.1) = 1.1$.
- If the bank pays interest rate twice a year then the 6-month interest rate is 5%, and the net value of deposit after the first year is $1(1+0.1)(1+0.1) = 1.10250$.
- If the bank pays interest rate m times a year then the interest for each payment is $10\%/m$ and the net value of deposit after the first year is $1(1+0.1/m)^m$. As m increases, this number approaches $1.1502 = \exp(0.1)$, where $0.1 = r = \frac{10\%}{100\%}$ is the interest rate per annum, which is the result of continuous compounding.

Table 1 gives results for commonly used time intervals on a deposit \$1 with interest rate 10% per annum.

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

Summarizing: if r is the interest rate per annum, C is the initial capital and n is the number if years then net value asset of continuous compounding is n

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

Hence, after one period, the investment P_{t-1} will become

$$P_t = P_{t-1}e^r.$$

Notice that

$$\ln P_t = \ln(P_{t-1}e^r) = \ln(P_{t-1}) + \ln(e^r) = \ln(P_{t-1}) + r.$$

Then $\quad r = \ln P_t - \ln(P_{t-1}).$

Log return. The natural logarithm of the simple gross return is called continuously compounded or log-return:

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

Note that for small R_t, it holds $\ln(1 + R_t) \sim R_t$. Then $r_t \sim R_t$.

Log-return enjoys some advantages over the simple return.

- First, the multiperiod return is the sum of one period returns:

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

- Secondly, statistical properties of log returns are more tractable.

Portfolio return. Suppose portfolio consists of 3 assets P_{1t}, P_{2t}, P_{3t}. These assets yield simple returns R_{1t}, R_{2t}, R_{3t}.

Let $w_1 100\%$, $w_2 100\%$, $w_3 100\%$ be percentage of the portfolio value of each asset. Then $w_1 + w_2 + w_3 = 1$.

Then simple return of portfolio at time t is

$$R_{portfolio} = w_1 R_{1t} + w_2 R_{2t} + w_3 R_{3t}.$$

Excess return of an asset at time t is the difference between the assets return R_t and the return of some reference asset R_{0t}.

Reference asset is often take to be riskless such as a short term ES Treasury bill returns. The simple excess return and log excess return are defined then as

$$Z_t = R_t - R_{0t}, \qquad z_t = r_r - r_{0t}.$$

In finance, this would be a payoff of portfolio which goes for long in an asset and short in the reference asset.

Remark (Long and Short).

- A long position means owning the asset.
- A short position means selling an asset which one does not own. An asset (shares) is borrowed from an investor who purchased it. At some time, the short seller must return shares to the lender. Cash dividends on the borrowed asset must be payed also to the lender.

Summary of relationships

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

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

- If returns are in percentages then

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

- Aggregation in time produces k-period returns

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

- If continuously compounded interest rate is r per annum, the the relation between present value c and future value A after time n is

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

Example

- If the monthly log return is $R = 4.46\%$ then the monthly simple return is
$$r = 100(\exp(4.46/100) - 1] = 4.56\%$$
- If monthly log returns of an asset within a quarter are 4, 46%, −7, 34% and 10.77%, then the quarterly log return is $(4,46 - 7,34 + 10.77)\% = 7.89\%$.

Transformations. Times series often have to be transformed to series which have nicer statistical properties.

For example, the prices P_t behaves as a non-stationary (increasing) stochastic trend. It is more realistic to expect that changes in prices from a stationary process. Indeed, log-returns r_t which are differences $\ln(P_t) - \ln(P_{t-1})$ of log prices $\ln(P_t)$ tend to behave as a stationary process:

P_t (non-stationary) $\rightarrow \log P_t$ (non-stationary) $\rightarrow r_t = \log P_t - \log P_{t-1}$ (stationary)

See Figure 1.6 (a) where P_t is exchange rate between US dollar and Japanese yen (non-stationary), and (b), which shows log returns of the exchange rate, which behave as a stationary time series.

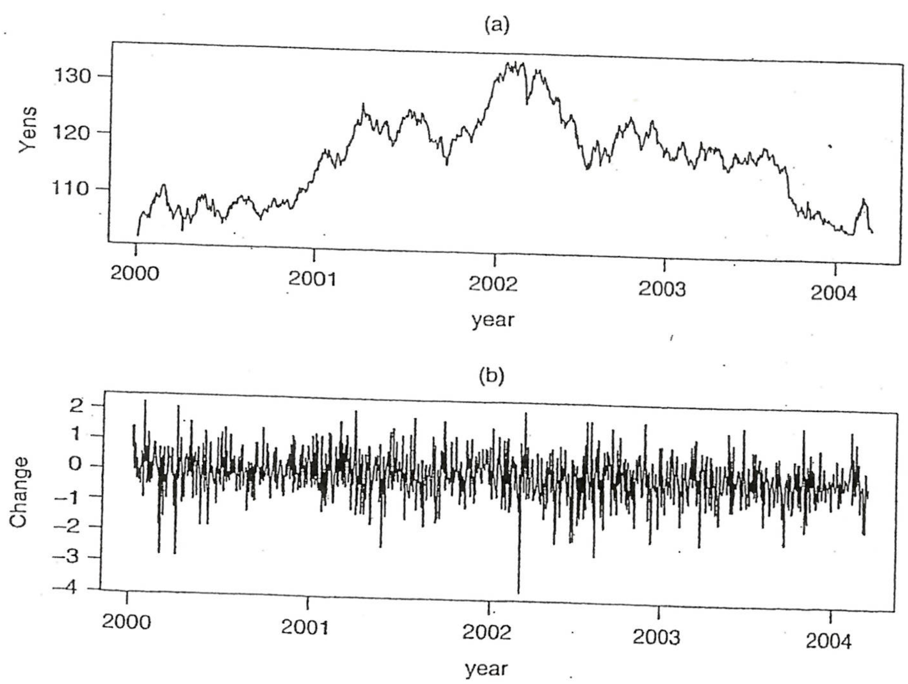


Figure 1.6. Time plot of daily exchange rate between U.S. dollar and Japanese yen from January 3, 2000 to March 26, 2004: (a) exchange rate and (b) changes in exchange rate.

3.2 Time series characteristics: statistical distribution and moments

Time series $X_1, \cdots, X_n$ is a collection of random variables.

Each X_j is a random variable. It is described by its distribution. It also can be partly characterized by moments.

Let X denotes the price of that asset at some fixed time t. This price is a random variable. So we have all information about X, if we know all probabilities

$$F(y) := P(X \leq y) \qquad \text{for all } y\text{'s.}$$

Function F is called distribution function of X.

Example:

$$P(X \geq K, \quad K \text{ is large}) = P(\text{event: "price X is very high"}) \sim 0,$$
$$P(X \leq E[X] \quad \text{average value of } X) \sim 0.5;$$
$$P(X \leq K, \quad K \text{ is small}) = P(\text{event: "price X is very small"}) \sim 0.$$

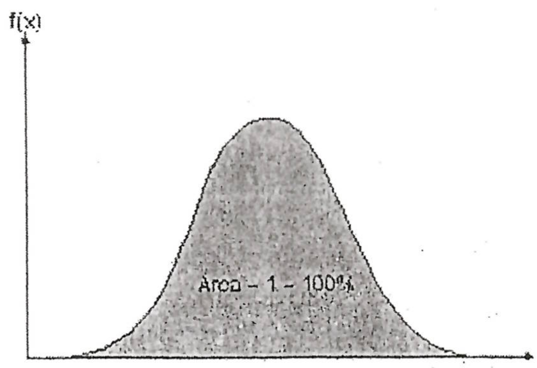
What is probability? It is a measure of random events.

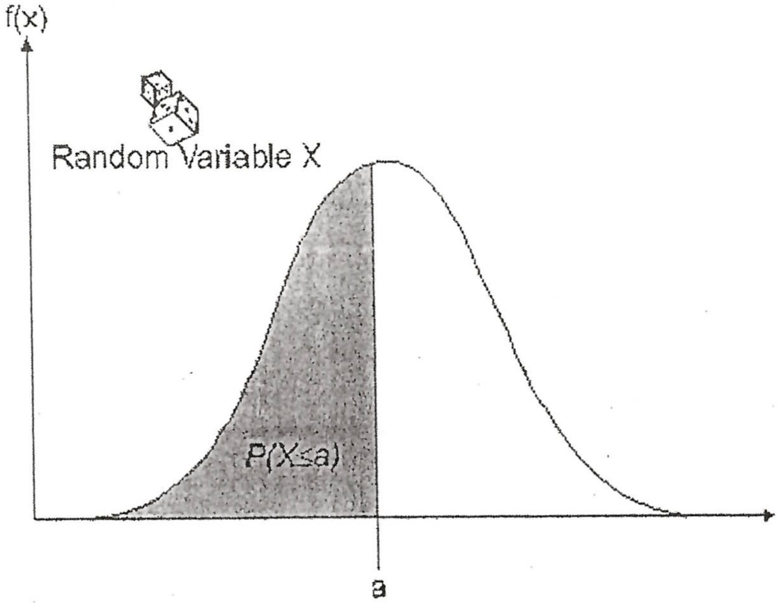
To a random event A it assigns a number $P(A) \in [0, 1]$. Higher $P(A)$ indicates that event A is more likely to happen.

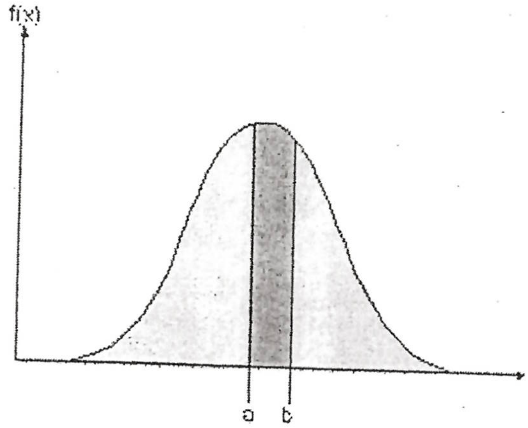
For two not-overlapping events A and B we have

$$P(A \cup B) = P(A) + P(B);$$
$$\text{In general, } P(A \cup B) \leq P(A) + P(B).$$

Probability density function f of a random variable X: it is a nonnegative functions, that completely describes distribution of X.







$$P(a \leq X \leq b) = \int_a^b f(x)dx \text{ and } f(x) \geq 0 \text{ for all } x$$

Definition. Gaussian or normal random variable $X \sim N(\mu, \sigma^2)$ with mean μ and variance σ^2 has the the distribution density

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp(-\frac{(x-\mu)^2}{2\sigma^2}), \quad x \in R.$$

Definition. Standard normal random variable $Z \sim N(0,1)$ has the distribution density

$$f(x) = \frac{1}{\sqrt{2\pi}} \exp(-\frac{x^2}{2}), \quad x \in R.$$

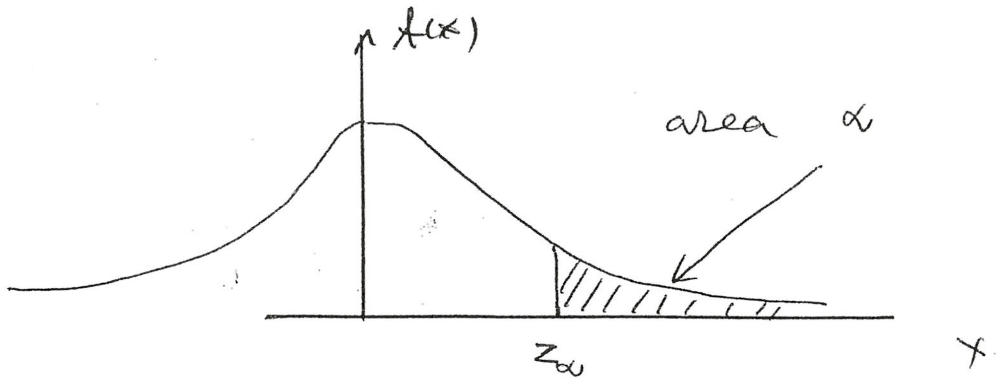
It has mean $\mu = 0$ and variance $\sigma^2 = 1$.
Standard deviation: σ is called deviation

If X has normal distribution $N(\mu, \sigma^2)$, then σ can be used as measuring unit, to describe the intervals of values taken by X.

$$Probab\Big(X \text{ takes values from interval } [\mu - \sigma, \quad \mu + \sigma]\Big) = 0.6827$$
$$Probab\Big(X \text{ takes values from interval } [\mu - 2\sigma, \quad \mu + 2\sigma]\Big) = 0.9545;$$
$$Probab\Big(X \text{ takes values from interval } [\mu - 3\sigma, \quad \mu + 3\sigma]\Big) = 0.9973.$$

Upper quantile of standard normal $Z \sim N(0,1)$ distribution. [It will appear in hypothesis testing]

Let $0 < \alpha < 1$. Then z_α is such that the area to the right equals to α.



In applications, usually we use
Foe $\alpha = 5\%$, $z_{5\%} = 1.64$,
For, $\alpha = 2.5\%$, $z_{2.5\%} = 1.96 \sim 2$.

Independence. We say that two random variables X and Y are independent, if Y does not carry any information about X. Then

$$E[XY] = E[X]E[Y],$$
$$E[X|Y] = E[X],$$

where $E[X|Y]$ is conditional expectation of X when we know Y.

3.3 Summary characteristics

- Instead of distribution function, random variable X can be characterized in part by its moment characteristics: mean, variance, kurtosis and skewness.

- in applications mean, variance, kurtosis and skewness can be estimated from the data. The estimates are called summary statistics

The mean. The first moment $\mu_X = E[X]$ is called the *mean* or expectation of X.

Theoretically, it can be computed as

$$E[X] = \int_{-\infty}^{\infty} y f(y) dy.$$

Note: we do not know probability density f.

Estimation: Let $X_1, .., X_N$ be as sample of X with N observations. We can estimate the mean by sample mean:

$$\hat{\mu}_X = \frac{1}{N} \sum_{j=1}^{N} X_j.$$

Variance. The variance of X is defined as

$$\sigma_X^2 = Var(X) = E(X - E[X])^2.$$

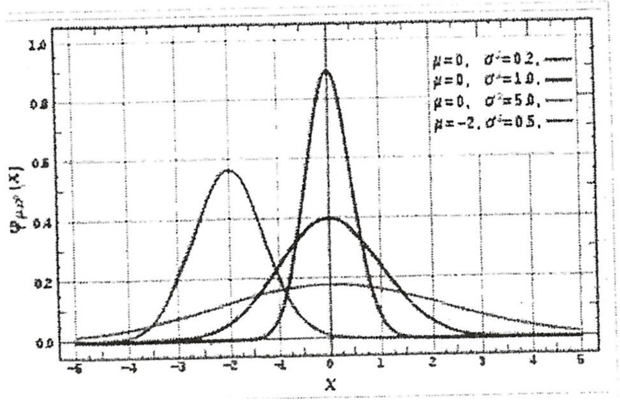
It measures the variability of X. Note that

$$\begin{aligned} Var(X) &= E(X - E[X])^2 = E(X^2 - 2XE[X] + E[X]^2) \\ &= E[X^2] - (E[X])^2. \end{aligned}$$

Standard deviation. The positive square root, σ_X, is called the standard deviation of X.

Estimation: We estimate σ_X^2 from that data $X_1, \cdots, X_N$ by sample variance

$$\hat{\sigma}_X^2 = \frac{1}{N} \sum_{j=1}^{N} (X_j - \hat{\mu}_X)^2.$$



Note: the first two moments μ_X and σ_X^2 define a normal distribution. Gaussian (normal) random variable with mean μ and variance σ^2 has the probability density

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp(-\frac{(x-\mu)^2}{2\sigma^2}).$$

For other distributions, higher order moments are also of interest.

Skewness. It measures asymmetry of X around the mean $E[X]$. Skewness is defined as the normalized 3rd moment

$$S(X) = E[\frac{(X-\mu_X)^3}{\sigma_X^3}].$$

In X has normal distribution, the $S(X) = 0$. Normally distributed variables are not skewed.

Estimation: We estimate skewness $S(X)$ from data $X_1, \cdots, X_N$ by sample skewness:

$$\hat{S}(X) = \frac{1}{N\hat{\sigma}_X^3} \sum_{j=1}^{N} (X_j - \hat{\mu}_X)^3.$$

Kurtosis. It measures the tail behaviour of X. The 4-th normalized central moment

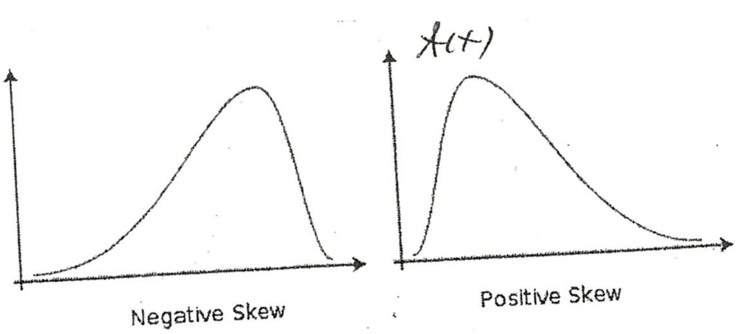
$$K(X) = E[\frac{(X-\mu_X)^4}{\sigma_X^4}]$$

is called the kurtosis of X. It measures the tail thickness of X distribution.

Estimation: We estimate kurtosis $K(X)$ from that data $X_1, \cdots, X_N$ by sample kurtosis:

$$\hat{K}(X) = \frac{1}{N\hat{\sigma}_X^4} \sum_{j=1}^{N} (X_j - \hat{\mu}_X)^4.$$

Note: Kurtosis of a normal random variable X is $K(X) = 3$.



Excess kurtosis: The quantity $K(X) - 3$ is called *excess kurtosis*.

- For a normal variable X, $K(X) - 3 = 0$.
- A distribution with a positive excess kurtosis is called *leptokurtic* or a distribution with *fat tails*. In practice such a distribution tends to contain more extreme values than a normal distribution.

Skewness and kurtosis are measures of shape distribution.

Tests for normality.

If X has normal distribution then $S(X) = 0$ and $K(X) = 3$. Let $X_1, \cdots, X_N$ be a random sample of X.

Under normality assumption on X, $\hat{S}(X)$ and $\hat{K}(X) - 3$ have asymptotically normal distribution:

$$\frac{\hat{S}(X)}{\sqrt{6/N}} \sim N(0,1), \qquad \frac{\hat{K}(X) - 3}{\sqrt{24/N}} \sim N(0,1).$$

These asymptotic properties can be used to test the normality of a time series, e.g. returns, i.e. to test that $S(X) = 0$ and $K(X) - 3 = 0$.

Test for symmetry. To test for skewness of distribution we test

$$\text{null-hypothesis } H_0 : S(X) = 0, \qquad \text{versus alternative } H_1 : S(X) \neq 0.$$

The *t- ratio statistics* of samples skewness is

$$t = \frac{\hat{S}(X)}{\sqrt{6/N}} \sim N(0,1).$$

Before testing

- select the significance level α (probability to reject true H_0). Usually $\alpha = 5\%$, $\alpha = 10\%$;
- find the upper -$\alpha/2$ quantile $z_{\alpha/2}$ of the standard normal distribution $N(0,1)$.

  For example, $z_{0.025} = 1.96 \sim 2$, $z_{0.1} = 1.64$.

Rule: reject H_0 at significance level α, if $|t| > z_{\alpha/2}$.
If $|t| \leq z_{\alpha/2}$, do not reject H_0 and normality.
Alternatively, one can compute the p value of the test statistic t and reject H_0 if p-value is less than α.

Test for heavy tails

Similarly, one can test the excess kurtosis, of time series, using the hypothesis

$$H_0 : K(X) - 3 = 0, \text{ versus } H_1 : K(X) - 3 \neq 0.$$

The test statistics is

$$t = \frac{\hat{K}(X) - 3}{\sqrt{24/N}} \sim N(0,1).$$

If H_0 is true, then $t \sim N(0,1)$ and testing can be done as above in case of skewness:

We reject H_0 (and normality) if the p-value is less than significance level α.

Jargue and Bera test combines the two prior tests and is based on statistics:

$$JB = \frac{\hat{S}^2(X)}{6/N} + \frac{(\hat{K}(X) - 3)^2}{24/N}$$

which under normality, has asymptotic non-normal distribution (chi-squared distribution with 2 degrees of freedom).

We reject hypothesis H_0 of normality if the p-value of JB statistic is less than significance level α.

Tables of summary statistics.

The summary statistics of mean, variance, skewness and kurtosis can be obtained easily using various statistical packages (see Table 1.2).

Example. Table 1.2 shows that excess kurtosis for daily simple returns for IBM stock is high, which implies that simple returns have heavy tails.

To test the symmetry of returns distribution, we use the test statistic

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

It gives a p-value of about 0.013, indicating that daily simple returns of IBM stock are significantly skewed to the right at 5% level and do not have normal distribution.

Comment: we would reject H_0 hypothesis of symmetry because p-values $p = 0.013 < 0.05$.

Definition 2 *p-value for t-ratio test statistic $t \sim N(0,1)$ taking value 2.49 is defined as probability $p = P\{|N(0,1)| > 2.49\}$.*

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.

Table 1.2 provides descriptive statistics of simple and log returns for selected US market indexes and individual stocks. It shows the following:

- Daily returns tend to have higher excess kurtosis the monthly returns.
- The mean of daily returns is close to zero while for monthly returns slightly higher.
- Monthly returns have higher standard deviation than daily returns
- skewness is not significant for daily and monthly returns.
- daily returns of individuals stocks have higher standard deviation than market indexes.
- difference between log and simple returns is not substantial
- The minimum and maximum of return series can be substantial. The negative extreme returns are important in risk management. Positive extreme returns are critical in holding a short position.

Figure 1.4 shows the empirical distribution density of monthly simple and log return of IBM stock. Dashes density is normal density with the sample mean and standard deviation from table 1.2. It shows the following:

- The plots indicate that normality assumption is questionable.
- Explanation: empirical density has higher peak and fatter tail than normal density.

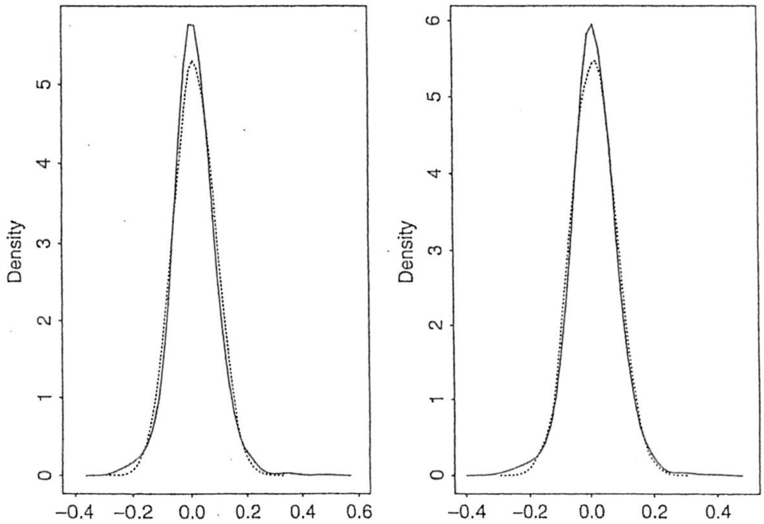


Figure 1.4 Comparison of empirical and normal densities for monthly simple and log returns of IBM stock. Sample period is from January 1926 to December 2008. Left plot is for simple returns and right plot for log returns. Normal density, shown by the dashed line, uses sample mean and standard deviation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

3.3.1 Empirical properties of returns

- Figure 1.2 shows time plots of monthly simple returns and log returns for IBM stock from Jan 1926 to Dec 2008.
- Figure 1.3 shows the same plots for monthly returns and log returns of value weighted market index.

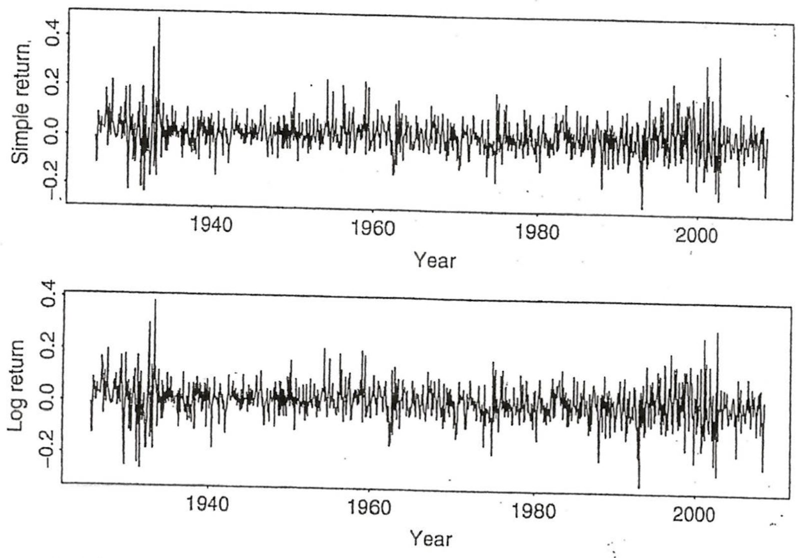


Figure 1.2 Time plots of monthly returns of IBM stock from January 1926 to December 2008. Upper panel is for simple returns, and lower panel is for log returns.

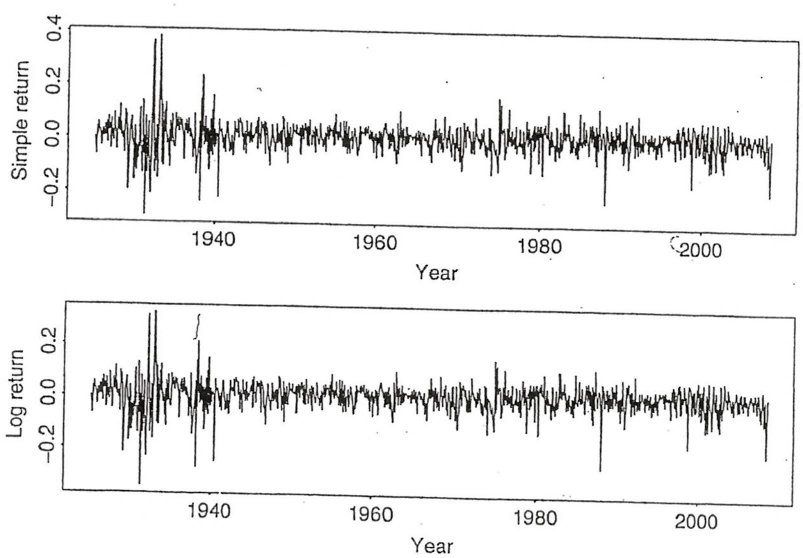


Figure 1.3 Time plots of monthly returns of value-weighted index from January 1926 to December 2008. Upper panel is for simple returns, and lower panel is for log returns.

Example. Figure 1.5 shows the time plot of two US monthly interest rates. They are 10 and 1 year Treasury constant maturity rates from 1954 to 2009.

Observation: Both interest rates move together but 1 year rate appears more volatile.

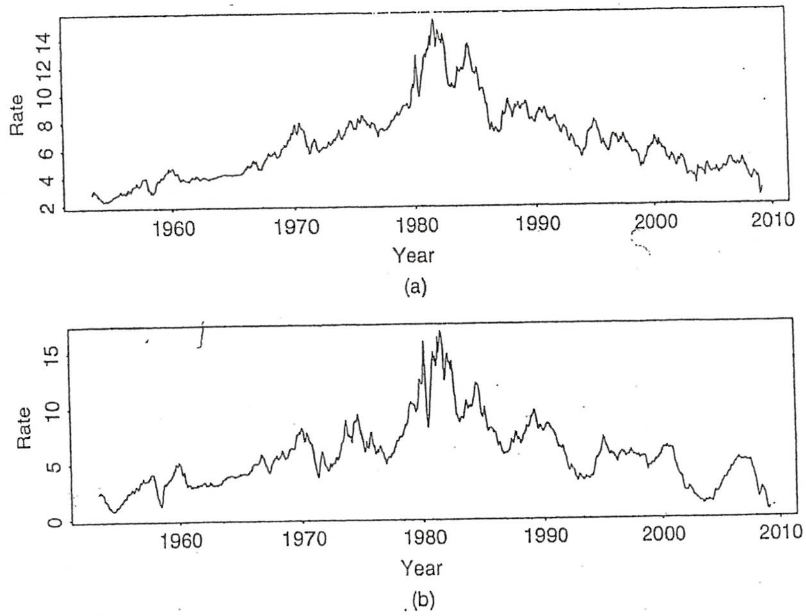


Figure 1.5 Time plots of monthly U.S. interest rates from April 1953 to February 2009: (a) 10-year Treasury constant maturity rate and (b) 1-year maturity rate.

Example. Table 1.3 provides some descriptive statistics for selected US financial time series. Observations:

- For interest rate series, the sample mean is proportional to the rate of maturity, but the standard deviation is inversely proportional.
- For bond rates, the SD is positively related to the time of maturity, while the sample mean remains stable for all maturities.
- most of series have positive excess kurtosis.

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.

3.4 Stationarity

In many applications in engineering and natural sciences processes are often described using deterministic models. Such models predict the output of the process exactly.

However deterministic model is not realistic when such process is affected by a number of random disturbances. For example, when pendulum movement is affected by random shocks, these unknown shocks have to be incorporated into model. Then it is impossible to come up with a comprehensive deterministic model and a stochastic model is more appropriate.

Such stochastic model will reflect reaction of the process to disturbances during observed period.

Prediction: it will be not possible to predict exactly the output of the process in the future. Instead, we may provide prediction interval and the probability that the future observation will lie in that interval.

Brockwell and Davis write: "If we wish to make predictions, we must assume that something does not vary with time". This brings us to the notion of stationarity.

Stationarity. The basis of time series analysis is stationary time series. For them we can develop models and forecasts.

However, in many applications in finance and economics, we have to deal with non-stationary time series. The explanation is the following: many processes behave as a stationary time series only taking control actions to keep the system in a stationary regime. Without deliberate action, left alone they start behaving as non-stationary processes x_t. Good news is, that for non-stationary processes x_t, often their changes $x_t - x_{t-1}$ may instead be stationary.

In that case we model changes, forecast their future values, and then create forecast of the original time series.

Therefore stationary time series play key role as the foundation for time series analysis. We now focuss on how to model stationary time series.

In most applications we deal with series of real data

$$X_0, X_1, X_2, X_3, ...$$

starting at time $t = 0$ and observed/recorded at discrete time periods $t = 0, 1, 2, ...$

In the strict sense, a time series is called stationary, if the joint probability distribution of any n observations

$$(X_{t+1}, X_{t+2}, ..., X_{t+n})$$

remains the same, if n observations are shifted by k units, that is

$$(X_{t+1+k}, X_{t+2+k}, ..., X_{t+n+k}).$$

For practical purposes we define a stationary time series as series which mean and variance are constant in time and correlation between observation form different points in time depends only on lag.

Definition 3 *We say that the time series (X_t) is weakly or covariance stationary, if*

(i) $E[X_t] = \mu_X$, *is independent of* t

(ii) $Var(X_t) = \sigma_X^2$, *is independent of* t

(iii) $Cov(X_t, X_{t+k}) \equiv \gamma_k$ *(for all t and k) depends on k only.*

Below we refer to stationarity as a weak or covariance stationarity.

Although we can conduct rigorous statistical test, often visual inspection of the time series plot will provide information whether it is stationary or not.

Stationary time series plot exhibits similar behavior: average value remains and spread around the mean remains stable.

If we observe this then we can proceed with fitting a stationary model.

Example. Figure 3.2 shows stationary and non-stationary time series. Here visual inspection is sufficient to confirm or deny stationarity.

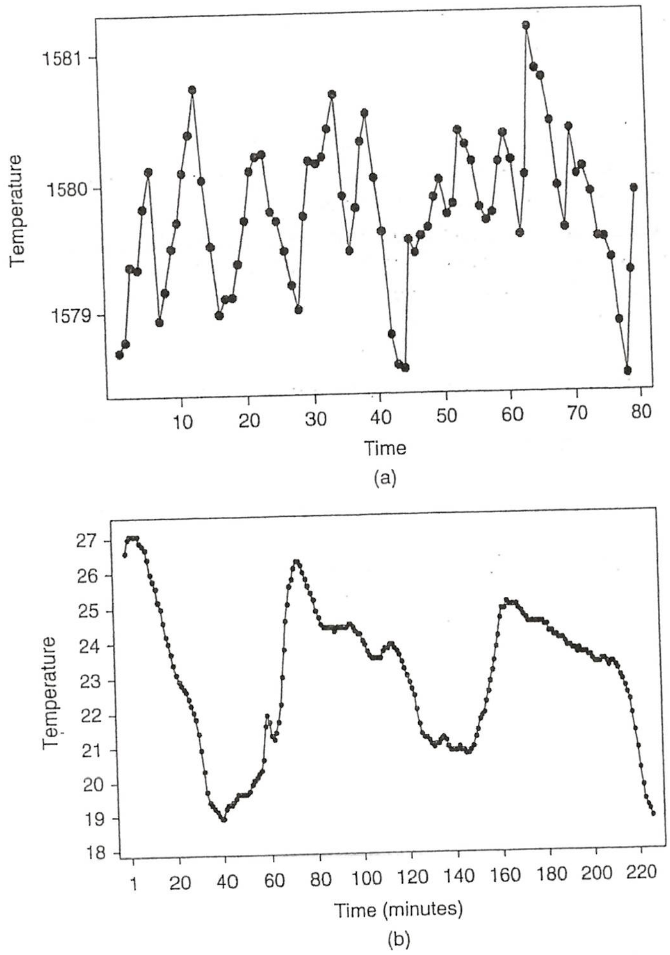


Figure 3.2 Temperature measurements for two different chemical processes. The process in (a) is tightly controlled to be around a certain target value (stationary behavior). The process in (b) is not controlled so that the temperature measurements do not vary around a target value (nonstationary behavior).

Caution: visual inspection can be also misleading. For short period non-stationary time series may give impression as being stationary. Therefore, beside visual inspection, use knowledge about the process.

Fore example, it makes sense that tightly controlled chemical process has constant mean and variance.

But should we expect the stock market "to remain in equilibrium about a constant mean level"? If yes, who would invest into it?

Hence, ask ... question: "Does it make sense...?" when selecting between stationarity and non-stationarity.

Autocorrelation function.

To further studies of time series we introduce the autocovariance and autocorrelation.

In general, we define the covariance between two two random variables X and Y as

$$Cov(X,Y) = E[(X - E[X])(Y - E[Y])]$$

Their correlation is their covariance scaled by standard deviations:

$$Corr(X,Y) = \frac{Cov(X,Y)}{\sqrt{Var(X)}\sqrt{Var(Y)}}.$$

We can visually investigate how correlated X and Y having a scatter plot of X versus Y.

How we define covariance in case of times series, when we have only observations $X_1, \cdots, X_n$?

We consider correlation between observations k lags apart, and call it autocovariance because we are dealing with the same data set.

Definition 4 *Autocovariance function of time series* (X_t)*: for* $k = 0, 1, 2, \cdots$

$$Cov(X_t, X_{t+k}) = E[(X_t - E[X_t])(X_{t+k} - E[X_{t+k}])].$$

Autocorrelation function

$$Corr(X_t, X_{t+k}) = \frac{Cov(X_t, X_{t+k})}{\sqrt{Var(X_t)}\sqrt{Var(X_{t+k})}}.$$

In case of a stationary time series we have

$$Cov(X_t, X_{t+k}) = Cov(X_0, X_k).$$

Therefore for a stationary time series we define the autocovariance function (ACF) as

$$\gamma_k = E[(X_t - E[X_t])(X_{t+k} - E[X_{t+k}]), \qquad k = 0, 1, 2, \cdots$$

which depends only on k which is called the lag (distance in time between observations).

- Note the the variance of time series is $Var(X_t) = \gamma_0$.
- Moreover, $\gamma_k = \gamma_{-k}$, $k = 1, 2, 3, \cdots$.

Following definition of correlation, we define autocorrelation of a stationary time series as

$$\rho_k = \frac{\gamma_k}{\sqrt{\gamma_0}\sqrt{\gamma_0}} = \frac{\gamma_k}{\gamma_0}.$$

Note: The ACF play extremely important role in identifying time series models.

Properties of ACF:

$$\begin{aligned}&\text{(i) } \rho_0 = 0\\ &\text{(ii) } |\rho_k| \leq 1 \quad \text{for all } k,\\ &\text{(iii) } \rho_{-k} = \rho_k \quad \text{for all } k.\end{aligned}$$

Note:

- Autocorrelation $corr(X_t, X_s)$ depend only on the distance $|t-s|$ in time (the lag).

  It is invariant with respect to scale of measurements.

- Autocorrelation captures the idea of stationarity, because the distribution of X_t remains unchanging through time.

In real life, we cannot know the true value of ACF, but we can estimate it from the data using

$$\hat{\gamma}_k = \frac{1}{N}\sum_{t=1}^{n-k}(X_{t+k} - \bar{X})(X_t - \bar{X})), \qquad \bar{X} = \frac{1}{N}\sum_{t=1}^{n} X_t$$

and

$$\hat{\rho}_k = \frac{\hat{\gamma}_k}{\hat{\gamma}_0}.$$

Note: As k increases we have fewer and fewer observations to estimate the autocovariance and autocorrelation.
A rule of thumb: the total number of observations N should be at least 50, and in the estimation of γ_k and ρ_k we should have $k \leq T/4$.