# 1

# Elementary financial calculus



## 1.1 Motivating examples

**Example 1.1.1** *Suppose a pension fund collecting contributions from workers intends to invest a certain fraction of the fund in a certain exchange-traded stock instead of buying treasury bonds. Whereas a bond yields a fixed interest known in advance, the return of a stock is volatile and uncertain. It may substantially exceed a bond's interest, but the pension fund is also exposed to the downside risk that the stock price goes down resulting in a loss. For the pension fund it is important to know what return can be expected from the investment and which risk is associated with the investment. It would also be useful to know the amount of the invested money that is under risk. In practice, investors invest their money in a portfolio of risky assets. Then the question arises: what can be said about the relationship? In modern finance, returns are modeled by random variables that have a distribution. Thus, we have to clarify how the return distribution and its mathematical properties are related to the economic notions* expected return, volatility, *and how one can define appropriate risk measures. Further, the question arises how one can estimate these quantities from historic time series.*

**Example 1.1.2** *In order to limit the loss due to the risky stock investment, the pension fund could ask a bank for a contract that pays the difference between a stop loss quote, $L$, and stock price, if that difference is positive when exercising the contract. Such financial instruments are called options. What is the fair price of such an option? And how can a bank initiate trades, which compensate for the risk exposure when selling the option?*

**Example 1.1.3** *Suppose a steel producer agrees with a car manufacturer to deliver steel for the production of 10 000 cars in one year. The steel production starts in one year and requires a large amount of oil. In order to calculate costs, the producer wants to fix the oil price at, say, $K$ dollars in advance. One approach is to enter a contract that pays the difference between the oil price and $K$ at the delivery date, if that difference is positive. Such contracts are named*

*Financial Statistics and Mathematical Finance: Methods, Models and Applications*, First Edition. Ansgar Steland.

*call options. Again, the question arises what is the fair price of such an agreement. Another possibility is to agree on a future/forward contract.*

**Example 1.1.4** *To be more specific and to simplify the exposition, let us assume that the steel producer needs 1 barrel whose current price at time $t = 0$ is $S_0 = 100$. To fix that price, he buys a call option with delivery price $K = 100$. The fixed interest rate is 1%. Further, suppose that the oil price, $S_1$, in one year at time $t = 1$ is distributed according to a two-point distribution,*

$$P(S_1 = 110) = 0.6, \qquad P(S_1 = 90) = 0.4.$$

*If $S_1 = 110$ one exercises the option right and the deal yields a profit of $G = 10$. Otherwise, the option has no value. Thus, the expected profit is given by*

$$E(G) = 10 \cdot 0.6 = 6.$$

*Because for the buyer of the option the deal has a non-negative profit and yields a positive profit with positive probability, he or she has to pay a premium to the bank selling the option. Should the bank offer the option for the expected profit 6? Surprisingly, the answer is no. Indeed, an oil dealer can offer the option for a lower price, namely $x = 5.45$ without making a loss. The dealer buys half of the oil when entering the contract at $t = 0$ for the current price of 50 and the rest when the contract is settled. His calculation is as follows. He finances the deal by the premium $x$ and a credit. At $t = 0$ his portfolio consists of a position in the money market, $x - 50$, and 0.5 units of oil. Let us anticipate that $x < 50$. Then at $t = 1$ the dealer has to pay back $1.01 \cdot |x - 50|$ to the bank. We shall now consider separately the cases of an increase or decreases of the oil price. If the oil price increases, the value of the oil increases to $0.5 \cdot 110 = 55$ and he receives 100 from the steel producer. He has to fix the premium $x$ such that the net income equals the price he has to pay for the remaining oil. This means, he solves the equation*

$$100 + 1.01 \cdot (x - 50) = 55$$

*yielding $x = 5.445545 \approx 5.45$. Now consider the case that the oil price decreases to 90. In this case the steel producer does not exercise the option but buys the oil at the spot market. The oil dealer has to pay back the credit, sells his oil at the lower price, which results in a loss of 5. The premium $x$ should ensure that his net balance is 0. This means, the equation*

$$0.5 \cdot 90 + 1.01(x - 50) = 0$$

*should hold. Solving for $x$ again yields $x = 5.445545$. Notice that both equations yield the same solution $x$ such that the premium is not random.*

## 1.2 Cashflows, interest rates, prices and returns

Let us now introduce some basic notions and formulas. To any financial investment initiated at $t = t_0$ with time horizon $T$ is attached a sequence of payments settled on a bank account that describe the investment from a mathematical point of view. Our standard notation is as follows: We denote the time points of the payments by $0 = t_0 < t_1 < \cdots < t_n = T$ and the associated payments by $X_1, \ldots, X_T$. Our sign convention will be as follows: Positive payments, $X_i > 0$, are deposits increasing the investor's bank account, whereas negative payments, $X_i < 0$, are charges.

From an economic point of view, there is a huge difference between a payment today or in the future. Thus, to compare payments, they either have to refer to the same time point $t^*$ or one has to take into account the effects of interest. As a result, to compare investments one has to cumulate the payments discounted or accumulated to a common time point $t^*$. If all payments are discounted to $t^* = t_0$ and then cumulated, the resulting quantity is called the **present value**. Alternatively, one can accumulate all payments to $t^* = T$.

In practice, one has to specify how to determine times and how to measure the economic distance between two time points $t_1$ and $t_2$. It is common practice to measure the time as a multiple of a year. At this point, suppose that the dates are given using the day-month-year convention, i.e. $t = (d, m, y)$. In what follows, we denote the economic time distance between two dates $t_1$ and $t_2$ by $\tau(t_1, t_2)$. Here are some market conventions for the calculation of $\tau(t_1, t_2)$.

(i) Actual/365: Each year has 365 days and the actual number of days is used.

(ii) Actual/360: Each year has 360 days and the actual number of days is used.

(iii) 30/360: Each month has 30 days, a year 360 days.

In the following we assume that all times have been transformed using such a convention.

If the fixed interest rate is $r$ per annum, interest is paid during the period without compound interest, the accumulated value of payments $X_1, \ldots, X_n$ at dates $t_1, \ldots, t_n$ is given by

$$V_T = \sum_{i=0}^{n} X_i(1 + \tau(t_i, 1)r).$$

The present value at $t = 0$ is calculated using the formula

$$V_0 = \sum_{i=0}^{n} X_i D(0, t_i), \qquad \text{with} \quad D(0, t_i) = \frac{1 + \tau(t_i, T)r}{1 + rT}.$$

Here $D(0, t_i)$ denotes the discount factor taking into account that the payment $X_i$ takes place at $t_i$.

Often, interest is paid at certain equidistant time points, e.g. quarterly or monthly. When decomposing the year into $m$ periods and applying the interest rate $r/m$ to each of them, an investment of one unit of currency grows during $k$ periods to

$$1 + \frac{r}{m}k.$$

When compound interest is taken into account, the value is

$$(1 + r/m)^k.$$

For $k = m \to \infty$ that discrete interest converges to continuous compounding

$$\lim_{m\to\infty} (1 + r/m)^m = \mathrm{e}^r.$$

Thus, the accumulation factor for an investment lasting for $t \in (0, \infty)$ years, i.e. corresponding to $tm$ periods, equals

$$\lim_{m\to\infty} (1 + r/m)^{mt} = \mathrm{e}^{rt}.$$

Let us now assume that the interest rate $r = r(t)$ is a function of $t$, such that for $r(t) > 0$, $t > 0$, the bank account, $S_0(t)$, increases continuously. There are two approaches to relate these quantities. Either start from a model or formula for $S_0(t)$ or start with $r(t)$. Let us first suppose that $S_0(t)$ is given. The annualized relative growth during the time interval $[t, t + h]$ is given by

$$\frac{1}{h}\frac{S_0(t+h) - S_0(t)}{S_0(t)}.$$

**Definition 1.2.1** *Suppose that the bank account $S_0(t)$ is a differentiable function. Then the quantity*

$$r(t) = \lim_{h\downarrow 0} \frac{1}{h}\frac{S_0(t+h) - S_0(t)}{S_0(t)},$$

*is well defined and is called* **instantaneous (spot) rate** *or simply* **short rate**.

We have the relationship

$$r(t) = \frac{S_0'(t)}{S_0(t)} \quad \Leftrightarrow \quad S_0'(t) = r(t)S_0(t).$$

As a differential:

$$\mathrm{d}B(t) = r(t)B(t)\mathrm{d}t.$$

It is known that this ordinary differential equation has the general solution $S_0(t) = C\exp(\int_0^t r(s)\,\mathrm{d}s)$, $C \in \mathbb{R}$. For our example the special solution

$$S_0(t) = \exp\left(\int_0^t r(s)\,\mathrm{d}s\right) \tag{1.1}$$

with starting value $S_0(0) = 1$ matters. In the special case $r(t) = r$ for all $t$, we obtain $S_0(t) = \mathrm{e}^{rt}$ as above.

Often, one starts with a model for the short rate. Then we define the bank account via Equation (1.1).

**Definition 1.2.2** (BANK ACCOUNT)
*A bank account with a unit deposit and continuous compounding according to the spot rate $r(t)$ is given by*

$$S_0(t) = \exp\left(\int_0^t r(s)\,\mathrm{d}s\right), \qquad t \geq 0.$$

When depositing $x$ units of currency into the bank account, the time $t$ value is $xS_0(t)$. Vice versa, for an accumulated value of 1 unit of currency at time $T$, one has to deposit $x = 1/S_0(T)$

at time $t = 0$. The value of $x = 1/S_0(T)$ at an arbitrary time point $t \in [0, T]$ is

$$xS_0(t) = \frac{S_0(t)}{S_0(T)}.$$

This means that the value at time $t = 0$ of a unit payment at the time horizon $T$ is given by $S_0(t)/S_0(T)$.

**Definition 1.2.3** *The* **discount factor** *between two time points $t \leq T$ is the amount at time $t$ that is equivalent to a unit payment at time $T$ and can be invested riskless at the bank. It is denoted by*

$$D(t, T) = \frac{S_0(t)}{S_0(T)} = \exp\left(-\int_t^T r(s)\,\mathrm{d}s\right).$$

### 1.2.1 Bonds and the term structure of interest rates

The basic insights of the above discussion can be directly used to price bonds and understand the term structure of interest rates.

A **zero coupon bond** pays a fixed amount of money, the **face value** or **principal** $X$ at a fixed future time point called **maturity**. Such a bond is also referred to as a **discount bond** or **zero coupon bond**. Here and in what follows, we assume that the bond is issued by a government such that we can ignore default risk. Measuring time in years and assuming that the interest rate $r$ applies in each year, we have learned that the present value of the payment $X$ equals

$$P_n(X) = \frac{X}{(1+r)^n}.$$

Notice that this simple formula determines a 1-to-1 correspondence between the bond price and the interest rate. The interest rate $r$ is the **discount rate** or **spot interest rate** for time to maturity $n$; *spot* rate, since that rate applies to a contract agreed on today.

Let us now consider a coupon bearing bond that pays coupons $C_1, \ldots, C_k$ at times $t_1, \ldots, t_k$ and the face value $X$ at the maturity date $T$. This series of payments is equivalent to $k + 1$ zero coupon bonds with face values $C_1, \ldots, C_k, X$ and maturity dates $t_1, \ldots, t_k, T$. Thus, its price is given by the **bond price equation**

$$P(t) = \sum_{i=1}^{k} C_i P(t, t_i) + X P(t, T),$$

or equivalently

$$P(t) = \sum_{i=1}^{k} C_i P(t, t + \tau_i) + X P(t, T),$$

if $\tau_j = t_j - t$ denotes the time to maturity of the $j$th bond. It follows that the price of the bond can be determined by the curve $\tau \mapsto P(t, t + \tau)$ that assigns to each maturity $\tau$ the time $t$ price for a zero coupon bond with unit principal $t$. It is called the **term structure of interest rates**.

There is a second approach to describe the term structure of interest rates. Let $P(t, t+m)$ denote the price at time $t$ of a zero coupon bond paying the principal $X = 1$ at the maturity date $t+m$. Given the yearly spot rate $r(t, t+m)$ applying to a payment in $m$ years, its price is given by

$$P(t, t+m) = \frac{1}{(1 + r(t, t+m))^m}.$$

If the coupon corresponding to the interest rate $r(t, t+m)$ is paid at $n$ equidistant time points with continuous compounding, the formula

$$P(t, t+m) = \frac{1}{(1 + r(t, t+m)/n)^{nm}}$$

applies, which converges to the formula for continuously compounding

$$P(t, t+m) = \mathrm{e}^{-r(t,t+m)m} \quad \Leftrightarrow \quad P(t, T) = \mathrm{e}^{-r(t,T)(T-t)},$$

using the substitution $T = t+m$. The continuously compounded interest rate $r(t, T)$ is also called **yield** and the function

$$t \mapsto r(t, T)$$

the **yield curve**.

Finally, one can also capture the term structure of interest rates by the **instantaneous forward rate** at time $t$ for the maturity date $T$ defined by

$$f(t, T) = \frac{-\frac{\partial}{\partial T} P(t, T)}{P(t, T)} = -\frac{\partial}{\partial T} \log P(t, T).$$

Here it is assumed that the bond price $P(t, T)$ is differentiable with respect to maturity. It then follows that

$$P(t, T) = \exp\left(-\int_0^\tau f(t, t+s)\,\mathrm{d}s\right), r(t, t+\tau) = -\frac{1}{\tau}\int_0^\tau f(t, t+s)\,\mathrm{d}s.$$

### 1.2.2 Asset returns

For fixed-income investments such as treasury bonds the value of the investment can be calculated in advance, since the interest rate is known. By contrast, for assets such as exchange-traded stocks the interest rates, i.e. returns, are calculated from the quotes that reflect the market prices.

Let $S_t$ be the price of a stock at time $t$. Since such prices are quoted at certain (equidistant) time points, it is common to agree that the time index attains values in the discrete set of natural numbers, $\mathbb{N}$. If an investor holds one share of the stock during the time interval from time $t-1$ to $t$, the asset price changes to

$$S_t = S_{t-1}(1 + R_t),$$

where

$$R_t = \frac{S_t - S_{t-1}}{S_{t-1}} = \frac{S_t}{S_{t-1}} - 1$$

is called the simple net return and

$$1 + R_t = \frac{S_t}{S_{t-1}}$$

are the gross returns. How are asset returns aggregated over time? Suppose an investor holds a share between $s$ and $t = s + k$, i.e. over $k$ periods, $s, t, k \in \mathbb{N}$ (or more generally $s, t, k \in [0, \infty)$). Define the $k$-period return

$$R_t(k) = \frac{S_t - S_s}{S_s} = \frac{S_t}{S_s} - 1.$$

One easily checks the following relationship between the simple returns $R_{s+1}, \ldots, R_t$ and the $k$-period return:

$$1 + R_t(k) = \frac{S_t}{S_s} = \prod_{i=s+1}^{t} \frac{S_i}{S_{i-1}} = \prod_{i=s+1}^{t} (1 + R_i).$$

When an asset is held for $k$ years, the annualized average return (effective return) is given by the geometric mean

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

A fixed-income investment with a annualized interest rate of $R_{t,k}$ yields the same accumulated value. Note that

$$R_{t,k} = \exp \left[ \frac{1}{k} \sum_{i=0}^{k-1} \log(1 + R_{t+i}) \right] - 1. \tag{1.2}$$

The natural logarithm of the gross returns,

$$r_t = \log(1 + R_t) = \log \frac{S_t}{S_{t-1}}$$

is called log return. Using Equation (1.2) we see that the $k$-period log return for the period from $s$ to $t = s + k$ can be calculated as

$$r_t(k) = \log(1 + R_t(k)) = \sum_{i=s+1}^{t} \log(1 + R_i) = \sum_{i=s+1}^{t} r_i.$$

Thus, in contrast to the returns $R_t$ the log returns possess the pleasant property of additivity w.r.t. time aggregation.

Using these definitions we obtain the following fundamental multiplicative decomposition of an asset price:

$$S_t = S_0 \prod_{i=1}^{t} (1 + R_i) = S_0 \prod_{i=1}^{t} \exp(r_i).$$

### 1.2.3 Some basic models for asset prices

When a security is listed on a stock exchange, there exists no quote before that time. Let us denote the sequence of price quotes, often the daily closing prices, by $S_0, S_1, \ldots$. Since $S_0 > 0$ denotes the first quote, it is often regarded as a constant. If one wants to avoid possible effects of the initial price, one puts formally $S_0 = 0$.

A first approach for a stochastic model is to assume that the price differences are given by

$$\Delta + u_n, \qquad n = 1, 2, \ldots$$

with a deterministic, i.e. nonrandom, constant $\Delta \in \mathbb{R}$ and i.i.d. random variables $u_n$, $n \in \mathbb{N}$, with common distribution function $F$ such that

$$E(u_n) = 0, \qquad \text{Var}\,(u_n) = \sigma^2 \in (0, \infty), \qquad \forall n \in \mathbb{N}.$$

In the present context, it is common to name the $u_n$ innovations. When referring to the sequence of innovations, we shall frequently write $\{u_n : n \in \mathbb{N}_0\}$ or, for brevity of notation, $\{u_n\}$ if the index set is obvious. The above model for the differences implies that the price process is given by

$$S_t = S_0 + \sum_{i=1}^{t}(\Delta + u_i) = S_0 + t\Delta + \sum_{i=1}^{t} u_i, \qquad t = 0, 1, \ldots$$

where we put $u_0 = 0$ and agree on the convention that $\sum_{i=1}^{0} a_i = 0$ for any sequence $\{a_n\}$. $S_t$ is called (arithmetic) **random walk** and **random walk with drift** if $\Delta \neq 0$. Obviously

$$E(S_t) = S_0 + \Delta\, t$$

and

$$\text{Var}\,(S_t) = t\sigma^2.$$

This particular model for an asset price dates back to the work of Bachelier (1900).

An alternative approach is based on the log returns. Let us denote

$$R_i := \log(S_i/S_{i-1}), \qquad i \geq 1.$$

Then

$$S_t = S_0 \prod_{i=1}^{t} S_i/S_{i-1} = S_0 \prod_{i=1}^{t} \exp(R_i).$$

The associated log price process is then given by

$$\log S_t = \log S_0 + \sum_{i=1}^{t} R_i, \qquad t = 0, 1, \ldots,$$

which is again a random walk.

A classic distributional assumption for the log returns $\{R_n\}$ is the normal one,

$$R_i \overset{i.i.d.}{\sim} N(\mu, \sigma^2)$$

with $\mu \in \mathbb{R}$ and $\sigma^2 > 0$. As a consequence, the log prices are normally distributed as well,

$$\log(S_t) = \log(S_0) + \sum_{i=1}^{t} R_i \sim N(\log(S_0) + t\mu, t\sigma^2).$$

Thus, $S_t$ follows a lognormal distribution. Let us summarize some basic facts about that distribution:

A random variable $X$ follows a **lognormal distribution** with parameters $\mu \in \mathbb{R}$ (**drift**) and $\sigma > 0$ (**volatility**) if $Y = \log(X) \sim N(\mu, \sigma^2)$. $X$ takes on values in the interval $(0, \infty)$ and

$$P(\log X \leq y) = \frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^{y} \mathrm{e}^{-(t-\mu)^2/2\sigma^2} \, \mathrm{d}t, \qquad y \in (0, \infty).$$

The change of variable $u = \mathrm{e}^t$ leads to

$$P(X \leq \mathrm{e}^y) = P(\log X \leq y) = \int_{-\infty}^{\mathrm{e}^y} \frac{1}{\sqrt{2\pi}\sigma u} \mathrm{e}^{-(\log u - \mu)^2/2\sigma^2} \, \mathrm{d}u.$$

By evaluating the right-hand side at $y = \log x$, we see that the density $f(x)$ of $X$ is given by

$$f(x) = \frac{1}{\sqrt{2\pi}x\sigma} \mathrm{e}^{-(\log x - \mu)^2/2\sigma^2} \mathbf{1}(x > 0), \qquad x \in \mathbb{R}. \tag{1.3}$$

Now it is easy to verify that mean and variance of $X$ are given by

$$E(X) = \mathrm{e}^{\mu + \sigma^2/2} \quad \text{and} \quad \mathrm{Var}\,(X) = \mathrm{e}^{2\mu + \sigma^2}(\mathrm{e}^{\sigma^2} - 1).$$

In order to model distributions that put more mass to extreme values than the standard normal distribution, one often uses the ***t*-distribution with *n* degrees of freedom** defined via the density function

$$f(x) = \frac{1}{n\pi} \frac{\Gamma((n+1)/2)}{\Gamma(n/2)} \left(1 + \frac{x^2}{n}\right)^{-\frac{n+1}{2}}, \qquad x \in \mathbb{R},$$

which is parametrized by $n \in \mathbb{N}$. By symmetry, its expectation is zero and the variance turns out to be $n/(n-2)$, if $n > 2$.

Several questions arise: Which of the above two models holds true or provides a better approximation to reality? Are returns and log returns, respectively, normally distributed? Are asset returns symmetrically distributed? How can we estimate important distributional parameters such as $\mu, \sigma^2$ or the skewness? Does the assumption of independent returns apply to real returns? Do price processes follow random walk models at all? What is the effect of changes of economic conditions on the distribution of returns? Can we test or detect such effects? How can we model the stochastic relationship between the return series of, say, $m$ securities?

There is some evidence that some financial variables have much heavier tails than a normal distribution.

A random variable $X$ has a **stable distribution** or is **stable**, if $X$ has a **domain of attraction**. The latter means that there exist i.i.d. random variables $\{\xi_n\}$ and sequences $\{\sigma_n\} \subset (0, \infty)$ and $\{\mu_n\} \subset \mathbb{R}$, such that

$$\frac{1}{\sigma_n} \sum_{i=1}^{n} \xi_i + \mu_n \stackrel{d}{\to} X,$$

as $n \to \infty$. The classic central limit theorem tells us that the $X \sim N(\mu, \sigma^2)$ is stable. By the Lévy–Khintchine formula, the characteristic function

$$\varphi(\theta) = E(e^{\imath\theta X}), \qquad \theta \in \mathbb{R},$$

where $\imath^2 = -1$, of a stable random variable $X$ has the representation

$$\varphi(\theta) = \begin{cases} \exp\left\{\imath\mu\theta - \sigma^\alpha|\theta|^\alpha \left(1 - \imath\beta(\operatorname{sgn}(\theta)) \tan \frac{\pi\alpha}{2}\right)\right\}, & \alpha \neq 1, \\ \exp\left\{\imath\mu\theta - \sigma|\theta| \left(1 + \imath\beta\frac{2}{\pi}(\operatorname{sgn}(\theta)) \log |\theta|\right)\right\}, & \alpha = 1, \end{cases}$$

where $0 < \alpha \leq 2$ is the **stability (characteristic) exponent**, $-1 < \beta < 1$ the **skewness parameter**, $\sigma > 0$ the **scale parameter** and $\mu \in \mathbb{R}$ the **location parameter**. For $\alpha = 2$ one obtains the normal distribution $N(\mu, \sigma^2)$, since then $\varphi(\theta) = \exp(\imath\mu\theta - \sigma^2\theta^2)$. The tails of a standard normal distribution decay exponentially fast,

$$P(|X| > x) \sim \sqrt{\frac{2}{\pi}} \frac{e^{-x^2/2}}{x}, \qquad x \to \infty \qquad (X \sim N(0, 1)).$$

By contrast, the **tails** of a stable random variable $X$ with characteristic exponent $0 < \alpha < 2$ decay as $x^{-\alpha}$, since

$$\lim_{x\to\infty} x^\alpha P(X > x) = C_\alpha \frac{1 + \beta}{2} \sigma^\alpha \tag{1.4}$$

and

$$\lim_{x\to\infty} x^\alpha P(X < -x) = C_\alpha \frac{1 - \beta}{2} \sigma^\alpha, \tag{1.5}$$

where $C_\alpha = \left(\int_0^\infty x^{-\alpha} \sin(x) \, dx\right)^{-1}$.

Stable distributions appear as a special case of infinitely divisible distributions. A random variable (or random vector) $X$ and its distribution are called **infinitely divisible**, if for every $n \in \mathbb{N}$ there exist independent and identically distributed random variables $X_{n1}, \ldots, X_{nn}$ such that

$$X \stackrel{d}{=} X_{n1} + \cdots + X_{nn}.$$

Those infinitely divisible distributions are exactly the distributions that can appear as limits of the distributions of sums $\sum_{k=1}^{n} X_{nk}$ of such arrays of row-wise i.i.d. random variables. Let $X$ be a $d$-dimensional random vector and again let $\varphi(\theta) = E(\exp(\imath\theta' X))$, $\theta \in \mathbb{R}^d$, be its characteristic function. Then, the **Lévy–Khintchine formula** asserts that

$$\varphi(\theta) = \exp\left\{\imath\theta' b - \frac{1}{2}\theta' C\theta + \int_{\mathbb{R}^d} \left(e^{\imath\theta' x} - 1 - \imath\theta' h(x)\right) d\nu(x)\right\}, \tag{1.6}$$

where

$$h(x) = x\mathbf{1}(|x| \le 1), \qquad x \in \mathbb{R}^d,$$

is a *truncation function*, $b \in \mathbb{R}^d$ and $C$ a symmetric and non-negative definite $(d \times d)$-matrix and $\nu$ a **Lévy measure**, that is a positive measure on the Borel sets of $\mathbb{R}^d$ such that $\nu(\{0\}) = 0$ and

$$\int_{\mathbb{R}^d} \left(|x|^2 \wedge 1\right) \mathrm{d}\nu(x) < \infty.$$

As a consequence, $\varphi(\theta)$ is characterized by the **triplet** $(b, C, \nu)$.

The characteristics of the normal distribution $N(\mu, \sigma^2)$ are $(b, C, \nu) = (\mu, \sigma^2, 0)$, of course. For a Poisson distribution with intensity $\lambda$, the characteristic function is

$$\varphi(\theta) = \exp(\lambda(e^{\imath\theta} - 1)),$$

which results, if we put $b = \lambda$, $C = 0$ and $\nu$ the one-point measure that assigns mass $\lambda$ to the single point 1.

## 1.3 Elementary statistical analysis of returns

We have seen that price processes can be build from returns $R_t$ that are modeled as random variables. For simplicity of our exposition, let us assume that $R_1, \ldots, R_T$ are independent and identically distributed. To simplify notation, let $R$ denote a generic return, i.e. $R \stackrel{d}{=} R_1$ which means that for any event $A$ we have $P(R \in A) = P(R_1 \in A)$.

But before focusing on returns, let us briefly review the most basic probabilistic quantities to which we will refer frequently in the following for an arbitrary random variable $X$. In general, the distribution of a random variable is uniquely determined by its **distribution function (d.f.)**

$$F_X(x) = P(X \le x), \qquad x \in \mathbb{R}.$$

If $f : \mathbb{R} \to \mathbb{R}$ is a density, i.e. non-negative function with $\int f(x)\,\mathrm{d}x = 1$, then the d.f. $F(x)$ can be calculated by

$$F(x) = \int_{-\infty}^{x} f(t)\,\mathrm{d}t, \qquad x \in \mathbb{R}.$$

A random variable $X$ that attains a density function $f$ is called a **continuous random variable**. Usually, it is assumed that returns are continuous random variables in that sense.

The first moment is defined by $\mu = E(X)$ and can be calculated for a continuous random variable via

$$\mu = E(X) = \int_{-\infty}^{\infty} x f(x)\,\mathrm{d}x.$$

$E(X)$ is also called the **expectation** or **mean** of $X$. If $X$ is a **discrete random variable**, that is $X$ takes values in some discrete set $\{x_1, x_2, \cdots\}$ of possible values with corresponding probabilities $p_1, p_2, \ldots$ such that

$$P(X = x_i) = p_i, \qquad i = 1, 2, \ldots,$$

then

$$E(X) = \sum_{i=1}^{\infty} x_i p_i.$$

More generally, the $k$**th moment** of $X$ is defined as $E(X^k)$ and $E|X|^k$ is referred to as the $k$**th absolute moment**. Assumptions on the existence of higher moments control the probability of **outliers**, that is extreme values. Indeed, by virtue of Markov's inequality, the probability that $X$ takes values larger than $c > 0$ in absolute value decays faster for increasing $c$, if higher moments exist, since

$$P(|X| > c) \leq \frac{E|X|^k}{c^k}.$$

Compare this inequality with the formulas (1.4) and (1.5) for the special class of stable distributions. As extreme values (outliers) of daily returns, usually negative ones, correspond to unexpected high-impact news such as a crash, the behavior of the **tail probabilities** $P(X < -c)$ and $P(X > c)$, $c > 0$, are of substantial interest, and moment assumptions automatically constrain them.

Suppose we are given a random sample $X_1, \ldots, X_T$ of sample size $T$. The **empirical distribution function** of the sample $X_1, \ldots, X_T$ is defined as

$$F_T(x) = \frac{1}{T} \sum_{t=1}^{T} \mathbf{1}(X_t \leq x), \qquad x \in \mathbb{R}.$$

Notice that $F_T(x)$ is the fraction of observations that are less or equal than $x$.

For a distribution function $F$ let

$$F^{-1}(y) = \inf\{x : F(x) \geq y\}$$

denote the left-continuous inverse called **quantile function**. Applying that definition to the empirical distribution function yields the **sample quantile function**

$$F_T^{-1}(p) = \inf\{x : F_T(x) \geq p\} = X_{(\lceil np \rceil)}, \qquad p \in (0, 1).$$

For a fixed $p$, $F_T^{-1}(p)$ is called the **sample** $p$**-quantile** or **empirical** $p$**-quantile**. Here $X_{(1)} \leq \cdots \leq X_{(T)}$ denotes the **order statistic** and $\lceil x \rceil$ is the smallest integer larger or equal to $x$. Notice that $X_{(\lceil np \rceil)} = X_{(\lfloor np \rfloor + 1)}$ where $\lfloor x \rfloor$ is the floor function, i.e. the largest integer that is less than or equal to $x$. Quantiles play an important role in characterizing a distribution. The sample 0.5-quantile is called the **median** and is also denoted by $x_{\text{med}}$. Together with the 0.25- and 0.75-quantiles,

$$Q_1 = F_T^{-1}(0.25), \qquad Q_3 = F_T^{-1}(0.75),$$

called **quartiles**, we get a picture where the lower (upper) fourth and the central 50% of the data are located. Augmenting these three statistics with the minimum and maximum defining the range of the data set, we obtain the so-called **five-point summary** $x_{\min}, Q_1, x_{\text{med}}, Q_3, x_{\max}$. Those five numbers already provide an informative view on the distribution of the data. The **boxplot (box and whiskers plot)** is a convenient graphical representation by a box symbolizing the central half of the data between $Q_1$ and $Q_3$ and straight lines connecting $x_{\min}$ and $Q_1$ as well as $Q_3$ and $x_{\max}$. It is also common to replace $(x_{\min}, x_{\max})$ by the quantiles $(F_T^{-1}(p), F_T^{-1}(1-p))$. Typical values for $p$ are $p = 0.01, 0.05$ and $0.1$.

Sample quantiles are asymptotically normal under fairly general conditions. Let $p \in (0, 1)$ and denote by $x_p = F^{-1}(p)$ the theoretical $p$-quantile. If $F$ attains a density that is positive in a neighborhood of $x_p$, then

$$\sqrt{T}(F_T^{-1}(p) - x_p) \stackrel{d}{\to} N(0, p(1-p)/f(x_p)^2), \tag{1.7}$$

as $T \to \infty$. The problem arises that the asymptotic variance depends on the unknown density, which has to be estimated by some appropriate estimator $\widehat{f}_T$. We shall discuss this issue in Section 1.3.4 and anticipate that such an estimator can be defined having nice mathematical properties under fairly weak regularity conditions that do not impose a constraint on the shape of the density $f$, which is of particular importance when analyzing financial data such as returns. Based on the large sample result (1.7), which still holds true when plugging in a consistent estimator, it is straightforward to construct the confidence interval for $x_p$,

$$\left[ F_T^{-1}(p) - z_{1-\alpha/2} \frac{\sqrt{p(1-p)}}{\widehat{f}_T(x_p)}, F_T^{-1}(p) + z_{1-\alpha/2} \frac{\sqrt{p(1-p)}}{\widehat{f}_T(x_p)} \right],$$

which attains the coverage probability $1 - \alpha$, if $T \to \infty$, where $z_{1-\alpha/2}$ denotes the $(1-\alpha/2)$-quantile of the standard normal distribution. We discuss the derivation of such confidence intervals in greater detail in the next section.

### 1.3.1 Measuring location

Measures of locations are usually defined in terms of moments or quantiles. The expectation is the most commonly used measure of location of a random variable.

Returning to our problem to analyze financial returns, the problem arises that the distribution of the returns is unknown to us. But then the mean return $\mu = E(R)$ is unknown as well. The best we can do is to use statistical estimators, i.e. functions of the data $R_1, \dots, R_T$, which output a value that is regarded as a good estimate for $\mu$. A standard approach to obtain such estimators for quantities that are defined in terms of expectations is to replace the averaging with respect to the distribution by averaging with respect to the so-called **empirical probability measure** that attaches equal mass $1/T$ to the values $R_1, \dots, R_T$. The expectation with respect to that discrete distribution is simply the **arithmetic mean**

$$\overline{R} = \overline{R}_T = \frac{1}{T} \sum_{t=1}^{T} R_t.$$

It is easy to check that $E(\overline{R}_T) = E(R_1) = \mu$, and this calculation holds true whatever the value $\mu$ attains. In statistics, an estimator satisfying that property is called an **unbiased estimator**. It tells us that, averaged over all possible scenarios $\omega$ corresponding to all possible values

$r = R(\omega)$ for the return and weighted with the corresponding probabilities, the estimator estimates the right value, namely $\mu$.

Suppose we have observed $T$ daily log returns $R_1, \dots, R_T$ and aim at testing the hypothesis that their common mean $\mu = E(R_1)$ equals some specified value $\mu_0$. The corresponding two-sided statistical testing problem is then given by

$$H_0 : \mu = \mu_0 \qquad \text{versus} \qquad H_1 : \mu \neq \mu_0.$$

Assuming that the returns are i.i.d. and follow a normal law suggest using the $t$-test that is based on the test statistic

$$Z = \sqrt{T}\frac{\overline{R}_T - \mu_0}{S_T} \tag{1.8}$$

with

$$S_T = \sqrt{\frac{1}{T-1}\sum_{t=1}^{T}(R_t - \overline{R}_T)^2};$$

the statistic $S_T$ will be discussed in greater detail in the next subsection. Under the null hypothesis $H_0$, the statistic $Z$ follows a $t$-distribution with $df = T - 1$ degrees of freedom. Consequently, we may reject $H_0$ at a significance level of $\alpha \in (0, 1)$, if

$$|T| > t(df)_{1-\alpha/2},$$

where $t(df)_{1-\alpha/2}$ denotes the $(1 - \alpha/2)-$quantile of the $t(df)$-distribution.

If the log returns are non-normal, one can often rely on the central limit theorem which asserts that the statistic $Z$ is asymptotically normal. Hence, the null hypothesis is then rejected, if $|Z| > z_{1-\alpha/2}$.

**Example 1.3.1** *For the FTSE log returns illustrated in Figure 1.1, one gets* $z = 2.340558$*, which exceeds the critical value* $1.959964$ *corresponding to the* $5\%$ *significance level, indicating that the mean log return differs from zero and is actually positive. However, this assertion is not valid on the* $1\%$ *significance level.*

Often, one is also interested to provide interval estimates for the mean. Again assuming i.i.d. normal returns, a **confidence interval** for the mean with coverage probability $1 - \alpha$, is an interval $[L, U]$ where $L = L(R_1, \dots, R_T)$ and $U = U(R_1, \dots, R_T)$ are functions of the sample such that

$$P(L \leq \mu \leq U) = 1 - \alpha$$

for any $\mu \in \mathbb{R}$. Such a confidence interval is given by

$$L = \overline{R}_T - t(df)_{1-\alpha/2}\frac{S_T}{\sqrt{T}}, \qquad U = \overline{R}_T + t(df)_{1-\alpha/2}\frac{S_T}{\sqrt{T}},$$

where, as above, $df = T - 1$. This can be easily established by noting that the event $L \leq \mu \leq U$ is equivalent to

$$-t(df)_{1-\alpha/2} \leq \sqrt{T}(\overline{R}_T - \mu)/S_T \leq t(df)_{1-\alpha/2}.$$

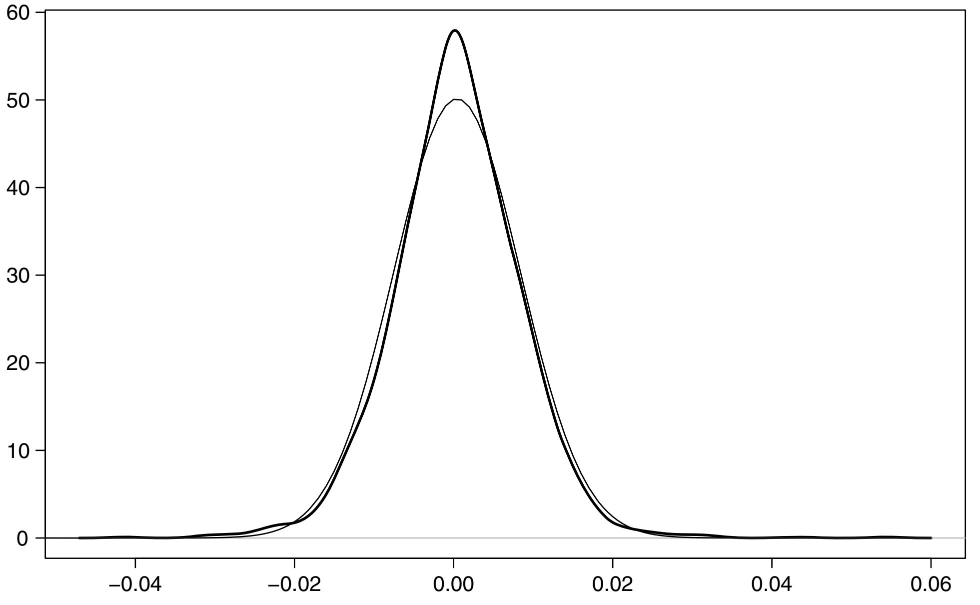


**Figure 1.1** Kernel density estimate of the FTSE daily log returns with cross-validated bandwidth choice.

But the latter event occurs with probability $1 - \alpha$, since

$$\sqrt{T}(\overline{R}_T - \mu)/S_T \sim t(df).$$

However, usually daily returns are not normal but affected by **stylized facts** such as asymmetry, peakedness (more mass around zero) and heavier tails than under a normal law. This can be easily seen from Figure 1.1. The famous central limit theorem asserts that the statistic $Z$ defined in Equation (1.8) is asymptotically standard normal, as long as the returns are i.i.d. with existing fourth moment.[1] Consequently, a valid asymptotic test is given by the decision rule

$$\text{reject } H_0 \text{ if } |Z| > z_{1-\alpha/2},$$

where $z_{1-\alpha/2} = \Phi^{-1}(1 - \alpha/2)$ denotes the $(1 - \alpha/2)$-quantile of the $N(0, 1)$-distribution. In the same vein, an asymptotic confidence interval for $\mu$ is obtained by replacing the quantiles of the $t(df)$-distribution in the formulas for $L$ and $U$ by the respective quantiles of the standard normal law.

Similarly, one may construct an asymptotic confidence interval for $\mu$ based on the central limit theorem. In this case, the probability of the event

$$-z_{1-\alpha/2} \leq \sqrt{T}(\overline{R}_T - \mu)/S_T \leq z_{1-\alpha/2},$$

[1] This result even remains true under the substantially weaker assumption that the log returns are a stationary martingale difference sequence.

which is equivalent to the event

$$L' = \overline{R}_T - z_{1-\alpha/2}\frac{S_T}{\sqrt{T}} \leq \mu \leq \overline{R}_T + z_{1-\alpha/2}\frac{S_T}{\sqrt{T}} = U',$$

converges to $1 - \alpha$, as $T \to \infty$. Thus, a confidence interval with **asymptotic coverage probability** $1 - \alpha$ is given by the random interval $[L', U']$.

**Example 1.3.2** *For the FTSE log returns one calculates the asymptotic* 95% *confidence interval* $[l, u] = [0.0000702, 0.000793]$ *for the mean log return.*

It is a general insight, supported by many empirical studies, that the statistical analysis of financial returns should not be based on procedures assuming the classic assumptions of normality and independent observations, since those assumptions are usually violated. Therefore, large sample theory forms the mathematical core for inferential procedures in finance.

### 1.3.2 Measuring dispersion and risk

The mean $\mu = E(R_t)$ tells us where the distribution is located; it is a measure for the center of the distribution. Then we can determine for each return $R_t$ its distance $|R_t - \mu|$ from the mean. The mean squared distance,

$$\sigma^2 = \text{Var}(R) = E(R - \mu)^2 = E(R^2) - \mu^2$$

is called the **variance** of $R$. Its square root,

$$\sigma = \sigma_R = \sqrt{\text{Var}(R)}$$

is called the **standard deviation**. Variance and standard deviation can be defined for any random variable $X$ with existing second moment. If $X$ and $Y$ are independent random variables with $EX^2 < \infty$ and $EY^2 < \infty$, then

$$\text{Var}(X + Y) = \text{Var}(X) + \text{Var}(Y)$$

yielding $\sigma_{X+Y} = \sqrt{\text{Var}(X+Y)} = \sqrt{\sigma_X^2 + \sigma_Y^2}$, whereas in the general case

$$\text{Var}(X + Y) = \text{Var}(X) + \text{Var}(Y) + 2\,\text{Cov}(X, Y).$$

Here

$$\text{Cov}(X, Y) = E(X - EX)(Y - EY)$$

is called the **covariance of** $X$ **and** $Y$.

When considering daily (log) returns, $\sigma$ is also frequently called **(actual) volatility**. When volatility of returns is addressed, it is important to be aware of the corresponding unit of time, e.g. yearly, monthly or daily. The **annualized volatility** $\sigma_{\text{an}}$ is the standard deviation of the yearly return, whereas **generalized volatility** addresses the volatility corresponding to the time horizon $\tau$ (in years) given by

$$\sigma_{\text{an}}\sqrt{\tau}.$$

Notice that the formula coincides with the standard deviation of the return $R(\tau)$ corresponding to the time period $\tau$, if $\tau$ is an integer and the yearly log returns are identically distributed and uncorrelated, since then the additivity of log returns gives $R(\tau) = \sum_{t=1}^{\tau} R'_t$ where $R'_1, \ldots, R'_\tau$ denote the $\tau$ yearly log returns. But then

$$\sigma_{R(\tau)} = \sqrt{\sum_{t=1}^{\tau} \mathrm{Var}\,(R'_t)} = \sqrt{\tau}\sigma',$$

where $\sigma'$ is the volatility of the yearly returns $R'_t$. However, usually the annualized volatility is determined from the actual volatility of the daily log returns. Since there are 252 trading days in a year, annualized volatility $\sigma_{\mathrm{an}}$ and actual volatility $\sigma$ are related by

$$\sigma_{\mathrm{an}} = \sigma\sqrt{252}.$$

The monthly volatility is then given by $\sigma_{\mathrm{m}} = \sigma\sqrt{252/12}$.

Estimation of the variance and standard deviation is usually based on the plug-in principle already explained in the previous subsection. Given a sample $R_1, \ldots, R_T$ of returns, it naturally leads to the the **empirical variance** or **sample variance**

$$V_T^2 = \frac{1}{T}\sum_{t=1}^{T}(X_t - \overline{R}_T)^2.$$

A tedious calculation shows that $E(V_T^2) = \frac{T-1}{T}\sigma^2$, i.e. $V_T^2$ is not an unbiased estimator of the variance. Thus, in practice the estimator

$$S_T^2 = \frac{1}{T-1}\sum_{t=1}^{T}(R_t - \overline{R}_T)^2$$

is used. The corresponding estimator for the standard deviation is the square root, $S_T = \sqrt{S_T^2}$, of that expression. Estimates of the various volatilities discussed above can be obtained by substituting $\sigma$ by $S_T$. For example, if the $R_t$s are daily log returns, annualized volatility is estimated by $S_T\sqrt{252}$.

#### 1.3.2.1 Value-at-risk

Another risk measure that has been become the de-facto standard in the financial industry is value-at-risk. Recall that the profit or loss (P&L) of any investment during a time period $[0, h]$ is uncertain and therefore represents a risk exposure, namely to suffer a loss. Roughly speaking, value-at-risk is a risk measure that represents the smallest loss we are exposed to with probability $\alpha$. Here the risk probability $\alpha$ is chosen by us; common values are 1% and 5%. Let $V_t$ denote the **marked-to-market** value of a long position at time $t$, i.e. the value is based on the current market value. Then the profit is $\Delta V = V_{t+h} - V_t$, where negative values are losses. Now let us consider the loss $L = -\Delta V$ and let $v$ be the fixed value satisfying

$$P(L > v) = \alpha.$$

This means, with a probability of $\alpha$ we suffer a loss exceeding $v$. That number $v$ (a loss) is called **value-at-risk (VaR)** at the probability level $\alpha$ and denoted by VaR or $\mathrm{VaR}_\alpha$. Roughly speaking, it is the *smallest* loss among the largest losses occurring with probability $\alpha$. By definition,

$$\mathrm{VaR}_\alpha = F_L^{-1}(1-\alpha),$$

where $F_L^{-1}$ denotes the quantile function associated to the loss distribution. That means, value-at-risk is the $(1-\alpha)$-quantile of the loss distribution. Notice that value-at-risk can be also defined by the $\alpha$-quantile of the P&L distribution,

$$\mathrm{VaR}_\alpha = -F_{\Delta V}^{-1}(\alpha).$$

Often, VaR is calculated on a daily basis. If the daily 1% value-at-risk of a position is 100 000, the probability that the value of the position will fall below −100 000 is 1%; with probability 1% we suffer a loss being larger than 100 000.

Since VaR is defined as a quantile, we may estimate it by the corresponding sample quantiles. If $L_1, \dots, L_T$ are i.i.d. losses corresponding to the time horizon $h$,

$$\widehat{\mathrm{VaR}}_\alpha = L_{\lceil n(1-\alpha)\rceil}.$$

Statistical tests and the calculation of confidence intervals can therefore be based on the large sample theory of quantiles discussed above. In the same vein, the asymptotic confidence intervals carry over to confidence intervals for value-at-risk.

#### 1.3.2.2 Expected shortfall, lower partial moments and coherent risk measures

VaR gives the smallest loss among the largest losses occurring with probability $\alpha$. It is natural to average those losses, that is to consider the conditional expectation of the profit or loss $L$ over a given period of time

$$S_\alpha(L) = E(L \mid L \le \mathrm{VaR}_\alpha)$$

is called the **expected shortfall** or **conditional value-at-risk**. One can show that

$$S_\alpha(X) = -\frac{1}{\alpha}\int_0^\alpha F_{\Delta V}^{-1}(x)\,\mathrm{d}x.$$

For this reason, $S_\alpha(X)$ is also called the **average value-at-risk**.

Clearly, we do not worry about realizations $l$ of $L$ with $l > E(L)$, but are concerned about the **downside risk**, that is losses below the expectation $E(L)$ of the position. If $L$ is symmetrically distributed, then $P(L < E(L)) = P(L > E(L))$ and the variance or standard deviation provide meaningful measures for the downside risk. But especially for asymmetric distributions it makes sense to consider the **semivariance** defined as

$$E\big(\min(0, L - EL)^2\big).$$

Often there exists a benchmark profit $b$ to which a portfolio is compared. If the portfolio does not outperform the given benchmark $b$, that is if $L \le b$, then $b - L$ is the loss we suffer when we have a long position in the portfolio. The $m$th moment of the corresponding random variable $(b-L)\mathbf{1}(L \le b)$,

$$LP_m(L) = E\left((b-L)^m \mathbf{1}(L \le b)\right)$$

is called the **lower partial moment of the order** $m$, provided it exists. Notice that

$$LP_0(L) = P(L \leq b)$$

is the probability that the portfolio does not outperform the benchmark, and $LP_1(L)$ is the expected underperformance.

All the quantities discussed above assign a real number to a random variable interpreted as the loss of a portfolio or position over some fixed period of time, and that number is interpreted as a quantitative measure of the risk. The question arises which properties (axioms) such a risk measure should satisfy. Generally, a **risk measure** or **risk functional** $\rho$ is a function defined on a sufficiently rich set $\mathcal{A}$ of random variables (random payment profiles) taking values in the real numbers. Given such a risk measure $\rho$, we may distinguish risky payments with non-negative risks and **acceptable payments** with negative risks.

A risk measure $\rho : \mathcal{A} \to \mathbb{R}$ is called **coherent**, if it satisfies the following four axioms:

(i) $X \leq Y$ implies that $\rho(X) \leq \rho(Y)$ for all $X, Y \in \mathcal{A}$ (monotonicity).

(ii) $\rho(X + Y) \leq \rho(X) + \rho(Y)$ for all $X, Y \in \mathcal{A}$ (subadditivity).

(iii) $\rho(aX) = a\rho(X)$ for $a > 0$ (positive homogeneity).

(iv) $\rho(X + a) = \rho(X) - a$ for any $X \in \mathcal{A}$ and $a \in \mathbb{R}$ (translational invariance).

Sometimes, a further axiom is considered

(v) If $X \stackrel{d}{=} X'$, then $\rho(X) = \rho(X')$ (distributional invariance).

Axiom (i) requires that the risk of a position increases, if the random payment profile increases for all states $\omega \in \Omega$. The second axiom addresses an important aspect of risk management: Risks associated to two positions may cancel when aggregating them. The standard deviation $\sigma(X) = \sqrt{\text{Var}\,(X)}$ satisfies axiom (ii) and (iii). To see (ii), use the inequality

$$\text{Cov}(X, Y) \leq \sigma(X)\sigma(Y)$$

to obtain

$$\begin{aligned}
\sigma(X + Y) &= 2\sqrt{\text{Var}\,\left(\frac{X}{2} + \frac{Y}{2}\right)} \\
&\leq 2\sqrt{\left(\frac{1}{2}\right)^2 \sigma_X^2 + \left(\frac{1}{2}\right)^2 \sigma_Y^2 + \sigma_X\sigma_Y} \\
&= 2\sqrt{\left(\frac{\sigma_X}{2} + \frac{\sigma_Y}{2}\right)^2} \\
&= \sigma_X + \sigma_Y
\end{aligned}$$

for *any* pair $(X, Y)$ of random variables with existing second moments and arbitrary correlation. This also implies that in the Gaussian world value-at-risk also satisfied axiom (ii). This can be seen as follows. Notice that value-at-risk for a random P&L $X \sim N(\mu, \sigma^2)$ is given by

$$\text{VaR}_\alpha(X) = \mu + \Phi^{-1}(\alpha)\sigma.$$

Further, for a random vector $(X, Y)$ distributed according to a bivariate normal distribution with marginals $N(\mu_X, \sigma_X^2)$, $N(\mu_Y, \sigma_Y^2)$ and covariance $\gamma$ the sum $X + Y$ is again Gaussian,

$$X + Y \sim N(\mu_X + \mu_Y, \sigma_{X+Y}^2), \qquad \sigma_{X+Y}^2 = \sigma_X^2 + \sigma_Y^2 + 2\gamma,$$

such that the $\alpha$-quantile of $X + Y$ is

$$\mathrm{VaR}_\alpha(X + Y) = \mu_X + \mu_Y + \Phi^{-1}(\alpha)\sigma_{X+Y}.$$

Hence, $\sigma_{X+Y} \leq \sigma_X + \sigma_Y$ immediately implies

$$\mathrm{VaR}_\alpha(X + Y) \leq \mathrm{VaR}_\alpha(X) + \mathrm{VaR}_\alpha(Y).$$

However, for general distributions axiom (ii) can be violated, such that in general value-at-risk is not a coherent risk measure, which is probably the main criticism against value-at-risk.

Axiom (iii) is a scaling property, which allows us to compare risks expressed in different currencies, for example. Finally, the fourth axiom means that when adding a fixed payment to the position, in order to compensate losses and reduce the risk in this way, the risk measure is also reduced by exactly that amount, and, by contrast, withdrawing cash increases the risk. Then $\rho(X)$ can be interpreted as the amount of capital needed to eliminate the risk and transform a position into a acceptable payment. Obviously, that axiom is not satisfied by the standard deviation.

One can show that the expected shortfall satisfies all axioms and is therefore a coherent risk measure. More generally, any risk measure allowing a representation

$$\rho(X) = \sup_{P \in \mathcal{P}} E_P(-X),$$

where $\mathcal{P}$ is a set of probability measures and $E_P$ indicates that the expectation is calculated under $P$, can be shown to be a coherent risk measure. For $S_\alpha(X)$ the set $\mathcal{P}$ is given by all densities that are bounded by $1/\alpha$.

### 1.3.3 Measuring skewness and kurtosis

The most common approach to measure skewness, i.e. departures from symmetry, is to consider the third standardized moment,

$$\mu_3^* = E\left(\frac{R_1 - \mu}{\sigma}\right)^3,$$

where $\mu = E(R_1)$ and $\sigma^2 = \mathrm{Var}\,(R_1)$. Notice that $\mu_3^* = 0$, if $R_1 - \mu \overset{d}{=} \mu - R_1$.[2]

Given a sample $R_1, \ldots, R_T$, one uses the estimator

$$\widehat{\mu}_3^* = \frac{1}{T}\sum_{t=1}^{T}\left(\frac{R_t - \overline{R}_T}{S_T}\right)^3.$$

[2] If $X \overset{d}{=} -X$ and $f$ is a function with $f(-x) = -f(x)$ and $Ef(X) \in \mathbb{R}$, then $Ef(X) = Ef(-X) = -Ef(X)$, which implies $Ef(X) = 0$.

The statistic $\widehat{\mu}_3^*$ is very sensitive with respect to outliers. An alternative measure based on quantiles is to compare the distance between the 0.75-quantile and the median and the distance between the median and the 0.25-quantile, expressed as a fraction of the maximum value, i.e.

$$\gamma = \frac{[F^{-1}(0.75) - F^{-1}(0.5)] - [F^{-1}(0.5) - F^{-1}(0.25)]}{F^{-1}(0.75) - F^{-1}(0.25)}.$$

The corresponding estimator based on $R_1, \dots R_T$ is

$$\widehat{\gamma}_T = \frac{[Q_3 - x_{\text{med}}] - [x_{\text{med}} - Q_1]}{Q_3 - Q_1}.$$

Since sample quantiles, particularly $Q_1$, $Q_3$ and $x_{\text{med}}$ are more robust than an arithmetic mean, $\widehat{\gamma}_T$ provides a reliable measure of skewness even for data sets from distributions with heavy tails.

A common approach to measure deviations from the shape of the Gaussian density is based on the fourth standardized moment,

$$\mu_4^* = E\left(\frac{R_1 - \mu}{\sigma}\right)^4,$$

also called **kurtosis**. Since for a normal distribution one obtains $\mu_4^* = 3$, it is common to consider the **excess kurtosis**,

$$\kappa = \mu_4^* - 3.$$

Distributions such as the normal one with an excess kurtosis equal to 0 are called **mesokurtic**. The standard interpretations when $\kappa \neq 0$ are as follows. A distribution with $\kappa > 0$ is called **leptokurtic**. It has a more pronounced peak compared to the normal law and lighter tails. A distribution with $\kappa < 0$ is called **platykurtic**. Such distributions have a flatter peak and heavier tails than the Gaussian density. Kurtosis and excess kurtosis are estimated by their sample analogs

$$\widehat{\mu}_4^* = \frac{1}{T}\sum_{t=1}^{T}\left(\frac{R_t - \overline{R}_T}{S_t}\right)^4$$

and

$$\widehat{\kappa}_T = \widehat{\mu}_4^* - 3,$$

respectively.

### 1.3.4 Estimation of the distribution

We have already discussed that financial returns for shorter time horizons tend to violate properties of the normal distribution. Taking for granted that the return distribution attains a density function[3] $f$ in the sense that the distribution function

[3] For some financial instruments that assumption is violated, since there are trading periods where the price remains constant such that the return is 0.

$F(x) = P(R_1 \leq x)$ can be represented as

$$F(x) = \int_{-\infty}^{x} f(t)\,\mathrm{d}t, \qquad x \in \mathbb{R},$$

the question arises how we can estimate the density $f$. Noticing that $f(x) = F'(x)$, we may approximate $f(x)$ by the difference ratio

$$f(x) \approx \frac{F(x+h) - F(x-h)}{2h},$$

for small $h > 0$. A natural approach is to estimate the right-hand side by plugging in the empirical distribution function $F_T(x) = T^{-1} \sum_{t=1}^{T} \mathbf{1}(R_t \leq x)$ of $T$ historical returns $R_1, \ldots, R_T$ and to regard the resulting expression as an estimate for $f(x)$. Noting that

$$\mathbf{1}(R_t \leq x + h) - \mathbf{1}(R_t \leq x - h) = \mathbf{1}(x - h < R_t \leq x + h),$$

this idea leads to the estimator

$$x \mapsto \frac{1}{Th} \sum_{t=1}^{T} \frac{1}{2} \mathbf{1}\left(-1 < \frac{R_t - x}{h} \leq 1\right), \qquad x \in \mathbb{R}.$$

Each of the $T$ summands corresponds to the density $K_0(z) = \frac{1}{2}\mathbf{1}(-1 < |z| \leq 1)$, $z \in \mathbb{R}$, of the uniform distribution on $(-1, 1]$ evaluated at the points $(x - R_t)/h$, $t = 1, \ldots, T$. Obviously, as a function of $x$ the above density estimator is discontinuous, which results in many spurious jumps. If we replace the discontinuous density $K_0$ by other density functions, we arrive at the **Rosenblatt–Parzen** kernel density estimator

$$\widehat{f}_{Th}(x) = \frac{1}{Th} \sum_{t=1}^{T} K([R_t - x]/h), \qquad x \in \mathbb{R}.$$

The parameter $h$ is called the **bandwidth**. It has a strong influence on the resulting estimator. If $h$ is chosen too small, there will be many spurious artifacts such as local extrema in the graph, whereas too large values for the bandwidth lead to oversmoothing. $K$, called the **smoothing kernel**, is usually chosen as an arbitrary unimodal density function with finite second moment that is symmetric around zero. Table 1.1 lists some smoothing kernels frequently used in practice.

**Table 1.1** Some commonly used smoothing kernels for nonparametric density estimation.

| Kernel | Definition |
|---|---|
| Triangular | $(1 - \lvert x\rvert)\mathbf{1}(\lvert x\rvert \leq 1)$ |
| Cosine | $(\pi/4)\cos(x\pi/2)$ |
| Gaussian | $(2\pi)^{-1}\exp(-x^2/2)$ |
| Epanechnikov | $(3/4)(1 - x^2)\mathbf{1}(\lvert x\rvert \leq 1)$ |
| Biweight | $(15/16)(1 - x^2)^2\mathbf{1}(\lvert x\rvert \leq 1)$ |
| Silverman | $(1/2)\exp(-\lvert x\rvert/\sqrt{2})\sin(\lvert x\rvert/\sqrt{3} + \pi/4)$ |

Notice that the estimator $\widehat{f}_{Th}(x)$ allows the following nice interpretation: If $K$ is a density that is symmetric around 0 with unit variance, then

$$x \mapsto \frac{1}{h} K\left(\frac{x-m}{h}\right), \qquad x \in \mathbb{R},$$

is a density with mean $m$ and standard deviation $h$ for any fixed $m \in \mathbb{R}$ and $h > 0$. Consequently, $\widehat{f}_{Th}(x)$ averages those $T$ densities $x \mapsto h^{-1}K([x - R_t]/h)$, $t = 1, \ldots, T$, associated to the observed values.

It is worth discussing some further basic properties of the kernel density estimator, in order to understand why it estimates any underlying density under fairly general conditions. Another issue we have to discuss is the question how to select the smoothing kernel and the bandwidth. First, notice that it is easy to check that $\widehat{f}_{Th}(x)$ indeed is a density function, if $K$ has that property. Further, $\widehat{f}_{Th}$ inherits its smoothness properties from $K$. In particular, we may estimate $f'(x)$ by $\widehat{f}'_{Th}(x)$. Provided the returns $R_1, \ldots, R_T$ form an i.i.d. sample, we obtain

$$E(\widehat{f}_{Th}(x)) = \int \frac{1}{h} K\left(\frac{z-x}{h}\right) f(z)\,\mathrm{d}z = (K_h \star f)(x),$$

where $K_h(z) = h^{-1}K(z/h)$ is the rescaled kernel and $\star$ denotes the convolution operator. It follows that the Parzen–Rosenblatt estimator is not an unbiased estimator for $f$; its bias equals

$$b_h(x) = E(\widehat{f}_{Th}(x)) - f(x) = (K_h \star f)(x) - f(x).$$

However, Bochner's lemma, cf. Lemma A.2.1, implies that the convolution $(K_h \star f)(x)$ converges to $f(x)$, as $h \to 0$. Thus, the bandwidth should be chosen as a decreasing function of the sample size $T$. Under the i.i.d. assumption, it is easy to verify that the variance equals

$$\sigma^2_{Th}(x) = \operatorname{Var}(\widehat{f}_{Th}(x)) = \frac{1}{Th}\left[(K_h^2 \star f)(x) - (K_h \star f)^2(x)\right],$$

where $K_h^2(z) = h^{-1}K^2(z/h)$, $z \in \mathbb{R}$. Again, Bochner's lemma implies that the expression in brackets converges to finite constant, such that the variance of $\widehat{f}_{Th}$ is of the order $1/Th$ and tends to 0, if $Th \to \infty$. Let us consider the mean squared error (MSE),

$$\mathrm{MSE}(\widehat{f}_{Th}(x); f(x)) = E(\widehat{f}_{Th}(x) - f(x))^2,$$

which can be decomposed into its two additive components, the variance $\sigma^2_{Th}(x)$ and the squared bias $b_h^2(x)$,

$$\mathrm{MSE}(\widehat{f}_{Th}(x); f(x)) = \sigma^2_{Th}(x) + b_h^2(x).$$

We see that the MSE converges to zero for any bandwidth choice satisfying

$$h \to 0 \qquad \text{and} \qquad Th \to \infty.$$

To get further insights, we need the following notion. A kernel $K$ is called the **kernel of the order** $r$, if

$$\int z^j K(z)\,\mathrm{d}z = \begin{cases} 1, & j = 0, \\ 0, & j = 1, \ldots, r-1, \\ c \neq 0, & j = r. \end{cases}$$

For example, the kernel $K(x) = \left(\frac{9}{8} - \frac{15}{8}x^2\right) \mathbf{1}(|x| \leq 1)$, $x \in \mathbb{R}$, is a kernel of order 4. Let us assume that the underlying density $f$ is $r$ times differentiable. Then one can easily establish the expansions

$$E(\widehat{f}_{Th}(x)) = f(x) + h^r f^{(r)}(x) \frac{(-1)^r}{r!} \int u^r K(u)\,\mathrm{d}u + o(h^r),$$
$$\mathrm{Var}\,(\widehat{f}_{Th}(x)) = \frac{1}{Th} f(x) \int K^2(z)\,\mathrm{d}z + o(1/Th),$$

which yield the following expansion for the MSE

$$\mathrm{MSE}(\widehat{f}_{Th}(x); f(x)) = \frac{f(x)R(K)}{Th} + h^{2r}[f^{(r)}(x)]^2 M_r^2 + o(h^{2r} + 1/Th),$$

where

$$M_r = \frac{(-1)^r}{r!} \int u^r K(u)\,\mathrm{d}u$$

and

$$R(g) = \int g^2(x)\,\mathrm{d}x$$

measures the roughness of a $L_2$ function $g$. These expansion show that higher-order kernels reduce the order of the bias, which is now $O(h^{2r})$.

A bandwidth choice is called **local asymptotically optimal bandwidth**, if it minimizes the dominating terms of the above expansion represented by the function

$$h \mapsto \frac{f(x)R(K)}{Th} + h^{2r}[f^{(r)}(x)]^2 M_r^2, \qquad h > 0.$$

It is easy to see that the optimal bandwidth is given by

$$h^*(x) = h_T^*(x) = \left( \frac{f(x)R(K)}{2rM_r^2[f^{(r)}(x)]^2 T} \right)^{1/(2r+1)}.$$

In particular, we see that for a second-order kernel the optimal bandwidth is of the order $O(T^{-1/5})$. Notice that this approach leads to a local bandwidth choice. In order to use that approach in practice, one needs pilot estimators of the density $f(x)$ and the derivative $f^{(r)}(x)$.

However, more common are global approaches based on the **integrated mean squared error (IMSE)**

$$\mathrm{IMSE}(\widehat{f}_{Th}; f) = \int \mathrm{MSE}(\widehat{f}_{Th}(x); f(x))\,\mathrm{d}x = \int E\big(\widehat{f}_{Th}(x) - f(x)\big)^2\,\mathrm{d}x.$$

For $r = 2$ one obtains the expansion

$$\text{IMSE}(\widehat{f}_{Th}; f) = \frac{R(K)}{Th} + \frac{1}{4}h^4 M_2^2 \int [f^{(2)}(x)]^2 \, \mathrm{d}x + o(h^4 + 1/Th).$$

Neglecting the remainder yields the **asymptotic integrated mean squared error (AMISE)**,

$$\text{AMISE}(h) = \frac{R(K)}{Th} + \frac{1}{4}h^4 M_2^2 \int [f^{(2)}(x)]^2 \, \mathrm{d}x,$$

which we now study as a function of the bandwidth $h$. The optimal bandwidth $h_{\text{opt}}$ that minimizes the AMISE and is easily shown to be

$$h_{\text{opt}} = C_0 T^{-1/5},$$

where

$$C_0 = M_2^{-2/5} R(K)^{1/5} \left[ \int f^{(2)}(x)]^2 \, \mathrm{d}x \right]^{-1/5}.$$

Unfortunately, the constant $C_0$ is unknown. The **normal reference rule-of-thumb** determines the constant for the standard normal distribution with mean zero and variance $\sigma^2$ as a reference model. When also using a normal kernel for smoothing, we obtain the optimal bandwidth

$$h^*_{\text{opt}} = (4\pi)^{-1/10} \left[ (3/8)\pi^{-1/2} \right]^{-1/5} \sigma \cdot T^{-1/5} \approx 1.06\sigma T^{-1/5}.$$

This choice if often used in practice with $\sigma$ estimated by the sample standard deviation of the data.

Clearly, an undesirable feature of the above approach is that the method is tuned to a fixed reference distribution, as it tries to estimate the asymptotically optimal bandwidth in this case, although the kernel density aims at estimating an arbitrary (smooth) density. Thus, fully automatic procedures that do not make such restrictions are usually applied. Widespread approaches are unbiased and biased least-squares cross-validation, which we shall briefly discuss here.

**Least squares unbiased cross-validation** minimizes a nonparametric estimator of the integrated squared error and therefore provides an optimal bandwidth tailord to all $x$ in the support instead of fixing some $x$. Since

$$\int \left[ \widehat{f}_{Th} - f(x) \right]^2 \mathrm{d}x = \int \widehat{f}^2_{Th}(x) \, \mathrm{d}x - 2 \int \widehat{f}_{Th}(x) f(x) \, \mathrm{d}x + \int f(x)^2 \, \mathrm{d}x,$$

minimizing the IMSE is equivalent to minimizing the first two terms on the right-hand side. Observe that

$$\int \widehat{f}_{Th}(x) f(x) \, \mathrm{d}x = E_R(\widehat{f}_{Th}(R)),$$

if $R \sim f$ is independent from $R_1, \ldots, R_T$ and $E_R$ denotes the expectation with respect to $R$. Thus, we may estimate $E_R(\widehat{f}_{Th}(R))$ by

$$\widehat{f}_{T,-i} = \frac{1}{(T-1)h} \sum_{t=1, t \neq i}^{T} K\left(\frac{R_t - R_i}{h}\right).$$

That estimate is called the **leave-one-out estimate of** $f(X_i)$**.** The first term is estimated by plugging in the kernel density estimate,

$$\begin{aligned} \int \widehat{f}_{Th}^2(x)\, \mathrm{d}x &= \frac{1}{T^2h^2} \sum_{t=1}^{T} \sum_{s=1}^{T} \int K\left(\frac{R_t - x}{h}\right) K\left(\frac{R_s - x}{h}\right) \mathrm{d}x \\ &= \frac{1}{T^2h} \sum_{t=1}^{T} \sum_{s=1}^{T} (K \star K)\left(\frac{R_t - R_s}{h}\right). \end{aligned}$$

Least squares cross-validation uses these estimators and minimizes the objective function

$$\mathrm{UCV}(h) = \frac{1}{T^2h} \sum_{t=1}^{T} \sum_{s=1}^{T} (K \star K)\left(\frac{R_t - R_s}{h}\right) - \frac{2}{T(T-1)h} \sum_{s=1}^{T} \sum_{t=1, t \neq s}^{T} K\left(\frac{R_t - R_s}{h}\right),$$

which has to be done numerically. Thus, the expectation of both terms yielding $\mathrm{UCV}(h)$ match the first two terms of the IMSE. One can show that, asymptotically, minimizing $\mathrm{CV}(h)$ is indeed equivalent to minimizing

$$B_1 h^4 + \frac{R(K)}{Th},$$

where

$$B_1 = \frac{M_2^2}{4} \left\{ \int [f^{(2)}(x)]^2\, \mathrm{d}x \right\}.$$

From here it is easy to see that the minimizer of the last display coincides with the minimizer of the IMSE. Moreover, one can even show that

$$\frac{h_{\mathrm{LCV}} - h_{\mathrm{opt}}}{h_{\mathrm{opt}}} \to 0,$$

as $T \to \infty$, in probability, a strong justification of the method.

**Biased least-squares cross-validation** minimizes another estimate of the asymptotic mean squared error (AMISE). Recall that

$$\mathrm{AMISE}(h) = \frac{R(K)}{Th} + \frac{1}{4} K_2^2 h^4 R(f'').$$

The optimal bandwidth is given by

$$h_0 = \left(\frac{R(K)}{M_2^2 T R(f'')}\right)^{1/5}.$$

A natural estimate for the only unknown quantity $R(f'')$ is $R(\widehat{f}''_T)$, where $\widehat{f}''_T$ is the second derivative of the kernel estimator $\widehat{f}_T$, but it turns out that

$$E(R(\widehat{f}''_T)) = R(f'') + \frac{R(K'')}{Th^5} + O(h^2).$$

One can do better by estimating the positive bias. This leads to the estimator $R(\widehat{f}''_T) - \frac{R(K'')}{Th^5}$. Noticing that

$$R(\widehat{f}''_T) = \frac{R(K'')}{Th^5} + \frac{2}{T^2h^5} \sum_{1 \le s < t \le T} \phi\left(\frac{X_t - X_s}{h}\right),$$

where

$$\phi(x) = \int K''(u)K''(u + x)\,\mathrm{d}u, \qquad x \in \mathbb{R},$$

This leads to the biased cross-validation function

$$\mathrm{BCV}(h) = \frac{R(K)}{Th} + \frac{K_2^2}{2T^2h} \sum_{1 \le s < t \le T} \phi\left(\frac{X_t - X_s}{h}\right),$$

which is then minimized.

Figure 1.1 illustrates the kernel density estimator for the daily log returns of the FTSE from 1991 to 1998. The bandwidth is selected by the biased least-squares cross-validation method.

### 1.3.5 Testing for normality

Asset returns are often non-normal, particularly returns corresponding to small time lags such as daily or intraday returns. In order to check the hypothesis that the returns are normal, many statistical tests have been proposed in the literature. At this point, we shall discuss those tests that are most widely used in practice.

Let $R_1, \ldots, R_T$ be an i.i.d. sample of returns with common d.f. $F$. We aim at testing the null hypothesis that $F$ is a normal distribution,

$$H_0 : F \in \{\Phi_{\mu,\sigma^2} : \mu \in \mathbb{R}, \sigma^2 > 0\}$$

against the alternative hypothesis

$$H_1 : F \notin \{\Phi_{(\mu,\sigma^2)} : \mu \in \mathbb{R}, \sigma^2 > 0\}.$$

Notice that $H_1$ means that for all $\mu \in \mathbb{R}$ and $\sigma^2 > 0$ there exists at least one $x \in \mathbb{R}$ such that $F(x) \neq \Phi_{(\mu,\sigma^2)}(x)$.

The **Jarque and Bera test** is given by

$$J_T = T\left(\frac{\widehat{\mu}_3^2}{6} + \frac{(\widehat{\mu}_4 - 3)^2}{24}\right),$$

where $\widehat{\mu}_3$ is the sample skewness and $\widehat{\mu}_4$ the sample kurtosis. Since $J_T$ is asymptotically $\chi^2(2)$-distributed, as $T \to \infty$, one rejects $H_0$, if $J_T > \chi^2(2)_{1-\alpha}$. However, the test should be used only for large data sets. Notice that the Jarque and Bera test measures the departure of the sample skewness and kurtosis from their theoretical values under the null hypothesis.

Another class of tests is based on the following idea. If the null hypothesis is true, we estimate the parameters $\mu$ and $\sigma^2$ by their sample analogs $\widehat{\mu}_T$ and $S_T^2$. The corresponding estimate of the distribution function is then $\Phi_{(\widehat{\mu}_T, S_T^2)}(x)$. If the alternative hypothesis is true, we may rely on the empirical distribution function, i.e.

$$\widehat{F}_T(x) = \frac{1}{T}\sum_{t=1}^{T} \mathbf{1}(R_t \le x), \qquad x \in \mathbb{R},$$

which provides a consistent estimator of $F(x)$ without assuming any specific shape of the distribution. Now we can compare those two estimates by calculating the maximum deviation. This motivates the **Lilliefors test statistic**

$$L = \sup_{t \in \mathbb{R}} |\widehat{F}_T(t) - \Phi_{(\widehat{\mu}_T, S_T^2)}(t)|.$$

The asymptotic distribution of $L$ is none of the standard distributions that have appeared so far. To conduct the test on the 5% significance level, one compares $L$ with the critical value $0.805/\sqrt{T}$. However, the test is implemented in standard statistical software.

Sometimes, one wants to test the simple null hypothesis $H_0 : F = \Phi_{(\mu_0, \sigma_0^2)}$ against the alternative hypothesis $H_1 : F \neq \Phi_{(\mu_0, \sigma_0^2)}$ for some known constants $\mu_0 \in \mathbb{R}$ and $\sigma_0^2 > 0$. In this case, one may calculate

$$KS = \sup_{t \in \mathbb{R}} |\widehat{F}_T(t) - \Phi_{(\mu_0, \sigma_0^2)}(t)|.$$

That test is called the **Kolmogorov–Smirnov test**.

## 1.4 Financial instruments

Before proceeding, we shall introduce some financial slang and basic financial instruments. From an economic point of view, a trade is an agreement between two parties, a buyer and a seller, to buy or sell a certain amount of an asset at a certain date. The buyer attains a **long position** in the asset and the seller a **short position**. Associated to each trade are payments. For a given party we agree on the following sign convention: If the party receives a payment, it gets a positive sign. If the party has to pay the amount, we assign a negative sign.

### 1.4.1 Contingent claims

The payments of many financial instruments depend on other instruments or variables, often securities such as stocks, stock indices, oil, energy prices, or commodities, which are then called the **underlying** of such an instrument. It is even possible to buy financial instruments whose payment depends on quantities such as the weather.

Derivatives and futures are used for hedging risks associated with the production and distribution of goods and services in the real economy and, indeed, they are needed for those purposes. But they are also used a lot for pure speculation. To some extent speculators are needed as counterparties for hedges, but some markets are dominated by excessive speculation leading to substantial bubbles. For example, the unethical speculation in agricultural commodities since 2005, when volatility increased due to extreme weather incidents and increasing demand, is regarded as a substantial factor for record highs of food prices in developing countries leading to social instability and starvation.

A financial instrument whose payoff depends on another quantity is called a **contingent claim**. We shall give a mathematical definition later. If the underlying is a security such as an exchange-traded stock, it is called **derivative asset**. In what follows, we introduce the most important derivatives and related instruments and contracts.

### 1.4.2 Spot contracts and forwards

**Definition 1.4.1** *A* **spot contract** *is an agreement to buy or sell an asset at the same day at a certain price called spot price that we shall denote by $S_t$. In the following, we shall assume that $t = 0$ stands for the time when a trade is initiated and $T$ denotes the time horizon when the trade is settled. By contrast,* **forward** *contracts are agreements to buy or sell an asset at a future time at a price that is fixed when the parties agree on the contract, i.e. today. A forward allows the holder of the long position to buy the asset at a future time point $T$, the delivery date, at a fixed delivery price $K$, which coincides with the forward price $F$. The payoff of a long forward contract is $S_T - K$ and $K - S_T$ for a short position.*

The markets where spot contracts are traded are called spot markets. Forwards are traded over-the-counter (OTC), usually between financial institutions such as banks and their clients, e.g. an enterprise or private investor. There are no cash payments in $t = 0$. A forward is settled at the delivery date $T$ when the seller has to deliver the asset to the buyer. However, often the parties agree on cash settlement. If the price at delivery, $S_T$, is higher than the delivery price $K$, the holder of a long position receives the payment $S_T - K$ and makes a profit. That additional payment hast the effect that he buys the asset for the forward price $F = K$, since $-S_T + (S_T - K) = -K$. But if the price is lower, he has to pay the difference to the seller. Again due to this additional payment, the net price of buying the asset is the delivery price.

### 1.4.3 Futures contracts

**Definition 1.4.2** *Futures are standardized forward contracts usually traded on an exchange.*

For instance, the NYMEX light sweet crude oil futures is a contract on the physical delivery of 1000 barrel during a specified month. Standardization and handling by exchanges allows market participants to actively trade the contracts. Thus, in contrast to forwards, which can be highly specialized nontradeable agreements, futures can be very liquid financial instruments. The exchange specifies in detail the asset, how many units will be delivered under one contract (the contract size), the delivery date and how and where the asset will be delivered. For many contracts physical delivery is not possible or inconvenient and cash settlement applies. Here an equivalent cash payment between the parties is initiated. A futures contract can be bought and sold at any time point until its delivery date. The corresponding price is the futures price. At each trading day a settlement price is quoted, usually the closing price immediately before the end of trading day. The settlement price is used to determine the margins that are required from any investor. The investor has to deposit funds in a margin account. When entering a contract, the so-called initial margin has to be paid. At each trading day the account is marked to market to adjust the possible losses and gains. When the futures price rises, the holder of a long position makes a profit that is exactly the loss of the holder of the short position. The broker of an investor who is short reduces the margin account by the loss and the exchange transfers the money to the broker of the counter party where it increases the margin account. This is called daily settlement. If the margin account falls below the maintenance margin, the investor receives a margin call to deposit further funds. Otherwise the broker will close-out the position, i.e. neutralizing the existing contract.

### 1.4.4 Options

Options are agreements that give the holder of a long position the right, but not the obligation, to buy or sell the underlying at a fixed price in the future under certain conditions. There are a vast number of options traded nowadays; the most basic options are described in the following definition.

**Definition 1.4.3** (European Call/Put Option, Basis Price, Expiration Date)
*A* **European call option** *gives the holder the right to buy the underlying at a specified price, the* **strike price** *or* **basis price** *$K$ at a fixed time point $T$ called* **maturity** *or* **expiration date**. *The holder of a* **European put option** *has the right to sell the underlying for the strike price at maturity. If $S_t$ stands for spot price of the underlying at time $t \in [0, T]$, we will denote the price (fair value) of a European call at time t by $C_e(S_t, K, t, T)$. Our notation for the price of a put will be $P_e(S_t, K, t, T)$. $T - t$ is called the* **time-to-maturity**.

Often, cash settlement applies. This means, the buyer does not get the underlying but the equivalent amount of money he would realize as a profit when buying the underlying for the strike price and selling it on the market. Denote by $C(S_t, K, t, T)$ the price of such an option at time $t$. At time $T$ it coincides with the payoff given by

$$s \mapsto C(s, K, T, T), \qquad s \in [0, \infty).$$

The holder of a European call exercises the option, if $S_T > K$. The profit is $S_T - K$. Thus,

$$C_e = C_e(S_T, K, T, T) = \begin{cases} S_T - K, & S_T > K, \\ 0, & S_T \le K, \end{cases}$$

which can be written in the form

$$C_e = \max(0, S_T - K) = (S_T - K)^+.$$

Similarly, for a European put option we have

$$P_e = P_e(S_T, K, T, T) = \max(0, K - S_T) = (K - S_T)^+.$$

The **internal value** of an option is its positive cashflow when one would exercise it. For a European call it is given by $(S_t - K)\mathbf{1}(S_t > K)$ and for a put equals $(K - S_t)\mathbf{1}(S_t < K)$. An option is **in the money**, when its internal value is positive ($S_t > K$ for a call, $S_t < K$ for a put), and it is called **out of the money** if the internal value is 0. ($S_t < K$ for a call, $S_t > K$ for a put). The ratio $K/S$ is called **moneyness**.

**Example 1.4.4** (Portfolio Insurance)
*European put options can be used to solve the problem discussed in Exercise 1.1.1. Suppose the pension funds intends to buy a portfolio of stocks, frequently called basket of stocks, whose current price is* $S_t = 110$. *Further, assume that the pension fund can buy a European put option on that basket. If the pension fund is willing to take a (downside) risk of at most* 10 *units of currency, a put with strike* 100 *has to be chosen.*

*The portfolio of the pension fund consists of the basket and one put option. Consider its value at maturity* $T$. *If* $S_T > 100$, *the put option is out of the money, i.e. its value is* 0, *such that the porfolio's value is* $S_T$. *In the case* $S_T \leq 100$, *the payoff of the put option is* $100 - S_T$ *such that the porfolio's value is* $V_T = S_T + (100 - S_T) = 100$. *It follows that the loss can not exceed* 10 *units of currency.*

### 1.4.5 Barrier options

The value of a barrier call option depends on whether the price of the underlying touches a certain value called barrier. Knock-out options die if the barrier is reached, whereas knock-in options are activated in this case.

**Definition 1.4.5** *A European barrier option with expiration date* $T$, *barrier* $B$, $B < S_0$ *and* $B < K$, *and strike price* $K$ *gives the option holder the right to buy the underlying at time* $T$, *if*

$$S_t > B \text{ for all } 0 \leq t \leq T \quad \text{(down-and-out)}$$

*and*

$$S_t < B \text{ for all } 0 \leq t \leq T \quad \text{(up-and-out)},$$

*respectively. For a knock-in option the right is activated when*

$$S_t \leq B \text{ for some } t \in [0, T] \quad \text{(down-and-in)}$$

*or*

$$S_t \geq B \text{ for some } t \in [0, T] \quad \text{(up-and-in)}.$$

*American-style options allow buying the underlying at an arbitrary time point provided they are activated.*

Barrier options are examples of **path-dependent options** whose payoff and value depends on the price trajectory $S_t,\ 0 \leq t \leq T$, during the lifetime of the contract.

**Definition 1.4.6** *An American* **average price call options** *is given by the payoff profile*

$$\max(0, \overline{S_t} - K), \quad t = 1, \ldots, T,$$

*where* $K$ *stand for the exercise price and*

$$\overline{S_t} = \frac{1}{t} \sum_{i=1}^{t} S_i, \quad t = 1, \ldots, T,$$

*denotes the average price. American* **strike call options** *have the payoff profile* $\max(0, S_t - \overline{S_t})$, $t = 1, \dots, T$. *Their exercise price is determined when the option is exercised. The corresponding European-style options are given by the payoffs* $\max(0, \overline{S_T} - K)$ *and* $\max(0, S_T - \overline{S_T})$ *at maturity, respectively.*

### 1.4.6 Financial engineering

By combining financial instruments, particularly derivatives, one can implement interesting payoff profiles. For example, a **long straddle** consists of a long position in a European call option and a long position in a European put with the same underlying and the same maturity, both in the money. For large increases of the stock price, the long positions provides a profit, whereas for large decreases of the stock price the put earns the money. In this way, one can create a position that makes a profit if the stock price changes, independent of the direction.

Basically, we shall see that the fair price $\pi$ of a derivative or a contingent claims can be calculated by an expectation $E^*(C^*)$ of the discounted payoff $C^*$ of the derivative under a certain probability measure. This automatically also allows us to price portfolios of contingent claims. Suppose such a portfolio consists of $n$ positions given by the discounted payoffs $C_1^*, \dots, C_n^*$ of each claim and the numbers of contracts $x_1, \dots, x_n$ we held. Since expectations are linear, the fair price of the portfolio is

$$E^* \left( \sum_{i=1}^{n} x_i C_i^* \right) = \sum_{i=1}^{n} x_i \pi_i,$$

where $\pi_i = E^*(C_i^*)$ is the fair price of the $i$th claim.

In financial engineering, artificial portfolios of derivatives are often constructed in order to generated certain payoff profiles, for example in order to simultaneously hedge risks and generate opportunities for a profit, or as a complex financial product for customers. If a given payoff profile, $Z$, can be constructed by a portfolio such that $Z = \sum_{i=1}^{n} x_i C_i$, then the above formula allows us to determine the fair price of such a complex product. What makes such products challenging and risky is the fact that the underlying instruments $C_1, \dots, C_n$ may have quite different risk exposures to risk factors such as interest rates, price changes of the underlying, volatility changes of the underlying or the risk that the issuer of the instrument defaults. Furthermore, the underlying portfolio is often unknown to the customer, which hinder his or her evaluation of the risk associated to such a product.

## 1.5 A primer on option pricing

This section is devoted to an introduction to some basic ideas and principles that lead to a powerful and elegant theory of option pricing. It is a matter of fact that they can be explained and understood in the simplest framework of a financial market with one asset and one European call option. We will obtain first convincing answers to the question on how to determine a fair price for a contingent claim, but simultaneously these answers give rise to various questions on how to extend them to more general and realistic frameworks.

### 1.5.1 The no-arbitrage principle

The no-arbitrage principle says that on an idealized financial market the prices do not allow for a riskless profit, i.e. there is no *free lunch*. Such arbitrage opportunities can arise if, for example, the prices in New York are higher than in London or the price of a bond is less than the fair value of its future payments. For what follows, we use the following mathematical definition.

**Definition 1.5.1** *An* **arbitrage opportunity** *is a transaction yielding a random payment* $X_1$ *in* $t = 1$ *with initial value* $x_0$ *in* $t = 0$ *such that*

$$x_0 \le 0 \qquad \textit{(no costs)}$$

*and*

$$X_1 \ge 0 \ P-a.s., \qquad \text{and} \qquad P(X_1 > 0) > 0.$$

**Example 1.5.2** *Let us apply the no-arbitrage principle to determine the fair value* $F_0$ *of a forward contract, i.e. the arbitrage-free price that applies at time* $t = 0$*. We claim that there is a unique no-arbitrage forward price, namely* $F_0 = S_0 e^{rT}$*, when assuming continuous compounding. Assume* $F_0 > S_0 e^{rT}$*. In this case, the seller can make a riskless profit by borrowing* $S_0$ *at time zero and buying the underlying. At maturity, he sells the underlying at the delivery price* $K$*, pays back* $S_0 e^{rT}$ *and earns* $F_0 - S_0 e^{rT} > 0$*. If* $F_0 < S_0 e^{rT}$*, the buyer sells the underlying aud puts the money to the bank. At maturity he receives* $S_0 e^{rT}$ *and pays* $F_0$ *for the underlying, leaving a profit* $S_0 e^{rT} - F_0$*. It is interesting and important to note that the forward price does not depend on the price of the underlying at maturity.*

The no-arbitrage principle also immediately leads to a simple formula that relates the price of an European call and European put. The idea is to set up a portfolio that leads to the same payoff as an European call option with maturity $T$ and strike price $K$. If we buy a stock and sell a zero bond with nominal $K$, the value at time $T$ is $S_T - K$. If we add a put to the portfolio, its value at maturity is zero, if $S_T > K$, but $K - S_T$, if $S_T \le K$. It follows that the value of the portfolio is 0, if $S_T \le K$, but $S_T - K$, if $S_T > K$. Its value at time 0 is

$$\pi(P_e) - Ke^{-rT} + S_0$$

and must be equal to the fair price of the call, which establishes the **put-call parity**

$$\pi(C_e) = \pi(P_e) - Ke^{-rT} + S_0.$$

The existence of arbitrage opportunities, which is ruled out by the no-arbitrage principle, means that the current prices of financial instruments are not balanced with their future payments. Many economists argue that on real financial markets arbitrage can at best exist temporarily, since they are discovered by market participants that then enter trades that quickly remove the arbitrage opportunity. If, for instance, the price of a financial instrument is too low and provides a free lunch, speculators will enter long positions such that its price will rise until the riskless profit disappears. We shall see that the no-arbitrage principle is a powerful and simple approach to determine fair prices.

### 1.5.2 Risk-neutral evaluation

The evaluation of a random (future) payment $X$ depends on the preferences that can be expressed via a probability measure on the underlying measure space. The crucial question is whether a fixed payment, i.e. the case $X(\omega) = x_0$, for all $\omega \in \Omega$ and some fixed $x_0$, is preferred to a risky payment that offers the chance that the event $\{X > x_0\}$ occurs, but usually at the risk that the event $\{X < x_0\}$ may occur as well.

For simplicity of exposition, let us assume that the uncertainty about the future payment is measured in terms of the volatility, i.e. the square root of the variance. Given two investment opportunities with equal means, a **risk-averse** investor prefers the alternative with the smaller variance. By contrast, if the investor is **risk neutral**, he has no preference at all, since he ignores the variance.

In a risk-neutral world of risk-neutral investors everybody just looks at the mean. Let us denote the probability measure corresponding to this risk-neural world by $P^*$. Under $P^*$ a stock is preferred to a riskless investment, if and only if its expected return is higher than the riskless return earned on a bank account. Denote the stock's price at time $t$ by $S_t$ and denote its random return by $R$. We assume that the price $S_0$ at $t = 0$ is a constant $S_0$ known to us. Then the random price at $t = 1$ is given by

$$S_1 = S_0(1 + R).$$

In a risk-neutral world the value of that payment is given by

$$E^*(S_1) = S_0(1 + E^*(R)).$$

Here and throughout, the symbol $E^*$ means that the expectation is calculated under the probability measure $P^*$. If we deposit the initial capital $S_0$ in a bank account, we obtain $S_0(1 + r)$. The principle of no-arbitrage implies that $E^*(S_1)$ and $S_0(1 + r)$ must coincide, i.e.

$$E^*(S_1) = S_0(1 + r) \qquad \Leftrightarrow \qquad E^*\left(\frac{S_1}{1 + r}\right) = S_0.$$

As a consequence, under risk-neutral pricing the (fair) price of the stock can be calculated as an expectation under the probability measure $P^*$. Can we calculate $P^*$ from the above equation?

To get first insights, we shall study a very simple one-period model for a financial market consisting of one stock and one European call option on that stock. To make the model as simple as possible, let us assume a binomial model for the stock price where the price can either go up or go down. In this case, we may choose the sample space $\Omega = \{+, -\}$ to represent the possible future states of our financial market, equipped with the power set sigma field. The real probability measure $P$ is uniquely determined by $P(\{+\}) = p$, $p \in (0, 1)$. Notice that we exclude the trivial cases $p = 0$ and $p = 1$. We model the stock price by

$$S_1(\omega) = \begin{cases} S_0 u, & \omega = +, \\ S_0 d, & \omega = -, \end{cases}$$

with constants $u$ (up factor) and $d$ (down factor) satisfying $0 < d < 1 + r < u$. The European call is given by its payoff

$$C_e = \begin{cases} S_1 - K, & S_1 > K, \\ 0, & S_1 \le K. \end{cases}$$

To avoid trivialities, we shall assume that the strike price $K$ ensures that $S_0 d < K < S_1 u$.

In the above simple model the risk-neutral probability measure $P^*$ is uniquely determined by $p^* = P^*(\{+\})$. The risk-neutral pricing formula $E^*(S_1) = S_0(1+r)$ is now equivalent to the equation

$$p^* S_0 u + (1 - p^*) S_0 d = S_0(1+r),$$

which has the unique solution

$$p^* = \frac{1 + r - d}{u - d}.$$

This means, given the model parameters $r, d$ and $u$ we can determine $P^*$. Relying on the principle of risk-neutral pricing, the fair value of any random payment $X_1$ at time $t = 1$ can be calculated by

$$\pi(X_1) = E^*(X_1/(1+r)).$$

In particular, for a European call option on a stock we obtain

$$\pi(C_e) = p^* \frac{S_0 u - K}{1 + r}.$$

**Example 1.5.3** *Recall Example 1.1.3 and Example 1.1.4, where the oil price was assumed to either go up by* 10% *or go down by* 10%. *This means that we have* $u = 1.1$ *and* $d = 0.9$. *The riskless rate was* $r = 0.01$. *Hence, the risk-neutral probability measure* $P^*$ *is given by*

$$p^* = \frac{1 + r - d}{u - d} = \frac{1.01 - 0.9}{0.2} = 0.55,$$

*yielding the risk-neutral option price*

$$E^*(C_e/(1+r)) = \frac{10}{1.01} 0.55 = 5.445545.$$

*This is exactly the lower price limit calculated by the oil trader.*

Let us slightly generalize our model to allow for a trinomial model for the stock price. We put $\Omega = \{+, \circ, -\}$ and assume that, given three factors $d < m < u$, the stock price at time $t = 1$ satisfies

$$S_1(\omega) = \begin{cases} S_0 u, & \omega = +, \\ S_0 m, & \omega = \circ, \\ S_0 d, & \omega = -. \end{cases}$$

The risk-neutral probability measure $P^*$ is now determined by $p_1^*, p_2^*, p_3^* \in [0, 1]$ such that $p_1^* + p_2^* + p_3^* = 1$. In this model, the pricing formula $E^*(S_1) = S_0(1+r)$ leads to

$$p_1^* u + p_2^* m + (1 - p_1^* - p_2^*) d = (1+r) \quad \Leftrightarrow \quad p_1^*(u - d) + p_2^*(m - d) = (1+r) - d.$$

This equation has infinite solutions. The special solution corresponding to $p_2^* = 0$ is the solution of the binomial model. In general, the solutions can be parameterized by $p_2^*$ yielding

$$p_1^* = \frac{1 + r - d + p_2^*(m - d)}{u - d}, \qquad p_2^* \in [0, 1], \; p_3^* = 1 - p_1^* - p_2^*.$$

It follows that pricing using the risk-neutral approach is not unique; there are infinitely many prices.

**Exercise 1.5.4** *Determine all risk-neutral probability measures. Which conditions on $d$, $m$, $u$ and $r$ are required?*

### 1.5.3 Hedging and replication

Options are usually written by banks that are interested in hedging the risk of such a deal. Again, we consider a European option $C_e$ on a stock $S_1$ that follows a binomial model. By introducing the notations $S_1(-)$, $S_1(+)$ and $C_e(-)$, $C_e(+)$, we shall see that the formulas we are going to derive hold for general options as well. The question arises whether it is possible to set up a portfolio that neutralizes any risk from the option deal. If we had a portfolio that exactly reproduces the option, we could buy that portfolio to neutralize the financial effect of selling the option to a customer. So, let us assume the bank holds a portfolio $(\theta_0, \theta_1)$, where $\theta_0$ is the amount of cash deposited in the bank account and $\theta_1$ stands for the shares. Denote the value of the portfolio at time $t$ by $V_t$. The portfolio neutralizes the option if it has the same value at $t = 0$ and $t = 1$. Obviously,

$$V_0 = \theta_0 + \theta_1 S_0,$$

and

$$V_1(\omega) = \begin{cases} \theta_0(1+r) + \theta_1 S_0 u, & \omega = +, \\ \theta_0(1+r) + \theta_1 S_0 d, & \omega = -. \end{cases}$$

The value $W_0$ of the option at time 0 is its price $\pi(C_e)$, and at time 1

$$W_1(\omega) = \begin{cases} S_0 u - K, & \omega = +, \\ 0, & \omega = -. \end{cases}$$

The portfolio replicates the option if $V_t(\omega) = W_t(\omega)$ holds true for all $\omega \in \Omega$ and all $t \in \{0, 1\}$. This leads to the equations

$$V_0 = \pi(C_e) \tag{1.9}$$

and

$$\theta_0(1+r) + \theta_1 S_0 u = S_0 u - K \tag{1.10}$$

$$\theta_0(1+r) + \theta_1 S_0 d = 0 \tag{1.11}$$

Substitute $\theta_0(1+r) = -\theta_1 S_0 d$ (Equation (1.11)) into Equation (1.10) to obtain

$$-\theta_1 S_0 d + \theta_1 S_0 u = S_0 u - K \Leftrightarrow \theta_1(S_0 u - S_0 d) = S_0 u - K \tag{1.12}$$

$$\Leftrightarrow \theta_1(S_0 u - S_0 d) = C_e(+) - C_e(-). \tag{1.13}$$

Thus, noting that $S_0 u - S_0 d = S_1(+) - S_1(-)$, we arrive at

$$\theta_1 = \frac{C_e(+) - C_e(-)}{S_1(+) - S_1(-)}.$$

This ratio, the number of shares needed to replicate (exactly!) the option, is called the **hedge ratio**. In Example 1.1.4 the hedge ratio is $\theta_1 = 10/20 = 1/2$. Indeed, the oil trader bought half of the oil at time $t = 0$, i.e. he constructed the hedge portfolio. For $\theta_0$ we obtain the formula

$$\theta_0 = C_e(-) - \frac{C_e(+) - C_e(-)}{u - d}\frac{d}{1 + r}.$$

For our example, we obtain $\theta_0 = 0 - \frac{10}{1.1-0.9} \cdot \frac{0.9}{1.01} \approx -44.55$. This means, the oil trader borrows the amount 44.554 from the bank. Since he receives the premium 5.45, he can buy the oil to hedge the option. The initial costs for the hedge, the **replication costs**, are $V_0 = \theta_0 + \theta_1 S_0$. These replication costs should be equal to the fair price of the option.

**Exercise 1.5.5** *Show that* $V_0 = E^*\left(\frac{C_e}{1+r}\right)$, *if* $P^*$ *is the probability measure given by* $p^* = \frac{1+r-d}{u-d}$.

## 1.5.4 Nonexistence of a risk-neutral measure

Consider a financial market with two stocks following a binomial model with up factors $u_1, u_2$ and down factors $d_1, d_2$. Risk-neutral evaluation now leads us to two equations, namely

$$p^* u_1 + (1 - p^*)d_1 = 1 + r,$$
$$p^* u_2 + (1 - p^*)d_2 = 1 + r,$$

for the free parameter $p^*$. Depending on the parameters $r, d_1, d_2, u_1, u_2$, there may be no solution. Consequently, there may be no risk-neutral probability measure at all.

## 1.5.5 The Black–Scholes pricing formula

We shall now discuss the famous Black–Scholes option pricing formula, although we have to anticipate some results derived later in this book.

Suppose we have a risk-neutral pricing measure $P^*$ at our disposal and consider a European call option on a stock with price $S_t$ and strike $K$. The payoff at maturity $T$ is $C = \max(S_T - K)$. Suppose that a fixed interest is paid in each period and let us express the corresponding discount factor in the form $e^{-r}$ for some $r > 0$. Then the discounted payoff is $C^* = e^{-rT}\max(S_T - K, 0)$. In a risk-neutral world, we must have $E^*(C) = C_0$, where $C_0$ denotes the fair price at time $t = 0$ of the random payment $C$, or, equivalently,

$$C_0 = E^*(C^*) = e^{-rT} E^*(\max(S_T - K, 0)).$$

This means, we may calculate the fair price of the European call option by evaluating the expression on the right-hand side, which requires determination of the distribution of $S_T$ under $P^*$.

The famous Black–Scholes model assumes that under the real probability measure log prices are normally distributed, say, with drift parameter $\mu \in \mathbb{R}$ and volatility $\sigma > 0$. Then it turns out that under $P^*$ the log price $S_T$ at maturity follows a lognormal distribution with

$$\text{drift } \log S + (r - \sigma^2/2)T \quad \text{and volatility } \sigma\sqrt{T}.$$

Here $S = S_0$ denotes today's stock price, which is the basis to determine the fair price of the option.

We will apply the following result: Suppose that $X$ follows a lognormal distribution with parameters $m \in \mathbb{R}$ and $s > 0$. Then

$$E(X-K)^+ = \mathrm{e}^{m+s^2/2}\Phi\left(\frac{m-\log K}{s}+s\right) - K\Phi\left(\frac{m-\log K}{s}\right). \tag{1.14}$$

We will give a sketch of the derivation and encourage the reader to work out the details. To check that nice result, first notice that for $x \geq 0$ we have $(X-K)^+ \geq x \Leftrightarrow X-K \geq x$. Hence, denoting the density of $X$ by $f(x)$,

$$\begin{aligned} E(X-K)^+ &= \int_0^\infty P(X \geq K+x)\,\mathrm{d}x \\ &= \int_0^\infty \int_{K+x}^\infty f(t)\,\mathrm{d}t\,\mathrm{d}x \\ &= \int_K^\infty \int_x^\infty f(t)\,\mathrm{d}t\,\mathrm{d}z, \end{aligned}$$

where we made the change of variable $z = x + K$. If we plug in Equation (1.3), the formula for the density of a lognormal distribution, we arrive at

$$E(X-K)^+ = \int_K^\infty \int_x^\infty \frac{1}{\sqrt{2\pi}st}\mathrm{e}^{-(\log t-m)^2/2s^2}\,\mathrm{d}t\,\mathrm{d}x.$$

Substituting $z = (\log t - m)/s$, such that $\mathrm{d}z = \mathrm{d}t/st$, leads to the integral

$$\int_K^\infty \int_{(\log x-m)/s}^\infty \varphi(z)\,\mathrm{d}z\,\mathrm{d}x,$$

where $\varphi(x) = 1/\sqrt{2\pi}\mathrm{e}^{-x^2/2}$ denotes the density of the standard normal distribution. Apply the integration by parts rule $\int uv' = uv| - \int u'v$ with $u(x) = \int_{(\log x-m)/s}^\infty \varphi(z)\,\mathrm{d}z$ and $v'(x) = 1$ to obtain that

$$E(X-K)^+ = \int_K^\infty \varphi\left(\frac{\log x - m}{s}\right)\mathrm{d}x - K\Phi\left(\frac{m-\log K}{s}\right),$$

where $\Phi(x) = \int_{-\infty}^x \varphi(t)\,\mathrm{d}t$ denotes the d.f. of the standard normal distribution. Finally, using the substitution $z = \log x$ one easily verifies Equation (1.14).

Now let us apply formula (1.14) with $m = \log S_0 + (r - \sigma^2/2)T$ and $s = \sigma\sqrt{T}$:

$$\begin{aligned} E^*(S_T-K)^+ = \mathrm{e}^{\log S + rT}\,&\Phi\left(\frac{\log(S/K)+(r-\sigma^2/2)T}{\sigma\sqrt{T}}+\sigma\sqrt{T}\right) \\ &- K\Phi\left(\frac{\log(S/K)+(r-\sigma^2/2)T}{\sigma\sqrt{T}}\right). \end{aligned}$$

Therefore, the fair price of a European call option is given by

$$\pi(C_e) = E^*(C^*) = \mathrm{e}^{-rT}E^*(S_T-K)^+ = S_0\Phi(d_1) - K\Phi(d_2)\mathrm{e}^{-rT}, \tag{1.15}$$

where

$$d_1 = \frac{\log(S/K) + (r + \sigma^2/2)T}{\sigma\sqrt{T}},$$
$$d_2 = \frac{\log(S/K) + (r - \sigma^2/2)T}{\sigma\sqrt{T}}.$$

It turns out that, by virtue of the call-put parity the price of an European put option is then given by

$$\pi(C_p) = \pi(C_e) + K\mathrm{e}^{-rT} - S.$$

Further, in order to obtain the time $t$ value of such options with time to maturity $\tau = T - t$, one only has to replace $T$ by $\tau$ and let $S$ denote the time $t$ price of the underlying.

### 1.5.6 The Greeks

The Black–Scholes price formula explicitly shows on which quantities the fair arbitrage-free price of a European call option depends: Besides the option parameters $K$, $T$ and the initial price $S_0$, which are fixed in the contract, the formula depends on the risk-free interest rate, $r$, and the volatility $\sigma$ of the the log stock price. For risk management it is essential to know how sensitive a position is with respect to those quantities. If, for example, the volatility of the underlying increases, this will affect immediately the value of a position in a European option.

#### 1.5.6.1 First-order Greeks

We shall now introduce the first-order greeks by referring to a European call priced within the Black–Scholes model. However, these definitions apply to any derivative.

In order to allow easy interpretation, we would like to define the sensitivity with respect to the stock price as the rate of the option's price $V$ if the stock price changes by one unit of currency, i.e. as the ratio $\frac{\Delta\pi(C_e)}{\Delta S}$. Having an explicit formula for $\pi(C_e)$, obviously a differentiable function of $S, T, \sigma$ and $r$, which are now regarded as variables, we can provide a rigorous definition of the sensitivity with respect to the changes of the stock price, called **Delta**, in terms of the partial derivative

$$\Delta = \frac{\partial\pi(C_e)}{\partial S}.$$

In the same vein, we may introduce the sensitivity with respect to a change of the expiration date $T$, which is called **Theta**,

$$\Theta = \frac{\partial\pi(C_e)}{\partial T}.$$

The parameter **Vega** (or **Kappa**) measures the rate of the option's price with respect to changes of the volatility and is defined as the corresponding partial derivative

$$\nu = \frac{\partial\pi(C_e)}{\partial\sigma}.$$

**Table 1.2** Greeks for European options

| Greek | Call Option | Put Option |
|---|---|---|
| $\Delta = \frac{\partial \pi(C_e)}{\partial S}$ | $\Phi(d_1)$ | $\Phi(d_1) - 1$ |
| $\Theta = \frac{\partial \pi(C_e)}{\partial T}$ | $-\frac{S\varphi(d_1)\sigma}{2\sqrt{T}} - rK\mathrm{e}^{-rT}\Phi(d_2)$ | $-\frac{S\varphi(d_1)\sigma}{2\sqrt{T}} + rK\mathrm{e}^{-rT}\Phi(-d_2)$ |
| $\nu = \frac{\partial \pi(C_e)}{\partial \sigma}$ | $S\varphi(d_1)\sqrt{T}$ | $S\varphi(d_1)\sqrt{T}$ |
| $\rho = \frac{\partial \pi(C_e)}{\partial r}$ | $KT\mathrm{e}^{-rT}\Phi(d_2)$ | $-KT\mathrm{e}^{-rT}\Phi(-d_2)$ |
| $\Gamma = \frac{\partial^2 \pi(C_e)}{\partial S^2}$ | $\frac{\varphi(d_1)}{S\sigma\sqrt{T}}$ | $\frac{\varphi(d_1)}{S\sigma\sqrt{T}}$ |

Calculations assume the Black-Scholes model

Finally, **Rho** is the standard notation for the sensitivity with respect to changes of the interest rate and formally given by

$$\rho = \frac{\partial \pi(C_e)}{\partial r}.$$

Table 1.2 lists the resulting formulas assuming the Black–Scholes model.

The first-order greeks allow us to approximate the option's price by a linear function. For example, if the price of the underlying changes from $S$ to $\widetilde{S}$, knowing $\Delta = \frac{\partial \pi(C_e)}{\partial S}$ provides the approximation

$$\pi \approx \pi(C_e) + \frac{\partial \pi(C_e)}{\partial S}(\widetilde{S} - S),$$

which is accurate if $|\widetilde{S} - S|$ is small.

It is important to note that these partial derivatives are still functions of the remaining variables. Hence, their values depend on the values of those variables, the model parameters. If more than one parameter changes, it can not be seen from a single sensitivity measure how the option price reacts.

Observing that, given the strike price $K$, the variables $S, T, r, \sigma$ determine $\pi(C_e)$, it is clear that the above greeks form the gradient

$$\frac{\partial \pi(C_e)}{\partial \vartheta} = \left( \frac{\partial \pi(C_e)}{\partial S}, \frac{\partial \pi(C_e)}{\partial T}, \frac{\partial \pi(C_e)}{\partial \sigma}, \frac{\partial \pi(C_e)}{\partial r} \right)' = (\Delta, \Theta, \nu, \rho)',$$

where $\vartheta = (S, T, \sigma, r)'$. The corresponding linear approximation following from Taylor's theorem is then given by

$$\pi \approx \pi(C_e) + \frac{\partial \pi(C_e)}{\partial \vartheta}(\widetilde{\vartheta} - \vartheta) = \pi(C_e) + \Delta(\widetilde{S} - S) + \Theta(\widetilde{T} - T) + \nu(\widetilde{\sigma} - \sigma) + \rho(\tilde{r}, - r),$$

if the parameters change from $\vartheta$ to $\widetilde{\vartheta} = (\widetilde{S}, \widetilde{T}, \widetilde{\sigma}, \widetilde{r})'$.

#### 1.5.6.2 Second-order Greeks

The first-order greeks correspond to first-order partial derivatives yielding linear approximations of the option's price. The next step is to take into account second-order partial derivatives as well, which lead to quadratic approximations.

Of particular concern is the dependence of the option price on the price of the underlying. The second-order partial derivative

$$\Gamma = \frac{\partial^2 \pi(C_e)}{\partial S^2}$$

is called **Gamma**.

### 1.5.7 Calibration, implied volatility and the smile

In order to price options with the Black–Scholes pricing formula, one has to specify the interest rate $r$. Usually, one takes the yield of a treasury bill with a short maturity. Further, one needs to determine in some way the volatility $\sigma$, which is not directly observable. Basically, there are two approaches. The *statistical approach* is to estimate $\sigma$ from historical data as discussed in Section 1.3.2. Another approach frequently applied in finance is **calibration**, which means that an unknown parameter of a formula for some quantity is determined (calibrated) by matching the formula with real market data for that quantity. This has the advantage that the model reproduces current market data and is therefore often preferred by traders, analysts and bankers, since they tend to mistrust models and methods that seem to contradict markets.

In the case of option pricing by the Black–Scholes formula one calibrates the model by matching the prices predicted by the Black–Scholes formula with real market prices for options by varying the free parameter $\sigma$. Notice that equating Equation (1.15) to a actual price leads to a nonlinear equation for $\sigma$. The matching is done for a fixed strike price $K$ and a fixed time to maturity $T - t$. The volatility $\sigma$ determined in this way is called the **implied volatility**.

In theory, the volatility $\sigma$ of the underlying asset is constant across strike prices and maturities. However, when determining the implied volatility for different values of $K$ and $T$, one observes a dependence on those parameters. Sometimes the volatility is a decreasing function of $K$, a phenomenon called **volatility skew**. In other cases, particularly for options on foreign currencies, the volatility is lower for at-the-money options and gets larger as the option moves into the money or out of the money. This effect is called **volatility smile**. The dependence on $K$ is usually parametrized by the **moneyness** or **strike ratio**, $S/K$. If one calculates the implied volatility over a two-dimensional grid of values for the strike $K$ (or $K/S$) and the maturity $T$, one obtains a two-dimensional curve called the **volatility surface**. Figure 1.2 shows a volatility surface for SIEMENS AG.

### 1.5.8 Option prices and the risk-neutral density

There is an interesting and important relationship between option prices and the probability density of the risk-neutral probability measure used for pricing. The validity of this relationship is not restricted to the Black–Scholes model, but is an intrinsic structural properties of a financial market. It can be used to infer the risk-neutral probability from option prices.

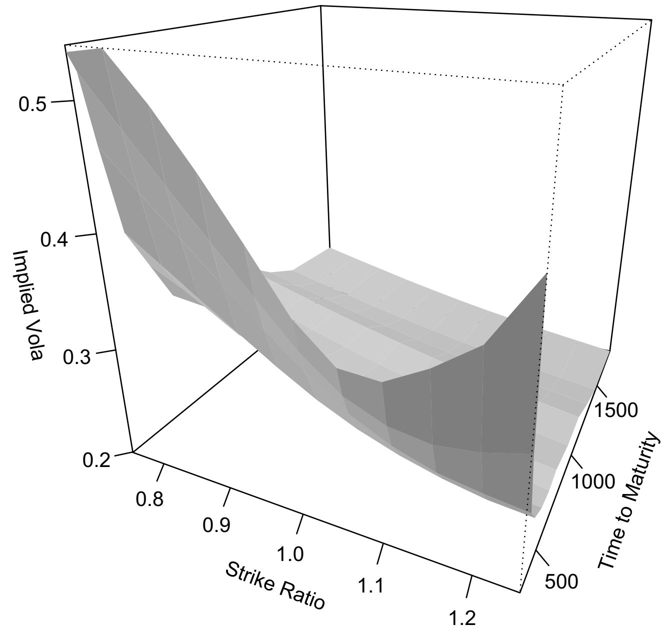


**Figure 1.2** Volatility surface at November 4th of European call options on SIEMENS AG for maturities ranging from November 2011 to December 2015. Time to maturity measured in days, data taken from DATASTREAM.

Recall the starting point of our derivation of the Black–Scholes formula, namely the equation

$$C_e(K) = \mathrm{e}^{-rT} E^*(S_T - K)^+, \tag{1.16}$$

which we now study as a function of the strike price $K$. We also denote the risk-neutral price by $C_e(K)$ to indicate that we do not refer to the Black–Scholes formula. At this point, it is only assumed that there exists a risk-neutral measure $P^*$ used to price random future payments. Let us also assume that the terminal stock stock price $S_T$ attains a probability density under the risk-neutral probability measure $P^*$, which we will denote by $\varphi_T^*(x)$. This means,

$$P^*(S_T \le x) = \int_{-\infty}^{x} \varphi_T^*(u)\,\mathrm{d}u, \qquad x \in \mathbb{R}.$$

Then we may rewrite Equation (1.16) as

$$C_e(K) = \mathrm{e}^{-rT} \int_{-\infty}^{\infty} (x - K)^+ \varphi_T^*(x)\,\mathrm{d}x = \mathrm{e}^{-rT} \int_{K}^{\infty} \varphi_T^*(x)\,\mathrm{d}x.$$

Apply the formula

$$\frac{d}{\mathrm{d}t} \int_{a(t)}^{b(t)} f(x,t)\,\mathrm{d}x = f(b(t),t)b'(t) - f(a(t),t)a'(t) + \int_{a(t)}^{b(t)} \frac{\partial f(x,t)}{\partial t}\,\mathrm{d}x$$

to obtain that the first derivative of the risk-neutral price of a European call satisfies

$$\frac{\partial C_e(K)}{\partial K} = -\int_K^{\infty} \varphi_T^*(x)\,\mathrm{d}x$$

and the second derivative

$$\frac{\partial^2 C_e(K)}{\partial K^2} = \varphi_T^*(K).$$

As a consequence, one may determine the risk-neutral probability measure by analyzing option prices for different strike prices $K$. Since any probability density function is non-negative, we also see that the option prices is a *convex* function of the strike price.

## 1.6 Notes and further reading

A popular text on options, futures and other derivatives avoiding mathematics is the comprehensive book of Hull (2009). It explains in great detail and accompanied by many examples the economic reasoning behind such financial instruments and how the corresponding markets operate, provides basic formulas for the valuation of such financial operations and sketches at an elementary level the mathematical theory behind it. We also refer to the introductions to mathematical finance of Baird (1992), Pliska (1997) and Buchanan (2006), which focus more or less on the discrete-time setting and finite probability spaces, respectively. For the theory of coherent risk measure we refer to the seminal work Artzner et al. (1999), the recent monograph Pflug and Römisch (2007) and the discussion Embrechts et al. (2002) of dependence measures and their properties. There are various text books on the general theory of statistics including estimation, optimal hypothesis testing and confidence intervals, for example Lehmann and Romano (2005) or Shao (2003). Financial statistics is discussed in Lai and Xing (2008). More on kernel smoothing methods and their properties can be found in the monographs Silverman (1986), Härdle (1990), Fan and Gijbels (1996) and Wand and Jones (1995). The problem how to select the bandwidth one may additional consult Scott and Terrell (1987) and Savchuk et al. (2010). For a recent approach using singular spectrum analyses, we refer to Golyandina et al. (2011).

## References

Artzner P., Delbaen F., Eber J.M. and Heath D. (1999) Coherent measures of risk. *Math. Finance* **9**(3), 203–228.

Bachelier L. (1900) Théorie de la spéculation. *Annales Scientifiques de l' É.N.S.* 3$^e$ **17**, 21–86.

Baird A.J. (1992) *Option Market Making: Trading and Risk Analysis for the Financial and Commodity Option Markets*. Wiley Finance, Wiley & Sons.

Buchanan J. (2006) *An Undergraduate Introduction to Mathematical Finance*. World Scientific, Singapore.

Embrechts P., McNeil A.J. and Straumann D. (2002) Correlation and dependence in risk management: properties and pitfalls *Risk Management: Value at Risk and Beyond (Cambridge, 1998)*. Cambridge Univ. Press Cambridge pp. 176–223.

Fan J. and Gijbels I. (1996) *Local Polynomial Modelling and its Applications*. vol. 66 of *Monographs on Statistics and Applied Probability*. Chapman & Hall, London.

Golyandina N., Pepelyshev A. and Steland A. (2011) New approaches to nonparametric density estimation and selection of smoothing parameters. **56**(7), 2206–2218.

Härdle W. (1990) *Applied Nonparametric Regression*. vol. 19 of *Econometric Society Monographs*. Cambridge University Press, Cambridge.

Hull J.C. (2009) *Options, Futures, and Other Derivatives*. 7th edn. Pearson Prentice Hall.

Lai T.L. and Xing H. (2008) *Statistical Models and Methods for Financial Markets*. Springer Texts in Statistics. Springer, New York.

Lehmann E.L. and Romano J.P. (2005) *Testing Statistical Hypotheses*. Springer Texts in Statistics third edn. Springer, New York.

Pflug G.C. and Römisch W. (2007) *Modeling, Measuring and Managing Risk*. World Scientific Publishing Co. Pte. Ltd., Hackensack, NJ.

Pliska S. (1997) *Introduction to Mathematical Finance*. Blackwell Publishing, Oxford.

Savchuk O.Y., Hart J.D. and Sheather S.J. (2010) Indirect cross-validation for density estimation. *J. Amer. Statist. Assoc.* **105**(489), 415–423. With supplementary material available online.

Scott D.W. and Terrell G.R. (1987) Biased and unbiased cross-validation in density estimation. *J. Amer. Statist. Assoc.* **82**(400), 1131–1146.

Shao J. (2003) *Mathematical Statistics*. Springer Texts in Statistics 2nd edn. Springer-Verlag, New York.

Silverman B.W. (1986) *Density Estimation for Statistics and Data Analysis*. Monographs on Statistics and Applied Probability. Chapman & Hall, London.

Wand M.P. and Jones M.C. (1995) *Kernel Smoothing*. vol. 60 of *Monographs on Statistics and Applied Probability*. Chapman and Hall Ltd., London.