2

Statistical Methods

We introduce in this chapter further fundamental results from probability theory and statistics which are important in quantitative finance. They are highly relevant for the empirical analysis of financial data. In particular, limit theorems are presented and confidence intervals constructed. Furthermore, the log-returns of a world stock index will be estimated pointing at a stylized empirical fact.

2.1 Limit Theorems

In this section some fundamental limit theorems are summarized. These include the Law of Large Numbers and the Central Limit Theorem.

Law of Large Numbers

In Sect. 1.1 we mentioned the intuitive idea of defining probabilities as limits of relative frequencies determined from many independent repetitions of a given probabilistic experiment. This idea can be given some theoretical justification from an asymptotic analysis of sequences of *independent* and *identically distributed* (i.i.d.) random variables X_1, X_2, ... An example would be a sequence of daily log-returns. Let us assume for the moment that these random variables have the same distribution as some random variable X with finite second moments. We then write for their mean

$$\mu = E(X_n) \tag{2.1.1}$$

and for their variance

$$\sigma^2 = \mathrm{Var}(X_n), \tag{2.1.2}$$

$n \in \mathcal{N}$. Since the random variables $X_1, X_2, \ldots$ are independent it follows that the *sample mean*

E. Platen, D. Heath, *A Benchmark Approach to Quantitative Finance*,
Springer Finance,

$$\hat{\mu}_n = \frac{1}{n} \sum_{i=1}^{n} X_i \tag{2.1.3}$$

has the mean

$$E(\hat{\mu}_n) = \mu \tag{2.1.4}$$

and the variance

$$\mathrm{Var}(\hat{\mu}_n) = \frac{\sigma^2}{n}. \tag{2.1.5}$$

Note that one does not need for (2.1.3) the independence of the random variables. The *Law of Large Numbers* (LLN) is one of the fundamental results of probability theory and statistics. To formulate this law we say, that a sequence of random variables $Y_1, Y_2, \ldots$ *converges in the mean square sense* to a random variable Y if

$$\lim_{n\to\infty} E\left(|Y_n - Y|^2\right) = 0. \tag{2.1.6}$$

In this case we write

$$Y \overset{\text{m.s.}}{=} \lim_{n\to\infty} Y_n. \tag{2.1.7}$$

The mean square version of the LLN using this mode of convergence is stated by the following result.

Theorem 2.1.1. (Mean-square LLN) *If the independent random variables X_1, $X_2, \ldots$ have the same finite first and second moments, then the sample mean $\hat{\mu}_n$ converges in the mean square sense to the mean μ, that is*

$$\mu \overset{m.s.}{=} \lim_{n\to\infty} \hat{\mu}_n. \tag{2.1.8}$$

To see this we can write, using the independence property of $X_1, X_2, \ldots$ and equations (2.1.3), (2.1.1) and (2.1.5), the relation

$$\begin{aligned} E\left((\hat{\mu}_n - \mu)^2\right) &= E\left(\left(\frac{1}{n}\sum_{i=1}^{n}(X_i - \mu)\right)^2\right) = \frac{1}{n^2}\sum_{i=1}^{n} E\left((X_i - \mu)^2\right) \\ &= \frac{1}{n}\left(\frac{1}{n}\sum_{i=1}^{n}\sigma^2\right) = \frac{1}{n}\sigma^2. \end{aligned} \tag{2.1.9}$$

Using this formula we see by (2.1.6) and (2.1.7) that (2.1.8) is established.

There exists also a strong LLN, which goes back to Kolmogorov. Here the sample mean converges *almost surely* (a.s.) and we write

$$\mu \overset{\text{a.s.}}{=} \lim_{n\to\infty} \hat{\mu}_n \tag{2.1.10}$$

for

$$P\left(\lim_{n\to\infty} \hat{\mu}_n = \mu\right) = 1. \tag{2.1.11}$$

Theorem 2.1.2. (Strong LLN, Kolmogorov) *For a sequence of independent random variables $X_1, X_2, \ldots$ with mean μ and*

$$\sum_{i=1}^{\infty} \frac{\mathrm{Var}(X_i)}{i^2} < \infty \tag{2.1.12}$$

it holds that

$$\mu \overset{a.s.}{=} \lim_{n\to\infty} \hat{\mu}_n. \tag{2.1.13}$$

To underline that there are different types of convergence let us also state a weak LLN, which is due to Markov. For this purpose we say that a sequence of random variables $Y_1, Y_2, \ldots$ *converges in probability* to a random variable Y if for any $\varepsilon > 0$

$$\lim_{n\to\infty} P(|Y_n - Y| > \varepsilon) = 0 \tag{2.1.14}$$

and we write

$$Y \overset{P}{=} \lim_{n\to\infty} Y_n. \tag{2.1.15}$$

Theorem 2.1.3. (Weak LLN, Markov) *For a sequence of uncorrelated random variables $X_1, X_2, \ldots$ with mean $E(X_i) = \mu,\ i \in \mathcal{N}$, and*

$$\lim_{n\to\infty} \frac{1}{n^2} \sum_{i=1}^{n} \mathrm{Var}(X_i) = 0 \tag{2.1.16}$$

one has

$$\mu \overset{P}{=} \lim_{n\to\infty} \hat{\mu}_n. \tag{2.1.17}$$

In the case of an i.i.d. sequence of log-returns one can therefore estimate via the sample mean the mean of the log-returns. In Fig. 2.1.2 we plot the sample mean for the log-returns of the S&P500 as it evolves for an increasing number of observations.

The link between relative frequencies, as discussed in Sect. 1.1, and corresponding probabilities can now be directly established by using the weak LLN. If A is an event and $\frac{N(A)}{N}$ the relative frequency of A occurring in $N \in \mathcal{N}$ independent, identical observations of A, then

$$P(A) \overset{P}{=} \lim_{N\to\infty} \frac{N(A)}{N}. \tag{2.1.18}$$

This is a fundamental result, which supports our empirical analysis and stochastic modeling in finance.

Empirical Moments

We have shown by the weak LLN under appropriate conditions that the sample mean $\hat{\mu}_n$, which is also the *first empirical moment*, approaches the true mean μ of uncorrelated random variables $X_1, X_2, \ldots, X_n$ in probability for increasing n. The sample mean $\hat{\mu}_n$ is therefore a reasonable estimate for the mean μ. This provides a method for estimating the mean of a sequence of uncorrelated random variables.

Consider i.i.d. random variables $X_1, X_2, \ldots$ under the conditions $E(X_i^2) < \infty$, $i \in \mathcal{N}$, one can also show that the *sample variance*

$$\hat{\sigma}_n^2 = \frac{1}{n-1} \sum_{i=1}^{n} (X_i - \hat{\mu}_n)^2, \tag{2.1.19}$$

which we define as the *second empirical central moment*, converges almost surely to the variance σ^2 of the i.i.d. random variables X_1, X_2, ...

Similarly, under the conditions $E(X_i^3) < \infty$ and $\mathrm{Var}(X_i) > 0$, $i \in \mathcal{N}$, the *sample skewness*

$$\hat{\beta}_n = \frac{1}{n} \sum_{i=1}^{n} \left(\frac{X_i - \hat{\mu}_n}{\hat{\sigma}_n} \right)^3 \tag{2.1.20}$$

approaches almost surely the skewness β_X, see (1.3.22). The *sample kurtosis*

$$\hat{\kappa}_n = \frac{1}{n} \sum_{i=1}^{n} \left(\frac{X_i - \hat{\mu}_n}{\hat{\sigma}_n} \right)^4 \tag{2.1.21}$$

provides under the conditions $E(X_i^4) < \infty$ and $\mathrm{Var}(X) > 0$, $i \in \mathcal{N}$, an a.s. converging estimate for the kurtosis κ_X, see (1.3.30), for i.i.d. random variables X_1, X_2, ... To obtain useful estimates for these empirical moments one has, therefore, only to ensure that the corresponding moments are finite if one has i.i.d. observations.

Let us consider a simulated sequence of independent identically $N(-1, 1)$ Gaussian distributed random variables. These have by Table 1.3.1 mean $\mu = -1$, variance $\sigma^2 = 1$, skewness $\beta = 0$ and kurtosis $\kappa = 3$. Figure 2.1.1 displays linearly interpolated graphs for the resulting sample mean, sample variance, sample skewness and sample kurtosis for increasing values of the sample size $n \in \{10, 11, \ldots, 1000\}$.

As suggested by the weak LLN, for increasing sample sizes we see that the empirical moments appear to converge towards the respective values of the moments; in this case the mean $\mu = -1$, variance $\sigma^2 = 1$, skewness $\beta = 0$ and kurtosis $\kappa = 3$. One notes that for higher order moments one needs more observations to stabilize the corresponding empirical sample moment.

As another illustration, let us calculate the empirical moments from observations of daily log-returns of the S&P500 index covering the twenty year period from 1977 until 1997. For these S&P500 log-returns we obtain the empirical moments

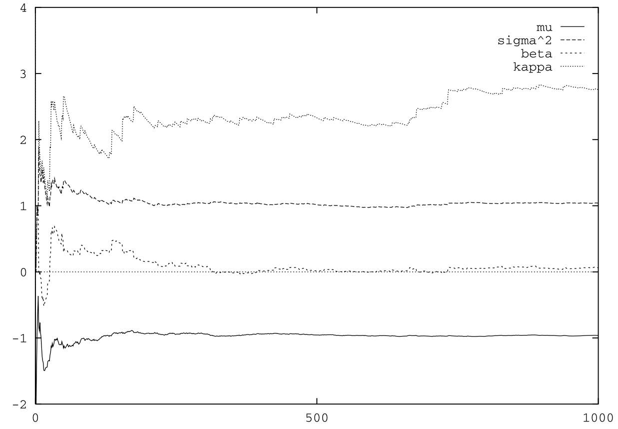

Fig. 2.1.1. Empirical moments from a simulation

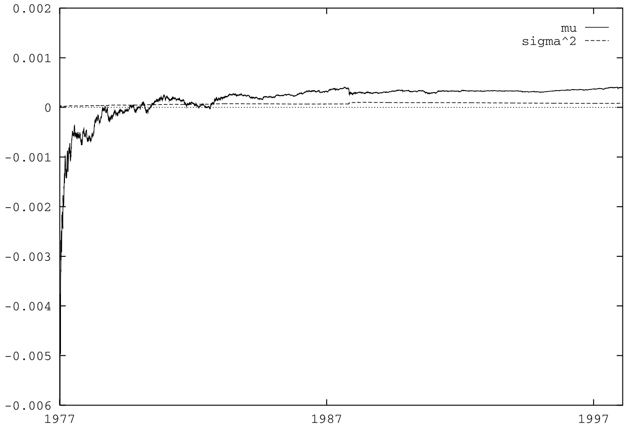

Fig. 2.1.2. Sample mean and variance for S&P500 log-returns

$$\hat{\mu}_n = 0.00040, \quad \hat{\sigma}_n^2 = 0.000082, \quad \hat{\beta}_n = -2.22, \quad \hat{\kappa}_n = 58.43 \tag{2.1.22}$$

for sample size $n = 5478$. In Fig. 2.1.2, we show the corresponding sample mean and sample variance as they evolve over time in dependence on time. These converge reasonably well towards the corresponding values shown in (2.1.22). We then display the resulting evolution of the sample skewness and sample kurtosis in Fig. 2.1.3. It is apparent that these are not very stable estimates. In particular, the values jump considerably at the 1987 stock market crash. If we remove from our sample the largest absolute log-return that occurred at the October 1987 market crash, then we obtain with the remaining $n = 5477$ observations the empirical moments

$$\hat{\mu}_n = 0.00044, \quad \hat{\sigma}_n^2 = 0.000074, \quad \hat{\beta}_n = -0.098, \quad \hat{\kappa}_n = 11.06. \tag{2.1.23}$$

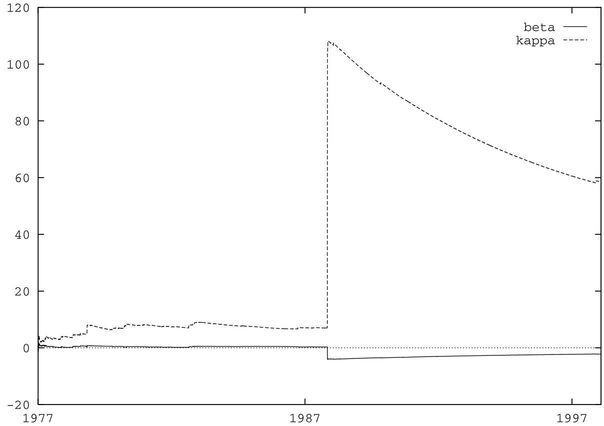

Fig. 2.1.3. Sample skewness and kurtosis for S&P500 log-returns

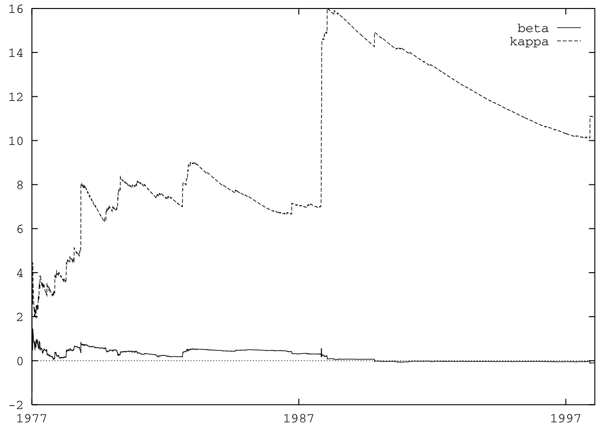

Fig. 2.1.4. Sample skewness and kurtosis for S&P500 log-returns without 1987 crash

This calculation shows that the estimated sample kurtosis changed dramatically after removing the most extreme log-return. In Fig. 2.1.4 we show the corresponding empirical skewness and kurtosis for the reduced sample. A comparison of Fig. 2.1.3 and Fig. 2.1.4 indicates that the fourth empirical moment is extremely sensitive with respect to this data set. The kurtosis κ might not even be finite for log-returns of the S&P500 index if fitted to a reasonable class of models. We shall show later in this chapter that typical parameter estimates of stock market index log-returns in the class of symmetric generalized distributions imply infinite kurtosis. For this reason, when estimating log-returns, it is recommended one uses a statistical approach that exploits the entire distribution and does not depend on any higher order empirical moments, such as the sample kurtosis.

Central Limit Theorem

To obtain more information regarding the asymptotics of the sample mean $\hat{\mu}_n$ one needs another fundamental result. To prepare its formulation we say that a sequence of random variables $X_1, X_2, \ldots$ *converges in distribution* to a random variable X if the distribution function $F_{X_n}(x)$ converges at each point x of continuity of $F_X(x)$, and we write

$$X \overset{\mathrm{d}}{=} \lim_{n\to\infty} X_n. \tag{2.1.24}$$

The Central Limit Theorem CLT states the following result.

Theorem 2.1.4. (CLT) *A standardized sample average*

$$\hat{Z}_n = \sqrt{n}\,\frac{(\hat{\mu}_n - \mu)}{\sigma}, \tag{2.1.25}$$

for a sequence X_1, X_2, ... of i.i.d. random variables with mean $\mu \in \Re$ and variance $\sigma^2 \in (0,\infty)$ converges in distribution, as $n \to \infty$, to a standard Gaussian random variable $Z \sim N(0,1)$, that is

$$\lim_{n\to\infty} \hat{Z}_n \overset{d}{=} Z. \tag{2.1.26}$$

This fundamental theorem states that

$$\lim_{n\to\infty} F_{\hat{Z}_n}(z) = F_Z(z) = N(z), \tag{2.1.27}$$

for all $z \in \Re$, where $N(\cdot)$ denotes the standard Gaussian distribution function, see (1.2.7). One can show that one only needs the existence of the third absolute moment $E(|X_i|^3) < \infty,\ i \in \mathcal{N}$, of the i.i.d. distributed random variables to achieve Gaussianity for the standardized sample average together with the *Berry-Esseen inequality*

$$\sup_{x\in\Re}\left|P\left(\hat{Z}_n < x\right) - N(x)\right| \le n^{-\frac{1}{2}}\,\frac{0.8}{\sigma^2}\,E\left(|X_1 - E(X_1)|^3\right). \tag{2.1.28}$$

As a consequence of the CLT the independence of the random variables $X_1, X_2, \ldots$ and the existence of second moments guarantee a Gaussian limit for $\hat{Z}_n$. This provides an explanation for the dominant role of the Gaussian distribution in probability and statistics and many areas of application including quantitative finance. For instance, one observes for most financial securities that for increasing periods of time the corresponding long term log-returns seem to approach Gaussian random variables. In view of the CLT this is not a surprising observation if one interprets short term log-returns as i.i.d. random variables.

Bernoulli Trials

Let us provide a simple illustration of the Law of Large Numbers and also the Central Limit Theorem. *Bernoulli trials* are independent repetitions of an experiment with two basic outcomes which might occur with probabilities p and $1-p$, respectively. If we set $X_n = 1$ for a positive log-return and $X_n = 0$ for a non-positive log-return, then we can model these log-returns by using an i.i.d. sequence of random variables $X_1, X_2, \ldots$ with mean

$$\mu = E(X_n) = p \tag{2.1.29}$$

and variance

$$\sigma^2 = \text{Var}(X_n) = p\,(1-p). \tag{2.1.30}$$

The sum

$$H_n = X_1 + X_2 + \cdots + X_n = n\,\hat{\mu}_n, \tag{2.1.31}$$

see (2.1.3), counts the number of positive log-returns occurring out of n observed trials. In a Bernoulli trial one is typically interested in the number of outcomes that correspond to a given specific event. Furthermore, $\hat{\mu}_n = \frac{H_n}{n}$ measures the relative frequency of observing such an event, in our example the occurrence of positive log-returns. The LLNs tell us, as $n \to \infty$, that the random variables $\hat{\mu}_n$ converge in a meaningful sense to the value p, which in our example is the probability of having a positive log-return for a single observation. Additionally, by the CLT we know that, as $n \to \infty$, the standardized sample mean

$$\hat{Z}_n = \sqrt{n}\,\frac{\left(\frac{H_n}{n} - p\right)}{\sqrt{p\,(1-p)}},$$

see (2.1.25)–(2.1.31), converges in distribution to a standard Gaussian random variable Z.

Binomial Distribution

We remark that the probability for the event $H_n = m$ is the same as the *binomial probability* for m successes out of n trials, that is

$$P(H_n = m) = p^m\,(1-p)^{n-m}\,\frac{n!}{(n-m)!\,m!}, \tag{2.1.32}$$

where we recall that $k! = 1 \cdot 2 \cdot \ldots \cdot k$ and $0! = 1$. Figure 2.1.5 shows for $p = 0.5$ and $n = 10$ the resulting binomial probabilities, when these are interpolated. These probabilities resemble the corresponding bell shaped Gaussian density function, as indicated by the CLT, which is also included in Fig. 2.1.5 for comparison. The Gaussian density is the curve with the slightly larger value at the mean. This means, for p asymptotically not vanishing for large n that the binomial probabilities tend asymptotically to the values of the Gaussian density.

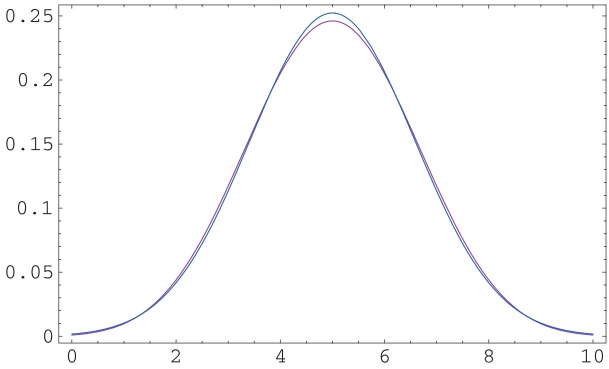

Fig. 2.1.5. Binomial probabilities for $p = 0.5$ and $n = 10$ and Gaussian density

Binomial probabilities appear in finance, for example, in random walk and binomial tree approximations of continuous time asset price models, as will be discussed later. However, binomial distributions are also linked to the Poisson distribution as the following statement shows. If n is large and p is small in (2.1.32), such that $\lambda = np > 0$, then it holds asymptotically for $n \to \infty$ that

$$P(H_n = m) \approx \exp\{-\lambda\} \frac{\lambda^m}{m!} \tag{2.1.33}$$

for $m \in \{0, 1, \ldots\}$. This means that for large n and $\lambda = np$ the binomial probabilities approach Poisson probabilities, see (1.1.30).

2.2 Confidence Intervals

To analyze empirically market data one needs sophisticated statistical tools. With the sample mean that we introduced in the previous section we have an estimate for the mean. We may ask, what is the number of observations needed to obtain a reasonably accurate estimate and how correct is this estimate? This question can be answered in different ways, as we shall see below.

Basic Confidence Intervals

Let us again consider the Bernoulli trials introduced previously. These are formed by a sequence of i.i.d. random variables $X_1, X_2, \ldots$ taking the value 1 with probability p and the value 0 with probability $1 - p$, where we now assume that p is unknown to us. We recall that $\mu = E(X_n) = p$ and $\sigma^2 = \text{Var}(X_n) = p\,(1 - p)$ and so it follows

$$E(\hat{\mu}_n) = p$$

and by (1.4.26) and (2.1.30)

$$\mathrm{Var}(\hat{\mu}_n) = \frac{p\,(1-p)}{n}.$$

As outlined previously, the sample mean $\hat{\mu}_n$ converges by the LLN, see (2.1.8), in a mean square sense to p. We can apply the Chebyshev inequality (1.3.58) to the random variable $\hat{\mu}_n - p$ and may then use the inequality $\sigma^2 = p\,(1-p) \leq \frac{1}{4}$ for $p \in [0,1]$ to obtain

$$P\left(|\hat{\mu}_n - p| \geq a\right) = P(|\hat{\mu}_n - \mu| \geq a) \leq \frac{\sigma^2}{n\,a^2} = \frac{p\,(1-p)}{n\,a^2} \leq \frac{1}{4\,n\,a^2} \tag{2.2.1}$$

for $a > 0$ and so

$$P\left(|\hat{\mu}_n - p| < a\right) = 1 - P\left(|\hat{\mu}_n - p| \geq a\right) \geq 1 - \frac{1}{4\,n\,a^2}. \tag{2.2.2}$$

Thus, for any $0 < \alpha < 1$ and $a > 0$ we can conclude that the unknown mean p lies in the interval $(\hat{\mu}_n - a,\, \hat{\mu}_n + a)$ with at least probability $1 - \alpha$ when

$$n \geq n(a, \alpha) = \frac{1}{4\,\alpha\,a^2}. \tag{2.2.3}$$

In statistical terminology we say that the hypothesis that p belongs to the interval $(\hat{\mu}_n - a,\, \hat{\mu}_n + a)$ is acceptable at a $100(1-\alpha)\%$ *level of confidence* if

$$P(|\hat{\mu}_n - p| < a) = 1 - \alpha \tag{2.2.4}$$

and call

$$(\hat{\mu}_n - a, \hat{\mu}_n + a)$$

the $100(1-\alpha)\%$ *confidence interval*. In our example, $(\hat{\mu}_n - 0.1,\, \hat{\mu}_n + 0.1)$ is at least a 95% confidence interval when $n \geq n(0.1, 0.05) = 500$. If n were fixed, then we would obtain from (2.2.3) in this example the inequality

$$a \geq \frac{1}{2\sqrt{\alpha\,n}}. \tag{2.2.5}$$

We note that the length of the confidence interval only decreases proportionally to $n^{-\frac{1}{2}}$ for increasing number of observations n. This is a general phenomenon, as we shall see later on.

Gaussian Confidence Interval and VaR

For a Gaussian random variable X with known mean μ and known variance σ^2 the $100(1-\alpha)\%$ confidence interval with

$$P\left(\left|\frac{X-\mu}{\sigma}\right| < p_{1-\frac{\alpha}{2}}\right) = 1 - \alpha \tag{2.2.6}$$

is given in the form

$$(\mu - \sigma\, p_{1-\frac{\alpha}{2}}, \mu + \sigma\, p_{1-\frac{\alpha}{2}}). \tag{2.2.7}$$

Here $p_{1-\alpha}$ is the $100(1-\alpha)\%$ quantile of the standard Gaussian distribution. For instance, a 99% confidence interval requires one to choose $p_{1-\alpha} \approx 2.58$.

The confidence interval (2.2.7) is a *two-sided confidence interval*. Sometimes however, in particular, in the computation of *Value at Risk* (VaR), one is interested in determining a critical maximum loss $\mathrm{VaR}((1-\alpha)\%)$ so that one can assert with $(1-\alpha)\%$ confidence that X is at least as large as $-\mathrm{VaR}((1-\alpha)\%)$, that is

$$P\left(X \geq -\mathrm{VaR}((1-\alpha)\%)\right) = 1-\alpha \tag{2.2.8}$$

or equivalently

$$P\left(\frac{X-\mu}{\sigma} < -z_\alpha\right) = \alpha, \tag{2.2.9}$$

with

$$z_\alpha = \frac{\mathrm{VaR}((1-\alpha)\%) + \mu}{\sigma} \tag{2.2.10}$$

Then (2.2.9) corresponds to the *one sided confidence interval*

$$(-\infty, -z_\alpha), \tag{2.2.11}$$

where the random variable $\frac{X-\mu}{\sigma}$ can be found with $\alpha\%$ probability in that interval. According to (2.2.8) and (2.2.10) X will thus be with $(1-\alpha)\%$ probability above the level

$$-\mathrm{VaR}((1-\alpha)\%) = -(z_\alpha\, \sigma - \mu). \tag{2.2.12}$$

For example, for a Gaussian random variable X and $\alpha = 0.01$ we have the $\alpha\%$ percentile with value $z_{0.01} \approx 2.35$, which determines according to (2.2.12) the corresponding maximum critical loss $\mathrm{VaR}(99\%) = z_{0.01}\, \sigma - \mu$.

Student *t* Confidence Interval

Often we do not know the variance σ^2 or do not have a reasonable estimate for it. However, in these cases we can use the sample variance $\hat{\sigma}_n^2$, see (2.1.19), to construct appropriate confidence intervals.

Let X_1, X_2, …, X_n denote n i.i.d. $N(\mu, \sigma^2)$ Gaussian random variables with known mean μ and unknown variance σ^2. As described in (2.1.3) and (2.1.19) we have the sample mean

$$\hat{\mu}_n = \frac{1}{n}\sum_{j=1}^{n} X_j \tag{2.2.13}$$

and sample variance

Table 2.2.1. Quantiles for the Student t distribution

n	10	20	30	40	60	100	200
$t_{0.9,n-1}$	1.83	1.73	1.70	1.68	1.67	1.66	1.65
$t_{0.99,n-1}$	3.25	2.86	2.76	2.70	2.66	2.62	2.58

$$\hat{\sigma}_n^2 = \frac{1}{n-1} \sum_{j=1}^{n} (X_j - \hat{\mu}_n)^2 . \tag{2.2.14}$$

Then for $n > 3$ it can be shown that the random variable

$$T_n = \frac{\hat{\mu}_n - \mu}{\sqrt{\frac{\hat{\sigma}_n^2}{n}}} \tag{2.2.15}$$

is Student t distributed, see (1.2.17), with $n-1$ degrees of freedom, that is $T_n \sim t(n-1)$.

We have

$$P\left(|\hat{\mu}_n - \mu| < a\right) = P\left(|T_n| < t\right) = 2\left(F_{T_n}(t) - 0.5\right), \tag{2.2.16}$$

where

$$t = a \sqrt{\frac{n}{\hat{\sigma}_n^2}} \tag{2.2.17}$$

and $F_{T_n}(x)$ is the value of the Student t distribution function with $n-1$ degrees of freedom evaluated at $x \in \Re$.

Thus, for a given $100\alpha\%$ confidence level, we can check whether or not the test variable

$$T_n^0 = \frac{\hat{\mu}_n - \mu_0}{\sqrt{\frac{\hat{\sigma}_n^2}{n}}}$$

with hypothesized mean μ_0 satisfies the inequality

$$|T_n^0| < t_{1-\alpha,n-1}. \tag{2.2.18}$$

Here $t_{1-\alpha,n-1}$ is the $100\,(1-\alpha)\%$ *quantile* of the Student t distribution with $n-1$ degrees of freedom for which

$$2\left(F_{T_n^0}(t_{1-\alpha,n-1}) - 0.5\right) = 1 - \alpha.$$

Some values of the quantile $t_{1-\alpha,n-1}$ are given in Table 2.2.1. If the relation (2.2.18) is not fulfilled, then we reject the *null hypothesis* H_0 that $\mu = \mu_0$. Otherwise, we accept it on the basis of this test. In addition, we can form the corresponding $100(1-\alpha)\%$ confidence interval

$$(\hat{\mu}_n - a, \hat{\mu}_n + a),$$

see (2.2.17), with

$$a = t_{1-\alpha,n-1}\sqrt{\frac{\hat{\sigma}_n^2}{n}}. \tag{2.2.19}$$

We call this interval the *Student t confidence interval.* It contains all of the values μ_0 for which the null hypothesis would not be rejected by this test. Applications of this result can be found, for instance, in the treatment of errors in Monte Carlo simulations.

We note also that the length of the above confidence interval is proportional to $n^{-\frac{1}{2}}$. This means, to obtain a ten times smaller confidence interval requires approximately hundred times as many observations. We face this phenomenon in the application of Monte Carlo methods.

The above described procedure requires $X_1, X_2, \ldots$ to be Gaussian. When this is not the case we note by the CLT that sample means of sufficiently large groups or batches of these i.i.d. random variables will be approximately Gaussian. Consequently, with such a construction we can approximately apply the above methodology to these sample means. To be more precise we take n batches of m i.i.d. random variables $X_1^{(j)}$, $X_2^{(j)}$, …, $X_m^{(j)}$ for $j \in \{1, 2, \ldots, n\}$. Then we form the sample means

$$\hat{\mu}_m^{(j)} = \frac{1}{m}\sum_{\ell=1}^{m} X_\ell^{(j)}$$

and apply the above Student t methodology to these sample means rather than the original random variables $X_i^{(j)}$. For practical applications it has often been found that the batches should consist of at least 15 random variables to provide a reasonable approximation.

In the case when the variance σ^2 is known the following test variable

$$\bar{T}_n = \sqrt{n}\,\frac{(\hat{\mu}_n - \mu)}{\sigma}$$

is Gaussian and we can construct similar confidence intervals as above, but based on the Gaussian distribution rather than the Student t distribution. Recall also that the Student t distribution asymptotically approaches a Gaussian distribution as the degrees of freedom tend to infinity.

VaR Analysis for Student t Log-Returns (*)

As outlined in regulatory recommendations the modeling of, so-called, *event risk* is of increasing importance in VaR analysis.

In the following version of a Student t log-return model we exploit the fact that symmetric generalized hyperbolic distributions admit a representation as a mixture of normal distributions. This means, if one chooses the variance of a conditionally Gaussian distribution as the inverse of a Gamma distributed random variable, then the resulting distribution is a Student t distribution, see also Sect. 1.2. For simplicity, we neglect here the impact of any asymmetry since for the short time intervals considered this is not relevant. We shall provide later in this chapter more details on *normal variance mixture models*.

To generate Student t distributed log-returns $Z^{(1)}$, ..., $Z^{(d)}$ for d securities at a fixed time we set

$$Z^{(k)} = \sqrt{\tau}\, Y^{(k)} \tag{2.2.20}$$

for $k \in \{1, 2, \ldots, d\}$, where τ denotes the conditional variance with

$$\tau = \left(1 - \frac{2}{n}\right) \left(\frac{1}{n} \sum_{\ell=1}^{n} \left(\psi^{(\ell)}\right)^2\right)^{-1} \tag{2.2.21}$$

and $n \in \{3, 4, \ldots\}$. Additionally to the independent standard Gaussian distributed random variables $Y^{(k)}$ that appear in (2.2.20) we employ further independent standard Gaussian random variables $\psi^{(\ell)}$. Hence, the random variable τ is chi-square distributed with n degrees of freedom, see Sect. 1.2. Consequently, the random variables $Z^{(k)}$, $k \in \{1, 2, \ldots, d\}$, are Student t distributed with unit variance and n degrees of freedom. The conditional variance τ can be interpreted as a measure of the random activity of the market during the time period of interest. Note that the conditional variance converges to one as the degrees of freedom n tend to infinity, which yields asymptotically normal log-returns.

In addition to the typical parameters of the lognormal model we have used here only the extra parameter n, which is sufficient to characterize the leptokurtosis of the Student t distribution. As will be shown later, a typical parameter choice for n is about four. Smaller degrees of freedom generate log-returns with more extreme movements.

An important feature of the resulting multivariate Student t distribution for log-returns is its copula, see Sect. 1.5. It realistically captures the estimated dependence of extreme asset price movements, as shown in Embrechts, McNeil & Straumann (2002) and McNeil, Frey & Embrechts (2005). Let $\boldsymbol{Y} = (Y^{(1)}, \ldots, Y^{(d)})^\top$ denote a vector of independent standard Gaussian distributed random variables and τ be an independent chi-square random variable. Note that the joint distribution of the random vector $\boldsymbol{X} = (X^{(1)}, \ldots, X^{(d)})^\top$ with

$$\boldsymbol{X} = \sqrt{\tau}\, \boldsymbol{D}\, \boldsymbol{Y}$$

is a multivariate Student t distribution with n degrees of freedom. Here $\boldsymbol{D}$ is the Cholesky decomposition of the covariance matrix $\mathrm{Cov}_{\boldsymbol{X}}$ of $\boldsymbol{X}$, see Sect. 1.4. Since $\boldsymbol{Y}$ is Gaussian and $\frac{1}{\tau}$ is independent chi-square distributed, the resulting multivariate Student t distribution of $\boldsymbol{X}$ belongs to the class of *elliptic distributions*. One can show that the calculation of VaR numbers is for this class of distributions analytically tractable, see Platen & Stahl (2003). More precisely, a theorem in Fang, Kotz & Ng (1990) yields the representation

$$\boldsymbol{a}^\top \boldsymbol{X} = |\boldsymbol{a}^\top \boldsymbol{D}|\, \zeta \tag{2.2.22}$$

for any given weight vector $\boldsymbol{a} = (a^1, a^2, \ldots, a^d)^\top$, where $|\cdot|$ is the Euclidean norm, $\boldsymbol{a} \in \Re^d$, $\boldsymbol{D}^\top \boldsymbol{D} = \mathrm{Cov}_{\boldsymbol{X}}$ and ζ denotes a Student t distributed scalar random variable with n degrees of freedom. The representation (2.2.22) significantly simplifies the VaR calculation for portfolios even if these have an extremely large number of constituents.

Since the multivariate Student t distribution is an elliptical distribution, it follows from Embrechts et al. (2002), that VaR is in this case a, so-called, *coherent risk measure*, see Artzner, Delbaen, Eber & Heath (1997). This fact is highly important for the consistent use of VaR as a risk measure for internal capital allocation to particular business lines. The property of coherent risk measures that sometimes creates problems for VaR is the additivity, where the risk measure for the sum of two risky securities should never be greater than the sum of their risk measures, see Föllmer & Schiedt (2002).

In order to calculate VaR for the given short term horizon we apply, the so-called, square root time rule, which is in line with regulatory recommendations. From (2.2.22) we obtain then the following formula for the VaR number of a given portfolio at the given time:

$$\mathrm{VaR}_h(V, \alpha) \approx V \sqrt{\boldsymbol{a}^\top \mathrm{Cov}_{\boldsymbol{X}}\, \boldsymbol{a}}\, \sqrt{h\, \varDelta}\; \tilde{t}_\alpha(n). \tag{2.2.23}$$

Here $\sqrt{\boldsymbol{a}^\top \mathrm{Cov}_{\boldsymbol{X}} \boldsymbol{a}}$ characterizes the total volatility of the portfolio, V denotes the market value of the portfolio at the given time, $\varDelta$ is the time step size for a trading day, h the number of trading days and $\tilde{t}_{1-\alpha}(n)$ the $100(1-\alpha)\%$-quantile of the Student t distribution with n degrees of freedom.

The product (2.2.23) generalizes a short hand formula, used in practice, to calculate VaR by including the *event factor*

$$\varphi = \frac{\tilde{t}_{1-\alpha}(n)}{p_{1-\alpha}}, \tag{2.2.24}$$

that is

$$\mathrm{VaR}_h(V, \alpha) \approx V \sqrt{\boldsymbol{a}^\top \mathrm{Cov}_{\boldsymbol{X}}\, \boldsymbol{a}}\, \sqrt{h\, \varDelta}\; p_{1-\alpha}\, \varphi. \tag{2.2.25}$$

Here $p_{1-\alpha}$ is the $100(1-\alpha)\%$-quantile of the standard Gaussian distribution. Consequently, the event factor φ adjusts the standard VaR formula to a level that captures the, so-called, event risk when one uses Student t log-returns. According to the quantiles of the Gaussian and Student t distribution one obtains by (2.2.23) the event factors shown in Table 2.2.2. Even for rather small degrees of freedom, say $n \approx 2$, the additional regulatory capital will not surpass 16%.

Table 2.2.2. Event factor φ in dependence on degrees of freedom n

n	∞	10	5	4	3	2
φ	1	1.06	1.11	1.12	1.14	1.16

Gibson (2001) performed an extensive study using an extremely large set of representative portfolios of US institutions, where he identified empirically an event factor of about $\hat{\varphi} \approx 1.12$. One notes that this is exactly the value of the event factor that matches in Table 2.2.2 the one for the degrees of freedom $n = 4$. We shall see later that this finding supports a model proposed in Platen (2002), the minimal market model, which will be derived later in Chap. 13. Also our inference later in this chapter will suggest a Student t distribution with four degrees of freedom as a realistic estimate for the log-return distribution of indices.

2.3 Estimation Methods

There exists a wide range of estimation techniques developed for various inference problems that have major importance in quantitative finance. The choice of a suitable estimation method depends on the available data and the assumed model. In this section we concentrate mainly on linear models.

Estimators

Assume that there are $n \in \mathcal{N}$ real valued observations $R_{t_1}, R_{t_2}, \ldots, R_{t_n} \in \Re$ of, say, log-returns. These observations contain information about the parameters $\theta_1, \theta_2, \ldots, \theta_q \in \Re$ that we wish to estimate. The observations can be represented as *observation vector* $\boldsymbol{R} = (R_{t_1}, R_{t_2}, \ldots, R_{t_n})^\top \in \Re^n$ and the parameters as *parameter vector* $\boldsymbol{\theta} = (\theta_1, \theta_2, \ldots, \theta_q)^\top \in \Theta \subseteq \Re^q$, where Θ specifies the set of allowable values for the parameters, $q \in \mathcal{N}$.

Generally, an *estimator* $\hat{\boldsymbol{\theta}} = (\hat{\theta}_1, \hat{\theta}_2, \ldots, \hat{\theta}_q)^\top \in \Theta$ is a function $\hat{\boldsymbol{\theta}} : \Re^n \to \Theta$ by which the parameters can be approximately identified from the observations, that is, by the *estimate*

$$\hat{\boldsymbol{\theta}} = \hat{\boldsymbol{\theta}}(R_{t_1}, R_{t_2}, \ldots, R_{t_n}). \tag{2.3.1}$$

For example, as discussed previously, two typical parameters that are often needed are the mean $\theta_1 = E(X)$ and the variance $\theta_2 = E((X-\mu)^2)$ of a Gaussian random variable $X \sim N(\theta_1, \theta_2)$, say Gaussian daily log-returns. Given an observation vector $\boldsymbol{R} = (R_{t_1}, R_{t_2}, \ldots, R_{t_n})^\top$ with components that consist of i.i.d. observations of X, these parameters can be estimated. According to (2.2.13) the sample mean

$$\hat{\theta}_1 = \frac{1}{n} \sum_{j=1}^{n} R_{t_j} \tag{2.3.2}$$

estimates the value of θ_1. By (2.2.14) the sample variance

$$\hat{\theta}_2 = \frac{1}{n-1} \sum_{j=1}^{n} (R_{t_j} - \hat{\theta}_1)^2 \tag{2.3.3}$$

provides an estimator for θ_2. In this example we have $q = 2$ and $\hat{\boldsymbol{\theta}} = (\hat{\theta}_1, \hat{\theta}_2)^\top \in \Re^2$. For Gaussian daily log-returns, $\hat{\theta}_1$ would be the estimate of the expected daily growth rate and $\hat{\theta}_2$ the estimate of the variance of the daily log-returns.

Unbiasedness and Consistency

The assessment of the quality of an estimate $\hat{\theta}_i$, $i \in \{1, 2, \dots, q\}$, of the ith parameter θ_i can be based on the *estimation error*

$$\tilde{\theta}_i(n) = \theta_i - \hat{\theta}_i. \tag{2.3.4}$$

Ideally, the estimation error should be close to zero almost surely. However, this is difficult to achieve for a finite set of observations and a general model. Therefore, less stringent requirements are typically used.

The first requirement, which is often formulated, is that the expected value of the estimation error should be zero. That is, by taking expectations we obtain from (2.3.4) the condition

$$0 = E(\tilde{\theta}_i(n)) = E(\theta_i) - E(\hat{\theta}_i) = \theta_i - E(\hat{\theta}_i) \tag{2.3.5}$$

for $i \in \{1, 2, \dots, q\}$. Estimators that satisfy relation (2.3.5) are called *unbiased* and it follows by (2.3.5) that

$$E(\hat{\theta}_i) = \theta_i \tag{2.3.6}$$

for $i \in \{1, 2, \dots, q\}$.

In the case when the estimator $\hat{\theta}_i$ does not meet the unbiasedness condition (2.3.5), then $\hat{\theta}_i$ is said to be *biased*. The *bias* $E(\tilde{\theta}_i(n))$ is defined as the expected value of the estimation error (2.3.4). If the bias tends to zero as the number n of observations increases, then the estimator is called *asymptotically unbiased*, that is

$$\lim_{n \to \infty} E(\tilde{\theta}_i(n)) = 0, \tag{2.3.7}$$

$i \in \{1, 2, \dots, q\}$. A reasonable requirement for an unbiased estimator $\hat{\theta}_i$ is that its estimation error should, for increasing number n of observations, converge in probability to zero, that is

$$\lim_{n \to \infty} \tilde{\theta}_i(n) \stackrel{P}{=} 0 \tag{2.3.8}$$

for $i \in \{1, 2, \dots, q\}$. An estimator satisfying the property (2.3.8) is called *consistent*.

In our previous example, where $X \sim N(\theta_1, \theta_2)$ is a Gaussian random variable and $R_{t_1}, R_{t_2}, \dots$ are independent observations of X, the expected value of the sample mean $\hat{\theta}_1$ given in (2.3.2) is

$$E(\hat{\theta}_1) = \frac{1}{n}\sum_{j=1}^{n} E(R_{t_j}) = \frac{1}{n}\, n\,\theta_1 = \theta_1. \tag{2.3.9}$$

Consequently, in this case the sample mean $\hat{\theta}_1$ is an unbiased estimator of the mean θ_1. We obtain for the estimation error $\tilde{\theta}_1(n)$ the variance

$$E(\tilde{\theta}_1(n)^2) = E\left(\left(\frac{1}{n}\sum_{j=1}^{n}(R_{t_j} - \theta_1)\right)^2\right) = \frac{1}{n^2}\sum_{j=1}^{n} E\left((R_{t_j} - \theta_1)^2\right) = \frac{1}{n}\,\theta_2. \tag{2.3.10}$$

The variance in (2.3.10) converges to zero as $n \to \infty$. This implies by the Chebyshev inequality (1.3.58) and equation (2.3.10) that $\tilde{\theta}_1(n)$ converges in probability to zero, that is

$$P(\tilde{\theta}_1(n) > \varepsilon) \leq \frac{1}{\varepsilon^2}\, E\left(\tilde{\theta}_1(n)^2\right) = \frac{\theta_2}{n\,\varepsilon^2}. \tag{2.3.11}$$

Therefore, the estimator $\hat{\theta}_1$ is by (2.3.8) consistent.

In (2.1.22) we have estimated for the S&P500 daily log-returns with mean $\theta_1 \approx 0.0004$ and variance $\theta_2 \approx 0.00008$. By (2.3.11) to obtain with about only a probability of $\frac{\theta_2}{n\,\varepsilon^2} \approx 0.1$ an estimation error $\tilde{\theta}_1(n) > \varepsilon \approx 0.1\theta_1$ one needs more than $n \approx \frac{\theta_2}{0.1\,\varepsilon^2} \approx 500,000$ observations. This is an enormous number of daily observations that is needed to get any rough idea about the daily expected growth of an underlying security, as the S&P500. It requires far more data than market history offers. Therefore, without the availability of any extra structure it is highly unrealistic to expect any reliably estimates of trend, drift or growth parameters for financial securities. In particular, it is unlikely that with the available data one can estimate equity risk premia realistically.

Efficiency

One calls an estimator, which yields the lowest variance estimate an *efficient estimator*. It uses optimally, in a least-square sense, the information contained in the observations. Therefore, a useful measure of the quality of an estimator with estimation error vector $\tilde{\boldsymbol{\theta}}(n) = (\tilde{\theta}_1(n), \ldots, \tilde{\theta}_q(n))^\top$ is given by the *error covariance matrix*

$$\mathrm{Cov}_{\tilde{\boldsymbol{\theta}}(n)} = E\left(\tilde{\theta}(n)\,\tilde{\theta}(n)^\top\right). \tag{2.3.12}$$

This matrix measures the errors of individual estimators also in relation to each other. One obtains a scalar error measure for the ith estimator by considering the ith diagonal element in (2.3.12), which is the ith *mean-square error*

$$\mathrm{Cov}^{i,i}_{\tilde{\boldsymbol{\theta}}(n)} = E\left((\tilde{\theta}_i(n))^2\right), \tag{2.3.13}$$

$i \in \{1, 2, \ldots, q\}$. An overall scalar error measure is obtained by summing up all the diagonal elements of $\mathrm{Cov}_{\tilde{\boldsymbol{\theta}}(n)}$, which yields the *mean-square error*

$$\mathrm{Mse}_{\tilde{\boldsymbol{\theta}}(n)} = E\left(\tilde{\boldsymbol{\theta}}(n)^\top \, \tilde{\boldsymbol{\theta}}(n)\right). \tag{2.3.14}$$

One calls a symmetric $q \times q$ matrix $\boldsymbol{B}$ *positive definite* if

$$\boldsymbol{a}^\top \boldsymbol{B} \, \boldsymbol{a} > 0 \tag{2.3.15}$$

for all q-vectors $\boldsymbol{a}$. A symmetric matrix $\boldsymbol{C}$ is said to be *smaller* than another symmetric matrix $\boldsymbol{A}$, or $\boldsymbol{C} < \boldsymbol{A}$, if the matrix $\boldsymbol{A} - \boldsymbol{C}$ is positive definite. This allows us to state that an estimator $\hat{\boldsymbol{\theta}}$, which provides the smallest error covariance matrix among all unbiased estimators, is the best estimator in the mean-square sense. Such an estimator is called an *efficient estimator*.

Fisher Information

It can be shown that there exists a lower bound for the error covariance matrix given in (2.3.12), see Mendel (1995). This bound involves the *Fisher information matrix* $\boldsymbol{J} = [J^{i,j}]_{i,j=1}^q$, where

$$J^{i,j} = E\left(\frac{\partial}{\partial \theta_i} \ln(F_{\boldsymbol{X}}(\boldsymbol{R})) \, \frac{\partial}{\partial \theta_j} \ln(F_{\boldsymbol{X}}(\boldsymbol{R}))\right) \tag{2.3.16}$$

and $F_{\boldsymbol{X}}(\boldsymbol{R})$ is the joint distribution of the vector $\boldsymbol{X} = (X_1, X_2, \ldots, X_n)^\top$ with the given parameter vector $\boldsymbol{\theta} \in \Theta$, when taken under the information given by the observations $\boldsymbol{R} = (R_{t_1}, R_{t_2}, \ldots, R_{t_n})^\top$. The term $\frac{\partial}{\partial \theta_i} \ln(F_{\boldsymbol{X}}(\boldsymbol{R}))$ is the partial derivative with respect to the parameter θ_i of the natural logarithm of the joint distribution $F_{\boldsymbol{X}}(\cdot)$ of the observed quantities $\boldsymbol{R}$. We assume that the partial derivatives exist and are absolutely integrable. If $\hat{\boldsymbol{\theta}}$ is any unbiased estimator of $\boldsymbol{\theta}$, then the error covariance matrix is bounded from below by the inverse of the Fisher information matrix $\boldsymbol{J}$, that is

$$\mathrm{Cov}_{\tilde{\boldsymbol{\theta}}(n)} \geq \boldsymbol{J}^{-1}. \tag{2.3.17}$$

The lower bound is called the *Cramér-Rao lower bound* and provides a useful measure for testing the efficiency of specific estimation methods.

In the context of estimating trend, drift or growth parameters in log-returns of financial securities it tells us that there is an objective lower bound for the error covariance matrix. As already indicated previously, this bound is so high that there are not enough data available to accurately estimate trend and growth parameters in security prices.

Method of Moments

A rather obvious and simple estimation method is the *method of moments*, see Hansen (1982) or Cochrane (2001). It often leads to computationally simple estimators and is intuitively satisfying. However, it has also some weaknesses, in particular, when moments that are involved in the derivation of the estimators are not certain to exist in reality. For instance, as we shall see later, it seems that empirical studies on log-return data indicate that the existence of the fourth moment could be questionable, see also Dacorogna, Müller, Pictet & De Vries (2001).

Assume that there are n i.i.d. observations $R_{t_1}, R_{t_2}, \ldots, R_{t_n} \in \Re$ that have the probability distribution function $F_{R,\boldsymbol{\theta}}$, which depends on the parameter vector $\boldsymbol{\theta} = (\theta_1, \theta_2, \ldots, \theta_q)^\top \in \Re^q$ for $q \in \mathcal{N}$. We know from Sect. 1.3 that the jth moment $m_j = m_j(\theta_1, \theta_2, \ldots, \theta_q)$ of the random variable R is obtained by the integral

$$m_j(\theta_1, \theta_2, \ldots, \theta_q) = E((R)^j) = \int_{-\infty}^{\infty} (r)^j \, dF_{R,\boldsymbol{\theta}}(r) \tag{2.3.18}$$

for $j \in \mathcal{N}$ as long as $m_j = m_j(\theta_1, \theta_2, \ldots, \theta_q) < \infty$. It is obvious that the moments m_j, $j \in \mathcal{N}$, are functions of the parameters $\theta_1, \theta_2, \ldots, \theta_q$.

By application of the strong Law of Large Numbers, see Sect. 2.1, one can estimate under appropriate assumptions the respective moments using the given observations $R_{t_1}, R_{t_2}, \ldots, R_{t_n}$. Let us denote by $\hat{m}_j$ the jth sample moment

$$\hat{m}_j = \frac{1}{n} \sum_{i=1}^{n} (R_{t_i})^j \tag{2.3.19}$$

for $j \in \mathcal{N}$. Typically, q equations for the first q moments are sufficient for identifying estimators for the q unknown parameters $\theta_1, \theta_2, \ldots, \theta_q$. The basic idea for the method of moments is therefore to equate the theoretical moments m_j with corresponding sample moments $\hat{m}_j$, that is

$$m_j(\hat{\theta}_1, \hat{\theta}_2, \ldots, \hat{\theta}_q) = \hat{m}_j \tag{2.3.20}$$

for $j \in \{1, 2, \ldots, q\}$, with $\hat{\theta}_1, \hat{\theta}_2, \ldots, \hat{\theta}_q$ denoting the resulting parameter estimators.

If the system of equations (2.3.20) has a solution that is acceptable, then we call $\hat{\theta}_1, \hat{\theta}_2, \ldots, \hat{\theta}_q$ the *method of moments estimators*.

Sometimes, it is recommended to use the jth central theoretical moments

$$\mu_j = \mu_j(\theta_1, \theta_2, \ldots, \theta_q) = E\left((R - m_1)^j\right) \tag{2.3.21}$$

and the respective jth *central sample moments*

$$\hat{\mu}_j = \frac{1}{n-1} \sum_{i=1}^{n} (R_{t_i} - m_1)^j \tag{2.3.22}$$

to form the q equations

$$\mu_j(\theta_1, \theta_2, \ldots, \theta_q) = \hat{\mu}_j \tag{2.3.23}$$

for $j \in \{1, 2, \ldots, q\}$. By solving the system of equations (2.3.23) one may obtain slightly different method of moment estimators $\hat{\theta}_1, \hat{\theta}_2, \ldots, \hat{\theta}_q$.

The theoretical justification for the method of moments relies on the fact that under appropriate assumptions the sample moments are consistent estimators of the respective theoretical moments, see (2.3.8). It is well-known that the method of moments is sometimes not very efficient. Furthermore, the unbiasedness of the method of moment estimators cannot be easily guaranteed.

Linear Least-Squares Estimation

The following well-known estimation method does, in principle, not require any information about the structure of the underlying distribution function of the observations. It uses only first and second order moments. In its basic form the *least-squares estimation method* assumes that the n-dimensional observation vector $\boldsymbol{R} = (R_{t_1}, R_{t_2}, \ldots, R_{t_n})^\top \in \Re^n$ satisfies the following linear model

$$\boldsymbol{R} = \boldsymbol{B}\,\boldsymbol{\theta} + \boldsymbol{\varepsilon}. \tag{2.3.24}$$

Here $\boldsymbol{B} = [b^{i,j}]_{i,j=1}^{n,q}$ is the (n,q)-*observation matrix*, $\boldsymbol{\varepsilon} = (\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_n)^\top \in \Re^n$ is the n-dimensional *observation error vector* and $\boldsymbol{\theta} = (\theta_1, \theta_2, \ldots, \theta_q)^\top \in \Theta$ denotes the q-dimensional parameter vector, $n, q \in \mathcal{N}$ with $q < n$. The observation matrix $\boldsymbol{B}$ is assumed to be known and to be of maximum rank q.

Note that for $\boldsymbol{\varepsilon} = (0, 0, \ldots, 0)^\top$ equation (2.3.24) has no solution. However, in reality the observation errors ε are random and unknown. Therefore, the best that one can achieve is to find an estimator $\hat{\boldsymbol{\theta}}$ that minimizes in a reasonable sense the effect of the observation errors. From a mathematical viewpoint it is convenient to use a *least-squares criterion* of the form

$$U_{\mathrm{LS}}(\boldsymbol{\theta}) = \frac{1}{2}\,\boldsymbol{\varepsilon}^\top\,\boldsymbol{\varepsilon} = \frac{1}{2}\,(\boldsymbol{R} - \boldsymbol{B}\,\boldsymbol{\theta})^\top\,(\boldsymbol{R} - \boldsymbol{B}\,\boldsymbol{\theta}). \tag{2.3.25}$$

One notes that no expectation is taken in the least-squares criterion (2.3.25). This criterion simply minimizes the observation error $\boldsymbol{\varepsilon}$. It does not directly minimize the absolute value of the estimation error $\boldsymbol{\theta} - \hat{\boldsymbol{\theta}}$.

Minimizing the quadratic form (2.3.25) with respect to the unknown parameter vector $\boldsymbol{\theta}$ yields by the corresponding first order conditions the, so-called, *normal equations* in the form

$$(\boldsymbol{B}^\top\,\boldsymbol{B})\,\hat{\boldsymbol{\theta}} = \boldsymbol{B}^\top\,\boldsymbol{R}. \tag{2.3.26}$$

This allows one to determine the least-squares estimator $\hat{\boldsymbol{\theta}}$, since we assumed that the matrix $\boldsymbol{B}$ has maximum rank $q < n$. Then we obtain

$$\hat{\boldsymbol{\theta}} = \boldsymbol{B}^{+} \boldsymbol{R} \tag{2.3.27}$$

with

$$\boldsymbol{B}^{+} = (\boldsymbol{B}^{\top} \boldsymbol{B})^{-1} \boldsymbol{B}^{\top}. \tag{2.3.28}$$

Statistically one can analyze the least-squares estimator by assuming that the measurement errors have zero mean, that is

$$E(\varepsilon_{t_j}) = 0 \tag{2.3.29}$$

for all $i \in \{1, 2, \ldots, n\}$. Obviously, the least-squares estimator is unbiased, since by (2.3.27) and (2.3.24) we obtain

$$E(\hat{\boldsymbol{\theta}}) = \boldsymbol{B}^{+} (\boldsymbol{B}\, \boldsymbol{\theta} + E(\boldsymbol{\varepsilon})) = \boldsymbol{\theta}. \tag{2.3.30}$$

Of great interest is the covariance matrix of the observation error vector, which has the form

$$\mathrm{Cov}(\boldsymbol{\varepsilon}) = E(\boldsymbol{\varepsilon}\, \boldsymbol{\varepsilon}^{\top}). \tag{2.3.31}$$

If this matrix is known, then one can use the relation

$$\hat{\boldsymbol{\theta}} - \boldsymbol{\theta} = \boldsymbol{B}^{+} \boldsymbol{R} - \boldsymbol{B}^{+} \boldsymbol{B}\, \boldsymbol{\theta} = \boldsymbol{B}^{+} (\boldsymbol{R} - \boldsymbol{B}\, \boldsymbol{\theta}) = \boldsymbol{B}^{+} \boldsymbol{\varepsilon}$$

to calculate the covariance matrix

$$\begin{aligned} \mathrm{Cov}(\boldsymbol{\theta} - \hat{\boldsymbol{\theta}}) &= E\left((\boldsymbol{\theta} - \hat{\boldsymbol{\theta}})\, (\boldsymbol{\theta} - \hat{\boldsymbol{\theta}})^{\top} \right) = E\left(\boldsymbol{B}^{+} \boldsymbol{\varepsilon}\, (\boldsymbol{B}^{+} \boldsymbol{\varepsilon})^{\top} \right) \\ &= \boldsymbol{B}^{+} E\left(\boldsymbol{\varepsilon}\, \boldsymbol{\varepsilon}^{\top} \right) \, (\boldsymbol{B}^{+})^{\top} = \boldsymbol{B}^{+} \, \mathrm{Cov}(\boldsymbol{\varepsilon})\, (\boldsymbol{B}^{+})^{\top} \end{aligned} \tag{2.3.32}$$

of the estimation error. One notes that this covariance matrix depends on the second moments of the observation errors and the observation matrix.

Curve Fitting

The linear least-squares estimation method is widely used, for instance, in linear regression analysis, that is, linear curve fitting. Linear and nonlinear curve fitting are common tasks in quantitative finance. Let us fit to given observations $R_{t_1}, R_{t_2}, \ldots, R_{t_n}$ the model

$$R_{t_k} = \sum_{i=1}^{q} \theta_i \, \phi_i(t_k) + \varepsilon_k \tag{2.3.33}$$

for $t_k \in \{t_1, t_2, \ldots, t_n\}$. Here $\phi_i : [0, \infty) \to \Re$ is the given ith basis function with $i \in \{1, 2, \ldots, q\}$, which can be a nonlinear function of the variable $t \in [0, \infty)$. The unknown parameter vector is again $\boldsymbol{\theta} = (\theta_1, \theta_2, \ldots, \theta_q)^{\top}$. The observation error vector $\boldsymbol{\varepsilon} = (\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_n)^{\top}$ is as before.

If we now assume that the observations $R_{t_1}, R_{t_2}, \ldots, R_{t_n}$ are available at the arguments $t_1, t_2, \ldots, t_n$, then by (2.3.33) the observation matrix $\boldsymbol{B} = [B_{\ell,i}]_{\ell,i=1}^{n,q}$ has the elements

$$B_{\ell,i} = \phi_i(t_\ell) \tag{2.3.34}$$

for $\ell \in \{1, 2, \ldots, n\}$ and $i \in \{1, 2, \ldots, q\}$.

By inserting the known values of the functions ϕ_i, $i \in \{1, 2, \ldots, q\}$ into the observation matrix and using the observation vector one obtains directly the least-squares estimate (2.3.27). Often it is convenient to choose the basis functions such that they are orthogonal, that is

$$\sum_{\ell=1}^{n} \phi_i(t_\ell)\, \phi_k(t_\ell) = \begin{cases} 1 & \text{for} \quad i = k \\ 0 & \text{otherwise.} \end{cases} \tag{2.3.35}$$

In this case one obtains $\boldsymbol{B}^\top \boldsymbol{B} = \boldsymbol{I}$, where $\boldsymbol{I}$ is the identity matrix. This simplifies the least-squares estimator (2.3.27) yielding the simple expression

$$\hat{\boldsymbol{\theta}} = \boldsymbol{B}^\top \boldsymbol{R}. \tag{2.3.36}$$

In this case, one obtains for the least-squares estimator of the ith parameter the formula

$$\hat{\theta}_i = \sum_{\ell=1}^{n} \phi_i(t_\ell)\, R_{t_\ell} \tag{2.3.37}$$

for $i \in \{1, 2, \ldots, q\}$. The covariance matrix of the estimation error reduces here to the matrix

$$\mathrm{Cov}(\boldsymbol{\theta} - \hat{\boldsymbol{\theta}}) = \boldsymbol{B}^\top E(\boldsymbol{\varepsilon}\, \boldsymbol{\varepsilon}^\top)\, \boldsymbol{B} = \boldsymbol{B}^\top \mathrm{Cov}(\boldsymbol{\varepsilon})\, \boldsymbol{B}, \tag{2.3.38}$$

which does not depend on the true parameter vector $\boldsymbol{\theta}$. The linear least-squares method is widely used because of its simplicity. Its application can be largely successful if the chosen model is reasonably accurate for the data.

Generalized Least-Squares Estimators

One can refine the previously given linear least-squares problem by adding a symmetric positive weighting matrix $\boldsymbol{W}$ into the least-squares criterion (2.3.25). The *generalized least-squares* criterion is then of the form

$$U_{\mathrm{GLS}}(\boldsymbol{\theta}) = \frac{1}{2}\, \boldsymbol{\varepsilon}^\top \boldsymbol{W}\, \boldsymbol{\varepsilon} = \frac{1}{2}\, (\boldsymbol{R} - \boldsymbol{B}\, \boldsymbol{\theta})^\top \boldsymbol{W}\, (\boldsymbol{R} - \boldsymbol{B}\, \boldsymbol{\theta}). \tag{2.3.39}$$

The optimal choice for the weighting matrix $\boldsymbol{W}$ is the inverse of the covariance matrix of the observation error, that is

$$\boldsymbol{W} = (\mathrm{Cov}(\boldsymbol{\varepsilon}))^{-1}. \tag{2.3.40}$$

This choice follows from the fact that the resulting generalized least-squares estimator

$$\hat{\boldsymbol{\theta}} = \left(\boldsymbol{B}^\top (\mathrm{Cov}(\boldsymbol{\varepsilon}))^{-1} \boldsymbol{B} \right)^{-1} \boldsymbol{B}^\top (\mathrm{Cov}(\boldsymbol{\varepsilon}))^{-1}\, \boldsymbol{R} \tag{2.3.41}$$

minimizes also the mean square error criterion

$$U_{\mathrm{MSE}}(\boldsymbol{\theta}) = E\left((\boldsymbol{\theta} - \hat{\boldsymbol{\theta}})^\top (\boldsymbol{\theta} - \hat{\boldsymbol{\theta}})\right). \tag{2.3.42}$$

The estimator (2.3.41) is often called the *best linear unbiased estimator*.

In some applications the generalized linear least-squares method is not sufficient for capturing the dependence between the observations $R_{t_1}, R_{t_2}, \ldots, R_{t_n}$ and the parameter vector $\boldsymbol{\theta}$. In such case one can consider a nonlinear model of the form

$$\boldsymbol{R} = \boldsymbol{G}(\boldsymbol{\theta}) + \boldsymbol{\varepsilon} \tag{2.3.43}$$

with a given nonlinear vector valued function $\boldsymbol{G} : \Theta \to \Re^n$.

Similarly as before, one can in this case minimize the observation error to obtain the criterion

$$U_{\mathrm{NLS}}(\boldsymbol{\theta}) = \frac{1}{2}\,(\boldsymbol{R} - \boldsymbol{G}(\boldsymbol{\theta}))^\top (\boldsymbol{R} - \boldsymbol{G}(\boldsymbol{\theta})). \tag{2.3.44}$$

By minimizing the criterion (2.3.44) one typically obtains a *nonlinear least-squares estimator*. However, one must note that this involves a nonlinear optimization, which can only be performed numerically and may not always yield unique estimators.

2.4 Maximum Likelihood Estimation

Maximum Likelihood Method

For proper financial modeling, it is essential to use an objective and reliable statistical methodology to distinguish between competing models. A key problem is the identification of a typical distribution for log-returns. For instance, one can try to use some moment based methods, as described previously. However, these may not say enough about the shape of the distribution. Alternatively, one could use the following *maximum likelihood methodology*, which appears to be reasonably objective. It is based on some hypothesized family of probability densities and does not require the use of higher order empirical moments.

In the following the maximum likelihood methodology will be explained in the context of observed sequences of i.i.d. log-returns. This framework can be used to identify a best fit of log-return distributions, for instance, for stock market index data as will be discussed later in detail. However, one must be aware of the fact that there is never a "true" distribution behind the random variables that one observes in practice. More realistic estimation techniques are, for instance, provided by the quasi-likelihood theory, as presented in Heyde (1997).

The maximum likelihood estimation method assumes that there is no prior information available on the parameters $\theta_1, \theta_2, \ldots, \theta_q$. What is needed for the maximum likelihood method is the probability density function f_R of the independent identically distributed observations R_{t_1}, R_{t_2}, …, R_{t_n}. The *maximum likelihood estimators* $\hat{\boldsymbol{\theta}} = (\hat{\theta}_1, \hat{\theta}_2, \ldots, \hat{\theta}_n)^\top$ have several theoretically highly desirable asymptotic optimality properties when the sample size n is large. For instance, in the case when there exists an estimator which satisfies the Cramer-Rao lower bound (2.3.17), then it can be constructed by using the maximum likelihood method. The maximum likelihood estimator $\hat{\theta}$ is consistent and asymptotically efficient, which means that it achieves asymptotically the Cramer-Rao lower bound for the estimation error.

Likelihood Function

We assume that the observed log-returns are denoted by R_{t_1}, R_{t_2}, … and form a sequence of i.i.d. random variables with some hypothesized, parameterized density $f_{\boldsymbol{R}}$. Note that the joint probability density function $f_{\boldsymbol{R}}$ of the random vector $\boldsymbol{R} = (R_{t_1}, R_{t_2}, \ldots, R_{t_n})^\top$ of log-returns can be written for these i.i.d. random variables, see (1.4.41), in the form

$$\mathcal{L}(\boldsymbol{\theta}) = f_{R_{t_1}, R_{t_2}, \ldots, R_{t_n}}(R_{t_1}, R_{t_2}, \ldots, R_{t_n}, \boldsymbol{\theta}) = \prod_{i=1}^{n} f_{\boldsymbol{R}}(R_{t_i}, \boldsymbol{\theta}). \tag{2.4.1}$$

Here $f_{\boldsymbol{R}}(\cdot, \boldsymbol{\theta})$ is the density of R_{t_i}, $i \in \{1, 2, \ldots, n\}$, given the parameter values $\boldsymbol{\theta} = (\theta_1, \theta_2, \ldots, \theta_q)^\top \in \Theta \subseteq \Re^q$, $q \in \mathcal{N}$. The set Θ specifies again the set of allowable values that the parameters can take. Our aim will be to find some best parameter estimates that fit the data. We call the above function (2.4.1) the *likelihood function* for the parameter $\boldsymbol{\theta} \in \Theta$.

Maximum Likelihood Estimate

To be able to optimize the choice of the parameter we need a criterion to identify a best fit. The maximum likelihood methodology uses the *maximum likelihood estimator* $\hat{\boldsymbol{\theta}} = (\hat{\theta}_1, \hat{\theta}_2, \ldots, \hat{\theta}_q)^\top$ as a best estimate of $\boldsymbol{\theta} \in \Theta$, where

$$\mathcal{L}^* = \mathcal{L}(\hat{\boldsymbol{\theta}}) = \sup_{\boldsymbol{\theta} \in \Theta} \mathcal{L}(\boldsymbol{\theta}). \tag{2.4.2}$$

Here $\mathcal{L}^* = \sup_{\boldsymbol{\theta} \in \Theta} \mathcal{L}(\boldsymbol{\theta})$ denotes the *supremum* of $\mathcal{L}(\boldsymbol{\theta})$, that is in our case the least upper bound of $\mathcal{L}(\boldsymbol{\theta})$ over all $\boldsymbol{\theta} \in \Theta$. This means, $\hat{\boldsymbol{\theta}}$ is the parameter that maximizes the likelihood function with respect to the set of permitted parameter values $\boldsymbol{\theta} \in \Theta$. Intuitively, this choice yields the parameter $\hat{\boldsymbol{\theta}}$ for which the observed log-returns are most likely chosen from the hypothesized density in the given parameterized family of probability densities. This means that $f_{\boldsymbol{R}}(\cdot, \hat{\boldsymbol{\theta}})$ represents the most probable density from the given class of densities having observed the log-returns $R_{t_1}, R_{t_2} \ldots, R_{t_n}$.

In practice, it is convenient to work with the *log-likelihood function*

$$\ell(\boldsymbol{\theta}) = \ln(\mathcal{L}(\boldsymbol{\theta})). \tag{2.4.3}$$

Under suitable conditions, when the true parameter is an interior point of Θ, the maximum likelihood estimate $\hat{\boldsymbol{\theta}}$ can be obtained as a root of the first order conditions

$$\frac{\partial \ell(\boldsymbol{\theta})}{\partial \theta_i} = 0 \tag{2.4.4}$$

for all $i \in \{1, 2, \ldots, q\}$, where $\frac{\partial}{\partial \theta_i}$ denotes the partial derivative with respect to θ_i. The system of equations (2.4.4) is called the system of *maximum likelihood equations*.

If one cannot explicitly solve the maximum likelihood equations, then a root finding method, for instance, a multi-dimensional Newton method, can be applied to solve the system of maximum likelihood equations. The above maximum likelihood approach only yields reliable estimates if the hypothesized model is suitable and a sufficient number of observations is available. For the identification of log-return distributions this means that the hypothesized distribution must be reasonably close to the true distribution and one needs a large number of observed log-returns that can be interpreted as being independent and identically distributed.

Likelihood Ratio Test

Now, let us suppose that we have a class of models that corresponds to a class of parameters characterizing some hypothesized density. Our goal will be to identify the parameters of the density which best fits our data set of observed log-returns using the maximum likelihood approach.

This can be achieved by the *likelihood ratio test* which is due to Neyman & Pearson (1928). We emphasize, that the maximum likelihood approach does not rely on certain higher moments, for instance the kurtosis, that might not even exist in a given situation, as we shall see later.

We define the *likelihood ratio* in the form

$$\Lambda = \frac{\mathcal{L}^*_{\text{model}}}{\mathcal{L}^*_{\text{general model}}}. \tag{2.4.5}$$

Here $\mathcal{L}^*_{\text{model}}$ represents the maximized likelihood function of a hypothesized model density, say, with q parameters. On the other hand, $\mathcal{L}^*_{\text{general model}}$ denotes the maximized likelihood function for the density of a more general model that has, say $q + \nu$ parameters and nests the hypothesized model density $q, \nu \in \mathcal{N}$.

Under appropriate conditions it can be shown, see Rao (1973), that the density of the *test statistic*

$$L_n = -2 \ln(\Lambda) \tag{2.4.6}$$

is asymptotically a chi-square density, or more generally a gamma density, see (1.2.9), for increasing number of observations $n \to \infty$. Here the degrees

Table 2.4.1. Quantiles for the chi-square-distribution

ν	1	2	30
$\chi^2_{0.99,\nu}$	6.635	9.210	50.9
$\chi^2_{0.95,\nu}$	3.841	5.991	43.8
$\chi^2_{0.90,\nu}$	2.706	4.605	40.3
$\chi^2_{0.20,\nu}$	0.064	0.446	23.4
$\chi^2_{0.10,\nu}$	0.0158	0.211	20.6
$\chi^2_{0.05,\nu}$	0.0039	0.103	18.5
$\chi^2_{0.01,\nu}$	0.000157	0.020	15.0
$\chi^2_{0.001,\nu}$	0.000002	0.002	11.6

of freedom ν equal the difference between the number of parameters in the general model density and the hypothesized model density. It can then be shown that as $n \to \infty$

$$P\left(L_n < \chi^2_{1-\alpha,\nu}\right) \approx F_{\chi^2(\nu)}\left(\chi^2_{1-\alpha,\nu}\right) = 1 - \alpha, \tag{2.4.7}$$

where $F_{\chi^2(\nu)}$ denotes the chi-square distribution with ν degrees of freedom and $\chi^2_{1-\alpha,\nu}$ is its $100(1-\alpha)\%$ quantile. In Table 2.4.1 we summarize some quantiles of the chi-square distribution.

As in Sect. 2.2 we can similarly check for a given $100\alpha\%$ confidence level whether or not the test statistic L_n is in the $100(1-\alpha)\%$ quantile of the chi-square distribution with ν degrees of freedom. If the relation

$$L_n < \chi^2_{1-\alpha,\nu} \tag{2.4.8}$$

is satisfied, then we cannot reject at the $100\alpha\%$ significance level the hypothesis that the suggested model is the true underlying model. Otherwise, we reject this hypothesis on the basis of this likelihood ratio test.

2.5 Normal Variance Mixture Models

Subordination

It is well-known that the log-return distributions of security prices are strongly leptokurtic. This means that they have larger kurtosis than the Gaussian distribution provides. The following simple modeling approach is called *subordination* and goes back to Bochner (1955) and Clark (1973). We used already some kind of subordination when generating Student t log-returns in (2.2.20). For capturing typical features of log-return distributions one can simply make the conditional variances themselves independent random variables. This yields a class of models with normal-variance mixture distributed log-returns, see Feller (1968). This kind of models allows us to keep the empirical analysis fairly simple. For simplicity we shall only consider here symmetric log-returns since any log-return mean is in reality extremely small and can be easily added to the model.

We now assume a normal-variance mixture density for the ith log-return Z_i by setting

$$Z_i = \sqrt{m_i}\,\xi_i. \tag{2.5.1}$$

Here we use for all $i \in \{0, 1, \ldots, n-1\}$ an independent identically distributed nonnegative *conditional variance* m_i, together with some independent, standard Gaussian distributed random variable $\xi_i \sim N(0,1)$. Note that each log-return Z_i can here be linked to a corresponding conditional variance m_i. The conditional variance m_i for the log-return at time t_i, $i \in \{0, 1, \ldots\}$, is assumed to be distributed according to a given density f_m. The generality of the resulting class of normal-variance mixture densities for log-returns follows from the freedom to adjust the density f_m. One obtains here the normal-variance mixture density function of the log-return Z_i, in the form

$$f_Z(x) = \frac{1}{\sqrt{2\,\pi}} \int_0^\infty \frac{1}{\sqrt{u}} \exp\left\{-\frac{x^2}{2\,u}\right\} f_m(u)\,du \tag{2.5.2}$$

for $x \in \Re$, as long as this integral exists. The ith log-return Z_i has then mean zero, variance $v_Z = v_m$, skewness zero and kurtosis

$$\kappa_Z = 3\left(1 + \frac{v_m}{(\mu_m)^2}\right). \tag{2.5.3}$$

Here the ith conditional variance m_i, $i \in \{0, 1, \ldots, n-1\}$ has mean μ_m and variance v_m. One notes for the case of nonzero variance of the conditional variance m_i that it follows by (2.5.3) that any normal-variance mixture density has kurtosis greater than three and is, therefore, *leptokurtic*. This can be used as an explanation for widely observed leptokurtic log-returns in practice. The *lognormal model* with constant m_i is an extremely useful modeling attempt that has as its justification mainly its mathematical simplicity.

Samuelson (1957), Osborne (1959) and subsequently many other authors have modeled asset price increments by lognormal random variables, where the resulting log-returns are Gaussian random variables. This has been an extremely important first step in quantitative finance. The corresponding Gaussian density for the lognormal model results from (2.5.2) when the density of the conditional variance degenerates to that of a constant $m_i = 1$ with $v_m > 0$ for $i \in \{0, 1, \ldots, n-1\}$. We emphasize that the Gaussian density has been clearly rejected in many studies as a suitable log-return density for most securities.

Let us remark that in a wider range of models, beyond the models that we consider here, one does not need the conditional variance for the log-returns to exist. This allows one to cover *logstable models* as suggested in Mandelbrot (1963, 1967), Mandelbrot & Taylor (1967), Fama (1965) and Hurst, Platen & Rachev (1999). These models have typically one additional parameter when compared to the lognormal model and generate log-returns that may have no conditional variance. We are not studying here any of these models since most log-returns in financial markets seem to have finite conditional variance.

SGH Models

We shall not consider any further the case of a Gaussian log-return density but study instead a rich class of analytically tractable densities that include the Gaussian one as a limit. Our aim will be to discriminate between a wide range of possible leptokurtic densities. The densities that we shall analyze can be classified as a class of normal-variance mixture densities. The Gaussian log-return density arises simply as limiting case for certain extreme parameters. As described in Sect. 1.2, it is noticeable that a large group of authors have proposed important models with log-returns that relate to the class of generalized hyperbolic densities. This class of densities was extensively examined by Barndorff-Nielsen (1977, 1978) and Barndorff-Nielsen & Blaesild (1981). We assume, for simplicity, zero skewness and consider in the following the *symmetric generalized hyperbolic* (SGH) density as a possible density for log-returns. This density results when the density of the conditional variance m_i, $i \in \{0, 1, \ldots\}$, is a *generalized inverse Gaussian density*. We call the resulting discrete time log-return models *SGH models*.

By (1.2.24) the SGH density function of a log-return Z_i is of the form

$$f_Z(x) = \frac{1}{\delta\, K_\lambda(\alpha\,\delta)} \sqrt{\frac{\alpha\,\delta}{2\,\pi}} \left(1 + \frac{x^2}{\delta^2}\right)^{\frac{1}{2}(\lambda - \frac{1}{2})} K_{\lambda - \frac{1}{2}}\left(\alpha\,\delta \sqrt{1 + \frac{x^2}{\delta^2}}\right) \tag{2.5.4}$$

for $x \in \Re$, where $\lambda \in \Re$ and $\alpha, \delta \geq 0$. We set $\alpha \neq 0$ if $\lambda \geq 0$ and $\delta \neq 0$ if $\lambda \leq 0$. We remark that the corresponding probability density function of a conditional variance m_i in the normal-variance mixture density (2.5.2) is here the *generalized inverse Gaussian density* of the form

$$f_m(x) = \frac{\alpha^\lambda}{2\,\delta^\lambda K_\lambda(\alpha\,\delta)}\, x^{\lambda-1} \exp\left\{-\frac{1}{2}\left(\frac{\delta^2}{x} + \alpha^2 x\right)\right\}, \tag{2.5.5}$$

where $K_\lambda(\cdot)$ is the modified Bessel function of the third kind with index λ, see (1.2.25).

The SGH density is a four parameter density. The two *shape parameters* are λ and $\bar{\alpha} = \alpha\,\delta$, defined so that they are invariant under scale transformations. The other parameters contribute to the scaling of the density. We define as in (1.2.27) the parameter c as the *scale parameter* such that $v_m = v_Z = c^2$, that is

$$c^2 = \begin{cases} \frac{2\,\lambda}{\alpha^2} & \text{if } \delta = 0 \text{ for } \lambda > 0,\ \bar{\alpha} = 0, \\ \frac{\delta^2\, K_{\lambda+1}(\bar{\alpha})}{\bar{\alpha}\, K_\lambda(\bar{\alpha})} & \text{otherwise.} \end{cases} \tag{2.5.6}$$

The variance of m_i is

$$v_m = c^4 \left(\frac{K_\lambda(\bar{\alpha})\, K_{\lambda+2}(\bar{\alpha})}{K_{\lambda+1}(\bar{\alpha})^2} - 1 \right). \tag{2.5.7}$$

Consequently, the log-return Z_i has mean zero, variance $v_Z = c^2$, skewness zero and kurtosis

$$\kappa_Z = \frac{3\, K_\lambda(\bar{\alpha})\, K_{\lambda+2}(\bar{\alpha})^2}{K_{\lambda+1}(\bar{\alpha})}. \tag{2.5.8}$$

Furthermore, it can be shown that as $\lambda \to \pm\infty$ or $\bar{\alpha} \to \infty$ the SGH density asymptotically approaches the Gaussian density.

To illustrate certain candidate densities for the log-returns within the class of SGH densities we recall in the following four special cases of the SGH density that coincide with the log-return densities of important asset price models, which we previously mentioned in Sect. 1.2.

Praetz (1972) and Blattberg & Gonedes (1974) suggested for log-returns a *Student t density* with degrees of freedom $n > 0$. This density follows from the above SGH density for the shape parameters $\lambda = -\frac{1}{2}\, n < 0$ and $\bar{\alpha} = 0$, that is $\alpha = 0$ and $\delta = \varepsilon\, \sqrt{n}$. Using these parameter values the Student t density function has then the form (1.2.28). Here the variance mixture density is an inverse gamma density.

Barndorff-Nielsen (1995) considered log-returns to follow a *normal-inverse Gaussian mixture distribution.* The corresponding density is obtained from the SGH density when the shape parameter $\lambda = -\frac{1}{2}$ is chosen. For this parameter value the conditional variance m_i is inverse Gaussian distributed and it follows by (2.5.4) that the probability density function (1.2.29) for Z_i.

Eberlein & Keller (1995) and Küchler et al. (1999) suggested models, where log-returns appear to be *hyperbolicly distributed.* This type of models is covered by the choice of the shape parameter $\lambda = 1$ in the SGH density. The probability density function of the Z_i is then of the form (1.2.30).

Madan & Seneta (1990), Geman, Madan & Yor (2001) and Carr, Geman, Madan & Yor (2003) assumed log-returns to be distributed with a *normal-variance gamma mixture distribution.* This particular case occurs when the shape parameters are such that $\lambda > 0$ and $\bar{\alpha} = 0$, that is, $\delta = 0$ and $\alpha = \frac{\sqrt{2\,\lambda}}{c}$. In this case the conditional variance m_i is gamma distributed and the probability density function of Z_i is given by (1.2.31)

This means the variance mixture density is here a gamma density. The model is also known as *variance gamma* (VG) model. Below we shall investigate which kinds of densities best fit observed log-returns.

2.6 Distribution of Index Log-Returns

Estimation of Log-Returns

In the literature a vast amount of empirical work has been accomplished estimating the distributions of log-returns of financial securities. Starting with papers by Mandelbrot (1963) and Fama (1963), and in a subsequent stream of literature, it became clear that the standard assumption that log-returns are Gaussian distributed is very crude and for certain risk management tasks even a dangerous assumption. It is now widely recognized that, in reality, extreme log-returns are more likely than suggested by the Gaussian distribution.

From the perspective of the benchmark approach, which we present later in this book, it is not surprising that empirical studies on log-returns of exchange rates and equity prices have so far not identified a particular distribution that fits the observed data well. The reason is that an exchange rate involves at least two major factors. One of these relates to the domestic economy and the second reflects the foreign economy. As we shall explain later, the benchmark approach suggests we analyze and model the denominations of the market portfolio in different currencies. Intuitively, one separates other impacts from that of the currency of interest. The market portfolio acts as the reference unit that is practically least disturbed by any particular security. The benchmark approach will suggest that one may more easily find some testable empirical evidence about the log-return distribution of an index than for an exchange rate or equity price.

In two papers, the Nobel Laureate Harry Markowitz together with Usmen; see Markowitz & Usmen (1996a, 1996b), analyzed daily S&P500 stock index log-returns for the period from 1963 until 1983 in a Baysian framework. Within the wide family of Pearson distributions, see Stuart & Ord (1994), they identified the Student t distribution with about $n = 4.5$ degrees of freedom as the best fit.

In an independent empirical analysis of stock index log-returns observed from 1982 until 1996 of the S&P500 and other stock indices, Hurst & Platen (1997) demonstrated that the Student t distribution provides the best fit in a range of normal-variance mixture distributions when employing a maximum likelihood methodology. In a recent study on log-returns of a world stock index, when denominated in different currencies, Fergusson & Platen (2006) showed via likelihood ratio tests by using daily data for the period from 1970 until 2004 that the Student t distribution provides clearly the best fit in the class of SGH distributions. In the following we provide more details on these findings.

A World Stock Index

We apply the previously described SGH model, where we use as security a *world stock index* (WSI), which we construct as a self-financing portfolio of stock market accumulation indices. The weights for the market capitalization that were chosen at the end of the observation period in 2004 are given in the last column of Table 2.6.1. Weights for earlier years reflect the market capitalization as obtainable from Thomson Financial.

We cover with this index more than 95% of the world stock market capitalization for the period from 1970 until 2004. We exclude in our study weekends

Table 2.6.1. Empirical moments for log-returns of WSI

Country	Currency	$\hat{m}_y$	$\hat{\sigma}_y$	$\hat{\beta}_y$	$\hat{\kappa}_y$	Weights
Argentina	ARS	0.000495	0.009705	4.971980	170.049000	0.0023
Australia	AUD	0.000403	0.008815	-0.372951	14.051762	0.0017
Austria	ATS, EUR	0.000367	0.008716	-0.489599	13.406629	0.0161
Belgium	BEF, EUR	0.000332	0.009387	-0.443605	13.231579	0.0061
Brazil	BRL	0.001392	0.011284	2.330662	79.025855	0.0013
Canada	CAD	0.000404	0.007831	-0.579593	18.075123	0.0251
Denmark	DKK	0.000356	0.009526	-0.394520	14.431344	0.0037
Finland	FIM, EUR	0.000396	0.010612	0.238200	70.189255	0.0032
France	FRF, EUR	0.000376	0.009332	-0.416397	13.655084	0.0302
Germany	DEM, EUR	0.000285	0.009417	-0.483043	13.632198	0.0342
Greece	GRD, EUR	0.000569	0.009454	0.841563	35.815507	0.0012
Hong Kong	HKD	0.000421	0.008000	-0.446156	17.986618	0.0231
Hungary	HUF	0.000476	0.008578	-0.184216	17.086200	0.0023
India	INR	0.000542	0.013561	-0.131696	78.892868	0.0047
Indonesia	IDR	0.000571	0.008906	-0.118789	19.917163	0.0201
Ireland	IRP, EUR	0.000373	0.009033	-0.507152	13.844893	0.0055
Italy	ITL, EUR	0.000373	0.008957	-0.535339	16.379118	0.0132
Japan	JPY	0.000238	0.009033	-0.638245	14.427542	0.1550
Korea	KRW	0.000439	0.009636	-0.139465	57.088739	0.0072
Malaysia	MYR	0.000398	0.008500	-0.616024	18.996454	0.0158
Mexico	MXN	0.001152	0.020686	3.860894	278.762521	0.0016
Netherlands	NLG, EUR	0.000300	0.009358	-0.472830	13.914414	0.0193
Norway	NOK	0.000373	0.009196	-0.340799	14.541824	0.0029
Philippines	PHP	0.000460	0.009009	-0.354508	19.339190	0.0041
Portugal	PTE, EUR	0.000391	0.008810	-0.452567	13.152174	0.0013
Singapore	SGD	0.000353	0.007991	-0.554040	17.215279	0.0079
Spain	ESP, EUR	0.000464	0.010334	1.079021	48.935156	0.0124
Sweden	SEK	0.000421	0.009195	0.214018	19.355551	0.0124
Switzerland	CHF	0.000239	0.010097	-0.407918	11.406303	0.0206
Taiwan	TWD	0.000357	0.007900	-0.558432	17.784127	0.0141
Thailand	THB	0.000450	0.009285	0.628255	36.231337	0.0049
Turkey	TRL	0.001029	0.011354	5.010563	155.199258	0.0018
UK	GBP	0.000414	0.009141	-0.482534	13.541676	0.0846
US	USD	0.000374	0.007724	-0.613548	18.808550	0.4301

and other nontrading days at the US and European exchanges. As shown in Platen (2004c, 2005b) and as will be discussed later in Sect. 10.6 under the benchmark approach, such a diversified portfolio is robust against variations in weightings as long as the weight of each contributing security remains reasonably small. The constructed portfolio is a proxy for the world stock portfolio or market portfolio.

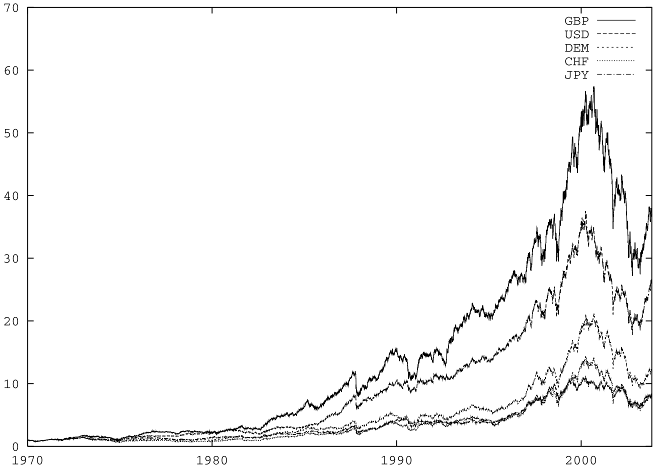

Fig. 2.6.1. WSI in units of different currencies

In Fig. 2.6.1 we plot the resulting WSI for the observation period when denominated in units of British Pound, US dollar, Deutsche Mark, Swiss Franc and Japanese Yen. For convenience we normalized the initial values to one.

In the following we shall study the distribution of log-returns of the WSI when denominated in 34 major currencies. This will provide some distributional characterization of the general market risk for the respective markets. The findings will be quite important for supporting the appropriateness of theoretically suggested financial market models.

We deliberately do not adjust for any changes over time, market crashes or other influences that may have affected the data. Some methods of data analysis discard extreme values of observations as outliers. But this would be inappropriate in a financial context because it is most important for risk management to capture the probability for extreme log-returns. For *daily log-returns* of the WSI for the period from 1970 until 2004 in 34 currency denominations, the first four empirical moments yield the average empirical mean $\hat{m}_y = 0.000486$, average standard deviation $\hat{\sigma}_y = 0.009789$, average skewness $\hat{\beta}_y = 0.460316$ and average kurtosis $\hat{\kappa}_y = 44.485182$.

For each currency let us centralize the log-returns of WSI denominations with respect to the mean. Furthermore, we scale these to obtain unit variance. The resulting transformed log-returns have then an estimated zero mean and unit variance. Now, we combine all observed centralized and scaled daily log-returns in one large sample. We show in Fig. 2.6.2 the logarithm of the resulting log-return histogram, that is, the relative frequencies. This figure shows also the logarithm of the Student t density with degrees of freedom 3.64, which appears to fit the data extremely well already by visual inspection. When comparing the shapes of the other densities covered in Fig. 1.2.8–1.2.10, then none of these match the shape shown in Fig. 2.6.2.

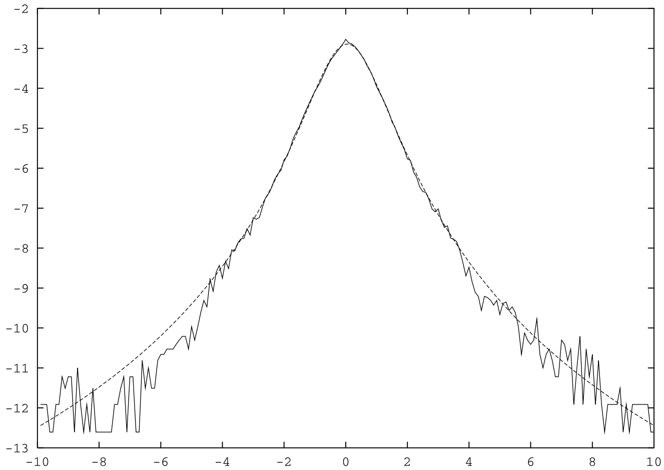

Fig. 2.6.2. Logarithm of WSI log-return histogram

By considering the above fit and the relatively small empirical skewness the empirical log-return density appears to be fairly symmetrical. This is also consistent with the findings in Markowitz & Usmen (1996a, 1996b) for the case of S&P500 log-returns and those in Hurst & Platen (1997) for stock index log-returns. Therefore, to simplify our analysis and to focus on the identification of the tail properties of log-return densities, we assume that the densities that we shall consider are symmetric, that is we assume zero mean and zero skewness. This assumption does not very much influence the empirical results that we obtain and allows us to use the SGH models described in Sect. 2.5. We emphasize that the following study focuses on the shape of the log-return densities. It avoids relying on any particular moment properties. Note that certain higher order moments may not exist. In particular, the kurtosis of Student t distributed log-returns with less than four degrees of freedom is *infinite*, see (1.3.37). This could become a problem in reality when taking into account that in Fig. 2.6.2 we fitted a Student t density with about 3.64 degrees of freedom.

Maximum Likelihood Estimation for SGH Densities

Let us now identify the typical distribution for log-returns of the WSI in the class of SGH distributions.

The class of SGH densities that we introduced in Sect. 1.2 represents a rich class of leptokurtic densities. To reject, on a given significance level, the assumption that a hypothetical SGH density is not the true underlying density we apply the maximum likelihood ratio test, as described in Sect. 2.4. We define the *likelihood ratio* in the form

$$\Lambda = \frac{\mathcal{L}^*_{\text{model}}}{\mathcal{L}^*_{\text{SGH}}}. \tag{2.6.1}$$

Here $\mathcal{L}^*_{\mathrm{model}}$ represents the maximized likelihood function of a given specific, nested log-return density, for instance, the Student t density. With respect to this density the maximum likelihood estimate for the parameters is calculated and then used to obtain the corresponding *likelihood function* $\mathcal{L}^*_{\mathrm{model}}$. On the other hand, $\mathcal{L}^*_{\mathrm{SGH}}$ denotes the maximized likelihood function for the SGH density, which is the nesting density that has been similarly obtained. We then calculate according to (2.4.6) the test statistic $L_n = -2\ln(\Lambda)$, which is for increasing number $n \to \infty$ of observations asymptotically chi-square distributed. Here the degrees of freedom equal the difference between the number of parameters of the nesting density and the nested density. The nesting density, which is the SGH density, is a four-parameter density and our nested densities are the Student t, normal-inverse Gaussian, hyperbolic and variance-gamma density, see Sect. 1.2. Each of these is a three-parameter density. Therefore, in the cases considered, the test statistic L_n is, for $n \to \infty$, asymptotically chi-square distributed with one degree of freedom.

According to (2.4.7) we have asymptotically as $n \to \infty$ that

$$P\left(L_n < \chi^2_{1-\alpha,1}\right) \approx 1 - \alpha, \tag{2.6.2}$$

where $\chi^2_{1-\alpha,1}$ is the $100(1-\alpha)\%$ quantile of the chi-square distribution $F_{\chi^2(1)}$ with one degree of freedom. One can then check, say, for a 99% *significance level* whether or not the test statistic L_n is in the 1% quantile of the chi-square distribution with one degree of freedom. By Table 2.4.1 it follows that if the relation

$$L_n < \chi^2_{0.01,1} \approx 0.000157 \tag{2.6.3}$$

is satisfied, then we cannot reject on a 99% significance level the hypothesis that the suggested density is the true underlying density. Similarly, if

$$L_n < \chi^2_{0.001,1} \approx 0.000002 \tag{2.6.4}$$

holds, then we cannot reject the hypothesis that the given density is the true underlying density on a 99.9% significance level.

The above maximum likelihood methodology offers a natural definition of a *best fit*. We call the density with the smallest test statistic L_n the best fit in the given class of SGH densities. This density maximizes the likelihood ratio Λ given in (2.6.1).

Log-Returns of World Stock Indices

Now let us study the log-returns of the WSI when denominated in units of each of the 34 currencies. By the above formulas together with the different SGH densities described in Sects. 1.2 and 2.5 one can construct the corresponding maximum likelihood estimators and test statistics. In Table 2.6.2 we display the test statistics L_n for log-returns of the WSI in the 34 different currency denominations, as given in Fergusson & Platen (2006). It is apparent that the

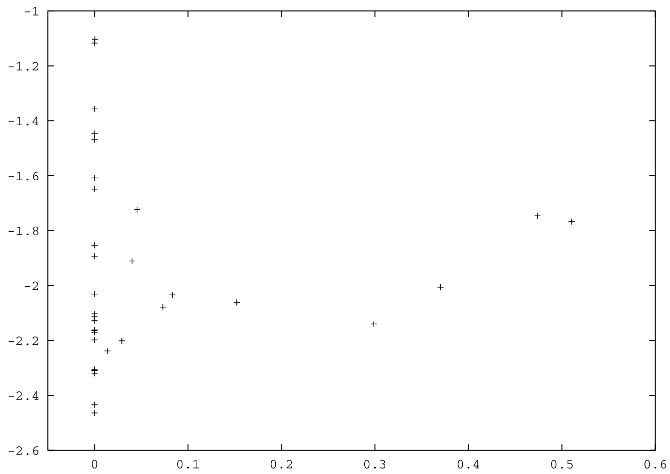

Fig. 2.6.3. $(\bar{\alpha}, \lambda)$-plot for log-returns of WSI in different currencies

Student t density shows in all cases the smallest test statistic. For 25 of the 34 currencies one cannot reject on a 99.9% significance level the hypothesis that the Student t density is the true density. The inverse Gaussian density seems to be the second best choice but can be for all stocks rejected on any reasonable significance level. In the last column of Table 2.6.2 one finds the estimated degrees of freedom for the Student t density for each of the currency denominations. One notes that the degrees of freedom are in the range of 2.2 to 4.9. We emphasize that the resulting arithmetic average of 3.9899 of the estimated degrees of freedom for log-returns of WSI currency denominations is very close to the number four. We shall find towards the end of the book a natural explanation for this empirical result.

Let us visualize the estimated shape parameters $\bar{\alpha}$ and λ of the SGH density in Fig. 2.6.3 in an $(\bar{\alpha}, \lambda)$-scatter plot for the log-returns of the 34 WSI currency denominations.

Interestingly, the estimated parameter points are scattered near the negative λ-axis with values between -2.5 and -1.0. It is the Student t density that is characterized by points directly located on the λ-axis with $n = -2\lambda$ degrees of freedom. This means that the large number of points near $\lambda = -2$ on the λ-axis indicate a *Student t density of about four degrees of freedom.* For comparison, the variance-gamma density, which would favor the Madan-Seneta variance gamma model, see Madan & Seneta (1990) and Geman et al. (2001), would need to have points on the positive λ-axis in Fig. 2.6.3. The hyperbolic density, suggested by Eberlein & Keller (1995) as log-return density corresponds to the shape parameter $\lambda = 1$ and any $\bar{\alpha} > 0$. Our findings in Fig. 2.6.3 do not support this type of model. The normal-inverse Gaussian distributed log-returns of a model proposed in Barndorff-Nielsen (1995) would generate points in the scatter plot of Fig. 2.6.3 for $\lambda = -\frac{1}{2}$ and $\alpha > 0$. However, there are also no points that come close to the line $\lambda = -\frac{1}{2}$. One

Table 2.6.2. The L_n test statistic for log-returns of the WSI in different currencies

Country	Student t	Inverse Gaussian	Hyperbolic	Variance Gamma	Degrees of Freedom
Argentina	0.000000	137.566726	377.265316	414.030946	3.215953
Australia	0.000000	24.403096	54.527486	73.900060	4.618187
Austria	1.272478	10.533450	43.923596	63.696828	4.177998
Belgian	0.000000	15.755730	41.664530	59.922270	4.639238
Brazil	0.000000	132.348986	430.843576	444.117392	2.937178
Canada	0.000000	42.829996	80.784298	110.357598	4.927523
Denmark	0.000000	29.607334	72.807340	96.852868	4.328256
Finland	0.000000	130.807532	286.692740	326.546792	3.707350
France	0.303708	12.578282	42.737860	61.753332	4.329017
Germany	0.000000	17.945998	45.546996	64.739532	4.620013
Greece	0.000000	60.072066	120.786456	150.578024	4.340990
Hong Kong	0.000000	31.399542	84.191740	111.611116	4.080899
Hungary	0.002812	33.488642	102.407366	132.556744	3.827598
India	0.000000	218.752194	1096.862148	962.798700	2.283747
Indonesia	0.000000	54.595328	121.131360	148.098694	4.062652
Ireland	0.031796	16.660834	53.818850	76.062630	4.187040
Italy	0.002606	19.207820	60.267448	83.332082	4.172936
Japan	0.000000	24.017652	60.214094	81.351358	4.392711
Korea S.	0.000000	129.955438	386.626152	425.311040	3.265655
Malaysia	0.000000	56.525498	149.299592	189.659002	3.786499
Mexico	0.000000	440.818300	2132.850298	1746.118774	2.207160
Netherlands	0.000000	15.802518	41.848016	60.873070	4.611005
Norway	0.000000	27.920608	71.785758	96.835862	4.256095
Philippines	0.017290	52.048754	167.407546	199.781080	3.458544
Portugal	1.582056	13.129484	54.154638	76.914946	4.071114
Singapore	0.000000	30.656496	73.326620	99.354034	4.396040
Spain	0.000000	70.602362	139.884600	165.163122	4.206288
Sweden	0.000000	66.852560	130.642934	166.827332	4.468983
Switzerland	0.144462	15.390592	46.179620	67.726852	4.471172
Taiwan	0.000000	33.290522	82.900518	110.434774	4.224246
Thailand	0.000000	100.851126	282.814096	314.709524	3.298124
Turkey	0.000000	152.625500	493.285862	506.466162	2.893205
UK	0.000000	21.124390	47.980512	68.613654	4.868241
US	0.000000	31.352956	75.539480	102.259640	4.323809

must emphasize that all the above discussed log-return densities are strongly leptokurtic and already rather close to the underlying type of log-return density. Still, the above analysis points clearly in the direction of a Student t density as the best fitting log-return density. Given the high significance level the demonstrated Student t property of daily index log-returns establishes a stylized empirical fact that has to be explained in an advanced financial market model. We shall explain later in Sect. 13.2 the findings by the minimal market model, see Platen (2001, 2002), which describes in some sense the optimal dynamics of a financial market. To do this properly, we need to apply the theory of stochastic processes and use stochastic differential equations for modeling.

2.7 Convergence of Random Sequences

As we have already seen in the LLN and the CLT, questions about the convergence of random sequences arise naturally in both theory and applications. This is certainly true in the area of quantitative finance, where uncertainty is the key feature that has to be modeled and large numbers of random variables arise.

Different Types of Convergence

In contrast to a deterministic setting one faces in a stochastic environment several different types of convergence, which is sometimes confusing. For this reason, we summarize several different types of convergence that are commonly used. Some of these we have already introduced previously. In what follows we assume that the random variables X_1, X_2, ... and X are all defined on the same probability space $(\Omega, \mathcal{A}, P)$.

I. The sequence X_1, X_2, ... *converges in probability* to X if for every $\varepsilon > 0$

$$\lim_{n\to\infty} P(|X_n - X| > \varepsilon) = 0. \tag{2.7.1}$$

In this case we write $X \stackrel{P}{=} \lim_{n\to\infty} X_n$.

II. The sequence X_1, X_2, ... *converges with probability one* to X if

$$P\left(\lim_{n\to\infty} X_n = X\right) = 1. \tag{2.7.2}$$

For this type of convergence we write $X \stackrel{\text{a.s.}}{=} \lim_{n\to\infty} X_n$ or say that the sequence converges *almost surely*, that is P-a.s. or a.s. to X.

III. For $p \in (0, \infty)$ the sequence X_1, X_2, ... *converges in mean order* p to X if

$$\lim_{n\to\infty} E\left(|X_n - X|^p\right) = 0. \tag{2.7.3}$$

Here we write $X \stackrel{L^p}{=} \lim_{n\to\infty} X_n$. For $p = 2$, this is convergence in the mean square sense, that is $X \stackrel{\text{m.s.}}{=} \lim_{n\to\infty} X_n$, see (2.1.7).

IV. The sequence X_1, X_2, ... *converges in distribution* to X if

$$\lim_{n\to\infty} E(f(X_n)) = E(f(X)) \tag{2.7.4}$$

for every bounded continuous function $f : \Re \to \Re$. In this case we write $X \stackrel{\mathrm{d}}{=} \lim_{n\to\infty} X_n$. It can be shown, see Shiryaev (1984), that this is equivalent to

$$\lim_{n\to\infty} F_{X_n}(x) = F_X(x) \tag{2.7.5}$$

for all continuity points of $x \in \Re$.

For the above types of convergence the following implications, denoted by $\Longrightarrow$, can be shown, see Shiryaev (1984):

$$X \stackrel{\text{a.s.}}{=} \lim_{n\to\infty} X_n \Longrightarrow X \stackrel{P}{=} \lim_{n\to\infty} X_n$$

$$X \stackrel{L^p}{=} \lim_{n\to\infty} X_n \Longrightarrow X \stackrel{P}{=} \lim_{n\to\infty} X_n$$

$$X \stackrel{P}{=} \lim_{n\to\infty} X_n \Longrightarrow X \stackrel{\mathrm{d}}{=} \lim_{n\to\infty} X_n \tag{2.7.6}$$

for $p \in (0, \infty)$. Furthermore, it can be shown, see Shiryaev (1984), that if $X \stackrel{P}{=} \lim_{n\to\infty} X_n$, then there exists a subsequence $X_{i_1}, X_{i_2}, \ldots$ with $X \stackrel{\text{a.s.}}{=} \lim_{k\to\infty} X_{i_k}$. By the implications (2.7.6) it follows if $X \stackrel{L^p}{=} \lim_{n\to\infty} X_n$, then there exists also such a subsequence.

Note that for $X \stackrel{P}{=} \lim_{n\to\infty} X_n$ and $Y \stackrel{P}{=} \lim_{n\to\infty} Y_n$ it follows that

$$X + Y \stackrel{P}{=} \lim_{n\to\infty} (X_n + Y_n) \quad \text{and} \quad X\,Y \stackrel{P}{=} \lim_{n\to\infty} X_n\,Y_n. \tag{2.7.7}$$

Similarly, for $X \stackrel{\text{a.s.}}{=} \lim_{n\to\infty} X_n$ and $Y \stackrel{\text{a.s.}}{=} \lim_{n\to\infty} Y_n$ one has

$$X + Y \stackrel{\text{a.s.}}{=} \lim_{n\to\infty} (X_n + Y_n) \quad \text{and} \quad X\,Y \stackrel{\text{a.s.}}{=} \lim_{n\to\infty} X_n\,Y_n. \tag{2.7.8}$$

In particular, the limits (2.7.8) allow us to work with a rich variety of equations that involve almost surely asymptotically determined random variables.

The *Borel-Cantelli Lemma* states for a sequence of events $A_1, A_2, \ldots$ in $\mathcal{A}$ that

(i) if $\sum_{k=1}^{\infty} P(A_k) < \infty$, then the event that consists of the realization of infinitely many of the events $A_1, A_2, \ldots$ has probability zero, that is

$$P\big(\omega : \text{ there exists a } j \in \mathcal{N} \text{ such that } \ \omega \in A_i \text{ for all } i \geq j\big) = 0$$

(ii) if $\sum_{k=1}^{\infty} P(A_k) = \infty$ and $A_1, A_2, \ldots$ are independent, then the event that consists of the realization of infinitely many of the events $A_1, A_2, \ldots$ has probability one, that is

$$P\big(\omega : \text{ there exists a } j \in \mathcal{N} \text{ such that } \ \omega \in A_i \text{ for all } i \geq j\big) = 1.$$

Limits under Expectation (*)

For many results in quantitative finance one needs to deal with limits and expectations. Therefore, let us formulate some fundamental results, which allow us to interchange limits and expectations, see Shiryaev (1984). First we mention the *Monotone Convergence Theorem.*

Theorem 2.7.1. (Monotone Convergence) *Let $Y, X, X_1, X_2, \ldots$ be random variables.*

(i) If $X_n \geq Y$ for all $n \in \mathcal{N}$, $E(Y) > -\infty$ and the sequence $(X_n)_{n \in \mathcal{N}}$ is monotone increasing, where $\lim_{n\to\infty} X_n \stackrel{a.s.}{=} X$, then

$$\lim_{n\to\infty} E(X_n) = E(X). \tag{2.7.9}$$

(ii) If $X_n \leq Y$ for all $n \in \mathcal{N}$, $E(Y) < \infty$ and the sequence $(X_n)_{n \in \mathcal{N}}$ is monotone decreasing, where $\lim_{n\to\infty} X_n \stackrel{a.s.}{=} X$, then

$$\lim_{n\to\infty} E(X_n) = E(X). \tag{2.7.10}$$

Let us denote by *lim inf* the lower limit and by *lim sup* the upper limit when a sequence has several limits. The following fundamental result on inequalities when interchanging limits and expectations is known as *Fatou's Lemma.*

Lemma 2.7.2. (Fatou) *Let $Y, X_1, X_2, \ldots$ be random variables.*

(i) If $X_n \geq Y$ for all $n \in \mathcal{N}$ and $E(Y) > -\infty$, then

$$E\left(\liminf_{n\to\infty} X_n\right) \leq \liminf_{n\to\infty} E(X_n). \tag{2.7.11}$$

(ii) If $X_n \leq Y$ for all $n \in \mathcal{N}$ and $E(Y) < \infty$, then

$$\limsup_{n\to\infty} E(X_n) \leq E\left(\limsup_{n\to\infty} X_n\right). \tag{2.7.12}$$

(iii) If $|X_n| \leq Y$ for all $n \in \mathcal{N}$ and $E(Y) < \infty$, then

$$E\left(\liminf_{n\to\infty} X_n\right) \leq \liminf_{n\to\infty} E(X_n) \leq \limsup_{n\to\infty} E(X_n) \leq E\left(\limsup_{n\to\infty} X_n\right). \tag{2.7.13}$$

We now mention *Lebesgue's Dominated Convergence Theorem* that underpins a range of practically important results in quantitative finance.

Theorem 2.7.3. (Lebesgue) *Let $Y, X, X_1, X_2, \ldots$ be random variables such that $|X_n| \leq Y$, $E(Y) < \infty$ and $\lim_{n\to\infty} X_n \stackrel{a.s.}{=} X$, then we have*

$$E(|X|) < \infty, \tag{2.7.14}$$

$$\lim_{n\to\infty} E(|X_n - X|) = 0 \tag{2.7.15}$$

and thus

$$\lim_{n\to\infty} E(X_n) = E(X). \tag{2.7.16}$$

Definition 2.7.4. *A family $(X_n)_{n\in\mathcal{N}}$ of random variables is said to be* uniformly integrable *if*

$$\lim_{q\to\infty} \left(\sup_{n\in\mathcal{N}} E\left(|X_n| \, I_{\{|X_n|>q\}} \right) \right) = 0. \tag{2.7.17}$$

Obviously, if $|X_n| \le Y$ for $n \in \mathcal{N}$ and $E(Y) < \infty$, then the family of random variables $(X_n)_{n\in\mathcal{N}}$ is uniformly integrable. This observation allows the proof of Fatou's Lemma and Lebesgue's Dominated Convergence Theorem by the following general result.

Theorem 2.7.5. *Let X, X_1, X_2, ... denote nonnegative random variables with $\lim_{n\to\infty} X_n \stackrel{a.s.}{=} X$ and $E(X_n) < \infty$ for all $n \in \mathcal{N}$, then it holds*

$$\lim_{n\to\infty} E(X_n) = E(X) \tag{2.7.18}$$

if and only if the family of random variables $(X_n)_{n\in\mathcal{N}}$ is uniformly integrable.

Extreme Value Theorem (*)

The understanding of the occurrence of extreme losses is the key to many risk management problems. It is important, for instance, in Value at Risk analysis and the evaluation of catastrophic insurance claims. We call a random variable or distribution *nondegenerate* if its probability is not concentrated in one single value. Let us assume that catastrophic insurance losses, say for hurricanes, are modeled by an i.i.d. sequence of nondegenerate random variables $X_1, X_2, \ldots$ with distribution function F_X. Given a number $n \in \{1, 2, \ldots\}$ of loss data $X_1, X_2, \ldots, X_n$ we are interested in the maximum

$$M_n = \max(X_1, X_2, \ldots, X_n) \tag{2.7.19}$$

of these losses. The minimum can be studied using a maximum by transforming the sequence such that

$$\min(X_1, X_2, \ldots, X_n) = -\max(-X_1, -X_2, \ldots, -X_n). \tag{2.7.20}$$

The following theorem is known as the *Extreme Value Theorem* or *Fisher-Tippett Theorem*, see Embrechts, Klüppelberg & Mikosch (1997) for further details.

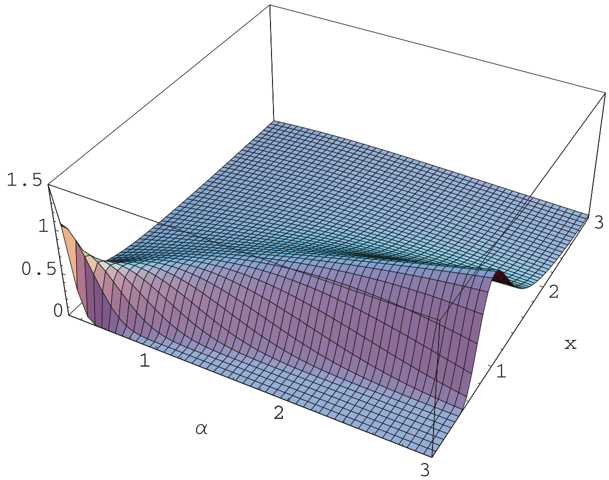

Fig. 2.7.1. Density of the Fréchet distribution in dependence on α

Theorem 2.7.6. (Fisher-Tippett) *If there exist sequences of norming constants $c_n > 0$, $d_n \in \Re$ for $n \in \{2, 3, \ldots\}$ and some random variable H with nondegenerate distribution function F_H such that*

$$\lim_{n \to \infty} (c_n \, M_n + d_n) \stackrel{d}{=} H, \tag{2.7.21}$$

then F_H belongs to the type of one of the following three distribution functions:

(i) *the* Fréchet distribution

$$F_H(x) = \begin{cases} 0 & \text{for} \quad x \leq 0, \\ \exp\{-x^{-\alpha}\} & \text{for} \quad x > 0, \end{cases} \tag{2.7.22}$$

with $\alpha > 0$,

(ii) *the* Weibull distribution

$$F_H(x) = \begin{cases} \exp\{-(-x)^{\alpha}\} & \text{for} \quad x \leq 0, \\ 1 & \text{for} \quad x > 0, \end{cases} \tag{2.7.23}$$

with $\alpha > 0$ or

(iii) *the* Gumbel distribution

$$F_H(x) = \exp\{-e^{-x}\} \tag{2.7.24}$$

for $x \in \Re$.

This is a remarkable and rather fundamental result. Whatever underlying loss distribution is given, there is only one of the above three limit distributions that is possible for its maxima. In Fig. 2.7.1, we show the density of the

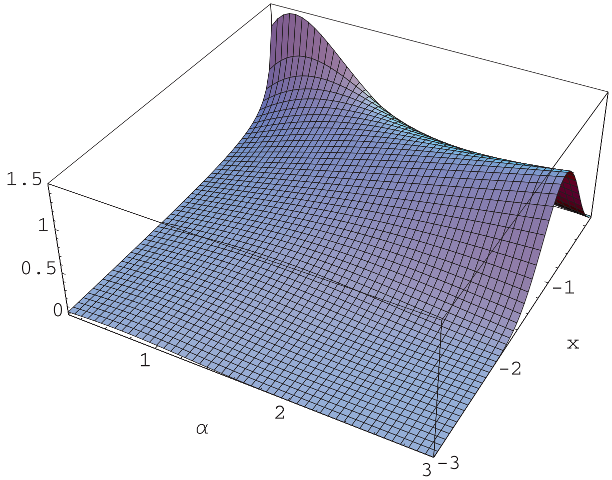

Fig. 2.7.2. Density of the Weibull distribution

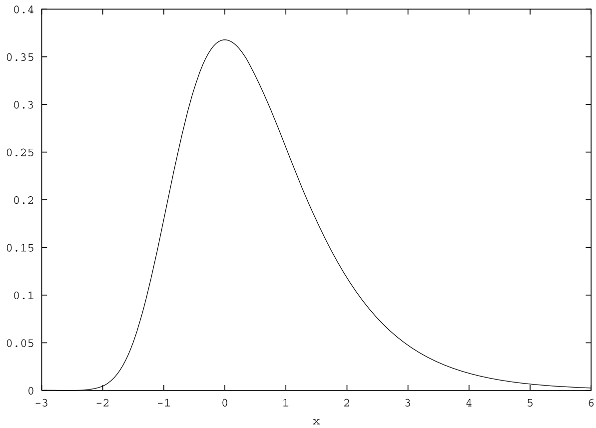

Fig. 2.7.3. Density of the Gumbel distribution

Fréchet distribution in dependence on the parameter α. Note that the maxima have here only positive values. Figure 2.7.2 displays the Weibull density. This density captures only negative maxima. For the Gumbel distribution the density is plotted in Fig. 2.7.3. In this case the maxima can be positive or negative.

The Extreme Value Theorem raises the question whether a distribution F_X of losses is in the domain of attraction of a given extreme value distribution. For a detailed answer we refer to Embrechts et al. (1997). However, let us mention that appropriately normalized maxima of the Cauchy distribution are, for instance, in the domain of attraction of the Fréchet distribution. Normalized minima of the Student t distribution correspond to the Weibull distribution. Adequately normalized maxima of the normal, gamma, lognormal and exponential distribution are captured by the Gumbel distribution.

It is important to know for risk management purposes which extreme value distribution attracts a given loss distribution. This knowledge is extremely useful, for instance, in insurance premium calculations or in Value at Risk analysis. One knows from Theorem 2.7.6 that one needs only to consider one of the described three extreme value distributions in a specific application when one deals with extreme values.

2.8 Exercises for Chapter 2

2.1. Consider a sequence of independent random variables $X_1, X_2, \ldots$ with mean μ and variance $\mathrm{Var}(X_i) \leq K < \infty$. What is the almost sure limit of the sample mean $\hat{\mu}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$ of that sequence?

2.2. For a sequence of independent identically distributed random variables $X_1, X_2, \ldots$ with mean $\mu \in \Re$ and variance $\sigma^2 > 0$ characterize for the sequence of random variables

$$\hat{Y}_n = \sqrt{n}\left(\frac{1}{n}\sum_{i=1}^{n} X_i - \mu\right)$$

for $n \to \infty$ the limit.

2.3. Consider a Gaussian random variable Z with known mean $E(Z)$ and known variance $\mathrm{Var}(Z)$. Provide the 99% confidence interval for $2Z$.

2.4. For the random variable Z in Exercise 2.3 calculate the one sided confidence interval with 99% confidence such that Z is at least as large as $-\mathrm{VaR}((1-\alpha)\%)$, that is

$$P(Z \geq -\mathrm{VaR}((1-\alpha)\%)) = 1 - \alpha.$$

http://www.springer.com/978-3-540-26212-1

A Benchmark Approach to Quantitative Finance
Platen, E.; Heath, D.
2006, XVI, 700 p. 199 illus., Hardcover
ISBN: 978-3-540-26212-1