Chapter 2
Statistics of Financial Time Series

The price of a stock as a function of time constitutes a financial time series, and as such it contains an element of uncertainty which demands the use of statistical methods for its analysis. However, the plot of the price history of a stock in general resembles an exponential curve, and a time series of exponential terms is mathematically hard to manipulate and has little information to give even from classical functional transformations (e.g., the derivative of an exponential is again exponential). Campbell et al. (1997 Sect. 1.4) give two solid reasons for preferring to focus the analysis on returns rather than directly on prices. One reason is that financial markets are almost perfectly competitive, which implies that price is not affected by the size of the investment; hence, what is left for the investor to gauge is the rate of benefit that could be derived from his investment, and for that matter it is best to work with a size–independent summary of the investment opportunity, which is what the return represents. The second reason is that returns have more manageable statistical properties than prices, such as stationarity and ergodicity, being the case more often than not of dynamic general equilibrium models that give non stationary prices but stationary returns. Therefore, this chapter's main concern is the study of returns and their fundamental statistical properties. We briefly review some of the fundamental concepts of statistics, such as: moments of a distribution, distribution and density functions, likelihood methods, and other tools that are necessary for the analysis of returns and, in general, financial time series.

2.1 Time Series of Returns

Begin by considering the curve drawn by the price history of the Dow Jones Industrial Average (DJIA) from 1960 to 2010 . This can be obtained with the following R commands.

```
> require(quantmod)
> getSymbols("DJIA",src="FRED") ##DJIA from 1896-May
> plot(DJIA['1960/2010'],main="DJIA")
```

A. Arratia, *Computational Finance,* Atlantis Studies in Computational Finance and Financial Engineering 1, DOI: 10.2991/978-94-6239-070-6_2,

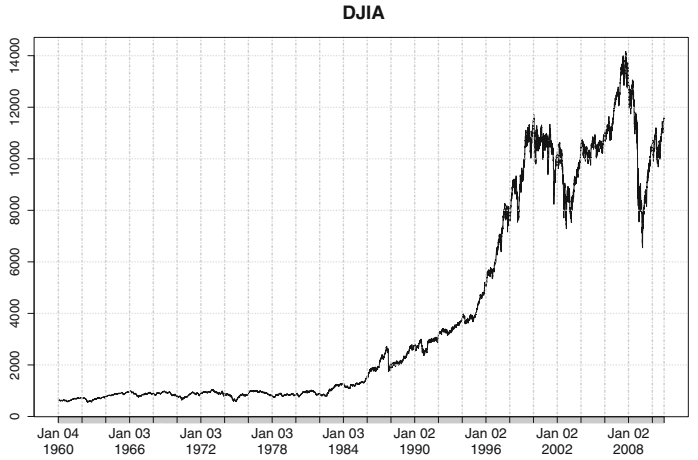

Fig. 2.1 Dow Jones Industrial Average from 1960 to 2010

The resulting picture can be seen in Fig. 2.1. One can observe that from 1960 to about 1999 the DJIA has an exponential shape; from 2000 to 2002 it has an exponential decay, to rise up exponentially again until 2008, and so on. This might suggest to fit an exponential function to this curve to forecast its future values. We show you how to do this in the R Lab 2.7.2. However, as argued in the introduction to this chapter it is best to turn to returns to analyze a stock's behavior through time. There are many variants of the definition of returns, according to whether we allow some extra parameters to be considered in their calculation, like dividends or costs of transactions. Thus, we first look at the most simple definition of returns where only the price is considered.

Definition 2.1 *(Simple return)* Let P_t be the price of an asset at time t. Given a time scale τ, the *τ-period simple return at time t*, $R_t(\tau)$, is the rate of change in the price obtained from holding the asset from time $t-\tau$ to time t:

$$R_t(\tau) = \frac{P_t - P_{t-\tau}}{P_{t-\tau}} = \frac{P_t}{P_{t-\tau}} - 1 \tag{2.1}$$

The *τ-period simple gross return at time t* is $R_t(\tau)+1$. If $\tau = 1$ we have a one-period simple return (respectively, a simple gross return), and denote it R_t (resp., $R_t + 1$). □

There is a practical reason for defining returns backwards (i.e. from time $t-\tau$ to t, as opposed to from t to $t+\tau$), and it is that more often than not we want to know the return obtained today for an asset bought some time in the past. Note that return values range from -1 to ∞; so, in principle, you can not loose more than what you've invested, but you can have unlimited profits.

Example 2.1 Consider the daily closing prices of Apple Inc. (AAPL:Nasdaq) throughout the week of July 16 to July 20, 2012[1]:

	1	2	3	4	5
Date	2012/07/16	2012/07/17	2012/07/18	2012/07/19	2012/07/20
Price	606.91	606.94	606.26	614.32	604.30

Let us refer to the dates in the table by their positions from left to right, i.e., as 1, 2, 3, 4 and 5. Then, the simple return from date 3 to date 4 is

$$R_4 = (614.32 - 606.26)/606.26 = 0.0133.$$

The return from day 1 to day 5 (a 4-period return) is $R_5(4) = (604.30 - 606.91)/606.91 = -0.0043$. The reader should verify that

$$1 + R_5(4) = (1 + R_2)(1 + R_3)(1 + R_4)(1 + R_5) \tag{2.2}$$

Equation (2.2) is true in general:

Proposition 2.1 *The τ-period simple gross return at time t equals the product of τ one-period simple gross returns at times $t - \tau + 1$ to t.*

Proof

$$\begin{aligned} R_t(\tau) + 1 &= \frac{P_t}{P_{t-\tau}} = \frac{P_t}{P_{t-1}} \cdot \frac{P_{t-1}}{P_{t-2}} \cdots \frac{P_{t-\tau+1}}{P_{t-\tau}} \\ &= (1 + R_t) \cdot (1 + R_{t-1}) \cdots (1 + R_{t-\tau+1}) \end{aligned} \tag{2.3}$$

For this reason these multiperiod returns are known also as *compounded* returns. □

Returns are independent from the magnitude of the price, but they depend on the time period τ, which can be minutes, days, weeks or any time scale, and always expressed in units. Thus, a return of 0.013, or in percentage terms of 1.3 %, is an incomplete description of the investment opportunity if the return period is not specified. One must add to the numerical information the time span considered, if daily, weekly, monthly and so on. If the time scale is not given explicitly then it is customary to assumed to be of one year, and that we are talking about an *annual rate of return*. For τ years returns these are re-scale to give a comparable one-year return. This is the *annualized* (or average) return, defined as

$$\textit{Annualized}\ (R_t(\tau)) = \left(\prod_{j=0}^{\tau-1} (1 + R_{t-j}) \right)^{1/\tau} - 1 \tag{2.4}$$

[1] *source* `http://finance.yahoo.com/q/hp?s=AAPL+Historical+Prices`

It is the geometric mean of the τ one-period simple gross returns. This can also be computed as an arithmetic average by applying the exponential function together with its inverse, the natural logarithm, to get:

$$\textit{Annualized}\,(R_t(\tau)) = \exp\left(\frac{1}{\tau}\sum_{j=0}^{\tau-1}\ln(1+R_{t-j})\right) - 1 \tag{2.5}$$

This equation expresses the annualized return as the exponential of a sum of logarithms of gross returns. The logarithm of a gross return is equivalent, in financial terms, to the continuous compounding of interest rates. To see the link, recall the formula for pricing a risk–free asset with an annual interest rate r which is continuously compounded (Chap. 1, Eq. (1.3)): $P_n = P_0 e^{rn}$, where P_0 is the initial amount of the investment, P_n the final net asset value, and n the number of years. Setting $n = 1$, then $r = \ln(P_1/P_0) = \ln(R_1 + 1)$. For this reason, the logarithm of (gross) returns is also known as *continuously compounded returns*.

Definition 2.2 *(log returns)* The *continuously compounded return* or *log return* r_t of an asset is defined as the natural logarithm of its simple gross return:

$$r_t = \ln(1+R_t) = \ln\left(\frac{P_t}{P_{t-1}}\right) = \ln P_t - \ln P_{t-1} \tag{2.6}$$

Then, for a τ-period log return, we have

$$\begin{aligned} r_t(\tau) &= \ln(1+R_t(\tau)) = \ln((1+R_t)(1+R_{t-1})\cdots(1+R_{t-\tau+1})) \\ &= \ln(1+R_t) + \ln(1+R_{t-1}) + \cdots + \ln(1+R_{t-\tau+1}) \\ &= r_t + r_{t-1} + \cdots + r_{t-\tau+1} \end{aligned} \tag{2.7}$$

which says that the continuously compounded τ-period return is simply the sum of τ many continuously compounded one-period returns.

Besides turning products into sums, and thus making arithmetic easier, the log returns are more amenable to statistical analysis than the simple gross return because it is easier to derive the time series properties of additive processes than those of multiplicative processes. This will become more clear in Sect. 2.2.

Example 2.2 Consider the same data from Example 2.1 and now, for your amazement, compute the continuously compounded, or log return, from date 3 to date 4: $r_4 = \ln(614.32) - \ln(606.26) = 0.0132$. And the 4-period log return

$$r_5(4) = \ln(604.3) - \ln(606.91) = -0.0043.$$

Observe how similar these values are to the simple returns computed on same periods. Can you explain why? (Hint: recall from Calculus that for small x, i.e., $|x| < 1$, $\ln(1+x)$ behaves as x.) □

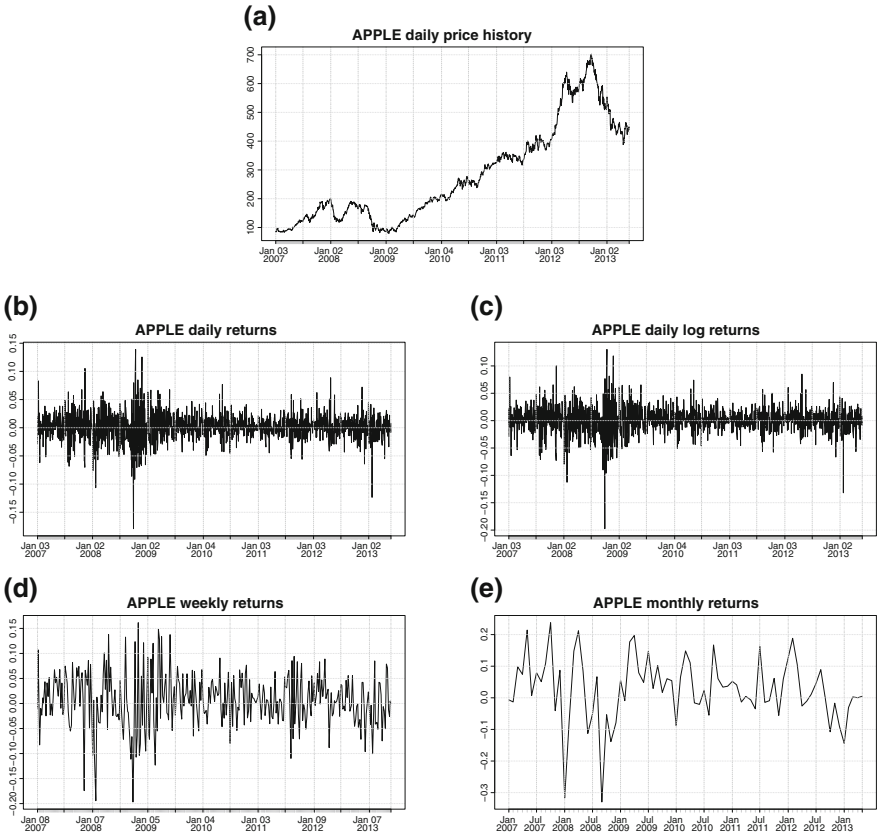

Fig. 2.2 Apple (Nasdaq) 2007–2013. **a** daily price; **b** daily return; **c** daily log return; **d** weekly return; **e** monthly return

Figure 2.2 shows different plots of Apple shares historical data from year 2007 to 2013. Again we can see an exponential behavior in the plot of price (a). Note the similarities between plots (b) and (c) corresponding to daily simple returns and log returns. Also observe that at any time scale the return series shows a high degree of variability, although this variability seems similar over several time periods.

The different period returns are easily computed with R's `quantmod` package. For example Fig. 2.2b can be obtained with

```
> getSymbols("AAPL",src="yahoo") ##data starts from 2007
> aplRd = periodReturn(AAPL,period="daily")
> plot(aplRd, main="APPLE daily returns")
```

Returns with dividends. If the asset pays dividends periodically then the simple return must be redefined as follows

$$R_t = \frac{P_t + D_t}{P_{t-1}} - 1 \tag{2.8}$$

where D_t is the dividend payment of the asset between dates $t-1$ and t. For multi-period and continuously compounded returns the modifications of the corresponding formulas to include dividends are similar and are left as exercises.

Remark 2.1 **(On adjusted close prices and dividends)** Some providers of financial data include in their stock quotes the adjusted close price (cf. Chap. 1, Sect. 1.1.2). Since this price is adjusted to dividend payments and other corporate actions on the stock, it is then more accurate to work with the adjusted close price when analyzing historical returns. □

Excess return. It is the difference between the return of an asset A and the return of a reference asset O, usually at a risk-free rate. The simple excess return on asset A would be then $Z_t^A = R_t^A - R_t^O$; and the logarithmic excess return is $z_t^A = r_t^A - r_t^O$. The excess return can be thought of as the payoff of a portfolio going long in the asset and short on the reference.

Portfolio return. Let $\mathscr{P}$ be a portfolio of N assets, and let n_i be the number of shares of asset i in $\mathscr{P}$, for $i = 1, \ldots, N$. Then, at a certain time t, the net value of $\mathscr{P}$ is $P_t^{\mathscr{P}} = \sum_{i=1}^N n_i P_t^i$, where P_t^i is the price of asset i at time t. Applying Eq. (2.3), we get that the τ-period simple return of $\mathscr{P}$ at time t, denoted $R_t^{\mathscr{P}}(\tau)$, is given by

$$R_t^{\mathscr{P}}(\tau) = \sum_{i=1}^{N} w_i R_t^i(\tau) \tag{2.9}$$

where $R_t^i(\tau)$ is the τ-period return of asset i at time t, and $w_i = (n_i P_{t-\tau}^i)/(\sum_{j=1}^N n_j P_{t-\tau}^j)$ is the weight, or proportion, of asset i in $\mathscr{P}$. Therefore, the simple return of a portfolio is a weighted sum of the simple returns of its constituent assets. This nice property does not holds for the continuously compounded return of $\mathscr{P}$, since the logarithm of a sum is not the sum of logarithms. Hence, when dealing with portfolios we should prefer returns to log returns, even though empirically when returns are measured over short intervals of time, the continuously compounded return on a portfolio is close to the weighted sum of the continuously compounded returns on the individual assets (e.g., look back to Example 2.2 where the log returns almost coincide with the simple returns). We shall be dealing more in depth with the financial analysis of portfolios in Chap. 8. In the remaining of this book we shall refer to simple returns just as returns (and use R_t to denote this variable); and refer to continuously compounded return just as log returns (denoted r_t).

2.2 Distributions, Density Functions and Moments

We depart from the fact that a security's time series of returns are random variables evolving over time, i.e., a *random process*, and that usually the only information we have about them is some sample observations. Therefore, in order to build some statistical model of returns, we must begin by specifying some probability distribution or, stepping further, by treating returns as continuous random variables, specify a probability density function, and from this function obtain a quantitative description of the shape of the distribution of the random values by means of its different moments. This will give us a greater amount of information about the behavior of returns from which to develop a model.

2.2.1 Distributions and Probability Density Functions

The *cumulative distribution function* (CDF) of a random variable X is defined, for all $x \in \mathbb{R}$, as $F_X(x) = \mathbb{P}(X \leq x)$. F_X is said to be continuous if its derivative $f_X(x) = F'_X(x)$ exists; in which case f_X is the *probability density function* of X, and F_X is the integral of its derivative by the fundamental theorem of calculus, that is

$$F_X(x) = \mathbb{P}(X \leq x) = \int_{-\infty}^{x} f_X(t)dt \tag{2.10}$$

When F_X is continuous then one says that the random variable X is continuous. On the other hand, X is discrete if F_X is a step function, which then can be specified, using the *probability mass function* $p_X(x) = \mathbb{P}(X = x)$, as

$$F_X(x) = \mathbb{P}(X \leq x) = \sum_{\{k \leq x : p_X(k) > 0\}} p_X(k).$$

The CDF F_X has the following properties: is non-decreasing; $\lim_{x \to -\infty} F_X(x) = 0$ and $\lim_{x \to \infty} F_X(x) = 1$. We can then estimate $F_X(x)$ from an observed sample $x_1, \ldots, x_n$ by first ordering the sample as $x_{(1)} \leq x_{(2)} \leq \cdots \leq x_{(n)}$ and then defining the *empirical cumulative distribution function* ECDF:

$$F_n(x) = (1/n) \sum_{i=1}^{n} \mathbf{1}\{x_{(i)} \leq x\}$$

where $\mathbf{1}\{A\}$ is the indicator function, whose value is 1 if event A holds or 0 otherwise. Observe that the ECDF give the proportion of sample points in the interval $(-\infty, x]$. If it were the case that the CDF F_X is strictly increasing (and continuous) then it is invertible, and we can turn around the problem of estimating $F_X(x)$ and find, for a

given $q \in [0, 1]$, the real number x_q such that $q = F_X(x_q) = \mathbb{P}(X \leq x_q)$; that is, for a given proportion q find the value x_q of X such that the proportion of observations below x_q is exactly q. This x_q is called the *q-quantile* of the random variable X with distribution F_X. However, in general, F_X is not invertible so it is best to define the q quantile of the random variable X as the smallest real number x_q such that $q \leq F_X(x_q)$, that is, $x_q = \inf\{x: q \leq F_X(x)\}$.

When it comes to estimating quantiles one works with the inverse ECDF of a sample of X or other cumulative distribution of the order statistics. It is not a trivial task, apart from the 0.5 quantile which is estimated as the *median* of the distribution. But in R we have a good sample estimation of quantiles implemented by the function `quantile`.

R Example 2.1 Consider the series of returns for the period 06/01/2009–30/12/2009 of Allianz (ALV), a German company listed in Frankfurt's stock market main index DAX. Assume the data is already in a table named `daxR` under a column labelled `alvR`. The instructions to build this table are given in the R Lab 2.7.6. After loading the table in your work space, run the commands:

```
> alv=na.omit(daxR$alvR)
> quantile(alv,probs=c(0,1,0.25,0.5,0.75))
```

This outputs the extreme values of the series (the 0 and 100 % quantiles) and the *quartiles*, which divide the data into four equal parts. The output is:

```
          0%         100%          25%          50%          75%
-0.098800569  0.124268348 -0.013133724  0.001246883  0.017879936
```

The first quartile (at 25 %) indicates the highest value of the first quarter of the observations in a non-decreasing order. The second quartile (at 50 %) is the median, or central value of the distribution, and so on. □

Now, if we want to deal with two continuous random variables jointly, we look at their joint distribution function $F_{X,Y}(x, y) = \mathbb{P}(X \leq x; Y \leq y)$, which can be computed using the joint density function $f_{X,Y}(x, y)$ if this exists (i.e. the derivative of $F_{X,Y}(x, y)$ exists) by

$$F_{X,Y}(x, y) = \mathbb{P}(X \leq x; Y \leq y) = \int_{-\infty}^{y} \int_{-\infty}^{x} f_{X,Y}(s, t) ds dt \tag{2.11}$$

Also with the joint density function $f_{X,Y}$ we can single handle either X or Y with their *marginal probability densities*, which are obtained for each variable by integrating out the other. Thus, the marginal probability density of X is

$$f_X(x) = \int_{-\infty}^{\infty} f_{X,Y}(x, y) dy \tag{2.12}$$

and similarly, the marginal probability density of Y is $f_Y(y) = \int_{-\infty}^{\infty} f_{X,Y}(x, y) dx$.

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

$$F_{X|Y\leq y}(x) = \frac{\mathbb{P}(X \leq x; Y \leq y)}{\mathbb{P}(Y \leq y)}$$

Again, if the corresponding probability density functions exist, we have the *conditional density of X given $Y = y$*, $f_{X|Y=y}(x)$, obtained by

$$f_{X|Y=y}(x) = \frac{f_{X,Y}(x, y)}{f_Y(y)} \tag{2.13}$$

where the marginal density $f_Y(y)$ is as in Eq. (2.12). From (2.13), we can express the joint density of X and Y in terms of their conditional densities and marginal densities as follows:

$$f_{X,Y}(x, y) = f_{X|Y=y}(x) f_Y(y) = f_{Y|X=x}(y) f_X(x) \tag{2.14}$$

Equation (2.14) is a very important tool in the analysis of random variables and we shall soon see some of its applications. Right now, observe that X and Y are independent if and only if $f_{X|Y=y}(x) = f_X(x)$ and $f_{Y|X=x}(y) = f_Y(y)$. Therefore,

Two random variables X and Y are *independent* if and only if their joint density is the product of their marginal densities; i.e., for all x and y,

$$f_{X,Y}(x, y) = f_X(x) f_Y(y) \tag{2.15}$$

A more general notion of independence is the following:

X and Y are independent if and only if their joint distribution is the product of the CDF's of each variable; i.e., for all x and y,

$$F_{X,Y}(x, y) = F_X(x) F_Y(y) \tag{2.16}$$

Eq. (2.16) is stronger than Eq. (2.15) since it holds even if densities are not defined.

2.2.2 Moments of a Random Variable

Given a continuous random variable X with density function f, the *expected value* of X, denoted $E(X)$, is

$$E(X) = \int_{-\infty}^{\infty} x f(x) dx. \tag{2.17}$$

If X is discrete then (2.17) reduces to $E(X) = \sum_{\{x: f(x)>0\}} x f(x)$.

Let $\mu_X = E(X)$. Then μ_X is also called the mean or first moment of X. The geometric intuition is that the first moment measures the central location of the distribution.

More general, the *n-th moment* of a continuous random variable X is

$$E(X^n) = \int_{-\infty}^{\infty} x^n f(x)dx, \tag{2.18}$$

and the *n-th central moment* of X is defined as

$$E((X - \mu_X)^n) = \int_{-\infty}^{\infty} (x - \mu_X)^n f(x)dx \tag{2.19}$$

The second central moment of X is called the *variance* of X,

$$Var(X) = E((X - \mu_X)^2) \tag{2.20}$$

and most frequently denoted by σ_X^2. The variance measures the variability of X (quantified as the distance from the mean). Observe that $\sigma_X^2 = Var(X) = E(X^2) - \mu_X^2$. The term $\sigma_X = \sqrt{E((X - \mu_X)^2)}$ is the *standard deviation* of X.

The third and fourth central moments are known respectively as the *skewness* (denoted $S(X)$) and the *kurtosis* ($K(X)$), which measure respectively the extent of asymmetry and tail thickness (amount of mass in the tails) of the distribution. These are defined as

$$S(X) = E\left(\frac{(X - \mu_X)^3}{\sigma_X^3}\right) \quad \text{and} \quad K(X) = E\left(\frac{(X - \mu_X)^4}{\sigma_X^4}\right) \tag{2.21}$$

Estimation of moments. In practice one has observations of a random variable X and from these one does an estimation of X's moments of distribution. Given sample data $\boldsymbol{x} = \{x_1, \ldots, x_m\}$ of random variable X, the *sample mean* of X is

$$\widehat{\mu}_{\boldsymbol{x}} = \frac{1}{m}\sum_{t=1}^{m} x_t, \tag{2.22}$$

and the *sample variance* is

$$\widehat{\sigma}_{\boldsymbol{x}}^2 = \frac{1}{m-1}\sum_{t=1}^{m}(x_t - \widehat{\mu}_{\boldsymbol{x}})^2. \tag{2.23}$$

The *sample standard deviation* is $\widehat{\sigma}_{\boldsymbol{x}} = \sqrt{\widehat{\sigma}^2_{\boldsymbol{x}}}$. The *sample skewness* is

$$\widehat{S}_{\boldsymbol{x}} = \frac{1}{(m-1)\widehat{\sigma}^3_{\boldsymbol{x}}} \sum_{t=1}^{m} (x_t - \widehat{\mu}_{\boldsymbol{x}})^3, \tag{2.24}$$

and the *sample kurtosis*

$$\widehat{K}_{\boldsymbol{x}} = \frac{1}{(m-1)\widehat{\sigma}^4_{\boldsymbol{x}}} \sum_{t=1}^{m} (x_t - \widehat{\mu}_{\boldsymbol{x}})^4 \tag{2.25}$$

Remark 2.2 The above estimators constitute *the basic statistics* for X. The sample mean and variance are *unbiased* estimators of their corresponding moments, in the sense that for each estimator $\widehat{\theta}_{\boldsymbol{x}}$ and corresponding moment θ_X, we have $E(\widehat{\theta}_{\boldsymbol{x}}) = \theta_X$. This does not hold for sample skewness and kurtosis, and so these are said to be biased estimators. From now on, whenever X or its sample $\boldsymbol{x}$ are clear from context we omit them as subscripts. □

R Example 2.2 In R Lab 2.7.6 we guide the reader through the R instructions to compute some basic statistics for a group of German stocks. We chose to analyze the four companies Allianz (ALV), Bayerische Motoren Werke (BMW), Commerzbank (CBK) and Thyssenkrupp (TKA). The returns from dates 06/01/2009 to 30/12/2009 for these stocks are placed in a table labelled `daxR`. Using the `basicStats()` command, from the package `fBasics` we get the following results. (A note of caution: the kurtosis computed by `basicStats()` is the *excess kurtosis*, that is, $K(X) - 3$. We will explain later what this means. Hence, to get the real kurtosis add 3 to the values given in the table.)

```
> basicStats(na.omit(daxR[,2:5]))
                  alvR       bmwR       cbkR       tkaR
nobs        251.000000 251.000000 251.000000 251.000000
NAs           0.000000   0.000000   0.000000   0.000000
Minimum      -0.098801  -0.077849  -0.169265  -0.081459
Maximum       0.124268   0.148430   0.187359   0.162539
1. Quartile  -0.013134  -0.015040  -0.025762  -0.015535
3. Quartile   0.017880   0.019219   0.020863   0.020753
Mean          0.001339   0.001988   0.000844   0.002272
Median        0.001247  -0.000389  -0.001709   0.000437
Sum           0.336189   0.498987   0.211965   0.570381
SE Mean       0.001880   0.001866   0.003071   0.002162
LCL Mean     -0.002363  -0.001687  -0.005204  -0.001986
UCL Mean      0.005041   0.005663   0.006893   0.006531
Variance      0.000887   0.000874   0.002367   0.001173
Stdev         0.029780   0.029563   0.048652   0.034253
Skewness      0.236224   0.663122   0.392493   0.618779
Kurtosis      2.060527   2.911924   2.126179   2.038302
```

You can also get the particular statistics for each individual return series (and here the function `kurtosis()` gives the real kurtosis, not the excess).

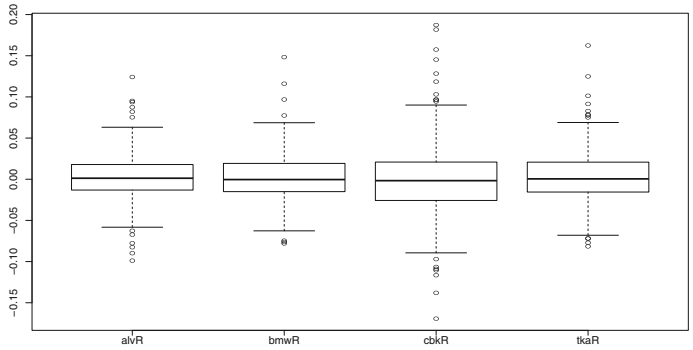

Fig. 2.3 Box plots for ALV, BMW, CBK and TKA (stocks listed in DAX)

```
> alvR <- na.omit(daxR$alvR)
> mean(alvR) ## the mean
> var(alvR) ## variance
> sd(alvR) ## standard deviation
> kurtosis(alvR, method="moment") #gives real kurtosis
> skewness(alvR)
```

We explain the terms in the table of `basicStats` that are not immediately clear: `nobs` is the number of observations; `Sum` is the sum of all `nobs` observations (hence, for example, note that `Mean` = `Sum`/`nobs`); `SE Mean` is the standard error for the mean, which is computed as the standard deviation (`Stdev`) divided by the square root of `nobs`; `LCL Mean` and `UCL Mean` are the Lower and Upper Control Limits for sample means, computed by the formulas:

$$\texttt{LCL} = \texttt{Mean} - 1.96 \cdot \texttt{SE} \quad \text{and} \quad \texttt{UCL} = \texttt{Mean} + 1.96 \cdot \texttt{SE} \tag{2.26}$$

and represent the lower and upper limit of a *confidence band*.[2]

A better way to "see" the results is through a *box plot*:

```
> boxplot(daxR[,2:5])
```

The resulting picture is in Fig. 2.3.

A box plot (proposed by John Tukey in 1977) gives a description of numerical data through their quartiles. Specifically, for the R function `boxplot( )`, the bottom and top of the box are the first quartile (Q_1) and the third quartile (Q_3), and the line across the box represents the second quartile (Q_2), i.e., the median. The top end of the whisker (represented by the dotted line) is the largest value $M < Q_3 + 1.5(Q_3 - Q_1)$; the low end of the whisker is the smallest value $m > Q_1 - 1.5(Q_3 - Q_1)$. The dots above and below the whisker (i.e. outside $[m, M]$) represent outliers. Inside the box

[2] Confidence refers to the following: if the probability law of the sample mean is approximately *normal*, then within these LCL and UCL bounds lie approximately 95 % of sample means taken over `nobs` observations. We shall discuss normality in the next pages.

there is 50 % of the data split evenly across the median; hence if, for example, the median is slightly up this is indication of major concentration of data in that region.

Viewing the box plots and the numeric values we can infer some characteristics of the behavior of stock's daily returns. We can see that the daily return of a stock presents:

- small (sample) median;
- more smaller (in value) returns than larger returns;
- some non null skewness, hence some gross symmetry;
- a considerable number of outliers (approx. 5 % shown in the box plots), and a high kurtosis. In fact, a positive excess kurtosis, for recall that `basicStats` gives $\widehat{K}_{\boldsymbol{x}} - 3$ for the `Kurtosis` entry.

These observations suggest that stocks' returns are distributed more or less evenly around the median, with decaying concentration of values as we move away from the median, and almost symmetrical. We can corroborate this suggestion by plotting a histogram of one of these return series and observing that the rectangles approximate a bell-shaped area (using the data from R Example 2.2):

```
> alvR <- na.omit(daxR$alvR)
> hist(alvR,probability=T,xlab="ALV.DE returns",main=NULL)
```

Consequently, as a first mathematical approximation to model the distribution of stock returns we shall study the benchmark of all bell-shaped distributions.

2.2.3 The Normal Distribution

The most important and recurred distribution is the *normal* or *Gaussian* distribution. The normal distribution has probability density function

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp(-(x-\mu)^2/2\sigma^2), \tag{2.27}$$

for $-\infty < x < \infty$, and with $\sigma > 0$ and $-\infty < \mu < \infty$. It is completely determined by the parameters μ and σ^2, that is by the mean and the variance, and so, we could write $f(x)$ as $f(x; \mu, \sigma^2)$ to make explicit this parameter dependance. Here the terminology acquires a clear geometrical meaning. If we plot $f(x) = f(x; \mu, \sigma^2)$ we obtain a bell shaped curve, centered around the mean μ, symmetrical and with tails going to infinity in both directions. Thus, the mean is the most likely or expected value of a random variable under this distribution. For the rest of the values we find that approximately 2/3 of the probability mass lies within one standard deviation σ from μ; that approximately 95 % of the probability mass lies within 1.96σ from μ; and that the probability of being far from the mean μ decreases rapidly. Check this by yourself in R, with the commands:

```
> x = seq(-4,4,0.01)
> plot(x,dnorm(x,mean=0.5,sd=1.3),type='l')
```

which plots the normal distribution of a sequence `x` of real values ranging from -4 to 4, and equally spaced by a 0.01 increment, with $\mu = 0.5$ and $\sigma = 1.3$.

The normal distribution with mean μ and variance σ^2 is denoted by $N(\mu, \sigma^2)$; and to indicate that a random variable X has normal distribution with such a mean and variance we write $X \sim N(\mu, \sigma^2)$.

Remark 2.3 In general, $X \sim Y$ denotes that the two random variables X and Y have the same distribution. $X|A \sim Y|B$ means X conditioned to information set A has same distribution of Y conditioned to information set B. □

The *standard normal distribution* is the normal distribution with zero mean and unit variance, $N(0, 1)$, and the standard normal CDF is

$$\Phi(x) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} \exp(-t^2/2)dt, \quad -\infty < x < \infty. \tag{2.28}$$

The normal distribution has skewness equal to 0 and kurtosis equal to 3. In view of this fact, for any other distribution the difference of its kurtosis with 3, namely $K(X) - 3$, is a measure of *excess of kurtosis* (i.e., by how much the kurtosis of the distribution deviates from the normal kurtosis), and the sign gives the type of kurtosis. A distribution with positive excess kurtosis ($K(X) - 3 > 0$), known as *leptokurtic*, has a more acute peak around the mean and heavier tails, meaning that the distribution puts more mass on the tails than a normal distribution. This implies that a random sample from such distribution tends to contain more extreme values, as is often the case of financial returns. On the other hand, a distribution with negative excess kurtosis ($K(X) - 3 < 0$), known as *platykurtic*, presents a wider and lower peak around the mean and thinner tails. Returns seldom have platykurtic distributions; hence models based on this kind of distribution should not be considered.

The normal distribution has other properties of interest of which we mention two:

P1 A normal distribution is invariant under linear transformations: If $X \sim N(\mu, \sigma^2)$ then $Y = aX + b$ is also normally distributed and further $Y \sim N(a\mu + b, a^2\sigma^2)$.

P2 Linear combinations of normal variables are normal: If $X_1,\ldots, X_k$ are independent, $X_i \sim N(\mu_i, \sigma_i^2)$, and $a_1,\ldots, a_k$ are constants, then $Y = a_1X_1 + \cdots + a_kX_k$ is normally distributed with mean $\mu = \sum_{i=1}^{k} a_i\mu_i$ and variance $\sigma^2 = \sum_{i=1}^{k} a_i^2\sigma_i^2$.

It follows from P1 that if $X \sim N(\mu, \sigma^2)$, then

$$Z = \frac{X - \mu}{\sigma} \sim N(0, 1) \tag{2.29}$$

Equation (2.29) is a very useful tool for "normalizing" a random variable X for which we do not know its distribution. This follows from the *Central Limit Theorem*.

Theorem 2.1 (Central Limit Theorem) If $X_1,\ldots, X_n$ is a random sample from a distribution of expected value given by μ and finite variance given by σ^2, then the limiting distribution of

$$Z_n = \frac{\sqrt{n}}{\sigma}\left(\frac{1}{n}\sum_{i=1}^{n} X_i - \mu\right) \tag{2.30}$$

is the standard normal distribution. □

Thus, if we have an observation x of random variable X then the quantity $z = (x-\mu)/\sigma$, known as *z-score*, is a form of "normalizing" or "standardizing" x. The idea behind is that even if we don't know the distribution of X (but know its mean and variance), we compare a sample to the standard normal distribution by measuring how many standard deviations the observations is above or below the mean. The z-score is negative when the sample observation is below the mean, and positive when above. As an application of the z-score there is the computation of the LCL and UCL thresholds in R Example 2.2 (Eq. (2.26)): We wish to estimate L and U bounds such that random variable X lies within (L, U) with a 95 % probability; that is, $\mathbb{P}(L < X < U) = 0.95$. We use the standardization Z of X, and hence we want L and U such that

$$\mathbb{P}\left(\frac{L-\mu}{\sigma} < Z < \frac{U-\mu}{\sigma}\right) = 0.95$$

From this equation it follows that $L = \mu - z\sigma$ and $U = \mu + z\sigma$, where z is the quantile such that $\mathbb{P}(-z < Z < z) = 0.95$. By the Central Limit Theorem

$$0.95 = \mathbb{P}(|Z| < z) \approx \Phi(z) - \Phi(-z)$$

where $\Phi(z)$ is given by Eq. (2.28). Computing numerically the root of $\Phi(z) - \Phi(-z) = 2\Phi(z) - 1 = 0.95$ one gets $z = 1.96$. Below we show how to do this computation in R.

R Example 2.3 In R the function `qnorm(x)` computes $\Phi^{-1}(x)$. Hence, for $x = (0.95 + 1)/2 = 0.975$, run the command `qnorm(0.975)` to get the quantile 1.96. □

For a proof of the Central Limit Theorem and more on its applications see Feller (1968).

2.2.4 Distributions of Financial Returns

Are returns normally distributed? Almost any histogram of an asset's return will present some bell-shaped curve, although not quite as smooth as the normal distribution. We show this empirical fact with a particular stock.

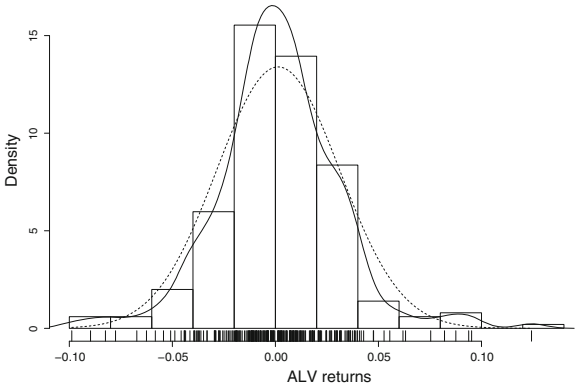

Fig. 2.4 Histogram of ALV returns from 06/01/2009 - 30/12/2009, with an estimate of its density from sample data (*solid line*), and adjusted normal distribution (*dashed line*)

R Example 2.4 Consider the sequence of returns for Allianz (ALV) obtained in R Example 2.1. We compute its histogram and on top we plot an estimate of its density from the sample (with a solid line). Then plot the normal density function with the sample mean and sample standard deviation of the given series (with a dashed line).

```
> alv <- na.omit(daxR$alvR); DS <- density(alv)
> yl=c(min(DS$y),max(DS$y)) #set y limits
> hist(alv,probability=T,xlab="ALV returns", main=NULL,ylim=yl)
> rug(alv); lines(DS); a=seq(min(alv),max(alv),0.001)
> points(a,dnorm(a,mean(alv),sd(alv)), type="l",lty=2)
> # if you rather have a red line for the normal distribution do:
> lines(a,dnorm(a,mean(alv), sd(alv)),col="red")
```

The output can be seen in Fig. 2.4.

The figure shows that a normal distribution does not fits well the sample estimate of the density of returns for the ALV stock. We can corroborate this empirical observation with one of many statistical test for the null hypothesis that a sample $R_1,\ldots,R_n$ of returns come from a normally distributed population. One popular such test is Shapiro-Wilk. If the p-value that the test computes is less than a given confidence level (usually 0.05) then the null hypothesis should be rejected (i.e. the data do not come from a normal distributed population). In R this test is implemented by the function `shapiro.test()`. Running this function on `alv` we get the following results:

```
> shapiro.test(alv)

Shapiro-Wilk normality test data:  alv W = 0.9701, p-value =
3.995e-05
```

The p-value is below 5×10^{-5}, hence the hypothesis of normality should be rejected.

In addition to these empirical observations and statistical tests, there are some technical drawbacks to the assumption of normality for returns. One is that an asset return has a lower bound in -1 and no upper bound, so it seems difficult to believe

in a symmetric distribution with tails going out to infinity in both directions (which are characteristics of normality). Another concern, more mathematically disturbing, is that multiperiod returns could not fit the normal distribution, since they are the product of simple returns, and the product of normal variables is not necessarily normal. This last mismatch between the properties of returns and normal variables steers our attention to log returns, since a multiperiod log return is the sum of simple log returns, just as the sum of normal variables is normal. Thus, a sensible alternative is to assume that the log returns are the ones normally distributed, which implies that the simple returns are *log-normally* distributed.

Example 2.3 **(The log-normal distribution)** A random variable X has the *log-normal distribution*, with parameters μ and σ^2, if $\ln X \sim N(\mu, \sigma^2)$. In this case we write $X \sim LogN(\mu, \sigma^2)$. The log-normal density function is given by

$$f_X(x) = \frac{1}{x\sigma\sqrt{2\pi}} \exp(-(\ln x - \mu)^2/2\sigma^2), \quad x > 0. \tag{2.31}$$

and the moments of the variable X are

$$E(X^n) = \exp\left(n\mu + \frac{1}{2}n^2\sigma^2\right), \; n > 0$$

Therefore, if we assume that the simple return series $\{R_t\}$ is log-normally distributed with mean μ_R and variance σ_R^2, so that the log return series $\{r_t\}$ is such that

$$r_t = \ln(R_t + 1) \sim N(\mu_r, \sigma_r^2)$$

with mean μ_r and variance σ_r^2, we have that the respective moments for both series are related by the following equations

$$E(R_t) = \mu_R = e^{\mu_r + \sigma_r^2/2} - 1, \qquad Var(R) = \sigma_R^2 = e^{2\mu_r + \sigma_r^2}(e^{\sigma_r^2} - 1) \tag{2.32}$$

As a first application of this hypothesis of normal distribution for the continuously compounded returns, we show how to compute bounds to the future price of the underlying asset with a 95 % precision.

Example 2.4 Let $\{P_t\}$ be the price series of a stock (or any other financial asset) and $\{r_t\}$ its corresponding log return series, and assume r_t are independent and normally distributed. The τ-period log returns $r_t(\tau)$ are also normally distributed, because of Eq. (2.7) and the fact that a finite sum of iid normal random variables is normal. Therefore the mean of $r_t(\tau)$ is $\mu_r\tau$ and the variance is $\sigma_r\tau$. Now, consider an initial observed price P_0 at time $t = 0$, and we want to estimate the price P_T at a later time $t = T$. Then $\ln(P_T/P_0)$ is the continuously compounded return over the period $\tau = T$, and by the previous observations

$$\ln\left(\frac{P_T}{P_0}\right) = r_1 + \cdots + r_T \sim N(\mu_r T, \sigma_r^2 T) \tag{2.33}$$

where μ_r and σ_r^2 are the mean and variance of $\{r_t\}$. Let $Z_T = \dfrac{\ln(P_T/P_0) - \mu_r T}{\sigma_r\sqrt{T}}$. Then $Z_T \sim N(0, 1)$, and we have seen as a consequence of the Central Limit Theorem that for the quantile $z = 1.96$ we have $\mathbb{P}(-z < Z_T < z) = 0.95$ (go back to Example 2.3). From this we have

$$\mu_r T - z\sigma_r\sqrt{T} < \ln\left(\frac{P_T}{P_0}\right) < \mu_r T + z\sigma_r\sqrt{T}$$

or, equivalently

$$P_0 \exp\left(\mu_r T - z\sigma_r\sqrt{T}\right) < P_T < P_0 \exp\left(\mu_r T + z\sigma_r\sqrt{T}\right) \tag{2.34}$$

These equations give bounds for the price at time $t = T$, with a 95 % probability if taking $z = 1.96$. On real data, one makes an estimate of μ_r and σ_r from a sample of the log returns $\{r_t\}$, and assumes these estimations of moments hold for the period $[0, T]$; that is, we must assume that the mean and variance of the log returns remain constant in time. Be aware then of all the considered hypotheses for these calculations to get the bounds in Eq. (2.34), so that estimations from real data should be taken as rough approximations to reality. □

Although the log-normal distribution model is consistent with the linear invariance property of log returns and with their lower bound of zero, there is still some experimental evidence that makes the distribution of returns depart from log-normality. For example, it is often observed some degree of skewness, where in fact negative values are more frequent than positive; also some excessively high and low values present themselves (suggesting some kurtosis), and these extreme values can not be discarded as outliers, for here is where money is lost (or gained) in tremendous amounts.

Which is then an appropriate model for the distribution of stock returns? Given a collection of log returns $\{r_{it} : i = 1, \ldots, n; t = 1, \ldots, m\}$ corresponding to returns r_{it} of n different assets at times $t = 1, \ldots, m$, each considered as random variable, we should start with the most general model for this collection of returns which is its joint distribution function:

$$G(r_{11}, \ldots, r_{n1}, r_{12}, \ldots, r_{n2}, \ldots, r_{1m}, \ldots, r_{nm}; \theta), \tag{2.35}$$

where to simplify notation we use $G(X, Y; \theta)$ instead of the usual $F_{X,Y}(x, y; \theta)$, with the additional explicit mention of θ as the vector of fixed parameters that uniquely determines G (e.g. like μ and σ determine the normal distribution). More than a model, (2.35) is a general framework for building models of distribution of returns

wherein we "*view financial econometrics as the statistical inference of* θ, *given* G *and realizations of* $\{r_{it}\}$".[3]

Therefore we should impose some restrictions to (2.35) to make it of practical use, but also to reflect our social or economic beliefs. For example, some financial models, such as the Capital Asset Pricing Model (to be studied in Chap. 8) consider the joint distribution of the n asset returns restricted to a single date t: $G(r_{1t}, \ldots, r_{nt}; \theta)$. This restriction implicitly assumes that returns are statistically independent through time and that the joint distribution of the cross-section of returns is identical across time. Other models focus on the dynamics of individual assets independently of any relations with other assets. In this case one is to consider the joint distribution of $\{r_{i1}, \ldots, r_{im}\}$,for a given asset i, as the product of its conditional distributions (cf. Eq. 2.14, and we drop θ for the sake of simplicity):

$$\begin{aligned} G(r_{i1}, \ldots, r_{im}) &= G(r_{i1})G(r_{i2}|r_{i1})G(r_{i3}|r_{i2}, r_{i1}) \cdots G(r_{im}|r_{im-1}, \ldots, r_{i1}) \\ &= G(r_{i1}) \prod_{i=2}^{m} G(r_{im}|r_{im-1}, \ldots, r_{i1}) \end{aligned} \tag{2.36}$$

This product exhibits the possible temporal dependencies of the log return r_{it}. So, for example, if we believe in the weak form of the efficient market hypothesis, where returns are unpredictable from their past, then in this model the conditional distributions should be equal to the corresponding marginal distributions, and hence (2.36) turns into

$$G(r_{i1}, \ldots, r_{im}) = G(r_{i1})G(r_{i2}) \cdots G(r_{im}) = \prod_{i=1}^{m} G(r_{im}) \tag{2.37}$$

We see then that under this framework issues of predictability of asset returns are related to their conditional distributions and how these evolve through time. By placing restrictions on the conditional distributions we shall be able to estimate the parameters θ implicit in (2.36). On the other hand, by focusing on the marginal or unconditional distribution we can approach the behavior of asset returns individually. This was the case of assuming a normal or log-normal distribution. From these basic distributions we can go on refining so as to capture, for example, the excess of kurtosis or non null skewness. Distributions such as *stable* distributions, or *Cauchy*, or a *mixture of distributions*, may capture these features found in returns better than log-normal distributions, but the downside is that the number of parameters in θ increases just as the computational difficulty to estimate them. In summary, fitting a model beyond the normal or log normal to the distribution of stock returns is a challenging albeit arduous problem.[4]

[3] Campbell et al. (1997 §1.4.2)

[4] For a more extensive discussion see Campbell et al. (1997 Chap. 1)

2.3 Stationarity and Autocovariance

Now, even if we do not believe that today's performance of returns is a reflection of their past behavior, we must believe that there exist some statistical properties of returns that remain stable through time; otherwise there is no meaningful statistical analysis of financial returns and no possible interesting models for their distribution at all. The invariance in time of the moments of the distribution of a random process is the stationarity hypothesis.

Definition 2.3 A random process $\{X_t\}$ is *strictly stationary* if for any finite set of time instants $\{t_1, \ldots, t_k\}$ and any time period τ the joint distribution of $\{X_{t_1}, \ldots, X_{t_k}\}$ is the same as the joint distribution of $\{X_{t_1+\tau}, \ldots, X_{t_k+\tau}\}$, i.e.,

$$F_{X_{t_1},\ldots,X_{t_k}}(x_1, \ldots, x_k) = F_{X_{t_1+\tau},\ldots,X_{t_k+\tau}}(x_1, \ldots, x_k). \tag{2.38}$$

One interesting property of strictly stationary processes is that once we have one we can produce many other strictly stationary processes by applying any "regular" operation on subsequences; e.g. moving averages, iterative products, and others. We state this important fact informally, and recommend the reading of Breiman (1992 Chap. 6, Prop. 6.6) for the precise mathematical details.[5]

Proposition 2.2 *Let $\{X_t\}$ be a strictly stationary process and Φ a function from $\mathbb{R}^{h+1}$ to $\mathbb{R}$. Then the process $\{Y_t\}$ defined by $Y_t = \Phi(X_t, X_{t-1}, \ldots, X_{t-h})$ is strictly stationary.* □

Stationary processes obtained as $Y_t = \Phi(X_t, X_{t-1}, \ldots, X_{t-h})$ can be classified by the scope of their variable dependency, in the following sense.

Definition 2.4 A random process $\{X_t\}$ is m-dependent, for an integer $m > 0$, if X_s and X_t are independent whenever $|t - s| > m$. □

For example, an iid sequence is 0-dependent; Y_t obtained as in Prop. 2.2 is h-dependent.

Example 2.5 A *white noise* $\{W_t\}$ is a sequence of iid random variables with finite mean and variance. This is a strictly stationary process: independence implies that $F_{W_1,\ldots,W_k}(w_1, \ldots, w_k) = \prod_{i=1}^{k} F_{W_i}(w_i)$ (cf. Eq. (2.15)), while being identically distributed implies that $F_{W_i}(w) = F_{W_i+\tau}(w) = F_{W_1}(w)$, for all i; hence, both hypotheses give Eq. (2.38). Next, consider the processes

[5] We remark that in many classical textbooks of probability and stochastic processes, such as Breiman (1992), strictly stationary is simply termed stationary, but in the modern literature of time series and financial applications, such as Brockwell and Davis (2002); Tsay (2010), it is more common to distinguish different categories of stationarity, in particular strict (as in Def. 2.3) and weak (to be defined later). We adhere to this latter specification of different levels of stationarity, and when using the term by itself –without adverb– is to refer to any, and all, of its possibilities.

- $Y_t = \alpha_0 W_t + \alpha_1 W_{t-1}$, with $\alpha_0, \alpha_1 \in \mathbb{R}$ (moving average);
- $Z_t = W_t W_{t-1}$

By Prop. 2.2 these are strictly stationary (and 1-dependent). □

It is not obvious that financial returns verify the strictly stationary hypothesis. However, it is a convenient assumption to ensure that one can estimate the moments of the returns by taking samples of data from any time intervals. Looking at some plots of returns (e.g., go back to Fig. 2.2), one often encounters that the mean is almost constant (and close to zero) and the variance bounded and describing a pattern that repeats through different periods of time. Therefore, an assumption perhaps more reasonable to the invariability in time of all moments of returns could be that the first moment is constant (hence invariant) and the second moment is in sync with its past (hence it can be well estimated from past data). To formalize this hypothesis we need as a first ingredient the notion of *covariance*.

Definition 2.5 The covariance of two random variables X and Y is

$$Cov\,(X, Y) = E((X - \mu_X)(Y - \mu_Y))$$

Note that $Cov\,(X, Y) = E(XY) - E(X)E(Y) = E(XY) - \mu_X \mu_Y$, and $Cov\,(X, X) = Var\,(X) = \sigma_X^2$. To be consistent with the σ notation, it is customary to denote $Cov\,(X, Y)$ by $\sigma_{X,Y}$. □

Remark 2.4 The expected value of the product of continuous random variables X and Y with joint density function $f_{X,Y}$ is

$$E(XY) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} xy f_{X,Y}(x, y) dx dy \tag{2.39}$$

Now observe that if X and Y are independent then $Cov\,(X, Y) = 0$. This follows directly from the definition of covariance and Eqs. (2.39) and (2.15). The converse is not true, for we can have $Cov\,(X, Y) = 0$ and X and Y being functionally dependent. The popular example is to consider X *uniformly distributed* in $[-1, 1]$, and $Y = X^2$. Then X and Y are dependent, but $Cov\,(X, X^2) = E(X^3) - E(X)E(X^2) = 0 - 0 \cdot E(X^2) = 0$. Note that the dependency of X and $Y = X^2$ is non-linear.

We can estimate $Cov\,(X, Y)$ from a given sample $\{(x_i, y_i) : i = 1, \ldots, m\}$ by the sample covariance:

$$\widehat{Cov}\,(X, Y) = \frac{1}{m-1} \sum_{i=1}^{m} (x_i - \widehat{\mu}_X)(y_i - \widehat{\mu}_Y) \tag{2.40}$$

R Example 2.5 In R the function `cov(x,y,...)` computes the sample covariance of vectors `x` and `y`, or if `x` is a matrix or table (and `y=NULL`) computes covariance

between the columns. We use it to compute the covariance between pairs of the four stocks considered in R Example 2.2. Execute in your R console the command

```
> cov(daxRlog[,2:5],use="complete.obs")
```

The resulting table is

```
             alvR         bmwR          cbkR         tkaR
alvR 0.0008816302 0.0005196871 0.0008677487 0.0006721131
bmwR 0.0005196871 0.0008570846 0.0006724442 0.0006361113
cbkR 0.0008677487 0.0006724442 0.0023419267 0.0008754300
tkaR 0.0006721131 0.0006361113 0.0008754300 0.0011487287
```

We observe positive values in all entries of the sample covariance matrix in the example above. What can we deduce from these values? Observe that if X and Y are linearly dependent, say $Y = aX + b$, then $Cov(X, aX + b) = aVar(X)$. Since $Var(X) \geq 0$ always, we see that the sign of covariance depends on the sign of the slope of the linear function of Y on X; hence positive (resp. negative) covariance means that Y and X move in the same (resp. opposite) direction, *under the assumption of a linear model for the dependency relation*. How strong could this co-movement be? In $Cov(X, aX+b) = aVar(X)$ we could have large covariance due to a large $Var(X)$, while the factor of dependency a be so small, so as to be negligible. Hence, we need a way to measure the strength of the possible (linear) co-movement signaled by a non null covariance. The obvious solution suggested by the previous observation is to factor out the square root of both the variances of X and Y from their covariance: $\frac{Cov(X,Y)}{\sqrt{Var(X)Var(Y)}}$. This is the *correlation coefficient* of X and Y, which we will study in more detail in the next chapter.

We move on to the problem of determining the possible dependence of the variance of a random process with its past. For that matter one uses the autocovariance function.

Definition 2.6 For a random process $\{X_t\}$ with finite variance, its *autocovariance* function is defined by $\gamma_X(s,t) = Cov(X_s, X_t)$. □

The autocovariance is the concept we need to narrow strict stationarity of returns to just some of its key moments. This gives us weak stationarity.

Definition 2.7 A random process $\{X_t\}$ is *weakly stationary* (or *covariance stationary*) if it has finite variance ($Var(X_t) < \infty$), constant mean ($E(X_t) = \mu$) and its autocovariance is time invariant: $\gamma_X(s,t) = \gamma_X(s+h, t+h)$, for all $s, t, h \in \mathbb{Z}$. In other words, the autocovariance only depends on the time shift, or *lag*, $|t - s|$, and not on the times t or s. Hence, we can rewrite the autocovariance function of a weakly stationary process as

$$\gamma_X(h) = Cov(X_t, X_{t+h}) = Cov(X_{t+h}, X_t), \quad h = 0, \pm 1, \pm 2, \ldots. \tag{2.41}$$

$\gamma_X(h)$ is also called the lag-h autocovariance. □

Remark 2.5 A strictly stationary process $\{X_t\}$ with finite second moments is weakly stationary. The converse is not true: there are weakly stationary processes that are not strictly stationary.[6] However, if $\{X_t\}$ is a weakly stationary *Gaussian* process (i.e., the distribution of $\{X_t\}$ is multivariate normal), then $\{X_t\}$ is also strictly stationary. □

Remark 2.6 Some obvious but important properties of the autocovariance function of any stationary process (strictly or weakly) are: $\gamma(0) \geq 0$ (since $\gamma(0) = Var(X)$), $|\gamma(h)| \leq \gamma(0)$ and $\gamma(h) = \gamma(-h)$, for all h. □

Example 2.6 Let $\{W_t\}$ be a white noise with mean zero and variance σ^2, and consider the following sequence: $S_0 = 0$, and for $t > 0$, $S_t = W_1 + W_2 + \ldots + W_t$. Note that $S_t = S_{t-1} + W_t$. The process $S = \{S_t\}$ is a *random walk*. Compute the autocovariance for S. For $h > 0$,

$$\gamma_S(t, t+h) = Cov\,(S_t, S_{t+h}) = Cov\left(\sum_{i=1}^{t} W_i, \sum_{j=1}^{t+h} W_j\right)$$
$$= Var\left(\sum_{i=1}^{t} W_i\right) = t\sigma^2$$

The third equality follows from $Cov\,(W_i, W_j) = 0$, for $i \neq j$. The autocovariance depends on t; hence, the random walk S is not weakly stationary (and not strictly stationary). □

All is well in theory, but in practice if one doesn't have a model fitting the data, one can still determine the possibility of the underlying random process being weakly stationary by applying the following empirical method: look at a plot of the return time series, and if the values do not seem to fluctuate with a constant variation and around a constant level, then one can conclude with high confidence that the process is not weakly stationary; otherwise we can suspect of stationarity. It might be the case that the data needs first to be massaged to remove some observable *trend* (e.g. bull or bear market periods) or *seasonality* (e.g. values repeating periodically), to reveal the noisy component with the possible stationary behavior (we will comment on this decomposition of time series in Chap. 4). Do the appropriate transformation first and then try different models for the transformed data. For more on the theory and applications of stationary time series see Brockwell and Davis (1991).

[6] The following counterexample is from Brockwell and Davis (1991): let $\{X_t\}$ be a sequence of independent random variables such that X_t is exponentially distributed with mean 1 when t is odd and normally distributed with mean 1 and variance 1 when t is even, then $\{X_t\}$ is weakly stationary, but X_{2k+1} and X_{2k} have different distributions for each $k > 0$, hence $\{X_t\}$ cannot be strictly stationary.

2.4 Forecasting

In the analysis of financial time series we are interested in designing models for predicting future values. We are going to formalize the framework for forecasting in this chapter and postpone the model design for Chap. 4. Given a stationary stochastic process $\{X_t\}$ with zero mean and autocovariance function γ, we want to predict X_{t+h}, the value of the series h time periods into the future, based on some known set of information Z about X_t. The information set Z is comprised of random observations, and it is in practice a finite set of recent past values of the series, that is, $Z = \{X_t, X_{t-1}, \ldots, X_{t-p}\}$. Given the set Z, we view an estimation $\widehat{X}_{t+h}$ of X_{t+h} as a function of Z, $\widehat{X}_{t+h} = F(Z)$, and we would like to know how good is $F(Z)$ as an estimator. The accepted criterion for measuring the performance of an estimator F of a random variable X, based on some information Z, is the following[7]

> The *best estimator* or *predictor* of a random variable X, on the basis of random information Z about X, is the function $F(Z)$ that minimizes $E(X - F(Z))^2$.

This minimum mean square error predictor can be characterized in statistical terms as follows. Assuming the conditional density function of X given $Z = z$ exists (cf. Eq. (2.13)), we can consider the *conditional expectation of* X *given* $Z = z$, denoted $E(X|Z = z)$ (or, for the sake of simplicity, $E(X|Z)$), and defined as

$$E(X|Z = z) = \int_{-\infty}^{\infty} x f_{X|Z=z}(x)dx \tag{2.42}$$

Then the best estimator of X given the information Z is $E(X|Z = z)$. We show this fact in the following theorem.

Theorem 2.2 Let X and Z be two random variables with $E(X^2) < \infty$ and joint density function $f_{X,Z}$, and let $\widehat{X} = E(X|Z)$ the conditional expectation of X given Z. Then, for any function $F = F(Z)$, we have

$$E(X - \widehat{X})^2 \leq E(X - F(Z))^2$$

Proof Let $F = F(Z)$ and consider the following equalities

$$\begin{aligned} E(X - F)^2 &= E([(X - \widehat{X}) + (\widehat{X} - F)]^2) \\ &= E(X - \widehat{X})^2 + 2E((X - \widehat{X})(\widehat{X} - F)) + E(\widehat{X} - F)^2 \end{aligned} \tag{2.43}$$

Note that $\widehat{X} - F$ is a function of Z, so observed values of this random variable are denoted z' (i.e., given values $\widehat{X} - F = z'$). We decompose the second term in (2.43), using linearity of E, as follows

[7] see Box and Jenkins (1976); Breiman (1992) on properties of conditional expectation and best predictor discussed in this section.

$$E((X-\widehat{X})(\widehat{X}-F)) = E(X(\widehat{X}-F)) - E(\widehat{X}(\widehat{X}-F))$$

and now resolve, using (2.39), (2.15) and (2.42),

$$\begin{aligned} E(X(\widehat{X}-F)) &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} xz' f_{X,Z}(x,z')dxdz' \\ &= \int_{-\infty}^{\infty} z' \left(\int_{-\infty}^{\infty} x f_{X|Z=z'}(x)dx \right) f_Z(z')dz' = \int_{-\infty}^{\infty} z' E(X|Z) f_Z(z')dz' \\ &= E(\widehat{X}(\widehat{X}-F)) \end{aligned}$$

Therefore, $E((X-\widehat{X})(\widehat{X}-F)) = 0$, and
$E(X-F)^2 = E(X-\widehat{X})^2 + E(\widehat{X}-F)^2 \geq E(X-\widehat{X})^2$.

This theorem applies to the particular case of interest in which the information set is $Z = \{X_t, X_{t-1}, \ldots, X_{t-p}\}$, a recent history of X_t, and we want to make estimation of X_{t+h}. Then the best predictor in this case is the conditional expectation $E(X_{t+h}|X_t, X_{t-1}, \ldots, X_{t-p})$. We are thus interested in determining the form of this conditional expectation. For Gaussian processes this is well determined as stated in the following proposition whose proof we leave as exercise.

Proposition 2.3 *If the variables X_{t+h}, X_t, ..., X_{t-p} have a normal distribution, i.e., the process $\{X_t\}$ is Gaussian, then*

$$E(X_{t+h}|X_t, X_{t-1}, \ldots, X_{t-p}) = \alpha_0 X_t + \alpha_1 X_{t-1} + \cdots + \alpha_p X_{t-p}$$

where α_0, ..., α_p are real numbers. □

In view of this proposition, the problem of building a best predictor for a Gaussian process is solved by forming a linear regression. For any other process, not necessarily normally distributed, it is still desirable to design a predictor as a linear combination of its past history, even if this does not coincides with the conditional expectation of the process given its past history. In this case we want a linear function L of $\{X_t, X_{t-1}, \ldots, X_{t-p}\}$ that minimizes the mean square error $E(X_{t+h}-L)^2$; that is, we want to find coefficients $\alpha_0, \alpha_1, \ldots, \alpha_p$ to form $L(X_{t+h}) = \sum_{j=0}^{p} \alpha_j X_{t-j}$ and such that their values minimize $F(\alpha_0, \ldots, \alpha_p) = E\left(X_{t+h} - \sum_{j=0}^{p} \alpha_j X_{t-j}\right)^2$. This F is a quadratic function bounded below by 0, and hence there exists values of $(\alpha_0, \ldots, \alpha_p)$ that minimizes F, and this minimum satisfies $\dfrac{\partial F(\alpha_0, \ldots, \alpha_p)}{\partial \alpha_j} = 0, j = 0, \ldots, p.$ Evaluating these derivatives we get:

$$E\left((X_{t+h} - \sum_{k=0}^{p} \alpha_k X_{t-k}) X_{t-j} \right) = 0, \quad j = 0, \ldots, p. \tag{2.44}$$

Since $E(X_t) = 0$, $\gamma(h+j) = E(X_{t+h}X_{t-j})$ for $j = 0, 1, \ldots, p$, and consequently we obtain from Eq. (2.44) the following system of equations:

$$\begin{aligned} \gamma(h) &= \alpha_0\gamma(0) + \alpha_1\gamma(1) + \cdots + \alpha_p\gamma(p) \\ \gamma(h+1) &= \alpha_0\gamma(1) + \alpha_1\gamma(0) + \cdots + \alpha_p\gamma(p-1) \\ &\vdots \qquad\qquad \vdots \\ \gamma(h+p) &= \alpha_0\gamma(p) + \alpha_1\gamma(p-1) + \cdots + \alpha_p\gamma(0) \end{aligned} \tag{2.45}$$

Therefore, in terms of conditional expectation, the best linear h-step forecaster for the process $\{X_t\}$ is

$$L(X_{t+h}) = \sum_{j=0}^{p} \alpha_j X_{t-j} \tag{2.46}$$

where $(\alpha_0, \alpha_1, \ldots, \alpha_p)$ satisfies Eq. (2.45). On the other hand, the mean square forecast error is given by the equation

$$E\left(X_{t+h} - L(X_{t+h})\right)^2 = \gamma(0) - 2\sum_{j=0}^{p} \alpha_j\gamma(h+j) + \sum_{i=0}^{p}\sum_{j=0}^{p} \alpha_i\alpha_j\gamma(i-j) \tag{2.47}$$

Note that if we don't know the autocovariance function γ we can make estimations using the sample covariance (Eq. (2.40)).

Thus, for a stationary process its best linear forecaster in terms of its past is completely determined by the covariances of the process. However, note that even though the system of linear Eq. (2.45) is computationally solvable, it can be time-consuming for large p. The solution of this system can be alleviated for processes which are themselves expressed by linear combinations of white noise, as it is the case of the ARMA processes to be studied in Chap. 4. For other arbitrary processes there are some very effective recursive procedures to compute the coefficients for a 1-step forecaster.[8]

2.5 Maximum Likelihood Methods

Once we have settled on a model for the distribution of returns, or for their second moments, an important next step is to estimate the parameters of the model from sample observations of the returns. The method of maximum likelihood is a much used technique in statistics for estimating the "most likely" values for the parameters of a model from observable data. It is an estimation method which is largely due, in its modern formulation, to R. A. Fischer.[9]

[8] One that we like very much is the Innovations Algorithm, see Brockwell and Davis (2002).

[9] see Hald (1999) for a historical account

Let X be a random variable for which we have sample observations $x_1, \ldots, x_m$, and let $\theta = (\theta_1, \ldots, \theta_p)$ be a vector of parameters, pertaining to some statistical model for X, and which we want to estimate. Assuming that the observations are made randomly and independently of each other, we can consider their joint distribution as the product of their conditional distributions (Eq. (2.36)). Then, if the distribution has density function $f(x; \theta)$, we define the likelihood of observing the m values of X, given θ, as

$$L(x_1, \ldots, x_m; \theta) = f(x_1; \theta) \cdot \prod_{i=2}^{m} f(x_i | x_{i-1}, \ldots, x_1; \theta) \tag{2.48}$$

The value of θ that maximizes this likelihood function is the *maximum–likelihood estimate* (MLE) of θ. This MLE might not be unique, and might even not exists. In the case that it exists, but is not unique, any of the points where the maximum is attained is an MLE of θ. A useful mathematical trick for simplifying the computation of an MLE is to find instead a value of θ that maximizes the logarithm of L, called the *log likelihood function*:

$$l(x_1, \ldots, x_m; \theta) = \ln(L(x_1, \ldots, x_m; \theta)) = \ln f(x_1; \theta) + \sum_{i=2}^{m} \ln(f(x_i | x_{i-1}, \ldots, x_1; \theta)) \tag{2.49}$$

Maximizing $\ln L$ is the same as maximizing L, because ln is a monotone function. The optimization of $l(x_1, \ldots, x_m; \theta)$ is often obtain only by numerical methods. In R this can be done with the function `mle()`, or by direct use of the optimization method `optim()`.

A further simplification to the likelihood computation is to assume that all random observations are made independently. In this case Eq. (2.48) and Eq. (2.49) turn into

$$L(x_1, \ldots, x_m; \theta) = \prod_{i=1}^{m} f(x_i; \theta) \quad \text{and} \quad l(x_1, \ldots, x_m; \theta) = \sum_{i=1}^{m} \ln(f(x_i; \theta)) \tag{2.50}$$

Although the independence assumption is quite strong for time series, it simplifies likelihood computation to analytical form. As an example of the MLE method we show how to obtain the best estimate of a constant variance from sample returns.

Example 2.7 Let $r_1, \ldots, r_m$ be a sequence of m independently observed log returns. We assume these returns have some constant variance σ^2, which we want to estimate under the normal distribution model, and a constant mean μ which we know. By the independence of the observations the log likelihood function is as in Eq. (2.50), with $\theta = (\sigma^2, \mu)$ and μ fixed, and the probability density function $f(r_i; \sigma^2, \mu)$ is given by the Gaussian density function (Eq. (2.27)). Therefore, the log likelihood of observing $r_1, \ldots, r_m$ is

$$l(r_1, \ldots, r_m; \sigma^2) = -\frac{1}{2} \sum_{i=1}^{m} \left[\ln(2\pi) + \ln(\sigma^2) + \frac{(r_i - \mu)^2}{\sigma^2} \right] \tag{2.51}$$

By the maximum likelihood method the value of σ^2 that maximizes (2.51) is the best estimate (or more likely) value of σ^2. Differentiating $l(r_1, \ldots, r_m; \sigma^2)$ with respect to σ^2 and setting the resulting derivative to zero, we obtain that the maximum likelihood estimate of σ^2 is given by

$$\sigma^2_{mle} = \frac{1}{m}\sum_{i=1}^{m}(r_i - \mu)^2$$

Note that this estimated value σ^2_{mle} is biased (cf. Remark 2.2), since

$$E(\sigma^2_{mle}) = \frac{m-1}{m}E(\widehat{\sigma}^2) = \frac{m-1}{m}\sigma^2$$

However, the bias of $-\sigma^2/m$ tends to 0 as $m \to \infty$, so the maximum likelihood estimate σ^2_{mle} is *asymptotically unbiased* for σ^2. Thus, taking m sufficiently large we have σ^2_{mle} a good estimator of the variance of log returns, sampled in a given interval of time, and which in fact, by construction, it is the most likely value that can be assigned to that variance. □

2.6 Volatility

One of the most important features of financial assets, and possibly the most relevant for professional investors, is the asset volatility. In practice volatility refers to a degree of fluctuation of the asset returns. However it is not something that can be directly observed. One can observe the return of a stock every day, by comparing the change of price from the previous to the current day, but one can not observe how the return fluctuates in a specific day. We need to make further observations of returns (and of the price) at different times on the same day to make an estimate of the way returns vary daily (so that we can talk about *daily volatility*), but these might not be sufficient to know precisely how returns will fluctuate. Therefore volatility can not be observed but estimated from some model of the asset returns. A general perspective, useful as a framework for volatility models, is to consider volatility as the conditional standard deviation of the asset returns.

Definition 2.8 Let r_t be the log-return of an asset at time t, and denote by F_{t-1} the information set available at time $t-1$. The conditional mean and variance of r_t given F_{t-1} are

$$\mu_t = E(r_t|F_{t-1}) \quad \text{and} \quad \sigma_t^2 = Var(r_t|F_{t-1}) = E((r_t - \mu_t)^2|F_{t-1}) \tag{2.52}$$

The volatility of the asset at time t is $\sigma_t = \sqrt{Var(r_t|F_{t-1})}$. □

Now, to compute $Var(r_t|F_{t-1})$ we need to make some assumptions on the nature of F_{t-1}, and according to these assumptions we obtain different categories of volatilities. A common assumption is to consider $F_{t-1} = \{r_{t-1}, \ldots, r_{t-m}\}$; that is, F_{t-1} consists of finitely many past returns. If we assume further that the returns $\{r_t\}$ are iid and normal, then we can take for estimation sample returns in any time interval, which implies that the variance is considered constant in time, and in this case the best estimate is given by the maximum likelihood estimate of the sample variance (cf. Example 2.7):

$$\widehat{\sigma}^2 = \frac{1}{m}\sum_{k=1}^{m}(r_{t-k} - \widehat{\mu})^2,$$

where $\widehat{\mu}$ is the sample mean of the series. Because we are interested in quantifying the variability of returns, we might as well consider a non-centered variance ($\mu = 0$), and thus take as the (constant) volatility the non-centered sample standard deviation of returns[10]:

$$\sigma_o = \sqrt{\frac{1}{m}\sum_{k=1}^{m} r_{t-k}^2} \tag{2.53}$$

This is a form of volatility known as *historical volatility*, since the standard deviation, estimated from m past observations of a time series of continuously compounded returns, measures the historical degree of fluctuation of financial returns.

Scaling the volatility. Because we have assumed $r_t = \ln P_t - \ln P_{t-1} \sim N(\mu, \sigma^2)$, then for $h > 0$ successive equally spaced observations of the price, $r_{t+h} \sim N(\mu h, \sigma^2 h)$. Therefore, to annualize the historical volatility estimate we must multiply it by the scaling factor $\sqrt{h}$, for h being the number of time periods in a year. Thus the annualized historical volatility with respect to h periods per year is given by

$$\sigma_{ann} = (\sqrt{h})\sigma_o \tag{2.54}$$

In practice, if daily data is used then we take $h = 252$; if weekly data is used (i.e. prices are sample every week), then $h = 52$; and for monthly data, $h = 12$.

Range-based volatility estimates. Note that in practice the return r_t is computed by taking the closing prices of successive days, i.e. $r_t = \ln(C_t/C_{t-1})$, where C_t is the Close price of day t. A sharper estimate of daily volatility would be obtained if one considers variation of prices within the same day. The best one can do with publicly available data is to consider the daily range of prices, which is defined as the difference between the High and Low prices of the day, i.e., $H_t - L_t$. One such range-based estimator was defined by Parkinson (1980) as follows:

[10] There other practical reasons to discard the mean: on real financial data the mean is often close to 0; also, a non-centered standard deviation gives a more accurate estimated volatility, even compared to other type of models (see Figlewski (1994)).

$$\sigma_p = \sqrt{\frac{1}{4\ln 2} \cdot \frac{1}{m} \sum_{k=1}^{m} \ln\left(\frac{H_k}{L_k}\right)^2} \tag{2.55}$$

Another range-based estimator was given by Garman and Klass (1980), defined as

$$\sigma_{gk} = \sqrt{\frac{1}{m} \sum_{k=1}^{m} \left[\frac{1}{2}\left(\ln \frac{H_k}{L_k}\right)^2 - (2\ln 2 - 1)\left(\ln \frac{C_k}{O_k}\right)^2\right]} \tag{2.56}$$

where O_k is the Open price (and C_k, H_k and L_k the Close, High and Low price).

You can easily program these estimators of historical volatility, or use the function `volatility` in the R package `TTR` as we show in the next R Example.

R Example 2.6 Load into your R console the packages `TTR` and `quantmod` and with the function `getSymbols` retrieve data for Apple Inc. (AAPL) from *yahoo finance*. We will estimate the historical volatility to the end of May 2009 for this stock, considering 40 past observations. We use the `TTR` function `volatility(ohlc, n=10, calc="close", N=260, ...)`, where `ohlc` is an object that is coercible to xts or matrix and contains Open-High-Low-Close prices; `n` is the length of the sample for the volatility estimate; `calc` is the type of estimator to use, where choices are: `"close"` for the Close-to-Close volatility (Eq. (2.53) with $r_t = \ln(C_t/C_{t-1})$); `"parkinson"` for the Parkinson estimator (Eq. (2.55)); `"garman.klass"` for the Garman and Klass estimator (Eq. (2.56)); and there are others (see Help documentation). Finally, `N` is the number of periods per year, as determined by the sample data. This is use to give the result in annualized form by multiplying the estimate by $\sqrt{N}$. Run the following commands:

```
> aapl=AAPL['2009-04/2009-05']; m=length(aapl\$AAPL.Close);
> ohlc <-aapl[,c("AAPL.Open","AAPL.High","AAPL.Low","AAPL.Close")]
> vClose <- volatility(ohlc, n= m,calc="close",N=252)
> vParkinson <- volatility(ohlc, n= m,calc="parkinson",N=252)
> vGK <- volatility(ohlc, n= m,calc="garman",N=252)
> vClose[m]; vParkinson[m]; vGK[m];
```

The resulting annualized volatility according to each estimator, as of 2009-05-29, is 0.356 for the Close-to-Close; 0.291 for the Parkinson; 0.288 for the Garman and Klass. □

For more on range-based volatility estimates see Tsay (2010), and for their effectiveness when continuous time models are considered see Alizadeh et al. (2002).

Time dependent weighted volatility. Now, volatility is not constant. We will argue on this theme in Chap. 4, Sect. 4.3, but right now we will give a first simple and intuitive approach to an estimation of volatility, in a scenario where returns are not necessarily iid, and their influence to the current variance is in some form diluted by how far away in the past they are. This idea can be modeled with the following weighted estimate of the variance

$$\sigma_w^2(t) = \sum_{k=1}^{m} \alpha_k r_{t-k}^2 \tag{2.57}$$

with $\sum_{k=1}^{m} \alpha_k = 1$ and so that $\alpha_1 > \alpha_2 > \ldots > \alpha_m$. In this way the values of returns that are further away in the past contribute less to the sum in Eq. (2.57). A standard assignment of weights that verify the strict decreasing property is to take $\alpha_k = (1-\lambda)\lambda^{k-1}$, for $k = 1, \ldots, m$, where $0 < \lambda < 1$ and is known as the decay factor. This is the *Exponential Weighted Moving Average* (EWMA) model for the variance

$$\sigma_{ewma}^2(t) = (1-\lambda) \sum_{k=1}^{m} \lambda^{k-1} r_{t-k}^2 \tag{2.58}$$

and it is this form which is usually considered at financial institutions. A good decay factor for developed markets has been estimated by J. P. Morgan's *RiskMetrics* methodology to be $\lambda = 0.94$. Observe that σ_{ewma}^2 can be obtained also from the following recursion, which is easily derived from Eq. (2.58)

$$\sigma_{ewma}^2(t) = \lambda \sigma_{ewma}^2(t-1) + (1-\lambda) r_{t-1}^2 \tag{2.59}$$

This recurrence has the computational advantage of only needing to keep in memory the previous day return r_{t-1} and the previous estimate of σ_{ewma}^2. Therefore it is also suitable for on-line computation of volatility. In the R Lab 2.7.10 we propose to do a EWMA estimate of the volatility of AAPL and compare to the estimates done with previous methods.

2.7 Notes, Computer Lab and Problems

2.7.1 Bibliographic remarks: There are many textbooks on financial time series analysis and general statistics that can be recommended for complementing this chapter. Our main and preferred sources are Campbell et al. (1997) and Tsay (2010); in particular we recommend reading Chaps. 1 and 2 of Campbell et al. (1997) for a comprehensive and detailed exposition of the econometrics of financial time series. For mathematical statistics and probability in general we use and recommend Feller (1968), an all-time classic; also Breiman (1992) and Stirzaker (2005), which are nice complements. An encyclopedic account of 40 major probability distributions can be found in Forbes et al. (2011). A good source for the traditional core material of computational statistics through programmed examples in R is the book by Rizzo (2008).

2.7.2 R Lab: We show how to fit an exponential curve to the DJIA from 1978 to 2001 (see Fig. 2.1). If P_t is the price variable of the DJIA, we want to fit a model of the form $P_t = e^{a+bt}$. Taking logarithms the problem reduces to fit $\ln(P_t) = a + bt$,

and we use regression analysis to estimate the coefficients a and b. R has the function `lm()` for fitting linear models using regression analysis, but it does not works well with `xts` objects (the data retrieved by `getSymbols` is encapsulated in that form), so we convert the data to numeric type. The R commands to do the `lm` fit and plot the results are the following:

```
> require(quantmod); getSymbols("DJIA",src="FRED")
> serie=DJIA["1978/2001"]
> price=as.numeric(serie) #extract numeric values of price
> time = index(serie) #extract the indices
> x=1:length(price)
> model=lm(log(price)~x)
> expo=exp(model$coef[1]+model$coef[2]*x)
> plot(x=time,y=price, main="Dow Jones",type="l")
> lines(time,expo,col=2,lwd=2)
```

2.7.3: In real life, every time you buy or sell an asset at a stock market you pay some fee for the transaction. Assume then, that the cost of a one-way transaction is some fixed percentage α of the asset's current price (for example, a reasonable transaction fee is in the range of 0.1 to 0.2 percent). Show that if we take into account this transaction fee then the k-period simple gross return is given by the equation $R_t[k] + 1 = \dfrac{P_t}{P_{t-k}} \cdot \dfrac{1-\alpha}{1+\alpha}$.

2.7.4 Binomial distribution: An important discrete distribution is the binomial distribution, which can be obtained as the probability that n iid Bernoulli trials with probabilities p of success and $1-p$ of failure result in k successes and $n-k$ failures. If S_n records the number of successes in the n Bernoulli trials, then S_n has the binomial distribution (written $S_n \sim Bin(n, p)$) with probability mass function

$$\mathbb{P}(S_n = k) = \left(\frac{n!}{(n-k)!k!}\right) p^k (1-p)^{n-k}$$

Try proving that $E(S_n) = np$ and $Var(S_n) = np(1-p)$. For further details on this particular distribution see Feller (1968). The binomial distribution plays an important role in the binomial tree option model to be studied in Chap. 4.

2.7.5: Let $\{\varepsilon_t\}$ be a time series with zero mean and covariance

$$Cov\,(\varepsilon_t, \varepsilon_s) = \begin{cases} \sigma^2 & \text{if } s = t \\ 0 & \text{if } s \neq t \end{cases}$$

ε_t is a weak form of white noise and is clearly weak stationary. Show that $X_t = a\varepsilon_t + b\varepsilon_{t-1}$, for constants a and b, defines a weak stationary time series.

2.7.6 R Lab: An R demonstration for doing some descriptive statistics of financial returns. We work with financial data from Allianz (ALV), Bayerische Motoren Werke (BMW), Commerzbank (CBK) and Thyssenkrupp (TKA), all German business trading in the Frankfurt Stock Exchange and listed in the main index DAX.

The data can be downloaded from the book's webpage, or if the reader wishes to build its own use the instructions in R Example 1.1. Each company's data is in a .csv file consisting of 5 columns labeled: Date, Open, High, Low, Close, Volume, AdjClose, and containing in each row a date, and for that date, the open price, highest price, lowest price, closing price, the total volume of transactions, and the adjusted close price. The data is ordered descending by date, beginning at 2009-12-30 down to 2003-01-02, and all files coincide by rows on the date.

```
> ### Part I: preprocessing the data ##########
> wdir="path-to-your-working-directory"; setwd(wdir)
>  # load the financial data from wdir
> ALV = read.csv(paste(wdir,"/ALV.csv", sep=""), header=T)
>  # extract 1 year of data from the AdjClose column
> alvAC= ALV$AdjClose[1:252]
>  ## repeat the previous instructs. with BMW, CBK, TKA.
> date= ALV$Date[1:252] # extract the column Date
> date <- as.Date(date) # and declare date as true date format
>  ##put all together into a data.frame
> dax =data.frame(date,alvAC,bmwAC,cbkAC,tkaAC)
>  # plot Adjusted prices vs date for ALV
> plot(dax$date,dax$alvAC, type="l",main="ALV.DE",
    +    xlab="dates",ylab="adj. close")
>  # Compute Returns. First define vectors of appropriate length
> alvR <- 1:252;  bmwR <- 1:252; cbkR <- 1:252; tkaR <- 1:252
> for (i in 1:252){alvR[i] <-(alvAC[i]/alvAC[i+1]) -1 }
>   #same with bmwR, cbkR, tkaR
>   # Remember dates are ordered descending. Make table Returns
> daxR =data.frame(dax$date,alvR,bmwR,cbkR,tkaR)
>  # Compute log returns (omit  column of dates)
> daxRlog <- log(daxR[2:5] +1)
> #plot returns and log returns (in red) and see coincidences:
> plot(dax$date,daxR$alvR, type="l",xlab="dates",ylab="returns")
> lines(dax$date,daxRlog$alvR, type="l",col="red")
> #### Part II: Basic statistics ############
> library(fBasics) ## load the library "fBasics"
> basicStats(daxRlog$alvR)
>  ## You can compute basic stats to a full data frame,
 + ## omitting non numeric data
> basicStats(na.omit(daxRlog[,2:5]))
>  ##Use a boxplot to help visualising and interpret results
> boxplot(daxRlog[,2:5])
> ##compute covariance matrix
> cov(daxRlog[,2:5],use="complete.obs")
> ####Extras: To save your table in your working directory
> write.table(dax,file="dax") ## or  as .csv use write.csv
>  ##To read the data saved in working directory
> dax = read.table("dax", header=T)
```

2.7.7: Prove Eq. (2.32). Use the definition of $E(X^n)$ for X log-normally distributed.

2.7.8 R Lab: With the data for one of the stocks analyzed in the previous R Lab, compute the bounds for the price of a stock one week ahead of some initial price P_0

chosen from the data, using the Eq. in (2.34). To estimate the mean and variance of the continuously compounded returns, use around 50 observations of prices previous to P_0. Compare your results with the real prices ahead of P_0.

2.7.9 R Lab: We explore in this lab the possibility of *aggregational normality* in stock returns. This refers to the fact that as one increases the time scale τ (daily, weekly and so on), the distribution of the τ-period returns looks more like a normal distribution. Execute in your R console the following commands (`library(quantmod)`):

```
> appl = getSymbols("AAPL",src="yahoo")
> apRd= periodReturn(appl,period="daily",type="log")
> dsd=density(apRd) #estimate density of daily log ret
> yl=c(min(dsd$y),max(dsd$y)) #set y limits
> plot(dsd,main=NULL,ylim=yl)
> ##plot the normal density with mean, stdv of apRd
> a=seq(min(apRd),max(apRd),0.001)
> points(a,dnorm(a,mean(apRd),sd(apRd)), type="l",lty=2)
```

Repeat the program with `period` $\in$ {`weekly`, `monthly`}. Comment your results.

2.7.10 R Lab: Implement in R the exponential weighted moving average (EWMA) model (Eq. (2.58)) to estimate the annualized volatility of AAPL by the end of May 2009, from its sample return. Use the data from R Example 2.6 and compare your estimate with the estimates done in that example.

http://www.springer.com/978-94-6239-069-0

Computational Finance
An Introductory Course with R
Arratia, A.
2014, X, 301 p. 41 illus., 26 illus. in color., Hardcover
ISBN: 978-94-6239-069-0
A product of Atlantis Press