1 Introduction and Basic Theoretical Concepts

Time series analysis is an integral part of every empirical investigation which aims at describing and modeling the evolution over time of a variable or a set of variables in a statistically coherent way. The economics of time series analysis is thus very much intermingled with macroeconomics and finance which are concerned with the construction of dynamic models. In principle, one can approach the subject from two complementary perspectives. The first one focuses on *descriptive statistics*. It characterizes the empirical properties and regularities using basic statistical concepts like mean, variance, and covariance. These properties can be directly measured and estimated from the data using standard statistical tools. Thus, they summarize the *external* (observable) or outside characteristics of the time series. The second perspective tries to capture the *internal data generating mechanism*. This mechanism is usually unknown in economics as the models developed in economic theory are mostly of a qualitative nature and are usually not specific enough to single out a particular mechanism.[1] Thus, one has to consider some larger class of models. By far most widely used is the class of autoregressive moving-average (ARMA) models which rely on linear stochastic difference equations with constant coefficients. Of course, one wants to know how the two perspectives are related which leads to the important problem of *identifying* a model from the data.

The observed regularities summarized in the form of descriptive statistics or as a specific model are, of course, of principal interest to economics. They can be used to test particular theories or to uncover new features. One of the main assumptions underlying time series analysis is that the regularities observed in the sample period

[1] One prominent exception is the random-walk hypothesis of real private consumption first derived and analyzed by Hall (1978). This hypothesis states that the current level of private consumption should just depend on private consumption one period ago and on no other variable, in particular not on disposable income. The random-walk property of asset prices is another very much discussed hypothesis. See Campbell et al. (1997) for a general exposition and Samuelson (1965) for a first rigorous derivation from market efficiency.

K. Neusser, *Time Series Econometrics*, Springer Texts in Business and Economics,
DOI 10.1007/978-3-319-32862-1_1

are not specific to that period, but can be extrapolated into the future. This leads to the issue of forecasting which is another major application of time series analysis.

Although its roots lie in the natural sciences and in engineering, time series analysis, since the early contributions by Frisch (1933) and Slutzky (1937), has become an indispensable tool in empirical economics. Early applications mostly consisted in making the knowledge and methods acquired there available to economics. However, with the progression of econometrics as a separate scientific field, more and more techniques that are specific to the characteristics of economic data have been developed. I just want to mention the analysis of univariate and multivariate integrated, respectively cointegrated time series (see Chaps. 7 and 16), the identification of vector autoregressive (VAR) models (see Chap. 15), and the analysis of volatility of financial market data in Chap. 8. Each of these topics alone would justify the treatment of time series analysis in economics as a separate subfield.

1.1 Some Examples

Before going into more formal analysis, it is useful to examine some prototypical economic time series by plotting them against time. This simple graphical inspection already reveals some of the issues encountered in this book. One of the most popular time series is the real gross domestic product. Figure 1.1 plots the data for the U.S. from 1947 first quarter to 2011 last quarter on logarithmic scale. Several observations are in order. *First*, the data at hand cover just a part of the time series. There are data available before 1947 and there will be data available after 2011. As there is no natural starting nor end point, we think of a time series as extending back into the infinite past and into the infinite future. *Second*, the observations are treated as the realizations of a random mechanism. This implies that we observe only one realization. If we could turn back time and let run history again, we would obtain a second realization. This is, of course, impossible, at least in the macroeconomics context. Thus, typically, we are faced with just one realization on which to base our analysis. However, sound statistical analysis needs many realizations. This implies that we have to make some assumption on the constancy of the random mechanism over time. This leads to the concept of stationarity which will be introduced more rigorously in the next section. *Third*, even a cursory look at the plot reveals that the mean of real GDP is not constant, but is upward *trending*. As we will see, this feature is typical of many economic time series.[2] The investigation into the nature of the trend and the statistical consequences thereof have been the subject of intense research over the last couple of decades. *Fourth*, a simple way to overcome this

[2]See footnote 1 for some theories predicting non-stationary behavior.

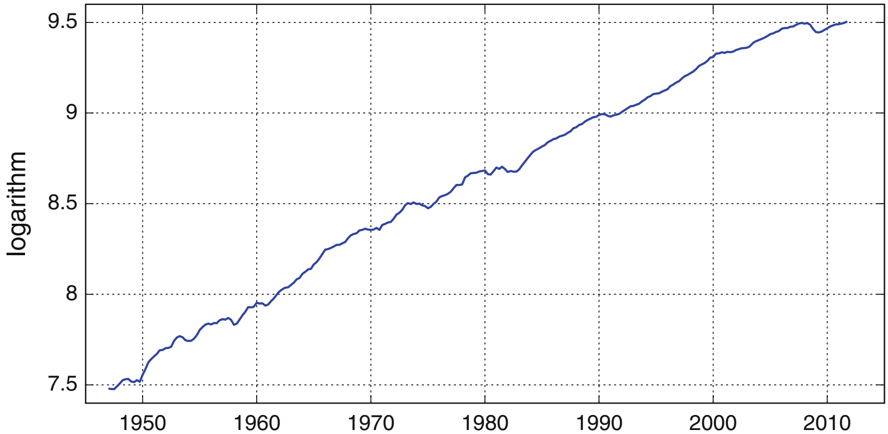

Fig. 1.1 Real gross domestic product (GDP) of the U.S. (chained 2005 dollars; seasonally adjusted annual rate)

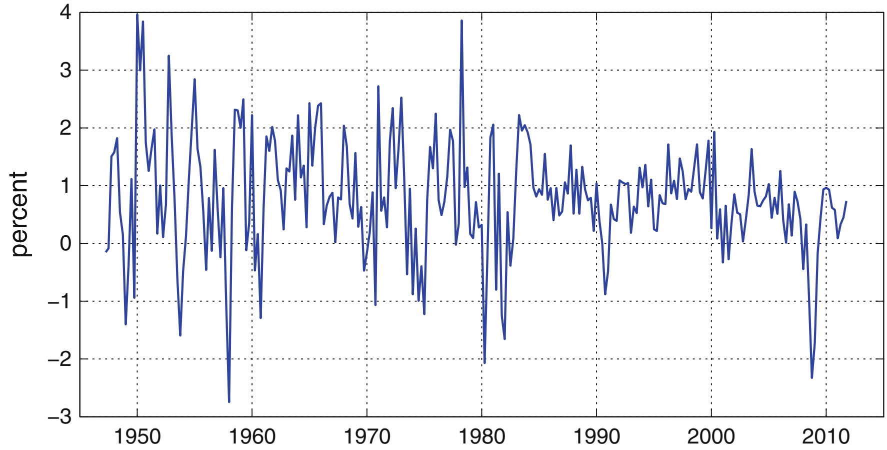

Fig. 1.2 Quarterly growth rate of U.S. real gross domestic product (GDP) (chained 2005 dollars)

problem is to take *first differences*. As the data have been logged, this amounts to taking growth rates.[3] The corresponding plot is given in Fig. 1.2 which shows no trend anymore.

Another feature often encountered in economic time series is *seasonality*. This issue arises, for example in the case of real GDP, because of a particular regularity within a year: the first quarter being the quarter with the lowest values, the second

[3]This is obtained by using the approximation $\ln(1+\varepsilon) \approx \varepsilon$ for small ε where ε equals the growth rate of GDP.

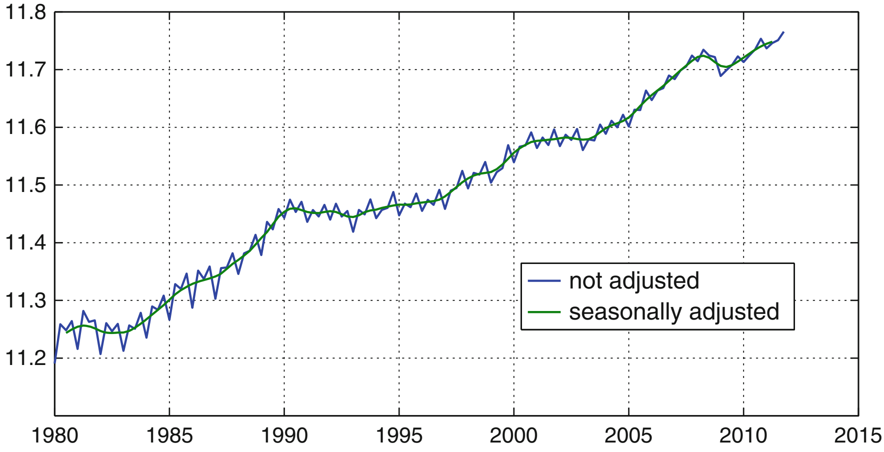

Fig. 1.3 Comparison of unadjusted and seasonally adjusted Swiss real gross domestic product (GDP)

and fourth quarter those with the highest values, and the third quarter being in between. These movements are due to climatical and holiday seasonal variations within the year and are viewed to be of minor economic importance. Moreover, these seasonal variations, because of there size, hide the more important business cycle movements. It is therefore customary to work with time series which have been adjusted for seasonality before hand. Figure 1.3 shows the unadjusted and the adjusted real gross domestic product for Switzerland. The adjustment has been achieved by taking a moving-average. This makes the time series much smoother and evens out the seasonal movements.

Other typical economic time series are interest rates plotted in Fig. 1.4. Over the period considered these two variables also seem to trend. However, the nature of this trend must be different because of the theoretically binding zero lower bound. Although the relative level of the two series changes over time—at the beginning of the sample, short-term rates are higher than long-terms ones—they move more or less together. This *comovement* is true in particular true with respect to the medium- and long-term.

Other prominent time series are stock market indices. In Fig. 1.5 the Swiss Market Index (SMI) in plotted as an example. The first panel displays the raw data on a logarithmic scale. One can clearly discern the different crises: the internet bubble in 2001 and the most recent financial market crisis in 2008. More interesting than the index itself is the return on the index plotted in the second panel. Whereas the mean seems to stay relatively constant over time, the volatility is not: in the periods of crisis volatility is much higher. This *clustering of volatility* is a typical feature of financial market data and will be analyzed in detail in Chap. 8.

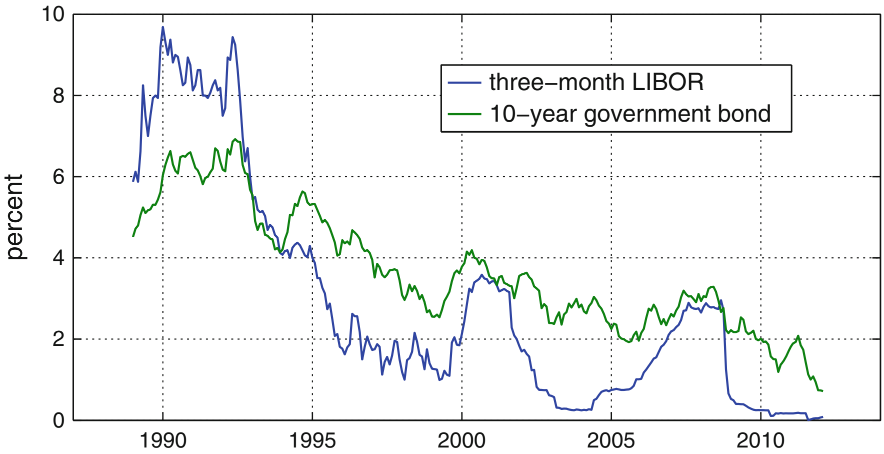

Fig. 1.4 Short- and long-term Swiss interest rates (three-month LIBOR and 10 year government bond)

Finally, Fig. 1.6 plots the unemployment rate for Switzerland. This is another widely discussed time series. However, the Swiss data have a particular feature in that the behavior of the series changes over time. Whereas unemployment was practically nonexistent in Switzerland up to the end of 1990's, several policy changes (introduction of unemployment insurance, liberalization of immigration laws) led to drastic shifts. Although such dramatic *structural breaks* are rare, one has to be always aware of such a possibility. Reasons for breaks are policy changes and simply structural changes in the economy at large.[4]

1.2 Formal Definitions

The previous section attempted to give an intuitive approach of the subject. The analysis to follow necessitates, however, more precise definitions and concepts. At the heart of the exposition stands the concept of a stochastic process. For this purpose we view the observation at some time t as the realization of random variable X_t. In time series analysis we are, however, in general not interested in a particular point in time, but rather in a whole sequence. This leads to the following definition.

Definition 1.1. *A stochastic process* $\{X_t\}$ *is a family of random variables indexed by* $t \in \mathcal{T}$ *and defined on some given probability space.*

[4]Burren and Neusser (2013) investigate, for example, how systematic sectoral shifts affect volatility of real GDP growth.

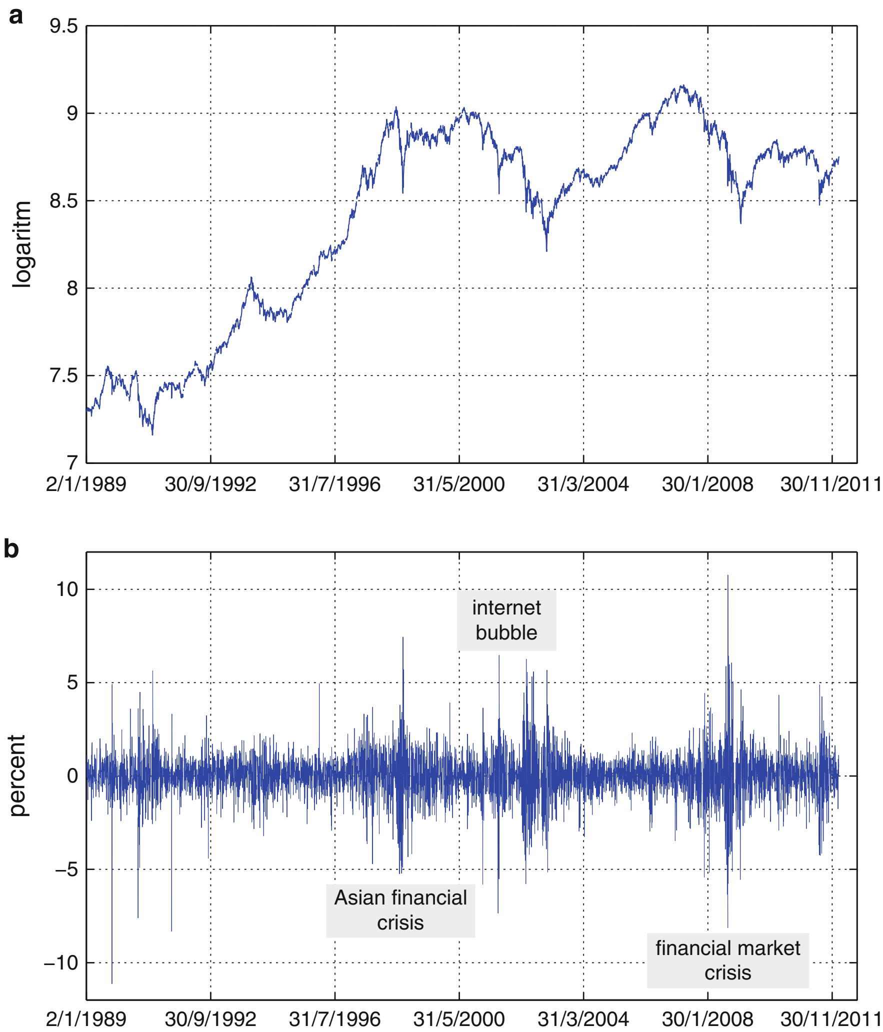

Fig. 1.5 Swiss Market Index (SMI). (**a**) Index. (**b**) Daily return

Thereby $\mathcal{T}$ denotes an ordered index set which is typically identified with time. In the literature one can encounter the following index sets:

$$\text{discrete time: } \mathcal{T} = \{1, 2, \ldots\} = \mathbb{N}$$

$$\text{discrete time: } \mathcal{T} = \{\ldots, -2, -1, 0, 1, 2, \ldots\} = \mathbb{Z}$$

$$\text{continuous time: } \mathcal{T} = [0, \infty) = \mathbb{R}^+ \text{ or } \mathcal{T} = (-\infty, \infty) = \mathbb{R}$$

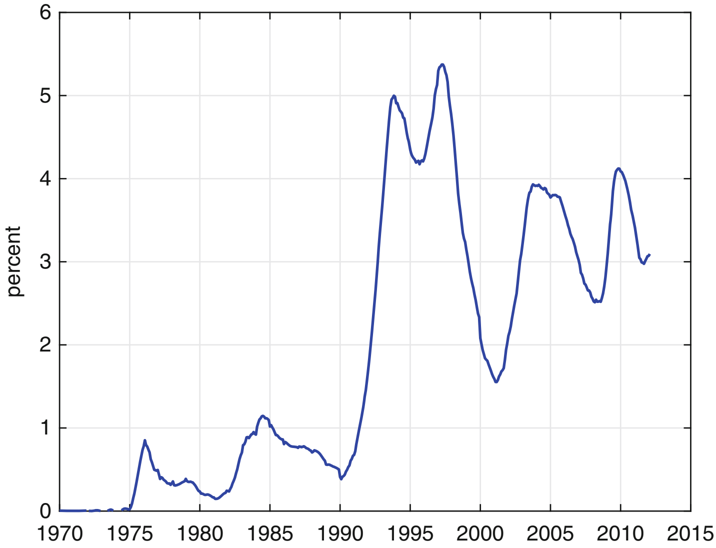

Fig. 1.6 Unemployment rate in Switzerland

Remark 1.1. Given that $\mathcal{T}$ is identified with time and thus has a direction, a characteristic of time series analysis is the distinction between past, present, and future.

For technical reasons which will become clear later, we will work with $\mathcal{T} = \mathbb{Z}$, the set of integers. This choice is consistent with the use of time indices in economics as there is, usually, no natural starting point nor a foreseeable endpoint. Although models in continuous time are well established in the theoretical finance literature, we will disregard them because observations are always of a discrete nature and because models in continuous time would need substantially higher mathematical requirements.

Remark 1.2. The random variables $\{X_t\}$ take values in a so-called *state space*. In the first part of this treatise, we take as the state space the space of real numbers $\mathbb{R}$ and thus consider only univariate time series. In part II we extend the state space to $\mathbb{R}^n$ and study multivariate times series. Theoretically, it is possible to consider other state spaces (for example, $\{0, 1\}$, the integers, or the complex numbers), but this will not be pursued here.

Definition 1.2. *The function $t \to x_t$ which assigns to each point in time t the realization of the random variable X_t, x_t, is called a realization or a trajectory of the stochastic process. We denote such a realization by $\{x_t\}$.*

We denominate by a *time series* the realization or trajectory (observations or data), or the underlying stochastic process. Usually, there is no room for misunderstandings. A trajectory therefore represents one observation of the stochastic process. Whereas in standard statistics a sample consists of several, typically, independent draws from the same distribution, a sample in time series analysis is just one trajectory. Thus, we are confronted with a situation where there is in principle just one observation. We cannot turn back the clock and get additional trajectories. The situation is even worse as we typically observe only the realizations in a particular time window. For example, we might have data on US GDP from the first quarter 1960 up to the last quarter in 2011. But it is clear, the United States existed before 1960 and will continue to exist after 2011, so that there are in principle observations before 1960 and after 2011. In order to make a meaningful statistical analysis, it is therefore necessary to assume that the observed part of the trajectory is typical for the time series as a whole. This idea is related to the concept of *stationarity* which we will introduce more formally below. In addition, we want to require that the observations cover in principle all possible events. This leads to the concept of *ergodicity*. We avoid a formal definition of ergodicity as this would require a sizeable amount of theoretical probabilistic background material which goes beyond the scope this treatise.[5]

An important goal of time series analysis is to build a model given the realization (data) at hand. This amounts to specify the *joint distribution* of some set of X_t's with corresponding realization $\{x_t\}$.

Definition 1.3 (Model)**.** *A time series model or a model for the observations (data) $\{x_t\}$ is a specification of the joint distribution of $\{X_t\}$ for which $\{x_t\}$ is a realization.*

The Kolmogorov existence theorem ensures that the specification of all *finite dimensional* distributions is sufficient to characterize the whole stochastic process (see Billingsley (1986), Brockwell and Davis (1991), or Kallenberg (2002)).

Most of the time it is too involved to specify the complete distribution so that one relies on only the first two moments. These moments are then given by the means $\mathbb{E}X_t$, the variances $\mathbb{V}X_t$, $t \in \mathbb{Z}$, and the covariances $\text{cov}(X_t, X_s) = \mathbb{E}(X_t - \mathbb{E}X_t)(X_s - \mathbb{E}X_s) = \mathbb{E}(X_t X_s) - \mathbb{E}X_t \mathbb{E}X_s$, respectively the correlations $\text{corr}(X_t, X_s) = \text{cov}(X_t, X_s)/(\sqrt{\mathbb{V}X_t}\sqrt{\mathbb{V}X_s})$, $t, s \in \mathbb{Z}$. If the random variables are jointly normally distributed then the specification of the first two moments is sufficient to characterize the whole distribution.

[5]In the theoretical probability theory ergodicity is an important concept which asks the question under which conditions the time average of a property is equal to the corresponding ensemble average, i.e. the average over the entire state space. In particular, ergodicity ensures that the arithmetic averages over time converge to their theoretical counterparts. In Chap. 4 we allude to this principle in the estimation of the mean and the autocovariance function of a time series.

Examples of Stochastic Processes

- $\{X_t\}$ is a sequence of independently distributed random variables with values in $\{-1, 1\}$ such that $\mathbf{P}[X_t = 1] = \mathbf{P}[X_t = -1] = 1/2$. X_t represents, for example, the payoff after tossing a coin: if head occurs one gets a Euro whereas if tail occurs one has to pay a Euro.
- The simple *random walk* $\{S_t\}$ is defined by

$$S_t = S_{t-1} + X_t = \sum_{i=1}^{t} X_i \quad \text{with} \quad t \geq 0 \text{ and } S_0 = 0,$$

  where $\{X_t\}$ is the process from the example just above. In this case S_t is the proceeds after t rounds of coin tossing. More generally, $\{X_t\}$ could be any sequence of identically and independently distributed random variables. Figure 1.7 shows a realization of $\{X_t\}$ for $t = 1, 2, \ldots, 100$ and the corresponding random walk $\{S_t\}$. For more on random walks see Sect. 1.4.4 and, in particular, Chap. 7.
- The simple branching process is defined through the recursion

$$X_{t+1} = \sum_{j=1}^{X_t} Z_{t,j} \quad \text{with starting value: } X_0 = x_0.$$

  In this example X_t represents the size of a population where each member lives just one period and reproduces itself with some probability. $Z_{t,j}$ thereby denotes the number of offsprings of the j-th member of the population in period t. In the simplest case $\{Z_{t,j}\}$ is nonnegative integer valued and identically and independently distributed. A realization with $X_0 = 100$ and with probabilities of one third each that the member has no, one, or two offsprings is shown as an example in Fig. 1.8.

1.3 Stationarity

An important insight in time series analysis is that the realizations in different periods are related with each other. The value of GDP in some year obviously depends on the values from previous years. This temporal dependence can be represented either by an explicit model or, in a descriptive way, by covariances, respectively correlations. Because the realization of X_t in some year t may depend, in principle, on all past realizations $X_{t-1}, X_{t-2}, \ldots$, we do not have to specify just a finite number of covariances, but infinitely many covariances. This leads to the concept of the *covariance function*. The covariance function is not only a tool for summarizing the statistical properties of a time series, but is also instrumental in

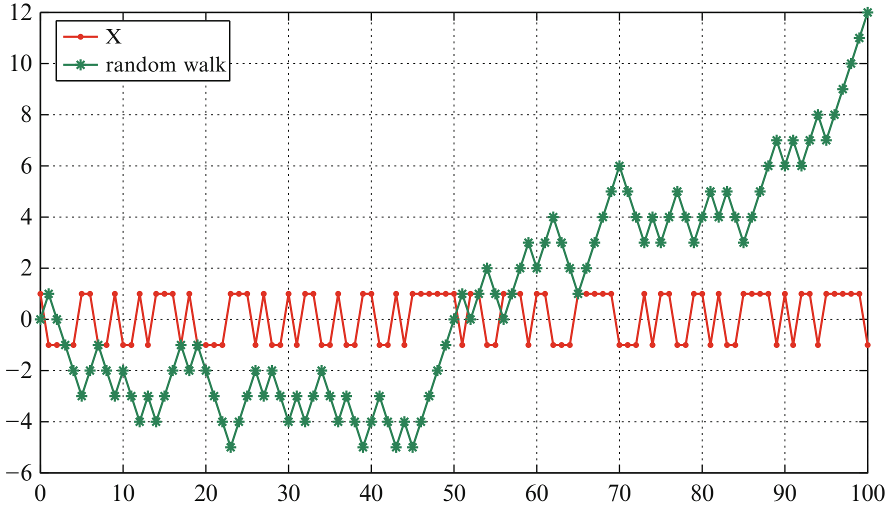

Fig. 1.7 Realization of a random walk

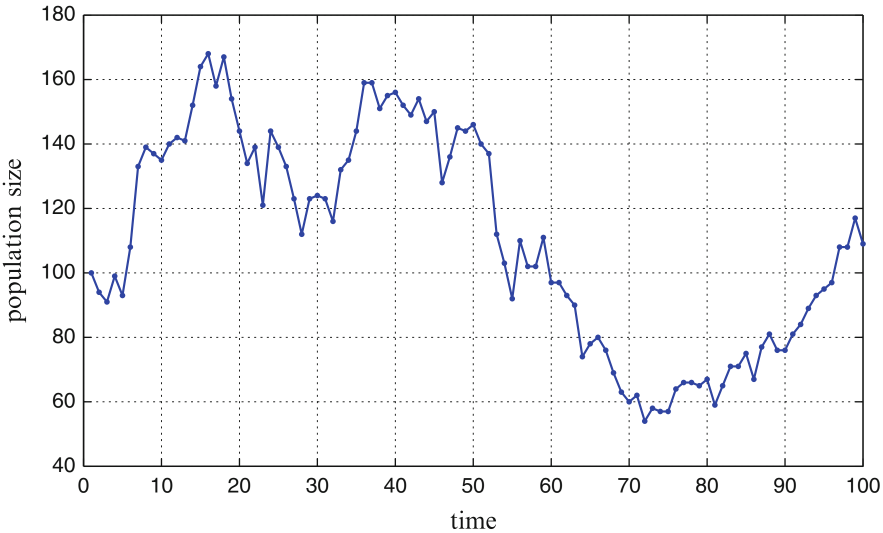

Fig. 1.8 Realization of a branching process

the derivation of forecasts (Chap. 3), in the estimation of ARMA models, the most important class of models (Chap. 5), and in the Wold representation (Sect. 3.2 in Chap. 3). It is therefore of utmost importance to get a thorough understanding of the meaning and properties of the covariance function.

Definition 1.4 (Autocovariance Function)**.** *Let $\{X_t\}$ be a stochastic process with $\mathbb{V}X_t < \infty$ for all $t \in \mathbb{Z}$ then the function which assigns to any two time periods t and s, $t, s \in \mathbb{Z}$, the covariance between X_t and X_s is called the* autocovariance function *of $\{X_t\}$. The autocovariance function is denoted by $\gamma_X(t, s)$. Formally this function is given by*

$$\gamma_X(t, s) = \text{cov}(X_t, X_s) = \mathbb{E}\left[(X_t - \mathbb{E}X_t)(X_s - \mathbb{E}X_s)\right] = \mathbb{E}X_tX_s - \mathbb{E}X_t\mathbb{E}X_s.$$

Remark 1.3. The acronym auto *emphasizes that the covariance is computed with respect to the same variable taken at different points in time. Alternatively, one may use the term* covariance function *for short.*

Definition 1.5 (Stationarity)**.** *A stochastic process $\{X_t\}$ is called* stationary *if and only if for all integers r, s and t the following properties hold:*

(i) $\mathbb{E}X_t = \mu$ *constant;*
(ii) $\mathbb{V}X_t < \infty$*;*
(iii) $\gamma_X(t, s) = \gamma_X(t + r, s + r)$.

Remark 1.4. Processes with these properties are often called weakly stationary, wide-sense stationary, covariance stationary, or second order stationary. As we will not deal with other forms of stationarity, we just speak of stationary processes, for short.

Remark 1.5. For $t = s$, we have $\gamma_X(t, s) = \gamma_X(t, t) = \mathbb{V}X_t$ which is nothing but the unconditional variance of X_t. Thus, if $\{X_t\}$ is stationary $\gamma_X(t, t) = \mathbb{V}X_t = \text{constant}$.

Remark 1.6. If $\{X_t\}$ is stationary, by setting $r = -s$ the autocovariance function becomes:

$$\gamma_X(t, s) = \gamma_X(t - s, 0).$$

Thus the covariance $\gamma_X(t, s)$ does not depend on the points in time t and s, but only on the number of periods t and s are apart from each other, i.e. from $t - s$. For stationary processes it is therefore possible to view the autocovariance function as a function of just one argument. We denote the autocovariance function in this case by $\gamma_X(h)$, $h \in \mathbb{Z}$. Because the covariance is symmetric in t and s, i.e. $\gamma_X(t, s) = \gamma_X(s, t)$, we have

$$\gamma_X(h) = \gamma_X(-h) \qquad \text{for all integers } h.$$

It is thus sufficient to look at the autocovariance function for positive integers only, i.e. for $h = 0, 1, 2, \ldots$. In this case we refer to h as the order of the autocovariance. For $h = 0$, we get the unconditional variance of X_t, i.e. $\gamma_X(0) = \mathbb{V}X_t$.

In practice it is more convenient to look at the autocorrelation coefficients instead of the autocovariances. The *autocorrelation function* (ACF) for stationary processes is defined as:

$$\rho_X(h) = \frac{\gamma_X(h)}{\gamma_X(0)} = \text{corr}(X_{t+h}, X_t) \quad \text{for all integers } h$$

where h is referred to as the order. Note that this definition is equivalent to the ordinary correlation coefficients $\rho(h) = \frac{\text{cov}(X_t, X_{t-h})}{\sqrt{\mathbb{V}X_t}\sqrt{\mathbb{V}X_{t-h}}}$ because stationarity implies that $\mathbb{V}X_t = \mathbb{V}X_{t-h}$ so that $\sqrt{\mathbb{V}X_t}\sqrt{\mathbb{V}X_{t-h}} = \mathbb{V}X_t = \gamma_X(0)$.

Most of the time it is sufficient to concentrate on the first two moments. However, there are situations where it is necessary to look at the whole distribution. This leads to the concept of *strict stationarity*.

Definition 1.6 (Strict Stationarity). *A stochastic process is called* strictly stationary *if the joint distributions of* $(X_{t_1}, \ldots, X_{t_n})$ *and* $(X_{t_1+h}, \ldots, X_{t_n+h})$ *are the same for all* $h \in \mathbb{Z}$ *and all* $(t_1, \ldots, t_n) \in \mathcal{T}^n$, $n = 1, 2, \ldots$

Definition 1.7 (Strict Stationarity). *A stochastic process is called* strictly stationary *if for all integers h and* $n \geq 1$ $(X_1, \ldots, X_n)$ *and* $(X_{1+h}, \ldots, X_{n+h})$ *have the same distribution.*

Remark 1.7. Both definitions are equivalent.

Remark 1.8. If $\{X_t\}$ is strictly stationary then X_t has the same distribution for all t (n=1). For $n = 2$ we have that X_{t+h} and X_t have a joint distribution which is independent of t. This implies that the covariance, if it exists, depends only on h. Thus, every strictly stationary process with $\mathbb{V}X_t < \infty$ is also stationary.[6]

The converse is, however, not true as shown by the following example:

$$X_t \sim \begin{cases} \text{exponentially distributed with mean 1 (i.e. } f(x) = e^{-x}\text{)}, & \text{t uneven;} \\ \text{N}(1, 1), & \text{t even;} \end{cases}$$

whereby the X_t's are independently distributed. In this example we have:

- $\mathbb{E}X_t = 1$
- $\gamma_X(0) = 1$ and $\gamma_X(h) = 0$ for $h \neq 0$

Thus $\{X_t\}$ is stationary, but not strictly stationary, because the distribution changes depending on whether t is even or uneven.

[6]An example of a process which is strictly stationary, but not stationary, is given by the IGARCH process (see Sect. 8.1.4). This process is strictly stationary with infinite variance.

Definition 1.8 (Gaussian Process)**.** *A stochastic process* $\{X_t\}$ *is called a* Gaussian process *if all finite dimensional distributions* $(X_{t_1}, \ldots, X_{t_n})$ *with* $(t_1, \ldots, t_n) \in \mathcal{T}^n$, $n = 1, 2, \ldots$*, are multivariate normally distributed.*

Remark 1.9. A Gaussian process is obviously strictly stationary. For all $n, h, t_1, \ldots, t_n$, $(X_{t_1}, \ldots, X_{t_n})$ and $(X_{t_1+h}, \ldots, X_{t_n+h})$ have the same mean and the same covariance matrix.

At this point we will not delve into the relation between stationarity, strict stationarity and Gaussian processes, rather some of these issues will be further discussed in Chap. 8.

1.4 Construction of Stochastic Processes

One important notion in time series analysis is to build up more complicated process from simple ones. The simplest building block is a process with zero autocorrelation called a *white noise* process which is introduced below. Taking moving-averages from this process or using it in a recursion gives rise to more sophisticated process with more elaborated autocovariance functions. Slutzky (1937) first introduced the idea that moving-averages of simple processes can generate time series whose motion resembles business cycle fluctuations.

1.4.1 White Noise

The simplest building block is a process with zero autocorrelation called a *white noise* process.

Definition 1.9 (White Noise)**.** *A stationary process* $\{Z_t\}$ *is called a* white noise *process if* $\{Z_t\}$ *satisfies:*

- $\mathbb{E}Z_t = 0$
- $\gamma_Z(h) = \begin{cases} \sigma^2 & h = 0; \\ 0 & h \neq 0. \end{cases}$

We denote this by $Z_t \sim \mathrm{WN}(0, \sigma^2)$.

The white noise process is therefore stationary and temporally uncorrelated, i.e. the ACF is always equal to zero, except for $h = 0$ where it is equal to one. As the ACF possesses no structure, it is impossible to draw inferences from past observations to its future development, at least in a least square setting with linear forecasting functions (see Chap. 3). Therefore one can say that a white noise process has no memory.

If $\{Z_t\}$ is not only temporally uncorrelated, but also independently and identically distributed, we write $Z_t \sim \text{IID}(0, \sigma^2)$. If in addition Z_t is normally distributed, we write $Z_t \sim \text{IIN}(0, \sigma^2)$. An $\text{IID}(0, \sigma^2)$ process is always a white noise process. The converse is, however, not true as will be shown in Chap. 8.

1.4.2 Construction of Stochastic Processes: Some Examples

We will now illustrate how complex stationary processes can be constructed by manipulating of a white noise process. In Table 1.1 we report in column 2 the first 6 realizations of a white noise process $\{Z_t\}$. Figure 1.9a plots the first 100 observations. We can now construct a new process $\{X_t^{(\text{MA})}\}$ by taking moving-averages over adjacent periods. More specifically, we take $X_t = Z_t + 0.9Z_{t-1}$, $t = 2, 3, \ldots$. Thus, the realization of $\{X_t^{(\text{MA})}\}$ in period 2 is $\{x_2^{(\text{MA})}\} = -0.8718 + 0.9 \times 0.2590 = -0.6387$.[7] The realization in period 3 is $\{x_3^{(\text{MA})}\} = -0.7879 + 0.9 \times -0.8718 = -1.5726$, and so on. The resulting realizations of $\{X_t^{(\text{MA})}\}$ for $t = 2, \ldots, 6$ are reported in the third column of Table 1.1 and the plot is shown in Fig. 1.9b. On can see that the averaging makes the series more smooth. In Sect. 1.4.3 we will provide a more detailed analysis of this moving-average process.

Another construction device is a recursion: $X_t^{(AR)} = \phi X_{t-1}^{(AR)} + Z_t$, $t = 2, 3, \ldots$, with starting value $X_1^{(AR)} = Z_1$. Such a process is called autoregressive because it refers to its own past. Taking $\phi = 0.9$, the realization of $\{X_t^{(AR)}\}$ in period 2 is $\{x_2^{(AR)}\} = -0.6387 = 0.9 \times 0.2590 - 0.8718$, in period 3 $\{x_3^{(AR)}\} = -1.3627 = 0.9 \times -0.6387 - 0.7879$, and so on. Again the resulting realizations of $\{X_t^{(\text{AR})}\}$ for $t = 2, \ldots, 6$ are reported in the fourth column of Table 1.1 and the plot is shown in Fig. 1.9c. On can see how the series becomes more persistent. In Sect. 2.2.2 we will provide a more detailed analysis of this autoregressive process.

Finally, we construct a new process by taking cumulative sums: $X_t^{(RW)} = \sum_{\tau=1}^{t} Z_\tau$. This process can also be obtained from the recursion above by taking $\phi = 1$ so that $X_t^{(RW)} = X_{t-1}^{(RW)} + Z_t$. It is called a random walk. Thus, the realization of $\{X_t^{(RW)}\}$ for period 2 is $\{x_2^{(RW)}\} = -0.6128 = 0.2590 - 0.8718$, for period 3 $\{x_3^{(RW)}\} = -1.4007 = -0.6128 - 0.7879$, and so on. Again the resulting realizations of $\{X_t^{(\text{RW})}\}$ for $t = 2, \ldots, 6$ are reported in the last column of Table 1.1 and the plot is shown in Fig. 1.9d. On can see how the series moves away from its mean of zero more persistently than all the other three process considered. In Sect. 1.4.4 we will provide a more detailed analysis of this so-called random walk process and show that it is not stationary.

[7] The following calculations are subject to rounding to four digits.

Table 1.1 Construction of stochastic processes assuming $Z_0 = X_0 = 0$

Time	White noise	Moving-average	Auto-regressive	Random walk
1	0.2590	0.2590	0.2590	0.2590
2	−0.8718	−0.6387	−0.6387	−0.6128
3	−0.7879	−1.5726	−1.3627	−1.4007
4	−0.3443	−1.0535	−1.5708	−1.7451
5	0.6476	0.3377	−0.7661	−1.0974
6	2.0541	2.6370	1.3646	0.9567
⋮	⋮	⋮	⋮	⋮

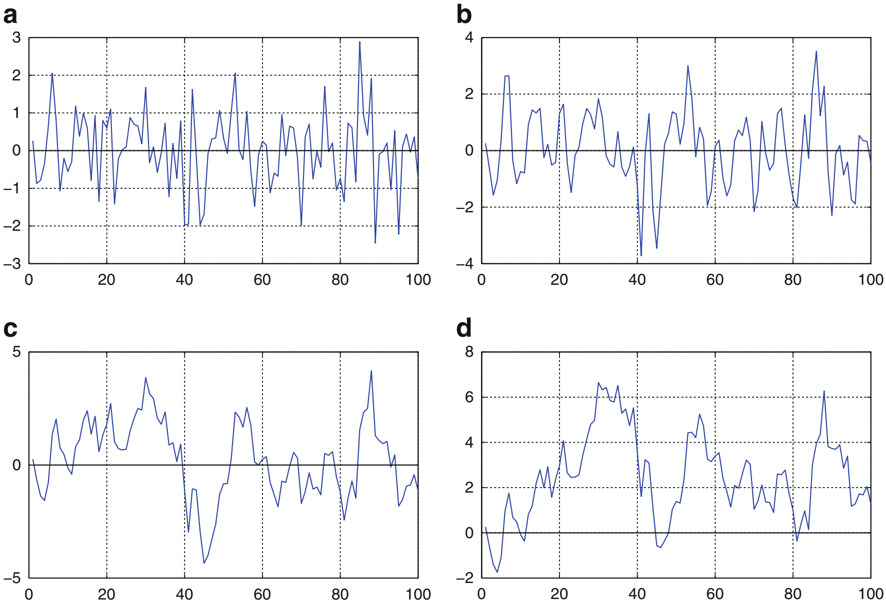

Fig. 1.9 Processes constructed from a given white noise process. (**a**) White noise. (**b**) Moving-average with $\theta = 0.9$. (**c**) Autoregressive with $\phi = 0.9$. (**d**) Random walk

1.4.3 Moving-Average Process of Order One

The white noise process can be used as a building block to construct more complex processes with a more involved autocorrelation structure. The simplest procedure is to take moving averages over consecutive periods.[8] This leads to the moving-average processes. The moving-average process of order one, MA(1) process, is defined as

[8]This procedure is an example of a filter. Section 6.4 provides a general introduction to filters.

$$X_t = Z_t + \theta Z_{t-1} \quad \text{with} \quad Z_t \sim \mathrm{WN}(0, \sigma^2).$$

Clearly, $\mathbb{E}X_t = \mathbb{E}Z_t + \theta \mathbb{E}Z_{t-1} = 0$. The mean is therefore constant and equal to zero.

The autocovariance function can be computed as follows:

$$\begin{aligned}\gamma_X(t+h, t) &= \mathrm{cov}(X_{t+h}, X_t) \\ &= \mathrm{cov}(Z_{t+h} + \theta Z_{t+h-1}, Z_t + \theta Z_{t-1}) \\ &= \mathbb{E}Z_{t+h}Z_t + \theta \mathbb{E}Z_{t+h}Z_{t-1} + \theta \mathbb{E}Z_{t+h-1}Z_t + \theta^2 \mathbb{E}Z_{t+h-1}Z_{t-1}.\end{aligned}$$

Recalling that $\{Z_t\}$ is white noise so that $\mathbb{E}Z_t^2 = \sigma^2$ and $\mathbb{E}Z_t Z_{t+h} = 0$ for $h \neq 0$, we therefore get the following autocovariance function of $\{X_t\}$:

$$\gamma_X(h) = \begin{cases} (1+\theta^2)\sigma^2 & h = 0; \\ \theta\sigma^2 & h = \pm 1; \\ 0 & \text{otherwise.} \end{cases} \tag{1.1}$$

Thus $\{X_t\}$ is stationary irrespective of the value of θ. The autocorrelation function is:

$$\rho_X(h) = \begin{cases} 1 & h = 0; \\ \frac{\theta}{1+\theta^2} & h = \pm 1; \\ 0 & \text{otherwise.} \end{cases}$$

Note that the newly created process now exhibits a dependence from its past as X_t is correlated with X_{t-1}. This correlation is restricted to the interval $[0, \frac{1}{2}]$, i.e. $0 \leq |\rho_X(1)| \leq \frac{1}{2}$. As the correlation between X_t and X_s is zero when t and s are more than one period apart, we call a moving-average process a process with finite memory or a process with finite-range dependence.

Remark 1.10. To motivate the name moving-average, we can define the MA(1) process more generally as

$$X_t = \theta_0 Z_t + \theta_1 Z_{t-1} \quad \text{with} \quad Z_t \sim \mathrm{WN}(0, \sigma^2) \text{ and } \theta_0 \neq 0.$$

Thus, X_t is a weighted average of Z_t and Z_{t-1}. If $\theta_0 = \theta_1 = 1/2$, X_t is just the arithmetic mean of Z_t and Z_{t-1}. This process is, however, (observationally) equivalent to the process

$$X_t = \tilde{Z}_t + \tilde{\theta} \tilde{Z}_{t-1} \quad \text{with} \quad \tilde{Z}_t \sim \mathrm{WN}(0, \tilde{\sigma}^2)$$

where $\tilde{\theta} = \theta_1/\theta_0$ and $\tilde{\sigma}^2 = \theta_0^2\sigma^2$. Both processes would generate the same first two moments and are therefore observationally indistinguishable from each other. Thus, we can set $\theta_0 = 1$ without loss of generality.

1.4.4 Random Walk

Let $Z_t \sim \mathrm{WN}(0, \sigma^2)$ be a white noise process then the new process $\{X_t\}$ defined as

$$X_t = Z_1 + Z_2 + \ldots + Z_t = \sum_{j=1}^{t} Z_j, \qquad t > 0, \tag{1.2}$$

is called a *random walk*. Note that, in contrast to $\{Z_t\}$, $\{X_t\}$ is only defined for $t > 0$. The random walk may alternatively be defined through the recursion

$$X_t = X_{t-1} + Z_t, \qquad t > 0 \text{ and } X_0 = 0.$$

If in each time period a constant δ is added such that

$$X_t = \delta + X_{t-1} + Z_t,$$

the process $\{X_t\}$ is called a *random walk with drift*.

Although the random walk has a constant mean of zero, it is a nonstationary process.

Proposition 1.1. *The random walk $\{X_t\}$ as defined in Eq. (1.2) is nonstationary.*

Proof. The variance of $X_{t+1} - X_1$ equals $\mathbb{V}(X_{t+1} - X_1) = \mathbb{V}\left(\sum_{j=2}^{t+1} Z_j\right) = \sum_{j=2}^{t+1} \mathbb{V}Z_j = t\sigma^2$.

Assume for the moment that $\{X_t\}$ is stationary then the triangular inequality implies for $t > 0$:

$$0 < \sqrt{t\sigma^2} = \mathrm{std}(X_{t+1} - X_1) \leq \mathrm{std}(X_{t+1}) + \mathrm{std}(X_1) = 2\,\mathrm{std}(X_1)$$

where "std" denotes the standard deviation. As the left hand side of the inequality converges to infinity for t going to infinity, also the right hand side must go to infinity. This means that the variance of X_1 must be infinite. This, however, contradicts the assumption of stationarity. Thus $\{X_t\}$ cannot be stationary. □

The random walk represents by far the most widely used nonstationary process in economics. It has proven to be an important ingredient in many economic time series. Typical nonstationary time series which are or are driven by random walks are stock market prices, exchange rates, or the gross domestic product

(GDP). Usually it is necessary to apply some transformation (filter) first to achieve stationarity. In the example above, one has to replace $\{X_t\}$ by its first difference $\{\Delta X_t\} = \{X_t - X_{t-1}\} = \{Z_t\}$ which is stationary by construction. Time series which become stationary after differencing are called integrated processes and are the subject of a more in depth analysis in Chap. 7. Besides ordinary differencing, other transformations are often encountered: seasonal differencing, inclusion of a time trend, seasonal dummies, moving averages, etc. Some of them will be discussed as we go along.

1.4.5 Changing Mean

Finally, here is another simple example of a nonstationary process.

$$X_t = \begin{cases} Y_t, & t < t_c; \\ Y_t + c, & t \geq t_c \text{ und } c \neq 0 \end{cases}$$

where t_c is some specific point in time. $\{X_t\}$ is clearly not stationary because the mean is not constant. In econometrics we refer to such a situation as a *structural change* which can be accommodated by introducing a so-called dummy variable. Models with more sophisticated forms of structural changes will be discussed in Chap. 18

1.5 Properties of the Autocovariance Function

The autocovariance function represents the directly accessible external properties of the time series. It is therefore important to understand its properties and how it is related to its inner structure. We will deepen the connection between the autocovariance function and a particular class of models in Chap. 2. The estimation of the autocovariance function will be treated in Chap. 4. For the moment we will just give its properties and analyze the case of the MA(1) model as a prototypical example.

Theorem 1.1. *The autocovariance function of a stationary process $\{X_t\}$ is characterized by the following properties:*

(i) $\gamma_X(0) \geq 0$*;*
(ii) $0 \leq |\gamma_X(h)| \leq \gamma_X(0)$*;*
(iii) $\gamma_X(h) = \gamma_X(-h)$*;*
(iv) $\sum_{i,j=1}^{n} a_i \gamma_X(t_i - t_j) a_j \geq 0$ *for all n and all vectors $(a_1, \ldots, a_n)'$ and $(t_1, \ldots, t_n)$. This property is called* non-negative definiteness.

Proof. The first property is obvious as the variance is always nonnegative. The second property follows from the Cauchy-Bunyakovskii-Schwarz inequality(see

Theorem C.1) applied to X_t and X_{t+h} which yields $0 \leq |\gamma_X(h)| \leq \gamma_X(0)$. The third property follows immediately from the definition of the covariance. Define $a = (a_1, \ldots, a_n)'$ and $X = (X_{t_1}, \ldots, X_{t_n})'$ then the last property follows from the fact that the variance is always nonnegative: $0 \leq \mathbb{V}(a'X) = a'\mathbb{V}(X)a = \sum_{i,j=1}^{n} a_i \gamma_X(t_i - t_j) a_j$. □

Similar properties hold for the correlation function ρ_X, except that we have $\rho_X(0) = 1$.

Theorem 1.2. *The autocorrelation function of a stationary stochastic process $\{X_t\}$ is characterized by the following properties:*

(i) $\rho_X(0) = 1$;
(ii) $0 \leq |\rho_X(h)| \leq 1$;
(iii) $\rho_X(h) = \rho_X(-h)$;
(iv) $\sum_{i,j=1}^{n} a_i \rho_X(t_i - t_j) a_j \geq 0$ *for all n and all vectors* $(a_1, \ldots, a_n)'$ *and* $(t_1, \ldots, t_n)$.

Proof. The proof follows immediately from the properties of the autocovariance function. □

It can be shown that for any given function with the above properties there exists a stationary process (Gaussian process) which has this function as its autocovariance function, respectively autocorrelation function.

1.5.1 Autocovariance Function of MA(1) Processes

The autocovariance function describes the external observable characteristics of a time series which can be estimated from the data. Usually, we want to understand the internal mechanism which generates the data at hand. For this we need a model. Hence it is important to understand the relation between the autocovariance function and a certain class of models. In this section, by analyzing the MA(1) model, we will show that this relationship is not one-to-one. Thus we are confronted with a fundamental *identification problem*.

In order to make the point, consider the following given autocovariance function:

$$\gamma(h) = \begin{cases} \gamma_0, \ h = 0; \\ \gamma_1, \ h = \pm 1; \\ 0, \quad |h| > 1. \end{cases}$$

The problem consists of determining the parameters of the MA(1) model, θ and σ^2, from the values of the autocovariance function. For this purpose we equate $\gamma_0 = (1 + \theta^2)\sigma^2$ and $\gamma_1 = \theta\sigma^2$ (see Eq. (1.1)). This leads to an equation system in the two unknowns θ und σ^2. This system can be simplified by dividing the second equation by the first one to obtain: $\gamma_1/\gamma_0 = \theta/(1+\theta^2)$. Because $\gamma_1/\gamma_0 = \rho(1) = \rho_1$

one gets a quadratic equation in θ:

$$\rho_1\theta^2 - \theta + \rho_1 = 0.$$

The two solutions of this equation are

$$\theta_{1,2} = \frac{1}{2\rho_1}\left(1 \pm \sqrt{1 - 4\rho_1^2}\right).$$

The solutions are real if and only if the discriminant $1 - 4\rho_1^2$ is positive. This is the case if and only if $\rho_1^2 \leq 1/4$, respectively $|\rho_1| \leq 1/2$. Note that one root is the inverse of the other. The identification problem thus takes the following form:

$|\rho_1| < 1/2$: there exists two observationally equivalent MA(1) processes corresponding to the two solutions θ_1 and θ_2.
$\rho_1 = \pm 1/2$: there exists exactly one MA(1) process with $\theta = \pm 1$.
$|\rho_1| > 1/2$: there exists no MA(1) process with this autocovariance function.

The relation between the first order autocorrelation coefficient, $\rho_1 = \rho(1)$, and the parameter θ of the MA(1) process is represented in Fig. 1.10. As can be seen, there exists for each $\rho(1)$ with $|\rho(1)| < \frac{1}{2}$ two solutions. The two solutions are inverses of each other. Hence one solution is absolutely smaller than one whereas the other is bigger than one. In Sect. 2.3 we will argue in favor of the solution smaller than one. For $\rho(1) = \pm 1/2$ there exists exactly one solution, namely $\theta = \pm 1$. For $|\rho(1)| > 1/2$ there is no solution. For $|\rho_1| > 1/2$, $\rho(h)$ actually does not represent a genuine autocorrelation function as the fourth condition in Theorem 1.1, respectively Theorem 1.2 is violated. For $\rho_1 > \frac{1}{2}$, set $a = (1, -1, 1, -1, \ldots, 1, -1)'$ to get:

$$\sum_{i,j=1}^{n} a_i\rho(i-j)a_j = n - 2(n-1)\rho_1 < 0, \quad \text{if } n > \frac{2\rho_1}{2\rho_1 - 1}.$$

For $\rho_1 = -\frac{1}{2}$ one sets $a = (1, 1, \ldots, 1)'$. Hence the fourth property is violated.

1.6 Exercises

Exercise 1.6.1. *Let the process $\{X_t\}$ be generated by a two-sided moving-average process*

$$X_t = 0.5Z_{t+1} + 0.5Z_{t-1} \qquad \textit{with} \quad Z_t \sim \mathrm{WN}(0, \sigma^2).$$

Determine the autocovariance and the autocorrelation function of $\{X_t\}$.

Exercise 1.6.2. *Let $\{X_t\}$ be the MA(1) process*

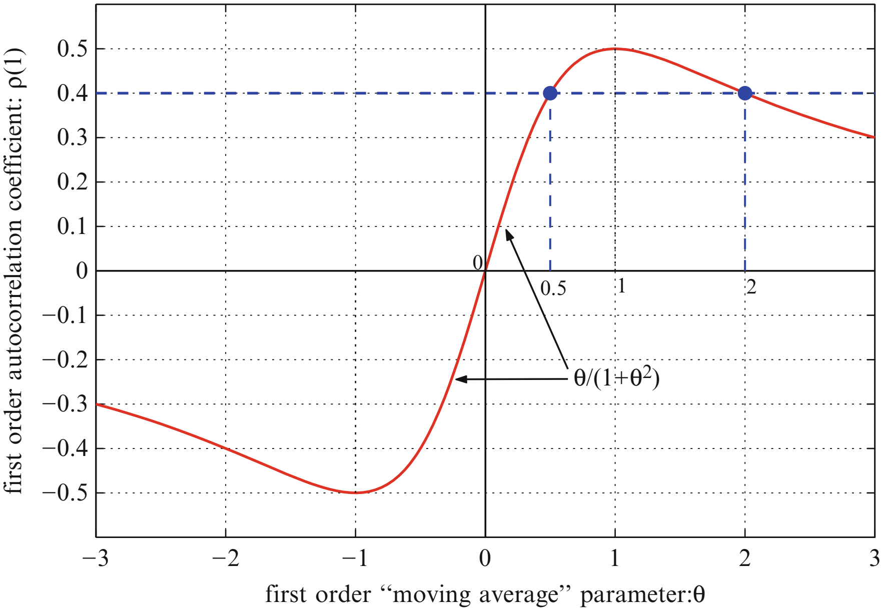

Fig. 1.10 Relation between the autocorrelation coefficient of order one, $\rho(1)$, and the parameter θ of a MA(1) process

$$X_t = Z_t + \theta Z_{t-2} \qquad \textit{with} \quad Z_t \sim \text{WN}(0, \sigma^2).$$

(i) *Determine the autocovariance and the autocorrelation function of* $\{X_t\}$ *for* $\theta = 0.9$.
(ii) *Determine the variance of the mean* $(X_1 + X_2 + X_3 + X_4)/4$.
(iii) *How do the previous results change if* $\theta = -0.9$?

Exercise 1.6.3. *Consider the autocovariance function*

$$\gamma(h) = \begin{cases} 4, & h = 0; \\ -2, & h = \pm 1; \\ 0, & \textit{otherwise.} \end{cases}$$

Determine the parameters θ *and* σ^2, *if they exist, of the first order moving-average process* $X_t = Z_t + \theta Z_{t-1}$ *with* $Z_t \sim \text{WN}(0, \sigma^2)$ *such that autocovariance function above is the autocovariance function corresponding to* $\{X_t\}$.

Exercise 1.6.4. *Let the stochastic process* $\{X_t\}$ *be defined as*

$$\begin{cases} Z_t, & \text{if } t \text{ is even;} \\ (Z_{t-1}^2 - 1)/\sqrt{2}, & \text{if } t \text{ is uneven,} \end{cases}$$

where $\{Z_t\}$ *is identically and independently distributed as* $Z_t \sim \mathrm{N}(0,1)$. *Show that* $\{X_t\} \sim \mathrm{WN}(0,1)$, *but not* $\mathrm{IID}(0,1)$.

Exercise 1.6.5. *Which of the following processes is stationary?*

(i) $X_t = Z_t + \theta Z_{t-1}$
(ii) $X_t = Z_t Z_{t-1}$
(iii) $X_t = a + \theta Z_0$
(iv) $X_t = Z_0 \sin(at)$

In all cases we assume that $\{Z_t\}$ *is identically and independently distributed with* $Z_t \sim \mathrm{N}(0, \sigma^2)$. θ *and* a *are arbitrary parameters.*