

# Chapter 1

## Definition of Long Memory

### 1.1 Historic Overview

A long time before suitable stochastic processes were available, deviations from independence that were noticeable far beyond the usual time horizon were observed, often even in situations where independence would have seemed a natural assumption. For instance, the Canadian–American astronomer and mathematician Simon Newcomb (Newcomb 1895) noticed that in astronomy errors typically affect whole groups of consecutive observations and therefore drastically increase the “probable error” of estimated astronomical constants so that the usual  $\sigma/\sqrt{n}$ -rule no longer applies. Although there may be a number of possible causes for Newcomb’s qualitative finding, stationary long-memory processes provide a plausible “explanation”. Similar conclusions were drawn before by Peirce (1873) (see also the discussion of Peirce’s data by Wilson and Hilferty (1929) and later in the book by Mosteller and Tukey (1977) in a section entitled “How  $\sigma/\sqrt{n}$  can mislead”). Newcomb’s comments were confirmed a few years later by Pearson (1902), who carried out experiments simulating astronomical observations. Using an elaborate experimental setup, he demonstrated not only that observers had their own personal bias, but also each individual measurement series showed persisting serial correlations. For a discussion of Pearson’s experiments, also see Jeffreys (1939, 1948, 1961), who uses the term “internal correlation”. Student (1927) observes the “phenomenon which will be familiar to those who have had astronomical experience, namely that analyses made alongside one another tend to have similar errors; not only so but such errors, which I may call semi-constant, tend to persist throughout the day, and some of them throughout the week or the month. . . . Why this is so is often quite obscure, though a statistical examination may enable the head of the laboratory to clear up large sources of error of this kind: it is not likely that he will eliminate all such errors. . . . The chemist who wishes to impress his clients will therefore arrange to do repetition analyses as nearly as possible at the same time, but if he wishes to diminish his real error, he will separate them by as wide an interval of time as possible.” Since, according to Student, it is difficult to remove the error even by careful statistical examination, simple trends are probably not what he had in mind. Instead, a second-order

property such as slowly decaying autocorrelations may come close to his notion of “semi-constant errors”. For spatial data, the Australian agronomer Smith (1938) found in so-called uniformity trials an empirical law for wheat yield variation across space that contradicts the assumption of independence or summable correlations since the standard deviation of the sample mean converges to zero at a slower rate than the square root of the plot size. These findings were later taken up by Whittle (1956, 1962), who proposed space-time models based on stochastic partial differential equations exhibiting hyperbolically decaying spatial correlations and thereby a possible explanation of Fairfield Smith’s empirical law. In hydrology, Hurst (1951) discovered an empirical law while studying the long-term storage capacity of reservoirs for the Nile (also see Hurst et al. 1965). Built on his empirical findings, Hurst recommended to increase the height of the planned Aswan High Dam far beyond conventional forecasts. Feller (1951) showed that Hurst’s findings are incompatible with the assumption of weak dependence or finite moments. Later Mandelbrot coined the terms “Noah effect” for long-tailed distributions and Joseph- or Hurst-effect for “long-range dependence”. The latter refers to Genesis 41, 29–30, where the “seven years of great abundance” and “seven years of famine” may be interpreted as an account of strong serial correlations. The approach of Mandelbrot and his coworkers lead to a new branch of mathematics that replaced conventional geometric objects by “fractals” and “self-similarity” (e.g. Mandelbrot 1965, 1967, 1969, 1971, 1977, 1983; Mandelbrot and van Ness 1968; Mandelbrot and Wallis 1968a, 1968b, 1969a, 1969b, 1969c) and popularized the topic in many scientific fields, including statistics. In economics, the phenomenon of long memory was discovered by Granger (1966). Simultaneously with Hosking (1981), Granger and Joyeux (1980) introduced fractional ARIMA models that greatly improved the applicability of long-range dependence in statistical practice. In geology, Matheron developed the field of geostatistics using, in particular, processes and statistical techniques for modelling spatial long memory (see e.g. Matheron 1962, 1973; Solo 1992). From the mathematical point of view, the basic concepts of fractals, self-similarity and long-range dependence existed long before the topic became fashionable; however, their practical significance had not been fully recognized until Mandelbrot’s pioneering work. For instance, the Hausdorff dimension, which plays a key role in the definition of fractals, was introduced by Hausdorff (1918) and studied in detail by Abram Samoilovitch Besicovitch (e.g. Besicovitch 1929; Besicovitch and Ursell 1937). In the 17th century, Leibnitz (1646–1716) considered recursive self-similarity, and about one hundred years later, Karl Weierstrass described a function that is continuous but nowhere differentiable. The first fractal is attributed to the Czech mathematician Bernard Bolzano (1781–1848). Other early fractals include the Cantor set (Cantor 1883; but also see Smith 1875; du Bois-Reymond 1880 and Volterra 1881), the Koch snowflake (von Koch 1904), Waclaw Sierpiński’s triangle (Sierpinksi 1915) and the Lévy curve (Lévy 1938). (As a precaution, it should perhaps be mentioned at this place that, although fractal behaviour is often connected with long-range dependence, it is by no means identical and can, in some situations, even be completely separated from the dependence structure; see Chap. 3, Sect. 3.6.) Mathematical models for long-memory type behaviour in physics have

been known for some time in the context of turbulence (see e.g. Kolmogorov 1940, 1941). Power-law correlations have been known to be connected with critical phenomena, for instance in particle systems such as the Ising model (Ising 1924) and the renormalization group (see e.g. Cassandro and Jona-Lasinio 1978, also see the review paper by Domb 1985 and references therein). The study of critical phenomena in physics goes even much further back in history (Berche et al. 2009), to Baron Charles Cagniard de la Tour (1777–1859), who called a critical point in the phase transition “l'état particulier”. With respect to unusual limit theorems for dependent observations, Rosenblatt (1961) seems to be among the first ones to derive a noncentral limit theorem where the limiting process is non-Gaussian due to nonsummable correlations and nonlinearity. This seminal paper led to further developments in the 1970s and 1980s (see e.g. Davydov 1970a, 1970b; Taqqu 1975, 1979; Dobrushin and Major 1979). The literature on statistical methods for long-memory processes until the early 1990s is summarized in Beran (1994a).

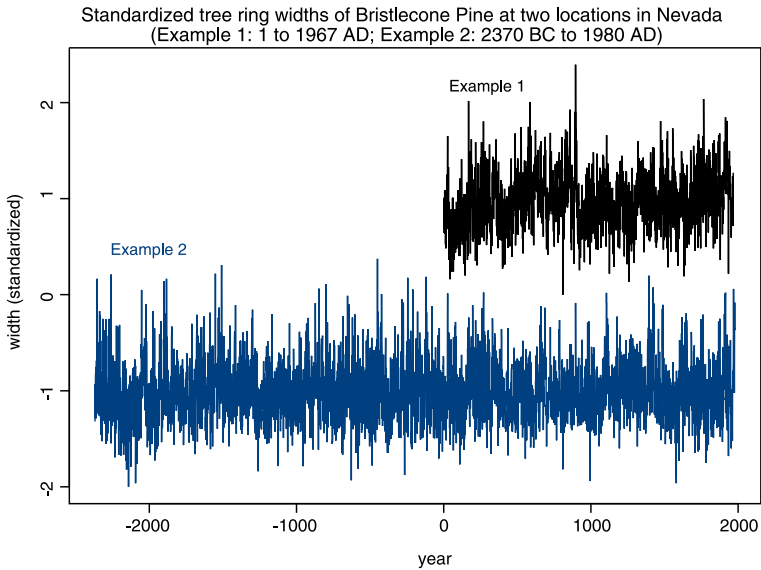
## 1.2 Data Examples

In this section we discuss some data examples with typical long-memory behaviour. On the way, a few heuristic methods for detecting and assessing the strength of long-range dependence will be introduced (see Sect. 5.4).

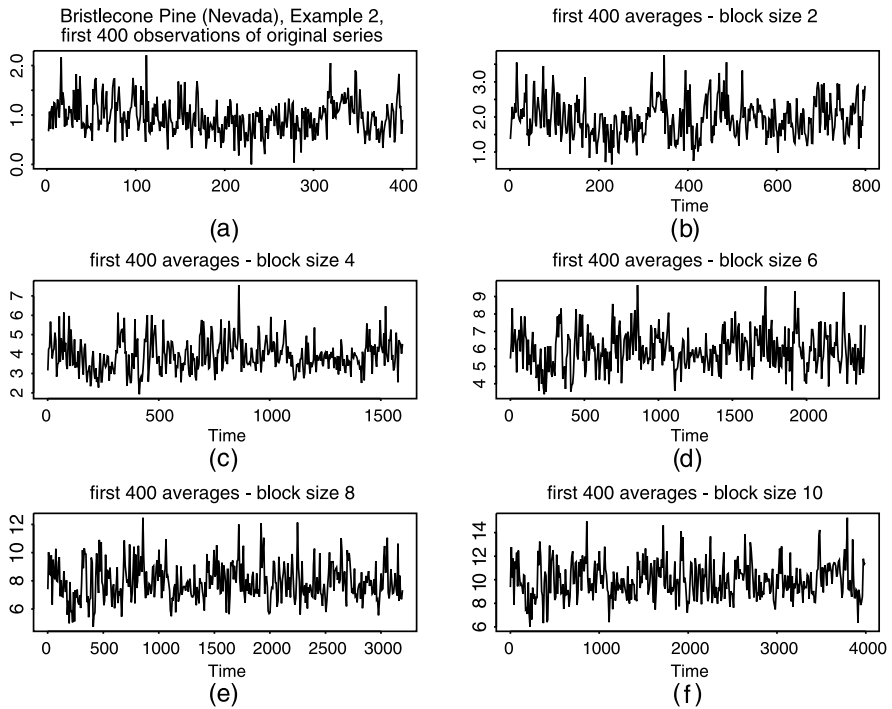
Classical areas where long-range dependence occurs frequently are dendrochronology and hydrology. We will therefore start with examples from these fields. Yearly tree ring measurements usually stretch over hundreds of years, and long memory often occurs in a rather ‘pure’ form, in the sense that a hyperbolic behaviour of the autocorrelations and the spectral density holds for almost all lags and frequencies respectively. Therefore, tree ring series are often used as prime examples of strong dependence and self-similarity. Consider for instance Fig. 1.1 (the data source is Hyndman, Time Series Data Library, <http://robjhyndman.com/TSDL>). The following typical features can be observed:

- (a) *Spurious trends and cycles, and self-similarity*: The observed series exhibit local trends and periodicities that appear to be spurious, however, because they disappear again and are of varying length and frequency. Furthermore, these features and the overall visual impression of the time series remain the same when considering aggregated data, with disjoint adjacent blocks of observations being averaged (see Fig. 1.2). This is an indication of stochastic ‘self-similarity’, which is the property that rescaling time changes the (joint) probability distribution by a scaling factor only.
- (b) *Slow hyperbolic decay*: The sample autocorrelations

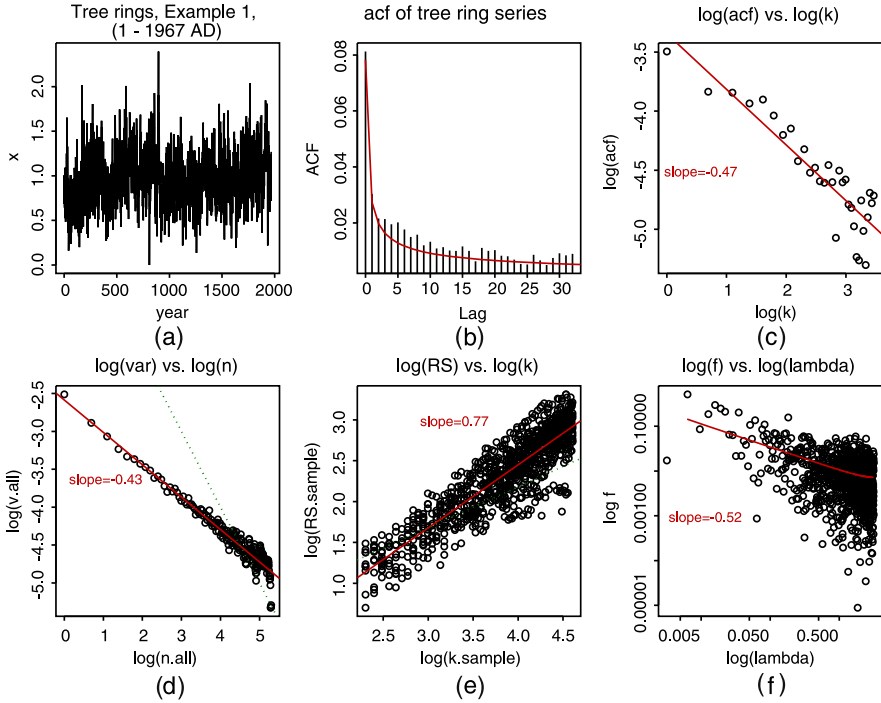
$$\hat{\rho}(k) = \frac{1}{n} \sum_{i=1}^{n-|k|} (x_i - \bar{x})(x_{i+|k|} - \bar{x})$$



**Fig. 1.1** Two typical tree ring series



**Fig. 1.2** (a) Tree ring series, Example 1; (b)–(f) aggregated series  $\bar{x}_t = m^{-1}(x_{(t-1)m+1} + \dots + x_{tm})$  ( $t = 1, 2, \dots, 400$ ) with blocks lengths equal to 2, 4, 6, 8 and 10 respectively



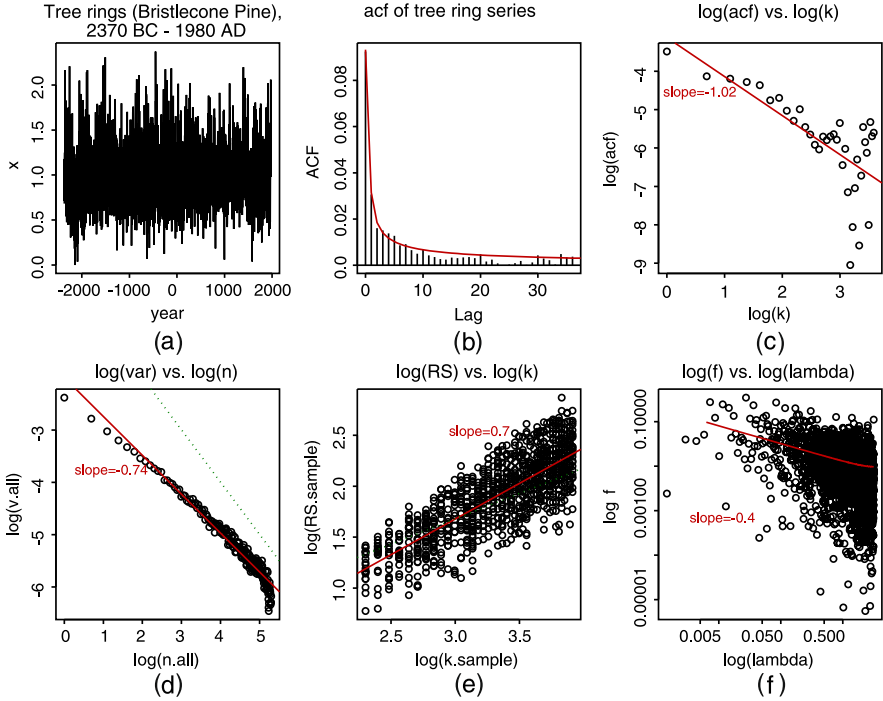
**Fig. 1.3** Tree ring example 1: (a) observed yearly series; (b) empirical autocorrelations  $\hat{\rho}(k)$ ; (c)  $\log \hat{\rho}(k)$  vs.  $\log k$ ; (d)  $\log s_m^2$  vs.  $\log m$ ; (e)  $\log R/S$  vs.  $\log k$ ; (f)  $\log I(\lambda)$  vs.  $\log \lambda$

(with  $\bar{x} = n^{-1} \sum x_i$ ) decay slowly with increasing lag  $k$ . More specifically, the decay of  $\hat{\rho}(k)$  appears to be hyperbolic with a rate  $k^{-\alpha}$  (for some  $0 < \alpha < 1$ ), implying nonsummability. This phenomenon is called long memory, strong memory, long-range dependence, or long-range correlations. This is illustrated in Fig. 1.3(c), where  $\log \hat{\rho}(k)$  is plotted against  $\log k$ . The points are scattered around a straight line of the form  $\log \hat{\rho}(k) \approx \text{const} + \beta_\rho \log k$  with  $\beta_\rho \approx -0.5$ . Similarly, the variance of the sample mean appears to decay to zero at a slower rate than  $n^{-1}$ . This can be seen empirically in Fig. 1.3(d) with  $\log s_m^2$  plotted against  $\log m$ , where  $s_m^2$  is the sample variance of means based on disjoint blocks of  $m$  observations, i.e.

$$s_m^2 = \frac{1}{n_m - 1} \sum_{i=1}^{n_m} (\bar{x}_{(i-1)m,m} - \bar{x})^2,$$

where

$$\bar{x}_{t,m} = \frac{1}{m} \sum_{j=1}^m x_{t+j}$$



**Fig. 1.4** Tree ring example 2: (a) observed yearly series; (b) empirical autocorrelations  $\hat{\rho}(k)$ ; (c)  $\log \hat{\rho}(k)$  vs.  $\log k$ ; (d)  $\log s_m^2$  vs.  $\log m$ ; (e)  $\log R/S$  vs.  $\log k$ ; (f)  $\log I(\lambda)$  vs.  $\log \lambda$

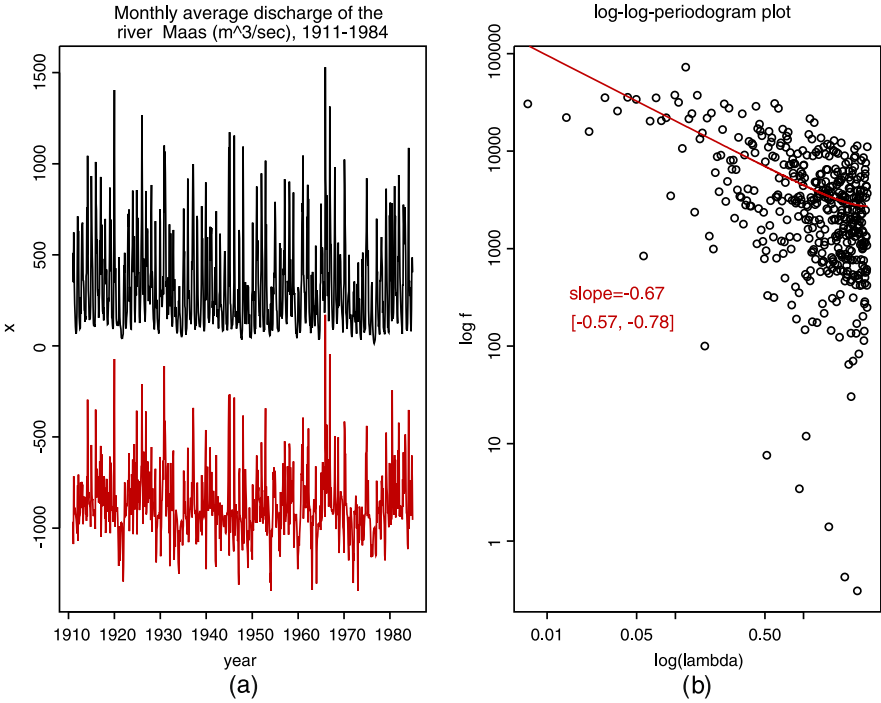
and  $n_m = \lceil n/m \rceil$ . The fitted slope in Fig. 1.3(d) is close to  $\beta_{s^2} = -0.4$ , suggesting  $s_m^2$  being proportional to  $m^{-0.4}$ , which is much slower than the usual rate of  $m^{-1}$ . A further statistic that is sometimes used to detect long-range dependence is the so-called  $R/S$ -statistic displayed in Fig. 1.3(f). The  $R/S$ -statistic is defined by

$$R/S(t, m) = \frac{R(t, m)}{S(t, m)},$$

where

$$R(t, m) = \max_{1 \leq i \leq m} \left( y_{t+i} - y_t - \frac{i}{m} (y_{t+m} - y_t) \right) - \min_{1 \leq i \leq m} \left( y_{t+i} - y_t - \frac{i}{m} (y_{t+m} - y_t) \right),$$

$$y_u = \sum_{i=1}^u x_i,$$

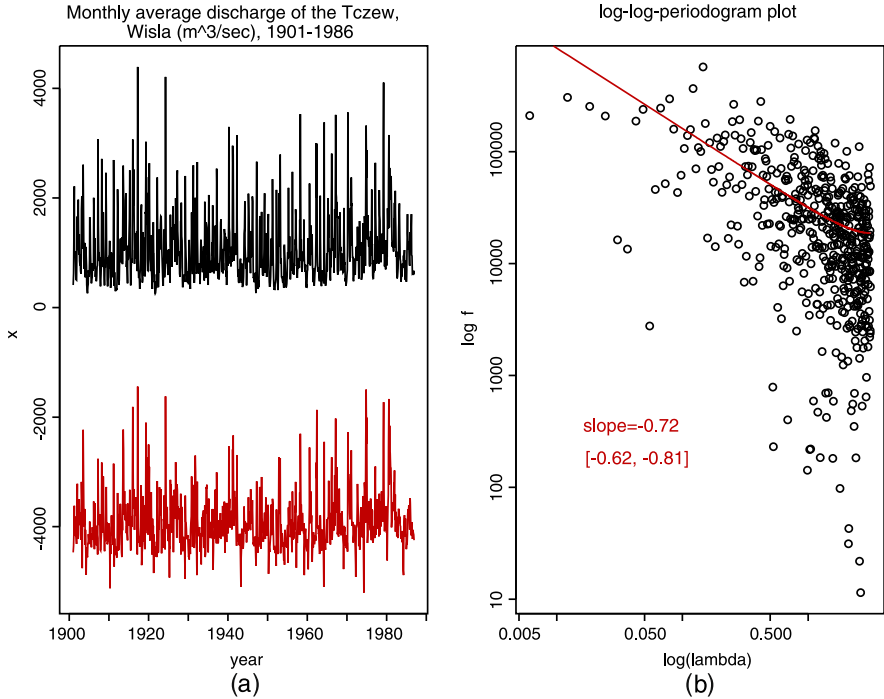


**Fig. 1.5** (a) Monthly average discharge of the river Maas (*upper series*: original; *lower series*: deseasonalized); (b) log-log-periodogram of the deseasonalized series in (a)

and

$$S(t, m) = \sqrt{\frac{1}{m} \sum_{i=t+1}^{t+m} (x_i - \bar{x}_{t,m})^2}.$$

This definition originates from hydrology (see e.g. Hurst 1951), where  $R$  corresponds to the optimal capacity of a reservoir when outflow is linear, with  $x_i$  denoting the inflow at time  $i$ . Figure 1.3(f) shows  $R/S(t, m)$  versus  $m$ , plotted in log-log-coordinates. Again, we see a linear relationship between  $\log R/S$  (as a function of  $m$ ) and  $\log m$ , with a slope close to  $\beta_{R/S} = 0.8$ . In contrast, under independence or short-range dependence, one expects a slope of 0.5 (see Sect. 5.4.1). Finally, Fig. 1.3(f) displays the logarithm of the periodogram  $I(\lambda)$  (as an empirical analogue of the spectral density  $f$ ) versus the log-frequency. Again an essentially linear relationship can be observed. The negative slope is around  $\beta_f = -0.5$ , suggesting the spectral density having a pole at the origin of the order  $\lambda^{-0.5}$ . Similar results are obtained for Example 2 in Figs. 1.4(a) through (f). The slopes for the log-log plots of  $\hat{\rho}(k)$ ,  $s_m^2$ ,  $R/S$  and  $I(\lambda)$  are this time  $\beta_\rho \approx -1$ ,  $\beta_{s^2} \approx -0.7$ ,  $\beta_{R/S} \approx 0.7$  and  $\beta_f \approx -0.4$  respectively.

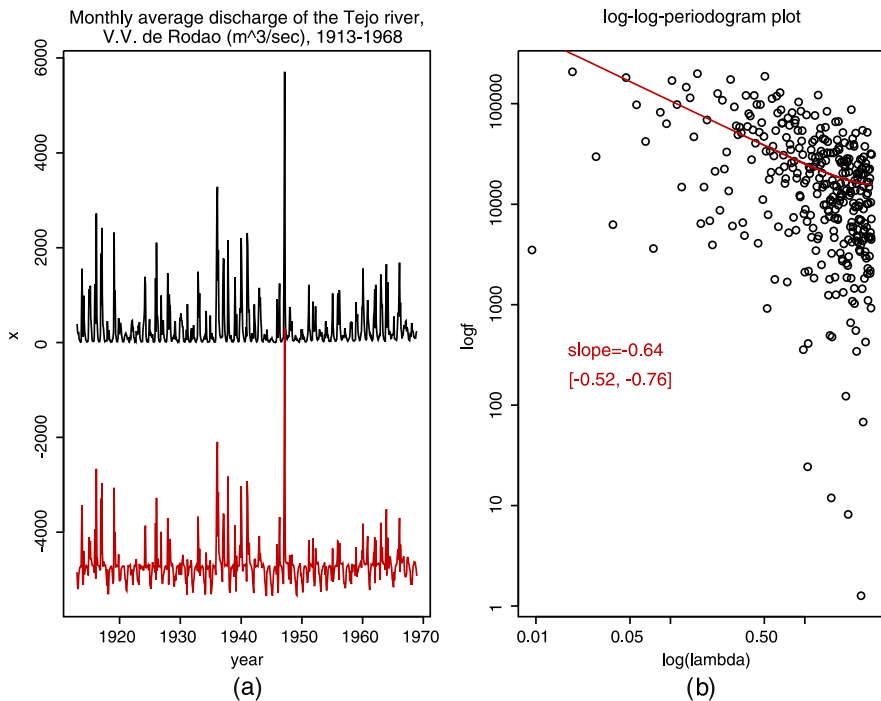


**Fig. 1.6** (a) Monthly average discharge of the river Wisła at Tczew (*upper series*: original; *lower series*: deseasonalized); (b) log-log-periodogram of the deseasonalized series in (a)

Next, we consider river flow data. Figures 1.5(a), 1.6(a), 1.7(a) and 1.8(a) show the average monthly river discharge (in  $\text{m}^3/\text{sec}$ ) for four rivers from different parts of the world: (1) Maas at the Lith station (The Netherlands); (2) Wisła at Tczew (Poland); (3) Tejo at V.V. de Rodao (Portugal) and (4) White River at Mouth Near Ouray, Utah (USA). The data are from the River Discharge Database of The Center of Sustainability and Global Environment, Gaylord Nelsen Institute for Environmental Studies, University of Wisconsin-Madison. Since these are monthly data, there is a strong seasonal component. To obtain an idea about the dependence structure for large lags, a seasonal effect is first removed by subtracting the corresponding monthly means (i.e. average January temperature, average February temperature etc.). The original and the deseasonalized data are shown in the upper and lower part of each time series picture respectively. For each of the deseasonalized series, the points in the log-log-periodogram (all figures (b)) are scattered nicely around a straight line for all frequencies.

The data examples shown so far may be somewhat misleading because one may get the impression that discovering long memory can be done easily by fitting a straight line to the observed points in an appropriate log-log-plot. Unfortunately, the situation is more complicated, even if one considers river flows only. For instance, Figs. 1.9, 1.10 and 1.11 show log-log-plots for the Danube at four different

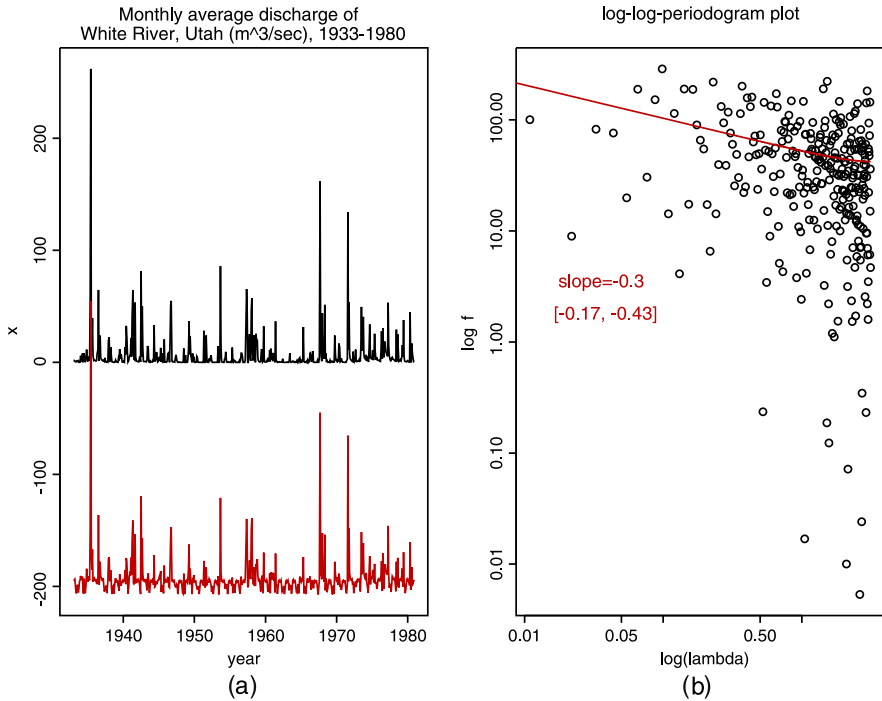




**Fig. 1.7** (a) Monthly average discharge of the river Tejo at V.V. de Rodao (*upper series*: original; *lower series*: deseasonalized); (b) log-log-periodogram of the deseasonalized series in (a)

stations: (1) Bratislava (Slovakia); (2) Nagymaros (Hungary); (3) Drobeta-Turnu Severin (Romania); (4) Ceatal Izmail (Romania). Consider first the measurements in Bratislava. The points in the log-log-plots no longer follow a straight line all the way. It is therefore not clear how to estimate the ‘ultimate’ slopes (i.e. the asymptotic slopes as  $m, k \rightarrow \infty$  and  $\lambda \rightarrow 0$  respectively). Fitting a straight line to all points obviously leads to a bad fit in the region of interest (i.e. for  $k$  and  $m$  large, and  $\lambda$  small). This is one of the fundamental problems when dealing with long-memory (and, as we will see later, also so-called antipersistent) series: the definition of ‘long memory’ is an asymptotic one and therefore often difficult to detect and quantify for finite samples. A substantial part of the statistical literature on long-memory processes is concerned with this question (this will be discussed in particular in Chap. 5). In contrast to the straight lines in Figs. 1.9(b) and (c), the fitted spectral density in Fig. 1.9(d) is based on a more sophisticated method that combines maximum likelihood estimation (MLE) with the Bayesian Information Criterion (BIC) for fractional ARIMA models. This and related data adaptive methods that allow for deviations from the straight line pattern will be discussed in Chap. 5 (Sects. 5.5 to 5.10) and Chap. 7 (Sects. 7.4.5 and 7.4.6).

Analogous observations can be made for the other Danube series. To save space, only the log-log-periodogram plots are shown (Figs. 1.10, 1.11). Note that the MLE

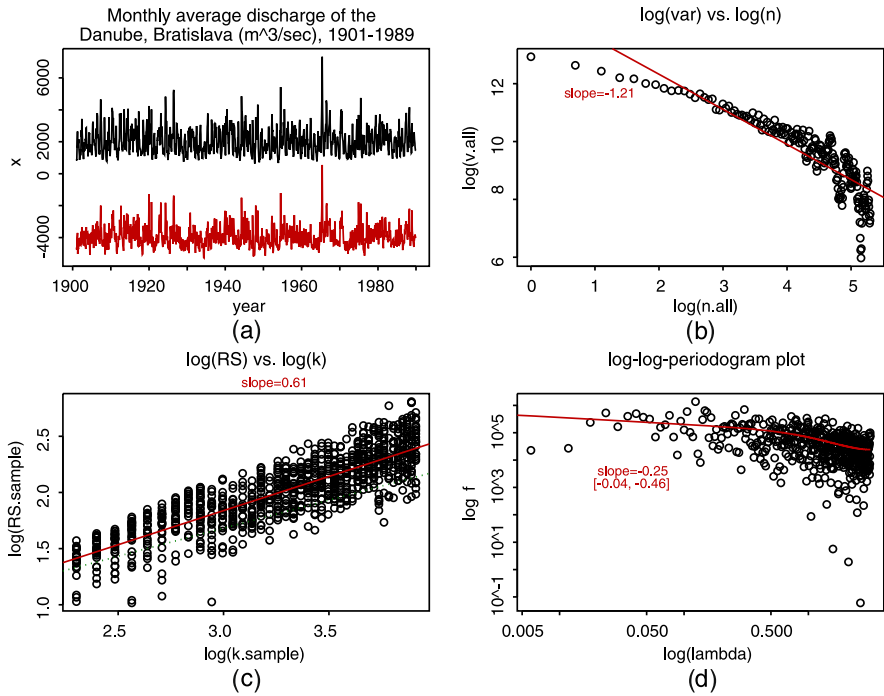


**Fig. 1.8** (a) Monthly average discharge of White River, Utah (*upper series*: original; *lower series*: deseasonalized); (b) log-log-periodogram of the deseasonalized series in (a)

estimates of  $\beta_f$  ( $-0.25$ ,  $-0.31$ ,  $-0.25$ ,  $-0.29$ ) are all very similar. It seems that a value around  $-0.25$  to  $-0.3$  is typical for the Danube in these regions. On the other hand, the slope changes as one moves upstream. For instance, at Hofkirchen in Germany (lower panel in Sect. 1.11), long memory appears to be much stronger with  $\beta_f \approx -0.75$ , and a straight line fits all the way.

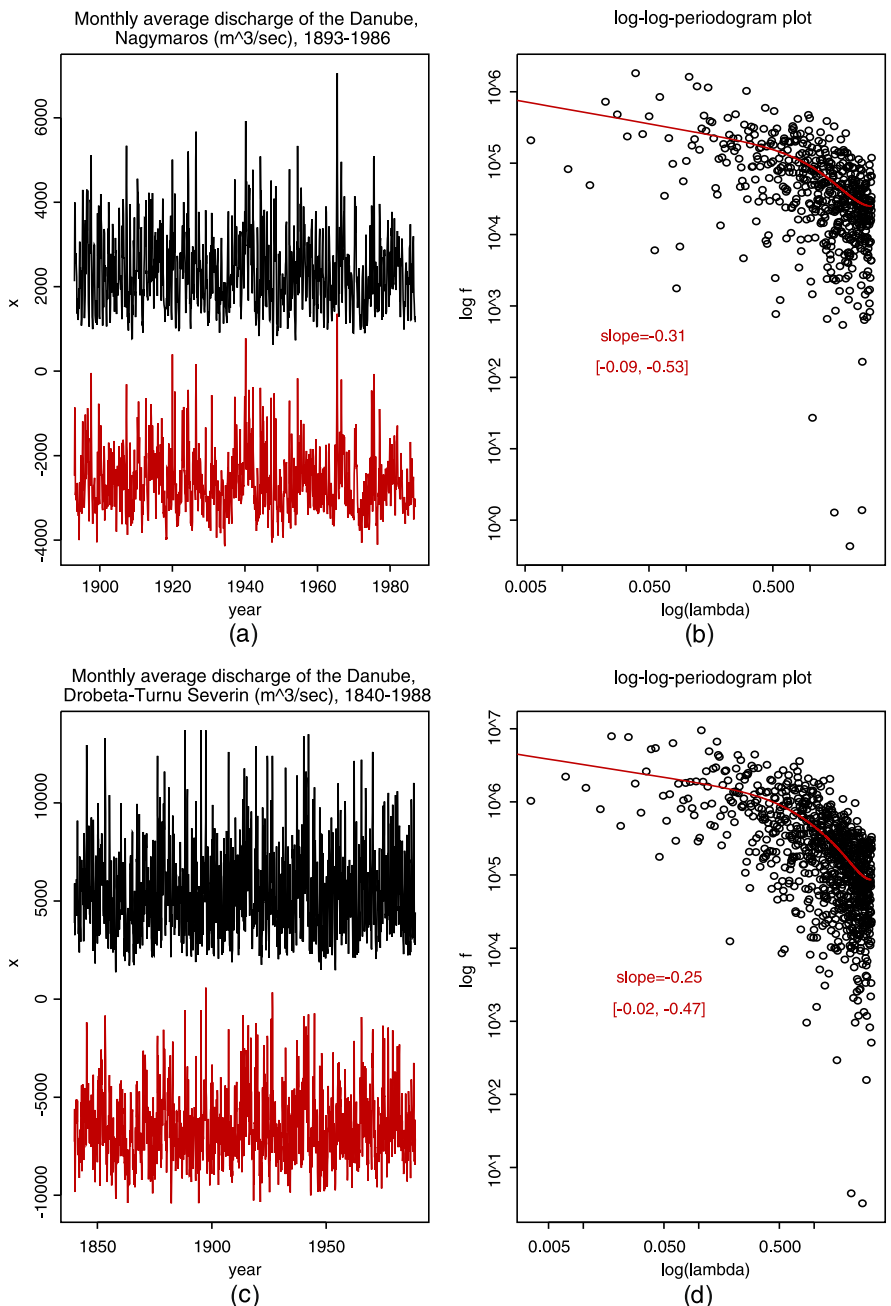
An even more complex river flow series are monthly measurements of the Nile river at Dongola in Sudan, displayed in Fig. 1.12. Seasonality is very strong here, and subtracting seasonal means does not remove all of it (see Figs. 1.12(a), (b)). A possible reason is that the seasonal effect may change over time; it may be non-linear, or it may be stochastic. The MLE fit combined with the BIC captures the remaining seasonality quite well. This model assumes seasonality (remaining after previous subtraction of the deterministic one) to be stochastic.

The data examples considered so far could be modelled by stationary processes. Often stationarity is not a realistic assumption, or it is at least uncertain. This makes identification of stochastic long memory even more difficult, because typical long-memory features may be confounded with nonstationary components. Identifying and assessing possible long-memory components is however essential for correct inference about the non-stationary components. A typical example is the assessment of global warming. Figure 1.13(a) shows yearly average temperatures in cen-

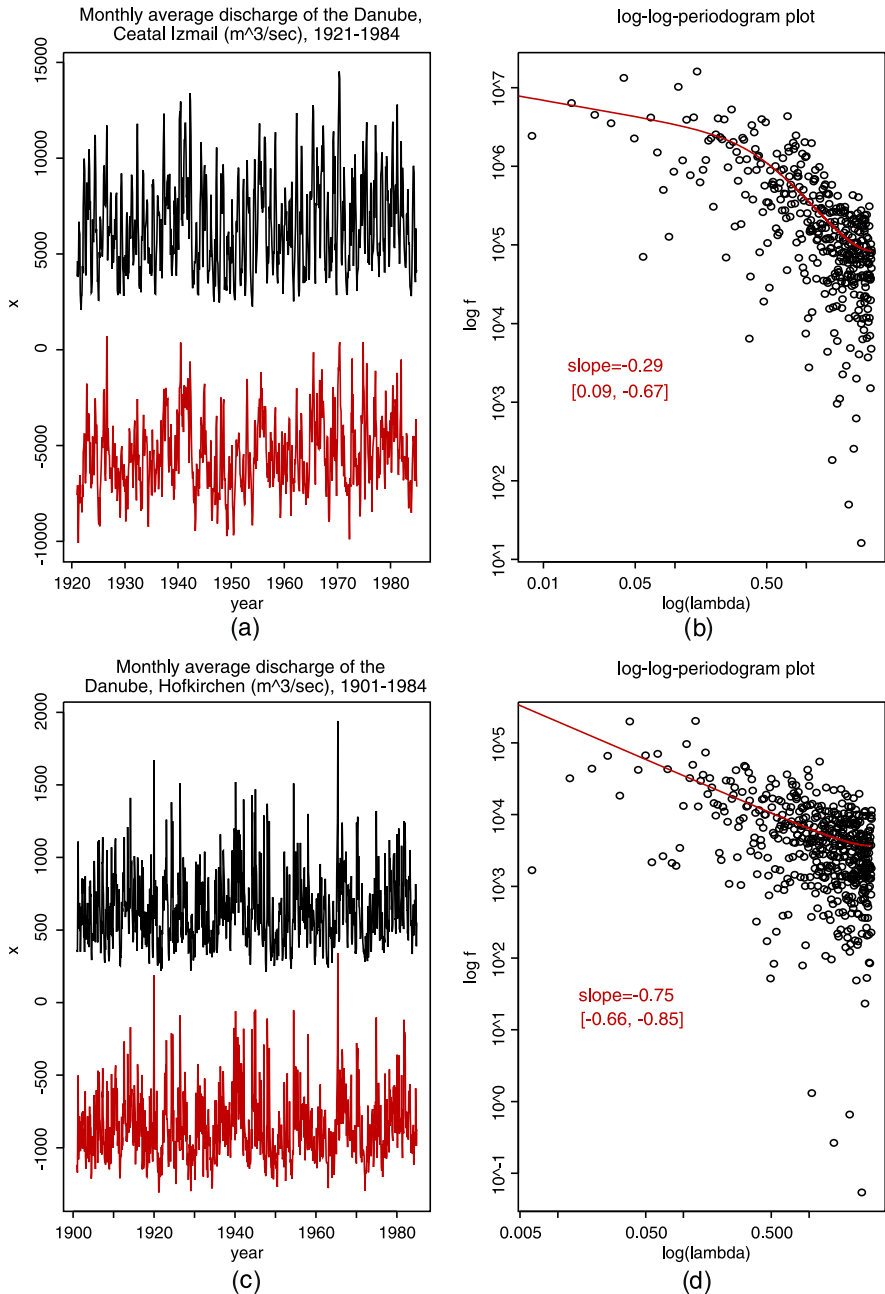


**Fig. 1.9** Monthly average discharge of the Danube at Bratislava (*upper series*: original; *lower series*: deseasonalized) and various log-log-plots for the deseasonalized series

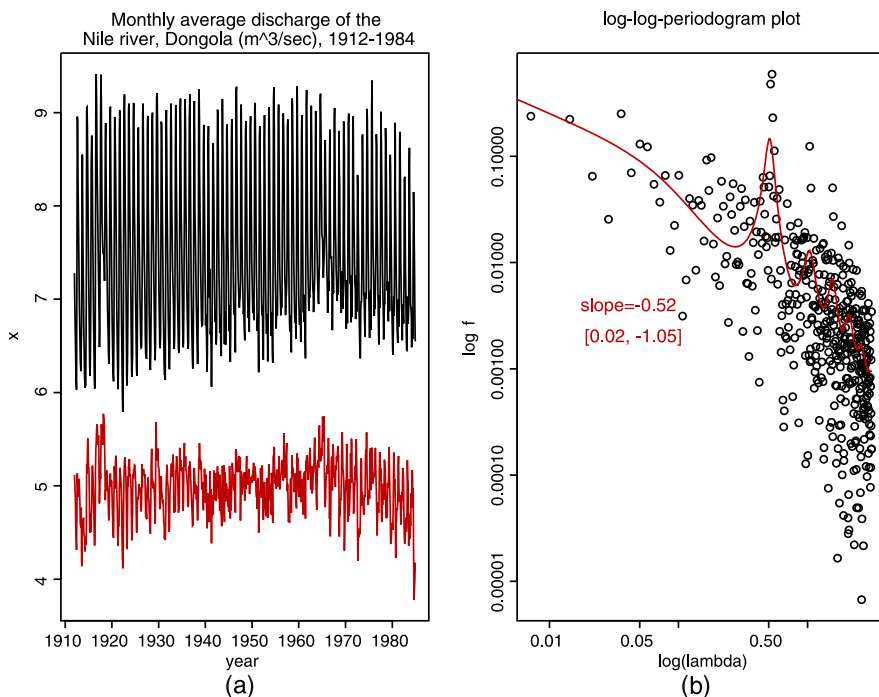
tral England for the years 1659 to 2010 (Manley 1953, 1974; Parker et al. 1992; Parker and Horton 2005). The data were downloaded using the Climate Explorer of the Royal Netherlands Meteorological Institute. The main question here is whether there is evidence for a systematic increase. The simplest way of answering this question is to fit a straight line and test whether the slope, say  $\beta_1$ , is positive. The dependence structure of the regression residuals has an influence on testing whether  $\beta_1$  is significantly larger than zero. As will be shown later, if the observations are given by  $y_t = \beta_0 + \beta_1 t + e_t$  with  $e_t$  being stationary with long-range dependence such that  $\rho(k) \sim c|k|^{2d-1}$  (as  $|k| \rightarrow \infty$ ) for some  $d \in (0, \frac{1}{2})$ , then the variance of the least squares estimator of  $\beta_1$  increases by a constant times the factor  $n^{2d}$  compared to the case of uncorrelated or weakly dependent residuals (see Sect. 7.1). This means that correct confidence intervals are wider by a factor proportional to  $n^d$ . The difference can be quite substantial. For example, the estimate of  $d$  for the Central England series is about 0.2. For the given data size, we thus have a factor of  $n^d = 704^{0.2} \approx 3.7$ . It is therefore much more difficult to obtain a significant result for  $\beta_1$  than under independence. Complicating the matter further, one may argue that the trend, if any, may not be linear so that testing for  $\beta_1$  leads to wrong conclusions. Furthermore, the observed series may even be nonstationary in the sense of random walk (or unit roots). As will be discussed in Chap. 7 (Sects. 7.4.5 and 7.4.6), there is a method (so-



**Fig. 1.10** (a) Monthly average discharge of the Danube at Nagymaros (*upper series*: original; *lower series*: deseasonalized); (b) log-log-periodogram of the deseasonalized series in (a); (c) monthly average discharge of the Danube at Drobeta-Turnu Severin (*upper series*: original; *lower series*: deseasonalized); (d) log-log-periodogram of the deseasonalized series in (c)



**Fig. 1.11** (a) Monthly average discharge of the Danube at Ceatal Izmail (*upper series*: original; *lower series*: deseasonalized); (b) log-log-periodogram of the deseasonalized series in (a); (c) monthly average discharge of the Danube at Hofkirchen (*upper series*: original; *lower series*: deseasonalized); (d) log-log-periodogram of the deseasonalized series in (c)

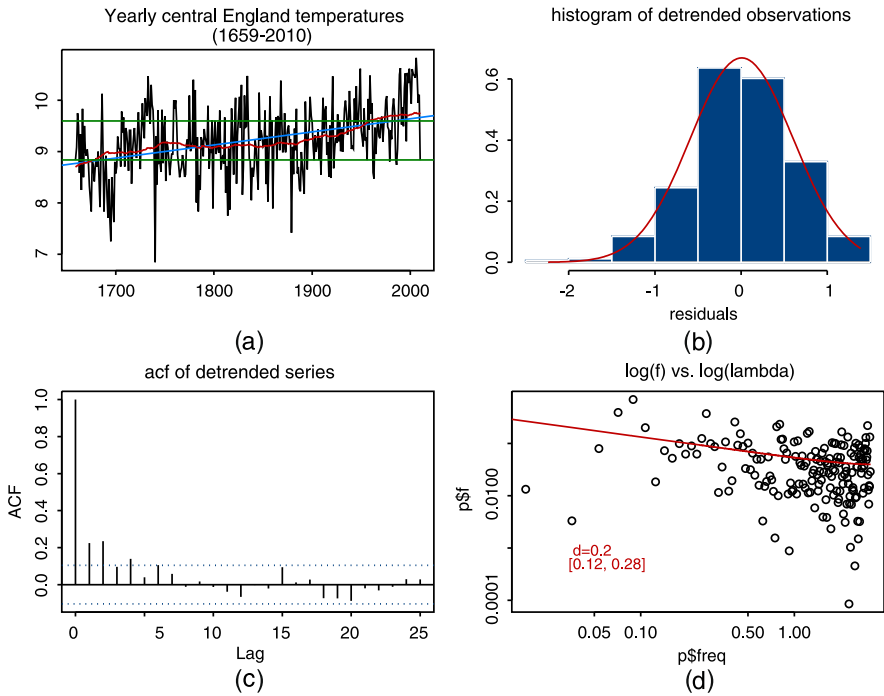


**Fig. 1.12** (a) Monthly average discharge of the Nile river at Dongola (*upper series*: original; *lower series*: deseasonalized); (b) log-log-periodogram of the deseasonalized series in (a)

called SEMIFAR models) that incorporates these possibilities using nonparametric trend estimation, integer differencing and estimation of the dependence parameters. Clearly, the more general a method is, the more difficult it becomes to obtain significant results. Nevertheless, the conclusion based on SEMIFAR models is that the trend is increasing and significantly different from a constant.

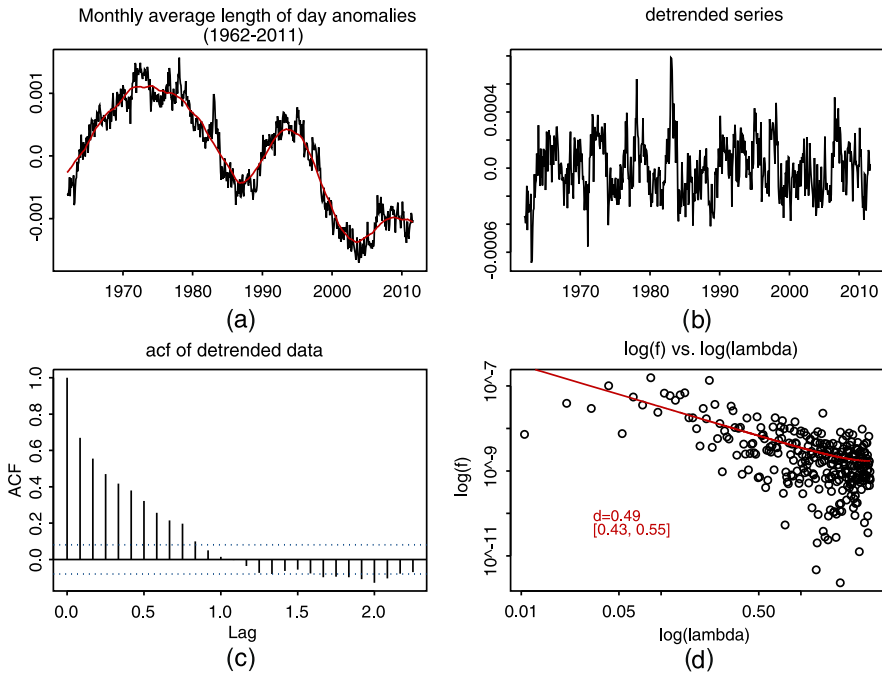
Another series with a clear trend function is displayed in Fig. 1.14. The measurements are monthly averaged length-of-day anomalies (Royal Netherlands Meteorological Institute). Overall, one can see that there is a slight decline together with a cyclic movement. The fitted line was obtained by kernel smoothing. As will be seen in Chap. 7, the crucial ingredient in kernel smoothing is the bandwidth. A good choice of the bandwidth depends on the dependence structure of the residuals. For the data here, the residuals have clear long memory. In fact, the estimated long-memory parameter is very close to the boundary of nonstationarity so that the possibility of a spectral density proportional to  $\lambda^{-1}$  (as  $\lambda \rightarrow 0$ ) cannot be excluded. Processes with this property are also called  $1/f$ -noise (which, in our notation, should actually be called  $1/\lambda$ -noise because  $f$  stands for frequency).

In the previous examples, the trend function is obviously smooth. Quite different time series are displayed in Figs. 1.15(a) and (d). The data were downloaded from the Physionet databank funded by the National Institute of Health (Goldberger et al.



**Fig. 1.13** (a) Yearly mean Central England temperatures together with a fitted least squares line and a nonparametric trend estimate; (b) histogram of residuals after subtraction of the nonparametric trend function; (c) acf of residuals; (d) log-log-periodogram of residuals

2000). The upper series in Fig. 1.15(a) shows consecutive stride intervals (stride-to-stride measures of footfall contact times) of a healthy individual, whereas the upper series in Fig. 1.15(d) was obtained for a patient suffering from Parkinson's disease. The complete data set consists of patients with Parkinson's disease ( $N = 15$ ), Huntington's disease ( $N = 20$ ) and amyotrophic lateral sclerosis ( $N = 13$ ), as well as a control group ( $N = 16$ ) (Hausdorff et al. 1997, 2000). Both series in Figs. 1.15(a) and (d) contain a spiky, somewhat periodic but also irregular, component. A natural approach to analysing such data is to decompose them into a 'spiky' component and the rest. Here, kernel smoothing is not appropriate because it tends to blur sharp peaks. Instead, wavelet thresholding (see e.g. Donoho and Johnstone 1995) separates local significant spikes from noise more effectively. The series plotted below the original ones are the trend functions fitted by standard minimax thresholding using Haar wavelets, the series at the bottom and, enlarged, in Figs. 1.15(b) and (e) are the corresponding residuals. The log-log-periodogram plots for the residual series and fitted fractional ARIMA spectral densities in Figs. 1.15(c) and (f) indicate long memory. A comparison of Figs. 1.15(c) and (f) shows that the slope  $\beta_f$  is less steep for the Parkinson patient. Indeed, using different techniques, Hausdorff et al. (1997, 2000) found evidence for  $\beta_f$  being closer to zero for patients suffering from Parkinson's disease (and other conditions such as Huntington's disease or



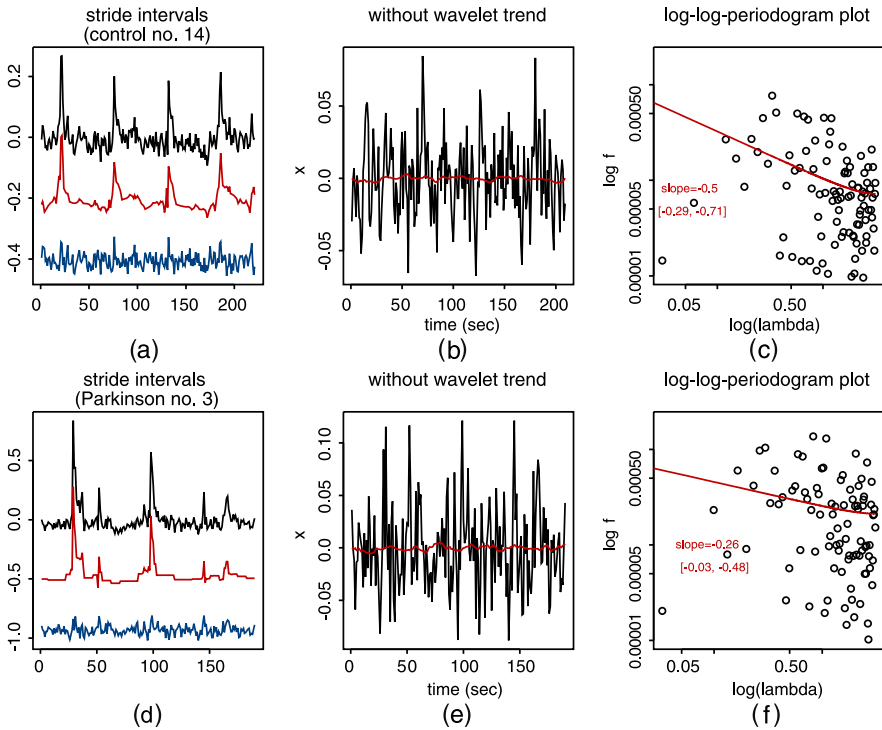
**Fig. 1.14** (a) Monthly averaged length-of-day anomalies (in seconds); (b) residuals after subtraction of the nonparametric trend function; (c) acf of residuals; (d) log-log-periodogram of residuals

Amyotrophic Lateral Sclerosis). Applying the approach described here to all available data confirms these findings. Boxplots of estimated values of  $\beta_f$  (Fig. 1.16) show a tendency for  $\beta_f$  to be closer to zero for the Parkinson patients. It should be noted, however, that the results may depend on the way tuning constants in wavelets thresholding were chosen. In view of the presence of long memory in the residuals, a detailed study of wavelet-based trend estimation under long-range dependence is needed. This will be discussed in more detail in Chap. 7 (Sect. 7.5).

A different kind of nonstationarity is typical for financial time series. Figure 1.17(a) shows daily values of the DAX index between 3 January 2000 and 12 September 2011. The series is nonstationary, but the first difference looks stationary (Fig. 1.17(b)), and the increments are uncorrelated (Fig. 1.17(c)). In this sense, the data resemble a random walk. However, there is an essential difference. Consider, as a measure of instantaneous volatility, the transformed series  $Y_t = |\log X_t - \log X_{t-1}|^{\frac{1}{4}}$  (see Ding and Granger 1996; Beran and Ocker 1999). Figure 1.17(d) shows that there is a trend in the volatility series  $Y_t$ . Moreover, even after removing the trend, the series exhibits very slowly decaying correlations and a clearly negative slope in the log-log-periodogram plot (Figs. 1.17(e) and (f)). This is very much in contrast to usual random walk.

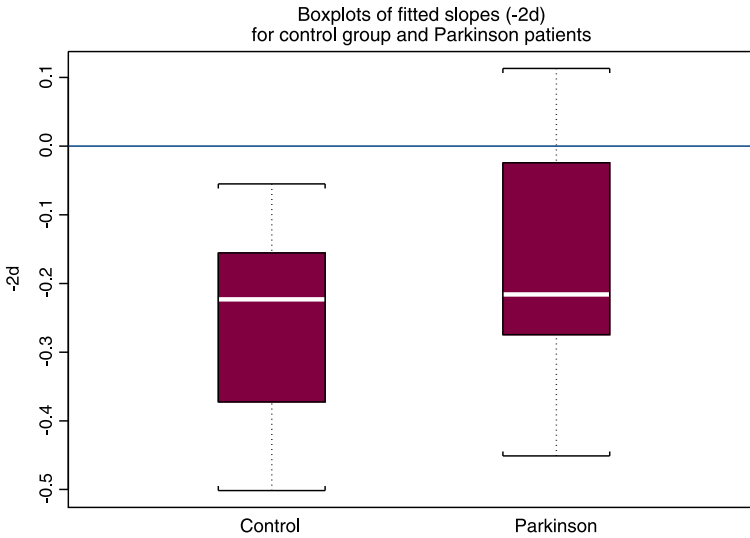
A completely different application where a trend and long memory are present is displayed in Figs. 1.18(a) through (d). These data were provided to us by Giovanni





**Fig. 1.15** Consecutive stride intervals for (a) a healthy individual and (d) a patient with Parkinson's disease. The original data are plotted *on top*, the trend functions fitted by minimax wavelet thresholding are given in the *middle*, and the series at the *bottom* correspond to the residuals. The residuals are also plotted separately in (b) and (e), the corresponding log-log-periodograms in Figs. (c) and (f) respectively

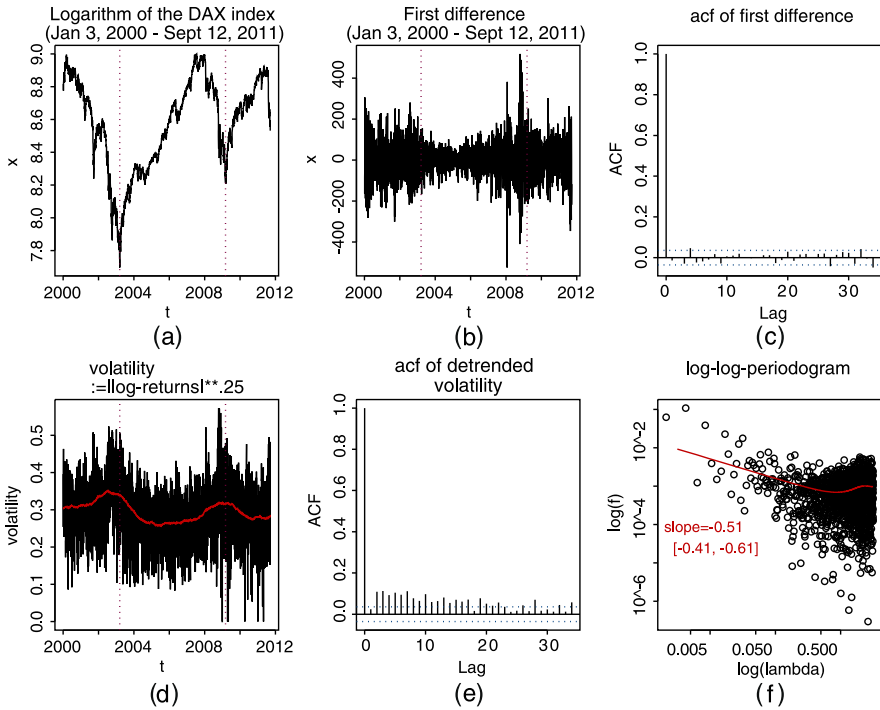
et al. (Department of Biology, University of Konstanz) and are part of a long-term project on olfactory coding in insects (see, Joerges et al. 1997; Galán et al. 2006; Galizia and Menzel 2001). The original observations consisted of optical measurements of calcium concentration in the antennal lobe of a honey bee. It is known that stimuli (odors) lead to characteristic activity patterns across spherical functional units, the so-called glomeruli, which collect the converging axonal input from a uniform family of receptor cells. It is therefore expected that, compared to a steady state, the between-glomeruli-variability of calcium concentration is higher during a response to an odor. This is illustrated in Fig. 1.18(a). For each time point  $t$  (with time rescaled to the interval  $[0, 1]$ ), an empirical entropy measure  $X_t$  was calculated based on the observed distribution of calcium concentration across the glomeruli. The odor was administered at the 30th of  $n = 100$  time points. The same procedure was carried out under two different conditions, namely without and with adding a neurotransmitter. The research hypothesis is that adding the neurotransmitter enhances the reaction, in the sense that the initial relative increase of the entropy curve is faster. Because of the known intervention point  $t_0$  and the specific shape of a typ-



**Fig. 1.16** Boxplots of slopes in the log-log-periodogram plot for the control group (*left*) and for a group of patients suffering from Parkinson’s disease (*right*)

ical response curve, a good fit can be obtained by a linear spline function with one fixed knot  $\eta_0$  at  $t_0$  and two subsequent free knots  $\eta_1, \eta_2 > t_0$ . The quantity to compare (between the measurements “without” and “with” neurotransmitter) is the slope  $\beta$  belonging to the truncated variable  $(t - \eta_0)_+$ . The distribution of the least squares estimate of  $\beta$  depends on the dependence structure of the residual process. For the bee considered in Fig. 1.18, the residuals exhibit clear long memory in the first case (no neurotransmitter), whereas long memory is not significant in the second case. For the collection of bees considered in this experiment, long memory, short memory and antipersistence could be observed. How to calculate confidence intervals for  $\beta$  and other parameters in this model will be discussed in Chap. 7 (Sect. 7.3).

An example of spatial long memory is shown in Fig. 1.19. The data in (a) correspond to differences between the maximal and minimal total column ozone amounts within the period from 1 to 7 January 2006, measured on a grid with a resolution of 0.25 degrees in latitude and longitude. The measurements were obtained by the Ozone Monitoring Instrument (OMI) on the Aura 28 spacecraft (Collection 3 OMI data; for details on the physical theory used in assessing ozone amounts, see e.g. Vasilkov et al. 2008; Ahmad et al. 2004; data source: NASA’s Ozone Processing Team, <http://toms.gsfc.nasa.gov>). Figures 1.19(c) and (d) display values of the periodograms in log-log-coordinates when looking in the horizontal (East–West) and vertical direction (North–South) of the grid respectively. Both plots indicate long-range dependence. The solid lines were obtained by fitting a fractional ARIMA lattice process (see Chap. 9, Sects. 9.2 and 9.3). This is a simple model that allows for different long-range, short-range and antipersistent dependence structures in the horizontal and vertical direction. A formal test confirms indeed that long-



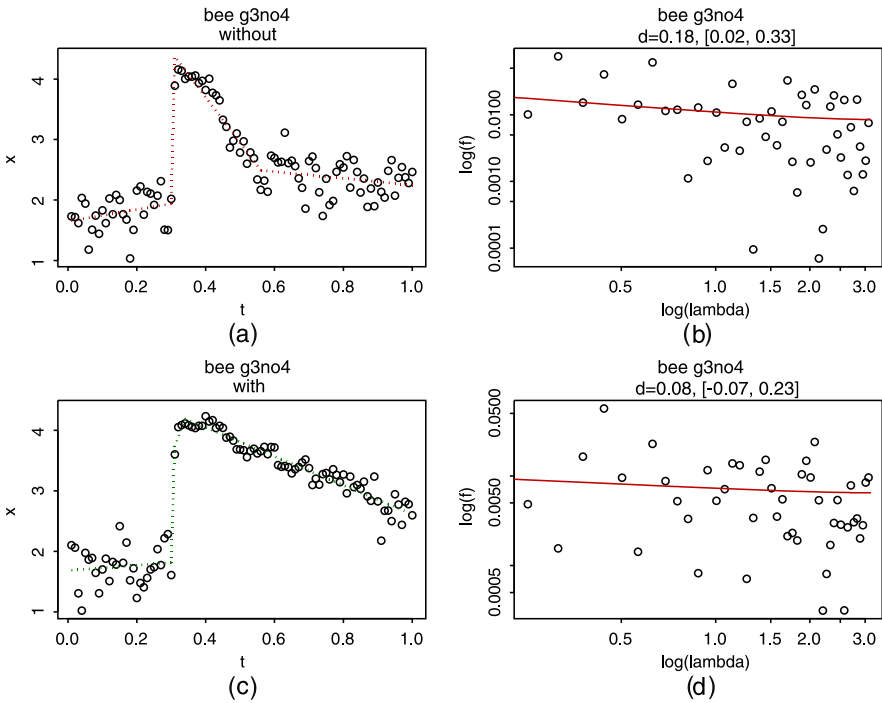
**Fig. 1.17** Daily values of the DAX index between 3 January 2000 and 12 September 2011: (a) logarithm of original series; (b) differenced series (log-returns); (c) acf of the series in (b); (d)  $Y_t = |\log X_t - \log X_{t-1}|^{\frac{1}{4}}$  together with a fitted nonparametric trend function; (e) acf of  $Y_t$  after detrending; (f) log-log-periodogram of  $Y_t$  after detrending

range dependence in the North–South direction is stronger than along East–West transects.

### 1.3 Definition of Different Types of Memory

#### 1.3.1 Second-Order Definitions for Stationary Processes

Consider a second-order stationary process  $X_t$  ( $t \in \mathbb{Z}$ ) with autocovariance function  $\gamma_X(k)$  ( $k \in \mathbb{Z}$ ) and spectral density  $f_X(\lambda) = (2\pi)^{-1} \sum_{k=-\infty}^{\infty} \gamma_X(k) \exp(-ik\lambda)$  ( $\lambda \in [-\pi, \pi]$ ). A heuristic definition of *linear* long-range dependence, short-range dependence and antipersistence is given as follows:  $X_t$  has (a) long memory, (b) short memory or (c) antipersistence if, as  $|\lambda| \rightarrow 0$ ,  $f_X(\lambda)$  (a) diverges to infinity, (b) converges to a finite constant, or (c) converges to zero respectively. Since  $2\pi f_X(\lambda) = \sum \gamma_X(k)$ , this is essentially (in a sense specified more precisely below) equivalent to (a)  $\sum \gamma_X(k) = \infty$ , (b)  $0 < \sum \gamma_X(k) < \infty$  and (c)  $\sum \gamma_X(k) = 0$ .



**Fig. 1.18** Empirical entropy of calcium concentrations in the antennal lobe of a honey bee exposed to hexanol: (a) original series without neurotransmitter and linear splines fit; (b) log-log-periodogram of residuals; (c) original series with neurotransmitter and linear splines fit; (d) log-log-periodogram of residuals

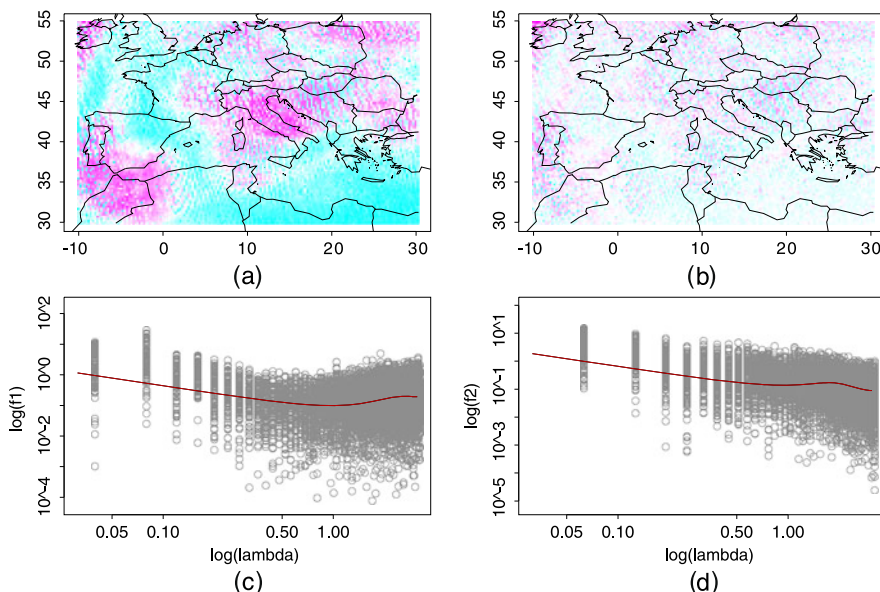
In the following more formal definitions will be given. First, the notion of slowly varying functions is needed (Karamata 1930a, 1930b, 1933; Bajšanski and Karamata 1968/1969; Zygmund 1968; also see e.g. Seneta 1976; Bingham et al. 1989; Sedletskii 2000). Here and throughout the book, the notation  $a_n \sim b_n$  ( $n \rightarrow \infty$ ) for two real- or complex-valued sequences  $a_n, b_n$  will mean that the ratio  $a_n/b_n$  converges to one. Similarly for functions,  $g(x) \sim h(x)$  ( $x \rightarrow x_0$ ) will mean that  $g(x)/h(x)$  converges to one as  $x$  tends to  $x_0$ .

First, we need to define so-called slowly varying functions. There are two slightly different standard definitions by Karamata and Zygmund respectively.

**Definition 1.1** A function  $L : (c, \infty) \rightarrow \mathbb{R}$  ( $c \geq 0$ ) is called slowly varying at infinity in Karamata’s sense if it is positive (and measurable) for  $x$  large enough and, for any  $u > 0$ ,

$$L(ux) \sim L(x) \quad (x \rightarrow \infty).$$

The function is called slowly varying at infinity in Zygmund’s sense if for  $x$  large enough, it is positive and for any  $\delta > 0$ , there exists a finite number  $x_0(\delta) > 0$  such



**Fig. 1.19** Daily total column ozone amounts from the Ozone Monitoring Instrument (OMI) on the Aura 28 spacecraft: **(a)** maximum minus minimum of observed ozone levels measured between 1–7 January 2006, plotted on a grid with a resolution of 0.25 degrees in latitude and longitude; **(b)** residuals after fitting a FARIMA lattice model; **(c)** and **(d)** log-log-periodogram of the data in **(a)** in the horizontal and vertical directions respectively

that for  $x > x_0(\delta)$ , both functions  $p_1(x) = x^\delta L(x)$  and  $p_2(x) = x^{-\delta} L(x)$  are monotone.

Similarly,  $L$  is called slowly varying at the origin if  $\tilde{L}(x) = L(x^{-1})$  is slowly varying at infinity.

A standard formal definition of different types of *linear* dependence structures is given as follows.

**Definition 1.2** Let  $X_t$  be a second-order stationary process with autocovariance function  $\gamma_X(k)$  ( $k \in \mathbb{Z}$ ) and spectral density

$$f_X(\lambda) = (2\pi)^{-1} \sum_{k=-\infty}^{\infty} \gamma_X(k) \exp(-ik\lambda) \quad (\lambda \in [-\pi, \pi]).$$

Then  $X_t$  is said to exhibit (linear) (a) long-range dependence, (b) intermediate dependence, (c) short-range dependence, or (d) antipersistence if

$$f_X(\lambda) = L_f(\lambda)|\lambda|^{-2d},$$

where  $L_f(\lambda) \geq 0$  is a symmetric function that is slowly varying at zero, and (a)  $d \in (0, \frac{1}{2})$ , (b)  $d = 0$  and  $\lim_{\lambda \rightarrow 0} L_f(\lambda) = \infty$ , (c)  $d = 0$  and  $\lim_{\lambda \rightarrow 0} L_f(\lambda) = c_f \in (0, \infty)$ , and (d)  $d \in (-\frac{1}{2}, 0)$  respectively.

Note that the terminology “short-range dependence” (with  $d = 0$ ) is reserved for the case where  $L_f(\lambda)$  converges to a finite constant  $c_f$ . The reason is that if  $L_f(\lambda)$  diverges to infinity, then the autocovariances are not summable although  $d = 0$ . This case resembles long-range dependence, though with a slower rate of divergence. For a discussion of models with “intermediate” dependence, see for instance Granger and Ding (1996). In principle, any of the usual notions of “slowly varying” may be used in the definition of  $L_f$ . The most common ones are the definitions by Karamata and Zygmund given above. The two theorems below show that Karamata’s definition is more general. First, we need the definition of regularly varying functions and two auxiliary results.

**Definition 1.3** A measurable function  $g : \mathbb{R}_+ \rightarrow \mathbb{R}$  is called regularly varying (at infinity) with exponent  $\alpha$  if  $g(x) \neq 0$  for large  $x$  and, for any  $u > 0$ ,

$$\lim_{x \rightarrow \infty} \frac{g(ux)}{g(x)} = u^\alpha.$$

The class of such functions is denoted by  $\text{Re}(\alpha)$ .

Similarly, a function  $g$  is called regularly varying at the origin with exponent  $\alpha$  if  $\tilde{g}(x) = g(x^{-1}) \in \text{Re}(-\alpha)$ . We will denote this class by  $\text{Re}_0(\alpha)$ .

Slowly varying functions are regularly varying functions with  $\alpha = 0$ . For regularly varying functions, integration leads to the following asymptotic behaviour.

**Lemma 1.1** Let  $g \in \text{Re}(\alpha)$  with  $\alpha > -1$  and integrable on  $(0, a)$  for any  $a > 0$ . Then  $\int_0^x g(t) dt \in \text{Re}(\alpha + 1)$ , and

$$\int_0^x g(t) dt \sim \frac{xg(x)}{\alpha + 1} \quad (x \rightarrow \infty).$$

Note that this result is just a generalization of the integration of a power  $x^{-\alpha}$ , where we have the exact equality  $\int_0^x t^{-\alpha} dt = x^{1-\alpha}/(\alpha + 1)$ . Lemma 1.1 is not only useful for proving the theorem below, but also because asymptotic calculations of variances of sample means can usually be reduced to approximations of integrals by Riemann sums. An analogous result holds for  $\alpha < -1$ :

**Lemma 1.2** Let  $g \in \text{Re}(\alpha)$  with  $\alpha < -1$  and integrable on  $(a, b)$  for any  $0 < a \leq b < \infty$ . Then  $\int_x^\infty g(t) dt \in \text{Re}(\alpha + 1)$ , and

$$\int_x^\infty g(t) dt \sim -\frac{xg(x)}{\alpha + 1} \quad (x \rightarrow \infty).$$

Now it can be shown that slowly varying functions in Karamata's sense can be characterized as follows.

**Theorem 1.1** *L is slowly varying at infinity in Karamata's sense if and only if*

$$L(x) = c(x) \exp \left\{ \int_1^x \frac{\eta(t)}{t} dt \right\} \quad (x \geq 1),$$

where  $c(\cdot)$  and  $\eta(\cdot)$  are measurable functions such that

$$\lim_{x \rightarrow \infty} c(x) = c \in (0, \infty),$$

$$\lim_{x \rightarrow \infty} \eta(x) = 0$$

and  $\eta(\cdot)$  is locally integrable.

*Proof* First, we show that the representation above yields a slowly varying function. Let  $s > 0$ ,  $s \in [a, b]$ , and write

$$\psi_s(x) := \frac{L(sx)}{L(x)} = \frac{c(sx)}{c(x)} \exp \left( \int_x^{sx} \frac{\eta(t)}{t} dt \right).$$

Since  $c(x) \rightarrow c$  and  $\eta(t) \rightarrow 0$ , we have for sufficiently large  $x$ , and arbitrary  $\varepsilon > 0$ ,

$$(1 - \varepsilon) \exp(-\varepsilon \max(|\log a|, |\log b|)) \leq \psi_s(x) \leq (1 + \varepsilon) \exp(\varepsilon \max(|\log a|, |\log b|)).$$

Letting  $\varepsilon \rightarrow 0$ , we obtain the slowly varying property.

Assume now that  $L$  is slowly varying. Define

$$\tilde{\eta}(s) := \frac{sL(s)}{\int_0^s L(t) dt}.$$

Then with  $U(s) = \int_0^s L(t) dt$ ,

$$\int_1^x \frac{\tilde{\eta}(s)}{s} ds = \int_1^x \frac{L(s)}{U(s)} ds = \int u^{-1} du = \log(cU(x)),$$

where the last integration is over  $(c = \int_0^1 L(t) dt, U(x) = \int_0^x L(t) dt)$ . Thus,

$$U(x) = c \exp \left( \int_1^x \frac{\tilde{\eta}(t)}{t} dt \right),$$

and consequently, taking derivatives on both sides of the latter expression, we have

$$L(x) = c \frac{\tilde{\eta}(x)}{x} \exp \left( \int_1^x \frac{\tilde{\eta}(t)}{t} dt \right) = c \tilde{\eta}(x) \exp \left( \int_1^x \frac{\tilde{\eta}(t) - 1}{t} dt \right).$$

Thus,  $L$  has the required representation. It remains to show that  $\eta(x) = \tilde{\eta}(x) - 1 \rightarrow 0$  and  $\tilde{\eta}(x) \rightarrow 1$ . This follows directly from Karamata's theorem (Lemma 1.1) and the definition of  $\tilde{\eta}(x)$ .  $\square$

On the other hand, for Zygmund's definition one can show the following:

**Theorem 1.2**  *$L$  is slowly varying in Zygmund's sense if and only if there is an  $x_0 \in [1, \infty)$  such that*

$$L(x) = c \exp \left\{ \int_1^x \frac{\eta(t)}{t} dt \right\} \quad (x \geq x_0),$$

where  $c$  is a finite positive constant, and  $\eta(\cdot)$  is a measurable function such that  $\lim_{x \rightarrow \infty} \eta(x) = 0$ .

In terms of regularly varying functions the definition of long-range dependence and antipersistence can be rephrased as follows: long memory and antipersistence means that  $f \in \text{Re}_0(-2d)$  with  $d \in (0, \frac{1}{2})$  and  $d \in (-\frac{1}{2}, 0)$  respectively. Since slowly varying functions are dominated by power functions,  $f(\lambda) = L_f(\lambda)|\lambda|^{-2d}$  implies that for  $d > 0$ , the spectral density has a hyperbolic pole at the origin, whereas it converges to zero for  $d < 0$ . In contrast, under short-range dependence,  $f(\lambda)$  converges to a positive finite constant. Alternative terms for long-range dependence are persistence, long memory or strong dependence. Instead of “(linear) long-range dependence”, one also uses the terminology “slowly decaying correlations”, “long-range correlations” or “strong correlations”. This is justified by the following equivalence between the behaviour of the spectral density at the origin and the asymptotic decay of the autocovariance function (see e.g. Zygmund 1968; Lighthill 1962; Beran 1994a; Samorodnitsky 2006):

**Theorem 1.3** *Let  $\gamma(k)$  ( $k \in \mathbb{Z}$ ) and  $f(\lambda)$  ( $\lambda \in [-\pi, \pi]$ ) be the autocovariance function and spectral density respectively of a second-order stationary process. Then the following holds:*

(i) *If*

$$\gamma(k) = L_\gamma(k)|k|^{2d-1},$$

where  $L_\gamma(k)$  is slowly varying at infinity in Zygmund's sense, and either  $d \in (0, \frac{1}{2})$ , or  $d \in (-\frac{1}{2}, 0)$  and  $\sum_{k \in \mathbb{Z}} \gamma(k) = 0$ , then

$$f(\lambda) \sim L_f(\lambda)|\lambda|^{-2d} \quad (\lambda \rightarrow 0)$$

with

$$L_f(\lambda) = L_\gamma(\lambda^{-1})\pi^{-1}\Gamma(2d)\sin\left(\frac{\pi}{2} - \pi d\right). \quad (1.1)$$



(ii) If

$$f(\lambda) = L_f(\lambda)|\lambda|^{-2d} \quad (0 < \lambda < \pi),$$

where  $d \in (-\frac{1}{2}, 0) \cup (0, \frac{1}{2})$ , and  $L_f(\lambda)$  is slowly varying at the origin in Zygmund's sense and of bounded variation on  $(a, \pi)$  for any  $a > 0$ , then

$$\gamma(k) \sim L_\gamma(k)|k|^{2d-1} \quad (k \rightarrow \infty),$$

where

$$L_\gamma(k) = 2L_f(k^{-1})\Gamma(1-2d)\sin\pi d. \quad (1.2)$$

Note that in the case of antipersistence the autocovariances are absolutely summable but  $|\gamma(k)|$  still converges at a hyperbolic rate that can be rather slow, compared for instance with an exponential decay. Also note that  $d = 0$  is not included in the theorem because the condition  $\gamma(k) = L_\gamma(k)|k|^{-1}$  would imply that  $\gamma(k)$  is not summable. In principle (possibly under additional regularity conditions), this would correspond to intermediate dependence with  $f(\lambda)$  diverging at the origin like a slowly varying function (see Definition 1.2). To obtain short-range dependence in the sense of Definition 1.2, the summability of  $\gamma(k)$  is a minimal requirement. For instance, an exponential decay defined by  $|\gamma(k)| \leq ca^k$  (with  $0 < c < \infty, 0 < a < 1$ ) together with  $\sum_{k \in \mathbb{Z}} \gamma(k) = c_f > 0$  implies  $f(\lambda) \sim c_f$  as  $\lambda \rightarrow 0$ . A general statement including all four types of dependence structures can be made however with respect to the sum of the autocovariances:

**Corollary 1.1** *If*

$$f(\lambda) = L_f(\lambda)|\lambda|^{-2d} \quad (0 < \lambda < \pi),$$

where  $d \in (-\frac{1}{2}, \frac{1}{2})$ , and  $L_f(\lambda) = L(\lambda^{-1})$  is slowly varying at the origin in Zygmund's sense and of bounded variation on  $(a, \pi)$  for any  $a > 0$ , then the following holds. For  $-\frac{1}{2} < d < 0$ ,

$$\sum_{k=-\infty}^{\infty} \gamma(k) = 2\pi f(0) = 0,$$

whereas for  $0 < d < \frac{1}{2}$ ,

$$\sum_{k=-\infty}^{\infty} \gamma(k) = 2\pi \lim_{\lambda \rightarrow 0} f(\lambda) = \infty.$$

Moreover, for  $d = 0$ , we have

$$0 < \sum_{k=-\infty}^{\infty} \gamma(k) = 2\pi f(0) = 2\pi c_f < \infty$$

if  $0 < \lim_{\lambda \rightarrow 0} L_f(\lambda) = c_f < \infty$  and

$$\sum_{k=-\infty}^{\infty} \gamma(k) = 2\pi \lim_{\lambda \rightarrow 0} f(\lambda) = \infty$$

if  $\lim_{\lambda \rightarrow 0} L_f(\lambda) = \infty$ .

From these results one can see that characterizing linear dependence by the spectral density is more elegant than via the autocovariance function because the equation  $f(\lambda) = L_f(\lambda)|\lambda|^{-2d}$  is applicable in all four cases (long-range, short-range, intermediate dependence and antipersistence).

*Example 1.1* Let  $X_t$  be second-order stationary with Wold decomposition

$$X_t = \sum_{j=0}^{\infty} a_j \varepsilon_{t-j},$$

where  $\varepsilon_t$  are uncorrelated zero mean random variables,  $\sigma_\varepsilon^2 = \text{var}(\varepsilon_t) < \infty$ , and

$$a_j = (-1)^j \binom{-d}{j} = (-1)^j \frac{\Gamma(1-d)}{\Gamma(j+1)\Gamma(1-d-j)}$$

with  $-1/2 < d < 1/2$ . Then  $a_j$  are the coefficients in the power series representation

$$A(z) = (1-z)^{-d} = \sum_{j=0}^{\infty} a_j z^j.$$

Therefore, the spectral density of  $X_t$  is given by

$$\begin{aligned} f_X(\lambda) &= \frac{\sigma_\varepsilon^2}{2\pi} |A(e^{-i\lambda})|^2 = \frac{\sigma_\varepsilon^2}{2\pi} |1 - e^{-i\lambda}|^{-2d} = \frac{\sigma_\varepsilon^2}{2\pi} |2(1 - \cos \lambda)|^{-d} \\ &\sim \frac{\sigma_\varepsilon^2}{2\pi} |\lambda|^{-2d} \quad (\lambda \rightarrow 0). \end{aligned}$$

Thus, we obtain short-range dependence for  $d = 0$  (and in fact uncorrelated observations), antipersistence for  $-\frac{1}{2} < d < 0$  and long-range dependence for  $0 < d < \frac{1}{2}$ . If the innovations  $\varepsilon_t$  are independent, then  $X_t$  is called a fractional ARIMA(0,  $d$ , 0) process (Granger and Joyeux 1980; Hosking 1981; see Chap. 2, Sect. 2.1.1.4).

*Example 1.2* Let  $X_t$  be second-order stationary with spectral density

$$f_X(\lambda) = \log \left| \frac{\pi}{\lambda} \right| = L_f(\lambda).$$

This is a case with intermediate dependence. The autocovariance function is given by

$$\text{var}(X_t) = \gamma_X(0) = 2 \left( \pi \log \pi - \int_0^\pi \log \lambda \, d\lambda \right) = 2\pi,$$

and for  $k > 0$ ,

$$\begin{aligned} \gamma_X(k) &= 2 \int_0^\pi \cos k\lambda \cdot (\log \pi - \log \lambda) \, d\lambda = -2 \int_0^\pi \cos k\lambda \cdot \log \lambda \, d\lambda \\ &= \frac{2}{k} \int_0^\pi \frac{\sin k\lambda}{\lambda} \, d\lambda = \frac{2}{k} Si(\pi k), \end{aligned}$$

where  $Si(\cdot)$  is the sine integral function. For  $k \rightarrow \infty$ , we obtain the Dirichlet integral

$$\lim_{k \rightarrow \infty} Si(\pi k) = \int_0^\infty \frac{\sin \lambda}{\lambda} \, d\lambda = \frac{\pi}{2},$$

so that

$$\gamma_X(k) \sim \pi k^{-1} \quad (k \rightarrow \infty),$$

$$\rho_X(k) \sim \frac{1}{2} k^{-1} \quad (k \rightarrow \infty),$$

and

$$\sum_{k=-(n-1)}^{n-1} \gamma_X(k) \sim 2\pi \log n \quad (n \rightarrow \infty).$$

The behaviour of the spectral density at the origin also leads to a simple universal formula for the variance of the sample mean  $\bar{x} = n^{-1} \sum_{i=1}^n X_i$ :

**Corollary 1.2** *Suppose that  $f(\lambda) \sim L_f(\lambda)|\lambda|^{-2d}$  ( $\lambda \rightarrow 0$ ) for some  $d \in (-\frac{1}{2}, \frac{1}{2})$ , where  $L_f(\lambda) = L(\lambda^{-1})$  is slowly varying at zero in Zygmund's sense and of bounded variation on  $(a, \pi)$  for any  $a > 0$ . Furthermore, assume that in the case of  $d = 0$  the slowly varying function  $L_f$  is continuous at the origin. Then*

$$\text{var}(\bar{x}) \sim v(d) f(n^{-1}) n^{-1} \quad (n \rightarrow \infty)$$

with

$$v(d) = \frac{2\Gamma(1-2d) \sin(\pi d)}{d(2d+1)} \quad (d \neq 0)$$

and

$$v(0) = \lim_{d \rightarrow 0} v(d) = 2\pi.$$

*Proof* We have

$$\begin{aligned}\text{var}(\bar{x}) &= n^{-1} \sum_{k=-(n-1)}^{n-1} \left(1 - \frac{|k|}{n}\right) \gamma(k) \\ &= n^{-1} \sum_{k=-(n-1)}^{n-1} \gamma(k) - n^{-1} \sum_{k=-(n-1)}^{n-1} \frac{|k|}{n} \gamma(k)\end{aligned}$$

with

$$\gamma(k) \sim L_\gamma(k) |k|^{2d-1}.$$

For  $0 < d < \frac{1}{2}$ , this implies

$$\begin{aligned}\text{var}(\bar{x}) &\sim 2L_\gamma(n) n^{-1} \left[ \sum_{k=1}^{n-1} k^{2d-1} - n^{-1} \sum_{k=1}^{n-1} k^{2d} \right] \\ &= 2L_\gamma(n) n^{2d-1} \left[ \sum_{k=1}^{n-1} \left(\frac{k}{n}\right)^{2d-1} n^{-1} - \sum_{k=1}^{n-1} \left(\frac{k}{n}\right)^{2d} n^{-1} \right] \\ &\sim 2L_\gamma(n) n^{2d-1} \left[ \int_0^1 x^{2d-1} dx - \int_0^1 x^{2d} dx \right] \\ &= 2L_\gamma(n) n^{2d-1} \left[ \frac{1}{2d} - \frac{1}{2d+1} \right] = \frac{L_\gamma(n) n^{2d-1}}{d(2d+1)}.\end{aligned}$$

Using Theorem 1.3, we can write this as

$$\frac{L_\gamma(n) n^{2d-1}}{d(2d+1)} = \frac{2\Gamma(1-2d) \sin(\pi d)}{d(2d+1)} L_f(n^{-1}) n^{2d-1} = \nu(d) L_f(n^{-1}) n^{2d-1}.$$

Thus,

$$\text{var}(\bar{x}) \sim \nu(d) L_f(n^{-1}) n^{2d-1} \sim \nu(d) f(n^{-1}) n^{-1}.$$

For  $d = 0$  and  $0 < L_f(0) = c_f < \infty$ , we have

$$0 < \sum_{k=-\infty}^{\infty} \gamma(k) = 2\pi f(0) < \infty,$$

so that  $|k|\gamma(k)$  is Cesaro summable with limit zero. Hence,

$$\lim_{n \rightarrow \infty} n^{-1} \sum_{k=-(n-1)}^{n-1} \frac{|k|}{n} \gamma(k) = 0$$

and

$$\text{var}(\bar{x}) \sim n^{-1} \sum_{k=-(n-1)}^{n-1} \gamma(k) \sim 2\pi f(0)n^{-1}.$$

Thus, we may write

$$\text{var}(\bar{x}) \sim v(0)L_f(0)n^{-1} \sim v(0)f(n^{-1})n^{-1},$$

where

$$v(0) = \lim_{d \rightarrow 0} v(d) = \lim_{d \rightarrow 0} \frac{2 \sin(\pi d)}{d} = 2\pi.$$

Finally, for  $-\frac{1}{2} < d < 0$ , we have  $\sum_{k \in \mathbb{Z}} \gamma(k) = 0$ , so that

$$\begin{aligned} \text{var}(\bar{x}) &= n^{-1} \sum_{k=-(n-1)}^{n-1} \gamma(k) - n^{-1} \sum_{k=-(n-1)}^{n-1} \frac{|k|}{n} \gamma(k) \\ &= -2n^{-1} \sum_{k=n}^{\infty} \gamma(k) - n^{-1} \sum_{k=-(n-1)}^{n-1} \frac{|k|}{n} \gamma(k) \\ &\sim 2L_\gamma(n)n^{-1} \left[ -\sum_{k=n}^{\infty} k^{2d-1} - n^{-1} \sum_{k=1}^{n-1} k^{2d} \right] \\ &= 2L_\gamma(n)n^{2d-1} \left[ -\sum_{k=n}^{\infty} \left(\frac{k}{n}\right)^{2d-1} n^{-1} - \sum_{k=1}^{n-1} \left(\frac{k}{n}\right)^{2d} n^{-1} \right] \\ &\sim 2L_\gamma(n)n^{2d-1} \left[ -\int_1^{\infty} x^{2d-1} dx - \int_0^1 x^{2d} dx \right] \\ &= 2L_\gamma(n)n^{2d-1} \left[ \frac{1}{2d} - \frac{1}{2d+1} \right] = v(d)L_f(n^{-1})n^{2d-1} \\ &\sim v(d)f(n^{-1})n^{-1}. \quad \square \end{aligned}$$

Corollary 1.2 illustrates that knowledge about the value of  $d$  is essential for statistical inference. If short memory is assumed but the actual value of  $d$  is larger than zero, then confidence intervals for  $\mu = E(X_t)$  will be too narrow by an increasing factor of  $n^d$ , and the asymptotic level of tests based on this assumption will be zero. This effect is not negligible even for small sample sizes. Table 1.1 shows simulated rejection probabilities (based on 1000 simulations) for the  $t$ -test at the nominal 5 %-level of significance. The numbers are based on 1000 simulations of a fractional ARIMA(0,  $d$ , 0) process with  $d = 0.1, 0.2, 0.3$  and  $0.4$  respectively (see Chap. 2, Sect. 2.1.1.4, for the definition of FARIMA models).

The second-order definitions of long-range dependence considered here can be extended to random fields with a multivariate index  $t$ . A complication that needs

**Table 1.1** Simulated rejection probabilities (under the null hypothesis) for the  $t$ -test at the nominal 5 %-level of significance. The results are based on 1000 simulations of a fractional ARIMA(0,  $d$ , 0) process with  $d = 0.1, 0.2, 0.3$  and  $0.4$  respectively

| $n$ | $d = 0.1$ | 0.2  | 0.3  | 0.4  |
|-----|-----------|------|------|------|
| 10  | 0.10      | 0.21 | 0.33 | 0.53 |
| 50  | 0.16      | 0.38 | 0.55 | 0.72 |
| 100 | 0.20      | 0.42 | 0.62 | 0.78 |

to be addressed for two- or higher-dimensional indices is however that dependence may not be isotropic (see e.g. Boissy et al. 2005; Lavancier 2006, 2007; Beran et al. 2009). This will be discussed in Chap. 9. A further important extension includes multivariate spectra with power law behaviour at the origin that may differ for the different components of the process (see e.g. Robinson 2008).

### 1.3.2 Volatility Dependence

The characterization of nonlinear long memory is more complicated in general since there are many ways in which nonlinearity can occur. In econometric applications, the main focus is on dependence in volatility in the sense that  $X_t$  are uncorrelated but the squares  $X_t^2$  are correlated. The definitions of long memory given above can then be carried over directly by simply considering  $X_t^2$  instead of  $X_t$ . A more difficult, and partially still open, issue is how to define concrete statistically convenient models that are stationary with existing fourth moments and long-range correlations in  $X_t^2$  (see e.g. Robinson 1991; Bollerslev and Mikkelsen 1996; Baillie et al. 1996a; Ding and Granger 1996; Beran and Ocker 2001; Giraitis et al. 2000a, 2004, 2006; Giraitis and Surgailis 2002). This is discussed in detail in Sect. 2.1.3. A very simple model that is well defined and obviously exhibits long-range dependence can be formulated as follows.

**Proposition 1.1** Let  $\varepsilon_t$  ( $t \in \mathbb{Z}$ ) be i.i.d. random variables with  $E(\varepsilon_t) = 0$  and  $\text{var}(\varepsilon_t) = 1$ . Define

$$X_t = \sigma_t \varepsilon_t$$

with  $\sigma_t = \sqrt{v_t}$ ,  $v_t \geq 0$  independent of  $\varepsilon_s$  ( $s \in \mathbb{Z}$ ) and such that

$$\gamma_v(k) = \text{cov}(v_t, v_{t+k}) \sim c \cdot |k|^{2d-1}$$

for some  $0 < d < \frac{1}{2}$ . Then for  $k \neq 0$ ,

$$\gamma_X(k) = 0,$$

whereas

$$\gamma_{X^2}(k) = \text{cov}(X_t^2, X_{t+k}^2) = \gamma_v(k) \sim c \cdot |k|^{2d-1} \quad (k \rightarrow \infty).$$

*Proof* Since  $E(X_t) = E(\sigma_t)E(\varepsilon_t) = 0$ , we have for  $k \neq 0$ ,

$$\gamma_X(k) = E(X_t X_{t+k}) = E(\sigma_t \sigma_{t+k})E(\varepsilon_t \varepsilon_{t+k}) = 0.$$

Moreover, for  $k \neq 0$ ,

$$\begin{aligned} \gamma_{X^2}(k) &= E(\sigma_t^2 \sigma_{t+k}^2)E(\varepsilon_t^2 \varepsilon_{t+k}^2) - E(\sigma_t^2 \varepsilon_t^2)E(\sigma_{t+k}^2 \varepsilon_{t+k}^2) \\ &= E(v_t v_{t+k}) - E(v_t)E(v_{t+k}) \\ &= \gamma_v(k) \sim c \cdot |k|^{2d-1} \quad (k \rightarrow \infty). \end{aligned} \quad \square$$

The main problem with this model is that  $\sigma_t$  and  $\varepsilon_t$  are not directly observable. One would however like to be able to separate the components  $\sigma_t$  and  $\varepsilon_t$  even though only their product  $X_s$  ( $s \leq t$ ) is observed. This is convenient, for instance, when setting up maximum likelihood equations for estimating parameters that specify the model (see e.g. Giraitis and Robinson 2001). One therefore often prefers to assume a recursive relation between  $v_t$  and past values of  $X_t$ . The difficulty that arises then is to prove the existence of a stationary solution and to see what type of volatility dependence is actually achieved. For instance, in the so-called ARCH( $\infty$ ) model (Robinson 1991; Giraitis et al. 2000a) one assumes

$$\sigma_t^2 = v_t = b_0 + \sum_{j=1}^{\infty} b_j X_{t-j}^2$$

with  $b_j \geq 0$  and  $\sum b_j < \infty$ . As it turns out, however, long-range dependence—defined in the second-order sense as above—cannot be obtained. This and alternative volatility models with long-range dependence will be discussed in Sect. 2.1.3.

### 1.3.3 Second-Order Definitions for Nonstationary Processes

For nonstationary processes, Heyde and Yang (1997) consider the variance

$$V_m = \text{var}(X_t^{(m)}) \tag{1.3}$$

of the aggregated process

$$X_t^{(m)} = X_{tm-m+1} + \cdots + X_{tm} \tag{1.4}$$

and the limit

$$V = \lim_{m \rightarrow \infty} D_m^{-1} V_m, \tag{1.5}$$

where

$$D_m = \sum_{i=tm-m+1}^{tm} E(X_i^2). \tag{1.6}$$

The process  $X_t^{(m)}$  ( $t \in \mathbb{Z}$ ) is then said to exhibit long memory if  $V = \infty$ . This definition is applicable both to second-order stationary processes and to processes that need to be differenced first. Note that the block mean variance  $m^{-2}V_m$  is also called Allan variance (Allan 1966; Percival 1983; Percival and Guttorp 1994).

### 1.3.4 Continuous-Time Processes

The definition of long memory and antipersistence based on autocovariances can be directly extended to continuous-time processes.

**Definition 1.4** Let  $X(t)$  ( $t \in \mathbb{R}$ ) be a stationary process with autocovariance function  $\gamma_X(u) = \text{cov}(X(t), X(t+u))$  and spectral density  $f_X(\lambda)$  ( $\lambda \in \mathbb{R}$ ). Then  $X(t)$  is said to have long memory if there is a  $d \in (0, \frac{1}{2})$  such that

$$\gamma_X(u) = L_\gamma(u)u^{2d-1}$$

as  $u \rightarrow \infty$ , or

$$f_X(\lambda) = L_f(\lambda)|\lambda|^{-2d}$$

as  $\lambda \rightarrow 0$ , where  $L_\gamma$  and  $L_f$  are slowly varying at infinity and zero respectively. Similarly,  $X(t)$  is said to be antipersistent if these formulas hold for some  $d \in (-\frac{1}{2}, 0)$  and, in case of the formulation via  $\gamma_X$ , the additional condition

$$\int_{-\infty}^{\infty} \gamma_X(u) du = 0$$

holds.

Note that, as in discrete time, the definition of long-range dependence given here implies  $\int \gamma_X(u) du = \infty$ . A more general definition is possible by using the conditions  $\int \gamma_X(u) du = \infty$  and  $\int \gamma_X(u) du = 0$  only. However, the first condition would then also include the possibility of intermediate dependence.

Finally note that an alternative definition can also be given in terms of the variance of the integrated process  $Y(t) = \int_0^t X(s) ds$ . This is analogous to a nonlinear growth of the variance of partial sums for discrete time processes.

**Definition 1.5** Let  $Y(t) = \int_0^t X(s) ds$  and assume that  $\text{var}(Y(t)) < \infty$  for all  $t \geq 0$ . Then  $Y$  (and  $X$ ) is said to have long-range dependence if

$$\text{var}(Y(t)) = L(t)t^{2d+1}$$

for some  $0 < d < \frac{1}{2}$ , where  $L$  is slowly varying at infinity. Moreover,  $Y$  (and  $X$ ) is said to be antipersistent if

$$\text{var}(Y(t)) = L(t)t^{2H} = L(t)t^{2d+1}$$



for some  $-\frac{1}{2} < d < 0$ , where  $L$  is slowly varying at infinity.

This definition means that the growth of the variance of  $Y(t)$  is faster than linear under long-range dependence and slower than linear for antipersistent processes. The connection between the two definitions is given by

$$\text{var}(Y(t)) = \int_0^t \left( \int_0^t \gamma_X(s-r) dr \right) ds = 2 \int_0^t (t-u) \gamma_X(u) du.$$

If  $\gamma_X(u) = \text{cov}(X(t), X(t+u)) \sim L_\gamma(u)|u|^{2d-1}$ , where  $d \in (0, \frac{1}{2})$  and  $L_\gamma$  is slowly varying at infinity (i.e.  $X(t)$  has long memory in the sense of Definition 1.4), then application of Lemma 1.1 leads to

$$\text{var}(Y(t)) \sim \frac{1}{d(2d+1)} L_\gamma(t) t^{2d+1}.$$

Thus,  $X(t)$  has also long memory in the sense of Definition 1.5. The analogous connection holds for antipersistence, taking into account the additional condition  $\int \gamma_X(u) du = 0$ .

For nonnegative processes, the expected value often grows at a linear rate. Typical examples are counting processes or renewal processes with positive rewards (see Sects. 2.2.4 and 4.9). Long-range dependence and antipersistence can therefore also be expressed by comparing the growth of the variance with the growth of the mean.

**Definition 1.6** Let  $Y(t) = \int_0^t X(s) ds \geq 0$  and assume that  $\text{var}(Y(t)) < \infty$  for all  $t \geq 0$ . Then  $Y$  (and  $X$ ) is said to have long-range dependence if

$$\lim_{t \rightarrow \infty} \frac{\text{var}(Y(t))}{E[Y(t)]} = +\infty.$$

Similarly,  $Y$  (and  $X$ ) is said to be antipersistent if

$$\lim_{t \rightarrow \infty} \frac{\text{var}(Y(t))}{E[Y(t)]} = 0.$$

### 1.3.5 Self-similar Processes: Beyond Second-Order Definitions

Another classical way of studying long memory and antipersistence is based on the relationship between dependence and self-similarity.

**Definition 1.7** A stochastic process  $Y(u)$  ( $u \in \mathbb{R}$ ) is called *self-similar* with self-similarity parameter  $0 < H < 1$  (or  $H$ -self-similar) if for all  $c > 0$ , we have

$$(Y(cu), u \in \mathbb{R}) \stackrel{d}{=} (c^H Y(u), u \in \mathbb{R}),$$

where  $\stackrel{d}{=}$  denotes equality in distribution.

Self-similar processes are a very natural mathematical object to look at because they are the only possible weak limits of appropriately normalized and centered partial sums  $S_n(u) = \sum_{t=1}^{\lfloor nu \rfloor} X_t$  ( $u \in [0, 1]$ ) based on stationary and ergodic sequences  $X_t$  ( $t \in \mathbb{Z}$ ) (Lamperti 1962, 1972). If a process  $Y(u)$  ( $u \in \mathbb{R}$ ) is  $H$ -self-similar with stationary increments (so-called  $H$ -SSSI), then the discrete-time increment process  $X_t = Y(t) - Y(t-1)$  ( $t \in \mathbb{Z}$ ) is stationary. Note also that  $Y(0) =_d c^H Y(0)$  for any arbitrarily large  $c > 0$ , so that necessarily  $Y(0) = 0$  almost surely.

To see how the self-similarity parameter  $H$  is related to long memory, we first consider a case where the second-order definition of long memory is applicable. If second moments exist, then the SSSI-property implies, for  $u \geq v > 0$ ,

$$\gamma_Y(u, u) = \text{var}(Y(u)) = u^{2H} \gamma_Y(1, 1) = u^{2H} \sigma^2$$

and

$$\text{var}(Y(u) - Y(v)) = \text{var}(Y(u - v)) = \sigma^2 (u - v)^{2H}.$$

Since  $\text{var}(Y(u) - Y(v)) = \gamma_Y(u, u) + \gamma_Y(v, v) - 2\gamma_Y(u, v)$ , this means that the autocovariance function is equal to

$$\gamma_Y(u, v) = \frac{\sigma^2}{2} [ |u|^{2H} + |v|^{2H} - |u - v|^{2H} ] \quad (u, v \in \mathbb{R}).$$

By similar arguments, the autocovariance function of the increment process  $X_t$  ( $t \in \mathbb{Z}$ ) is given by

$$\gamma_X(k) = \text{cov}(X_t, X_{t+k}) = \frac{\sigma^2}{2} [ |k-1|^{2H} + |k+1|^{2H} - 2|k|^{2H} ] \quad (k \in \mathbb{N}). \quad (1.7)$$

By Taylor expansion in  $x = k^{-1}$  around  $x = 0$  it follows that, as  $k$  tends to infinity,

$$\gamma_X(k) \sim \sigma^2 H(2H - 1) k^{2H-2}.$$

In the notation of Definition 1.2 we therefore have  $L_Y(k) = \sigma^2 H(2H - 1)$ ,

$$H = d + \frac{1}{2},$$

and  $X_t$  ( $t \in \mathbb{Z}$ ) has long memory if  $\frac{1}{2} < H < 1$ . Also note that for the variance of  $S_n = \sum_{t=1}^n X_t$ , self-similarity implies

$$\text{var}(S_n) = \text{var}(Y(n) - Y(0)) = n^{2H} \sigma^2,$$

so that, for  $H > \frac{1}{2}$ , the variance grows at a rate that is faster than linear. For  $H = \frac{1}{2}$ , all values of  $\gamma_X(k)$  are zero except for  $k = 0$ , so that  $X_t$  ( $t \in \mathbb{Z}$ ) is an uncorrelated sequence. For  $0 < H < \frac{1}{2}$ ,  $\gamma_X(k)$  is summable, so that, in contrast to the case with

$H > \frac{1}{2}$ , the sum over all covariances can be split into three terms,

$$\begin{aligned} & \sum_{k=-\infty}^{\infty} [ |k-1|^{2H} + |k+1|^{2H} - 2|k|^{2H} ] \\ &= \sum_{k=-\infty}^{\infty} |k-1|^{2H} + \sum_{k=-\infty}^{\infty} |k+1|^{2H} - 2 \sum_{k=-\infty}^{\infty} |k|^{2H} \\ &= \sum_{k=-\infty}^{\infty} |k|^{2H} + \sum_{k=-\infty}^{\infty} |k|^{2H} - 2 \sum_{k=-\infty}^{\infty} |k|^{2H} = 0. \end{aligned}$$

In other words,  $0 < H < \frac{1}{2}$  implies antipersistence. The simplest SSSI process with finite second moments is a Gaussian process, the so-called fractional Brownian motion (fBm), usually denoted by  $B_H$ . Note that  $B_H$  is the only Gaussian SSSI-process because apart from the variance  $\sigma^2$ , the first two moments are fully specified by the SSSI-property. The corresponding increment sequence  $X_t$  ( $t \in \mathbb{R}$  or  $\mathbb{Z}$ ) is called fractional Gaussian noise (FGN).

To see how to extend the relationship between the self-similarity parameter  $H$  and long-range dependence beyond Gaussian processes, we first look at an explicit time-domain representation of fractional Gaussian motion. The definition and existence of fBm follow directly from the definition of its covariance function. The difference between standard Brownian motion (with  $H = \frac{1}{2}$ ) and fractional Brownian motion with  $H \neq \frac{1}{2}$  can be expressed by a moving average representation of  $B_H(u)$  on the real line, which is a weighted integral of standard Brownian motion. For  $H \neq \frac{1}{2}$ , we have

$$B_H(u) = \int_{-\infty}^{\infty} Q_{u,1}(x; H) dB(x), \quad (1.8)$$

where

$$Q_{u,1}(x; H) = c_1 [(u-x)_+^{H-\frac{1}{2}} - (-x)_+^{H-\frac{1}{2}}] + c_2 [(u-x)_-^{H-\frac{1}{2}} - (-x)_-^{H-\frac{1}{2}}],$$

and  $c_1, c_2$  are deterministic constants. This representation is not unique since it depends on the choice of  $c_1$  and  $c_2$ . A causal representation of fBm is obtained if we choose  $c_2 = 0$  and

$$c_1 = \frac{\sqrt{\Gamma(2H+1) \sin(\pi H)}}{\Gamma(H+\frac{1}{2})} = \left\{ \int_0^{\infty} [(1+s)^{H-\frac{1}{2}} - s^{H-\frac{1}{2}}]^2 ds + \frac{1}{2H} \right\}^{-\frac{1}{2}}.$$

One can verify that the kernel  $Q_{u,1}(\cdot, H)$  has the following property: for all  $0 \leq v < u, x \in \mathbb{R}$ ,

$$Q_{u,1}(x; H) - Q_{v,1}(x; H) = Q_{u-v,1}(x-v; H), \quad (1.9)$$

$$Q_{cu,1}(cx; H) = c^{H-1/2} Q_{u,1}(x; H). \quad (1.10)$$

The first property reflects stationarity of increments. The second property leads to self-similarity with self-similarity parameter  $H$ . It should be mentioned at this point that representation (1.8) is not valid for an fBm on  $[0, 1]$ .

As we have seen above, if the second moments are assumed to exist, then the definition of self-similarity fully determines the autocorrelation structure. This leads to a direct definition of Gaussian self-similar processes. The existence and construction of *non-Gaussian* self-similar processes is less straightforward because the autocorrelation structure is not enough. One way of obtaining a large class of non-Gaussian self-similar processes is to extend the integral representation (1.8) to *multiple Wiener-Itô integrals* (see e.g. Major 1981). This can be done as follows. For  $q \geq 1$  and  $0 < H < 1$ , we define the processes

$$Z_{H,q}(u) = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} Q_{u,q}(x_1, \dots, x_q; H) dB(x_1) \cdots dB(x_q) \quad (1.11)$$

where the kernel  $Q_{u,q}$  is given by

$$Q_{u,q}(x_1, \dots, x_q) = \int_0^u \left( \prod_{i=1}^q (s - x_i)_+^{-\left(\frac{1}{2} + \frac{1-H}{q}\right)} \right) ds.$$

All kernels have the two properties guaranteeing stationarity of increments and self-similarity. The self-similarity property is of the form

$$Q_{cu,q}(cx_1, \dots, cx_q; H) = c^{H-\frac{q}{2}} Q_{u,q}(x_1, \dots, x_q; H).$$

The exponent  $-q/2$  instead of  $-1/2$  is due to the fact that  $dB$  occurs  $q$  times in the product. More explicitly, we can see that the scaling property of  $Q_{u,q}$  implies self-similarity with parameter  $H$  as follows:

$$\begin{aligned} Z_{H,q}(cu) &= \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} Q_{ct,q}(x_1, \dots, x_q; H) dB(x_1) \cdots dB(x_q) \\ &= \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} Q_{ct,q}\left(c\frac{x_1}{c}, \dots, c\frac{x_q}{c}; H\right) dB\left(c\frac{x_1}{c}\right) \cdots dB\left(c\frac{x_q}{c}\right) \\ &= \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} c^{H-\frac{q}{2}} Q_{u,q}(y_1, \dots, y_q; H) c^{\frac{q}{2}} dB(y_1) \cdots dB(y_q) \\ &= c^H \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} Q_{u,q}(y_1, \dots, y_q; H) dB(y_1) \cdots dB(y_q) = c^H Z_{H,q}(u). \end{aligned}$$

For  $q > 1$ , the process  $Z_{H,q}(u)$  ( $u \in \mathbb{R}$ ) is no longer Gaussian and is called *Hermite* process on  $\mathbb{R}$ . Sometimes one also uses the terminology *Hermite-Rosenblatt* process, though “Rosenblatt process” originally refers to the case with  $q = 2$  only (see Taqqu 1975).

Equation (1.11) also leads to a natural extension to self-similar processes with long memory and nonexistent second moments. This can be done by replacing

Brownian motion by a process whose second moments do not exist. Note that Brownian motion is just a special example of the much larger class of Lévy processes. These are defined by the property that they have stationary independent increments and vanish at zero almost surely. The nonexistence of second moments can be achieved by assuming that the Lévy process is a symmetric  $\alpha$ -stable (S $\alpha$ S) process  $Z_\alpha(\cdot)$  for some  $0 < \alpha < 2$ . This means that every linear combination  $Y = \sum_{j=1}^m c_j Z_\alpha(u_j)$  has a symmetric  $\alpha$ -stable distribution with characteristic function  $\varphi(\omega) = E[\exp(i\omega Y)] = \exp(-a|\omega|^\alpha)$ . In particular, S $\alpha$ S Lévy processes are self-similar with self-similarity parameter  $H_{Lévy} = 1/\alpha$ . Hence, we note that unlike in the Gaussian case of fBm, here self-similarity does not have anything to do with long memory. Furthermore, symmetric  $\alpha$ -stable Lévy processes arise as limits of appropriately standardized partial sums  $S_{[nu]} = \sum_{i=1}^{[nu]} X_t$ , where  $X_t$  are i.i.d. and have symmetric heavy tails with tail index  $\alpha$  in the sense that

$$\lim_{x \rightarrow -\infty} |x|^\alpha P(X < -x) = \lim_{x \rightarrow +\infty} x^\alpha P(X > x) = C_1 \quad (1.12)$$

for some  $0 < \alpha < 2$  and a suitable constant  $C_1$  (see e.g. Embrechts et al. 1997; Embrechts and Maejima 2002, and Sect. 4.3). In particular, the process  $S_{[nu]}$  has to be standardized by  $d^{-1}(n)$ , where  $d(n) = n^{H_{Lévy}} = n^{1/\alpha}$ . Therefore, for sequences  $X_t$  with tail index  $\alpha < 2$ , the self-similarity parameter  $H = H_{Lévy} = 1/\alpha$  is the analogue to  $H = \frac{1}{2}$  in the case of finite second moments. If, on the other hand, a nondegenerate limit of  $d^{-1}(n)S_{[nu]}$  is obtained for standardizations  $d(n)$  proportional to  $n^H$  with  $H > 1/\alpha$ , then the memory (in the sequence  $X_t$ ) is so strong that partial sums diverge faster than for Lévy processes. This is analogous to  $H > \frac{1}{2}$  in the case of finite second moments. Therefore, long memory is associated with the condition  $H > 1/\alpha$ . (Note that for  $\alpha = 2$ , we are back to finite second moments, so that we obtain the previous condition  $H > \frac{1}{2}$ .) In analogy to the case of finite second moments we may also define the fractional parameter  $d = H - 1/\alpha$ . Long memory is then associated with  $d > 0$ . Note also that, since the self-similarity parameter is by definition in the interval  $(0, 1)$ , long memory cannot be achieved for  $\alpha < 1$ .

As we will see in Sect. 4.3, in general the limit of  $d^{-1}(n)S_{[nu]}$  is a *Linear Fractional stable motion* defined by

$$\tilde{Z}_{H,\alpha}(u) = \int_{-\infty}^{\infty} Q_{u,1}(x; H, \alpha) dZ_\alpha(x) \quad (1.13)$$

with

$$Q_{u,1}(x; H, \alpha) = c_1[(u-x)_+^{H-1/\alpha} - (-x)_+^{H-1/\alpha}] + c_2[(u-x)_-^{H-1/\alpha} - (-x)_-^{H-1/\alpha}] \quad (1.14)$$

and  $H > 1/\alpha$ . This definition is obviously analogous to (1.8) for fractional Brownian motion. Moreover, the definition is valid for  $H \in (0, 1)$ ,  $H \neq 1/\alpha$ .

### 1.3.6 Other Approaches

#### 1.3.6.1 Different Dependence Measures

For processes with infinite second moments, long-range dependence has to be measured by other means than autocorrelations, the spectral density or the variance of cumulative sums. For instance, the variance  $V_m$  defined in (1.3) can be replaced by

$$\hat{V}_m = \frac{X_t^{(m)}}{\sum_{i=tm-m+1}^{tm} X_i^2} \quad (1.15)$$

(also see Hall 1997). An alternative dependence measure is for example the so-called codifference (Samorodnitsky and Taqqu 1994). Suppose that  $X_t$  ( $t \in \mathbb{Z}$ ) have a symmetric distribution. Then the codifference is defined by

$$\tau_X(k) = \log \frac{E[e^{i(X_{t+k}-X_t)}]}{E[e^{iX_{t+k}}]E[e^{-iX_t}]}. \quad (1.16)$$

Note that  $\tau_X$  can also be defined in continuous time. For Gaussian processes,  $\tau_X(k)$  coincides with the autocovariance function  $\gamma_X(k)$ .

#### 1.3.6.2 Extended Memory

Granger (1995) and Granger and Ding (1996) consider a different property characterizing long-term effects of observations from the remote past.

**Definition 1.8** Let  $X_t$  be a stochastic process defined for  $t \in \mathbb{Z}$  or  $t \in \mathbb{N}$  and such that  $E(X_t^2) < \infty$  for all  $t$ . Consider the prediction

$$\hat{X}_{t+k} = E[X_{t+k} | X_s, s \leq t].$$

Then  $X_t$  is said to have extended memory if there is no constant  $c \in \mathbb{R}$  such that  $\hat{X}_{t+k} \rightarrow_p c$  as  $k \rightarrow \infty$ .

*Example 1.3* Consider a random walk process defined by  $X_t = \sum_{s=1}^t \varepsilon_s$  ( $t \geq 1$ ) where  $\varepsilon_t$  are i.i.d.  $N(0, \sigma_\varepsilon^2)$  distributed with  $\sigma_\varepsilon^2 > 0$ . Then

$$\hat{X}_{t+k} = X_t$$

for all  $k \geq 1$ , so that  $\hat{X}_{t+k}$  does not converge to a constant but is instead  $N(0, t\sigma_\varepsilon^2)$ -distributed for all  $k$ . Thus, random walk has extended memory. Similarly, for  $Y_t = \exp(X_t) = \exp(\sum_{s=1}^t \varepsilon_s)$ , we have

$$\hat{Y}_{t+k} = Y_t E \left[ \exp \left( \sum_{j=1}^k \varepsilon_{t+j} \right) \right] = Y_t \exp \left( \frac{1}{2} \sigma_\varepsilon^2 k \right).$$

Again,  $Y_t$  does not converge to a constant, but instead  $P(\hat{Y}_{t+k} \rightarrow \infty) = 1$ . This illustrates that extended memory also captures nonlinear dependence. The reason is that the conditional expected value and not just the best linear forecast is considered. More generally, any strictly monotone transformation  $G(X_t)$  has extended memory (see e.g. Granger and Ding 1996 and references therein). In contrast, for  $|\varphi| < 1$ , the equation  $X_t = \varphi X_{t-1} + \varepsilon_t$  ( $t \in \mathbb{Z}$ ) has a unique stationary causal solution  $X_t = \sum_{j=0}^{\infty} \varphi^j \varepsilon_{t-j}$ , and

$$\hat{X}_{t+k} = \varphi^k X_t \xrightarrow{p} 0.$$

More generally, for a purely stochastic invertible linear second-order stationary process with Wold representation  $X_t = \sum_{j=0}^{\infty} a_j \varepsilon_{t-j}$  and i.i.d.  $\varepsilon_t$ , we have

$$\hat{X}_{t+k} = E[X_{t+k} | X_s, s \leq t] = \sum_{j=0}^{\infty} a_{j+k} \varepsilon_{t-j},$$

so that

$$\text{var}(\hat{X}_{t+k}) = E(\hat{X}_{t+k}^2) = \sum_{j=k}^{\infty} a_j^2 \xrightarrow{k \rightarrow \infty} 0.$$

Since  $\hat{X}_{t+k}$  converges to zero in the  $L^2$ -norm and in probability, the process  $X_t$  does not have extended memory.

### 1.3.6.3 Long Memory as Phase Transition

The approach in this section was initiated by G. Samorodnitsky, see Samorodnitsky (2004, 2006). Let  $\{P_\theta, \theta \in \Theta\}$  be a family of probability measures that describe the finite-dimensional distributions of a stationary stochastic process  $\mathbf{X} = (X_t)$  ( $t \in \mathbb{Z}$  or  $t \in \mathbb{R}$ ). We assume that as  $\theta$  varies over the parameter space  $\Theta$ , the marginal distribution of  $X_t$  does not change. Consider a measurable functional  $\phi = \phi(\mathbf{X})$ . Its behaviour may be different for different choices of  $\theta$ . Now, assume that the parameter space  $\Theta$  can be decomposed into  $\Theta_1 \cup \Theta_2$  such that the behaviour of the functional does not change too much as  $\theta \in \Theta_1$ , but changes significantly when we cross the boundary between  $\Theta_1$  and  $\Theta_2$ . Furthermore, the behaviour changes as  $\theta$  varies across  $\Theta_2$ . This way, we can view the models with  $\theta \in \Theta_1$  as short-memory models and those with  $\theta \in \Theta_2$  as long-memory models. One has to mention here that this notion of LRD does not look at one particular parameter (in contrast to the case of a finite variance where  $\theta$  can be thought of as an exponent of a hyperbolic decay of covariances). Instead, it is tied to each particular functional. It may happen that a particular model is LRD for one functional but not for another. In other words, if we have two functionals  $\phi_1$  and  $\phi_2$ , the decomposition of the parameter space may be completely different, i.e.  $\Theta = \Theta_1(\phi_1) + \Theta_2(\phi_1)$  and  $\Theta = \Theta_1(\phi_2) + \Theta_2(\phi_2)$  with  $\Theta_1(\phi_1) \neq \Theta_1(\phi_2)$ .

*Example 1.4 (Partial Sums)* Denote by  $L_f(\lambda)$  a function that is slowly varying at the origin in Zygmund's sense. Let  $\mathbf{X} = (X_t, t \in \mathbb{Z})$  be a stationary Gaussian sequence with spectral density  $f_X(\lambda) \sim L_f(\lambda)\lambda^{-2d}$  (as  $\lambda \rightarrow 0$ ) but  $f_X(0) \neq 0$ , and assume that  $d =: \theta \in [-\infty, \frac{1}{2})$ . (Here  $d = -\infty$  is interpreted as the case of i.i.d. random variables.) For the functional  $\phi_1(\mathbf{x}) = \sum_{t=1}^n x_t$ , the parameter space may be decomposed into  $(0, \frac{1}{2}) \cup \{0\} \cup (-\infty, 0]$ . For the sub-space  $(0, \frac{1}{2})$ , the rate of convergence changes for different choices of  $\theta$ . In other words, according to Samorodnitsky's definition,  $\mathbf{X}$  is  $\phi_1$ -LRD for  $\theta \in (0, \frac{1}{2})$  since then the partial sum has to be scaled by  $L_\gamma^{-1/2}(n)n^{-d-\frac{1}{2}}$  to obtain a nondegenerate limit. Otherwise, if  $\theta \in [-\infty, 0]$ , then the scaling is  $n^{-1/2}$ . If instead, we consider the functional  $\phi_2(\mathbf{x}) = \sum_{t=1}^n (x_t^2 - 1)$ , then the parameter space is decomposed into  $(\frac{1}{4}, \frac{1}{2}) \cup \{\frac{1}{4}\} \cup [-\infty, \frac{1}{4})$ . The process  $\mathbf{X}$  is  $\phi_2$ -LRD for  $\theta \in (\frac{1}{4}, \frac{1}{2})$ . We refer to Chap. 4 for a detailed discussion of limit theorems for partial sums.

*Example 1.5 (Maxima)* Let  $\mathbf{X} = (X_t, i \in \mathbb{Z})$  be as in Example 1.4, but we consider the functional  $\phi_3(\mathbf{x}) = \max_{t=1}^n x_t$ . The limiting behaviour of maxima of Gaussian sequences with nonsummable autocovariances or autocovariances that sum up to zero is the same as under independence. Thus, according to Samorodnitsky's definition,  $\mathbf{X}$  is not max-LRD. We refer to Sect. 4.10 for limit theorems for maxima.

However, the main reason to consider the “phase transition” approach is to quantify long-memory behaviour for stationary stable processes. In particular, if  $X_t = Z_{H,\alpha}(t) - Z_{H,\alpha}(t-1)$ , where  $Z_{H,\alpha}(\cdot)$  is a Linear Fractional Stable motion (1.13), then, due to self-similarity,  $n^{-H} \sum_{t=1}^n X_t$  equals in distribution  $Z_{H,\alpha}(1)$ , where  $H = d + 1/\alpha$ . On the other hand, if  $X_t$  are i.i.d. symmetric  $\alpha$ -stable, then  $n^{-1/\alpha} \sum_{t=1}^n X_t$  equals in distribution an  $\alpha$ -stable random variable. Hence, the phase transition from short memory to long memory occurs at  $H = 1/\alpha$ . A similar transition occurs in the case of ruin probabilities.

*Example 1.6 (Ruin Probabilities)* As in Mikosch and Samorodnitsky (2000), assume again that  $X_t = Z_{H,\alpha}(t) - Z_{H,\alpha}(t-1)$ , where  $Z_{H,\alpha}(\cdot)$  is a Linear Fractional Stable motion. The authors consider the rate of decay of ruin probabilities

$$\psi(u) = P\left(\sum_{t=1}^n X_t > cn + u \text{ for some } n \in \mathbb{N}\right)$$

as  $u$  tends to infinity. As it turns out, for  $H > 1/\alpha$ ,  $\psi(u)$  is proportional to  $u^{-(\alpha-\alpha H)}$ , whereas for  $0 < H \leq 1/\alpha$ , the decay is of the order  $u^{-(\alpha-1)}$ . Thus, for  $H > 1/\alpha$ , the decay is slower, which means that the probability of ruin is considerably larger than for  $H \leq 1/\alpha$ . Moreover, the decay depends on  $H$  for  $H > 1/\alpha$ , whereas this is not the case for  $H \leq 1/\alpha$ . It is therefore natural to say that  $X_t$  has long memory if  $H > 1/\alpha$  and short memory otherwise.

*Example 1.7 (Long Strange Segments)* Another possibility of distinguishing between short- and long-range dependent ergodic processes is to consider the rate at



which so-called long strange segments grow with increasing sample size (Ghosh and Samorodnitsky 2010; Mansfield et al. 2001; Rachev and Samorodnitsky 2001). Suppose that  $X_t$  is a stationary process with  $\mu = E(X_t) = 0$  and the ergodic property in probability holds (i.e. the sample mean converges to  $\mu$  in probability). Given a measurable set  $A$ , one defines

$$R_n(A) = \sup\{j - i : 0 \leq i < j \leq n, \bar{x}_{i:j} \in A\},$$

where

$$\bar{x}_{i:j} = (j - i)^{-1} \sum_{t=i+1}^j X_t$$

is the sample mean of observations  $X_{i+1}, \dots, X_j$ . In other words, the random number  $R_n(A) \in \mathbb{N}$  is the maximum length of a segment from the first  $n$  observations whose average is in  $A$ . Why such segments are called “strange” can be explained for sets  $A$  that do not include the expected value  $\mu = 0$ . Since the sample mean converges to zero, one should not expect too long runs that are bounded away from zero. It turns out, however, that for long-memory processes, the maximal length of such runs tends to be longer than under short memory, in the sense that  $R_n$  diverges to infinity at a faster rate.

The phase transition approach leads also to much more general stationary stable processes. It turns out that stationary stable processes can be decomposed into a dissipative and a conservative flow. The conservative flow part is usually associated with long memory. We refer to Samorodnitsky (2002, 2004, 2005, 2006), Racheva-Iotova and Samorodnitsky (2003), Resnick and Samorodnitsky (2004) for further details and examples.