SPRINGER BRIEFS IN PHYSICS

Călin Vamoş · Maria Crăciun

Automatic Trend Estimation

Springer

SpringerBriefs in Physics

For further volumes:
http://www.springer.com/series/8902

Călin Vamoş · Maria Crăciun

Automatic Trend Estimation

Springer

Călin Vamoş
"Tiberiu Popoviciu" Institute of Numerical Analysis
Romanian Academy
Cluj-Napoca
Romania

Maria Crăciun
"Tiberiu Popoviciu" Institute of Numerical Analysis
Romanian Academy
Cluj-Napoca
Romania

ISSN 2191-5423 ISSN 2191-5431 (electronic)
ISBN 978-94-007-4824-8 ISBN 978-94-007-4825-5 (eBook)
DOI 10.1007/978-94-007-4825-5
Springer Dordrecht Heidelberg New York London

Library of Congress Control Number: 2012942924

Printed on acid-free paper

Springer is part of Springer Science+Business Media (www.springer.com)

Preface

Huge amount of information is available as time series in many scientific fields: geophysics, astronomy, biophysics, quantitative finance, Internet traffic, etc. Processing so many time series is possible only by means of automatic algorithms usually designed in data mining. One of the critical tasks which has to be achieved by these algorithms is the automatic estimation of the trend contained in an arbitrary noisy time series. The aim of our book is to provide several automatic algorithms for nonmonotonic trend estimation. We do not intend to review the existing automatic trend estimation algorithms, but to present a thorough analysis for those presented in this book.

Obviously, an automatic algorithm is not able to work for all imaginable time series. By its automatic feature we mean that, without any subjective intervention, it efficiently processes time series of a well-defined type. The greater the diversity of the time series types, the more "automatic" the algorithm is. Therefore in designing a trend estimation algorithm an essential component is the method to evaluate its accuracy for a large diversity of time series. However, the algorithms are very often tested under unrealistic conditions and on too small number of time series. One reason for this situation is that the time series theory is dominated by stationary stochastic processes. The theoretical results for nonstationary time series containing a trend hold only under restrictive conditions, seldom satisfied by the real time series.

When the statistical theory is not applicable, Monte Carlo experiments can be used to evaluate the accuracy of the automatic algorithms. Even then the results are useful only if the members of the statistical ensemble have a diversity comparable with that of the real time series. The main difficulty is to generate realistic nonmonotonic trends. Usually, Monte Carlo simulations are performed on artificial time series much simpler than those encountered in practice, with monotonic (linear, power-law, exponential, and logarithmic) or periodic (sinusoidal) trends. The approach based on numerical Monte Carlo experiments in our book is much more general and the trends generated by our original algorithm are meaningful for real time series.

Chapter 1 contains fundamentals in probability theory, statistics, and time series theory which are used in the rest of the book. We analyze the autoregressive noise of order one denoted AR(1), which is a simple model depending only on two parameters: the variance and the constant of the serial correlation. Even for more complex noises an AR(1) model is a zero order approximation capturing their most important features. The noise serial correlation essentially influences the accuracy of the estimated trend because when it increases, the large-scale fluctuations of the noise cannot be distinguished from the trend variations.

In Chap. 2 we construct the statistical ensemble on which the Monte Carlo experiments are performed. There is no rigorous mathematical method to demonstrate that the variability of the obtained artificial time series is rich enough to simulate the variability of the real time series. In fact we construct an independent "numerical reality" on which we perform numerical experiments. Therefore, our approach is more typical to computational physics than to data mining or mathematical statistics. As examples of Monte Carlo experiments we evaluate the confidence interval for a method to estimate the serial correlation parameter of an AR(1) noise and we present a numerical method for testing if a time series is uncorrelated.

In Chaps. 3 and 4 we analyze in detail the accuracy of the classical algorithms of polynomial fitting and moving average in the case of arbitrary nonmonotonic trends. The quality of the estimated trend depends mainly on three parameters: the number of the time series values, the ratio between the amplitudes of the trend variations and the noise fluctuations, and the serial correlation of the noise. Our analysis shows that even in the case of the simplest trend estimation algorithms, due to the many parameters on which the artificial time series depend, a realistic evaluation of their performances is difficult and laborious.

In the last three chapters we present our original automatic algorithms for processing nonstationary time series containing a stationary noise superposed over a nonmonotonic trend. Their performances are tested by means of numerical experiments of the same type as those used in the previous chapters. The algorithms are designed to work on any time series, even if it has only a few values. Obviously, the best results are obtained for an AR(1) noise superposed over a deterministic trend with at least several hundreds of values. For other types of time series the outcomes of the algorithms have to be statistically analyzed by Monte Carlo experiments.

In Chap. 5 we design an automatic algorithm, called the averaged conditional displacement (ACD), to estimate a monotonic trend as a piecewise linear curve. The Monte Carlo experiments indicate that its accuracy is comparable with that of the classical methods, but it has the advantage to be automatic and to describe a much richer set of monotonic trend shapes. Applied to a time series with an arbitrary nonmonotonic trend, the ACD algorithm extracts one of the possible monotonic components which can be associated with the given trend. The probability that the estimated monotonic component is real can be estimated by a method based on surrogate time series.

In Chap. 6 we define the timescale of a local extremum of a time series such that it allows a classification of the local extrema with respect to their importance for the global shape of the time series. The local extrema with scales greater than a given value provide a partition of a noisy time series in segments which approximate the monotonic parts of the trend from a time series. The quality of this approximation is improved by first applying a moving average to the noisy time series. We use the monotonic component estimated by the ACD algorithm as a reference to measure the magnitude of the nonmonotonic variations of a time series. In this way we can build a criterion to stop the partition of a time series when the resulting segments may be considered monotonic.

In the last chapter we give an automatic form to the repeated central moving average (RCMA) analyzed in Chap. 4. In order to adjust the parameters of the RCMA algorithm to the characteristics of the processed time series, we have designed two simple statistical methods to estimate the noise serial correlation and the ratio between the amplitudes of the trend variations and of the noise fluctuations. The partitioning algorithm presented in Chap. 6 is used now to determine the local extrema of the estimated trend which corresponds to the real trend and not to the smoothed noise.

We illustrate the functioning of the analyzed algorithms by processing time series from astrophysics, finance, biophysics, and paleoclimatology. The examples of real time series are typical to the complex situations encountered in practice: data missing from the time series, superposition of several types of noises, long time series with tens of thousands of values, non-Gaussian probability distributions with fat tails, repeated values of the time series, additional conditions imposed on time series by the physical laws governing the studied phenomenon.

Our analysis is restricted to AR(1) noises superposed over nonmonotonic trends, but our methods can be applied to study other noise models. Such new applications could be: autoregressive noise of higher orders, long-range correlated noises, unevenly sampled time series, asymmetric probability distribution of the time series values. Obviously, the number of parameters could increase and the analysis of the accuracy of the estimated trend would become more burdensome.

We have limited our analysis to four methods of trend estimation: two classical (polynomial fitting and moving average) and two original and automatic (one for monotonic trends and the other for arbitrary nonmonotonic trends). Other trend estimation methods can be analyzed using the same type of Monte Carlo experiments. In order to obtain significant results, it is essential to use a statistical ensemble of artificial time series with a variety of trend shapes at least as rich as that generated by our algorithm presented in Chap. 2.

Even if the main definitions and theorems used in the book are briefly presented, nevertheless it is recommended that the reader has the knowledge of basic notions in probability, mathematical statistics, and time series theory. This book is of interest for researchers who need to process nonstationary time series. Detailed descriptions of all the numerical methods presented in the book allow the reader to reproduce the original automatic algorithms for trend estimation and time series partitioning. In addition, the source codes in MATLAB of the computer programs

implementing them are freely available on the web so that the researchers who merely apply trend estimation algorithms could successfully use them.

Cluj-Napoca, April 2012

Călin Vamoş
Maria Crăciun

Contents

Chapter 1
Introduction

A complete presentation of the theory of stochastic processes can be found in any treatise on the probability theory, e.g., [18] and for time series theory one can use [4]. In this introductory chapter we briefly present some basic notions which are used in the rest of the book. The main methods to estimate trends from noisy time series are introduced in Sect. 1.2. In the last section we discuss the properties of the order one autoregressive stochastic process AR(1) which has the serial correlation described by a single parameter and which is a good first approximation for many noises encountered in real phenomena.

1.1 Discrete Stochastic Processes and Time Series

At the occurrence of an event ω the *random variable* X takes the value $X(\omega) = x$. We follow the practice of denoting by small letters the realizations of the random variable denoted by the corresponding capital letters. Throughout this book we consider only continuous random variables with real values. If the random variable is absolutely continuous, then it has a *probability density function* (pdf) denoted $p(x)$. The *cumulative distribution function* (cdf) $F(x) = P(X \leq x)$ is the probability that the random variable X takes on a value less than or equal to x. We denote the *mean* of the random variable by $\mu = \langle X \rangle$ and its *variance* by $\sigma^2 = \langle (X - \langle X \rangle)^2 \rangle$.

The evolution in time of a random phenomenon is modeled by a *stochastic process*, i.e., a family of random variables $\{X(t),\ t \in I \subset \mathbf{R}\}$ defined on the same probability space and indexed by a set of real numbers I. In this book we study only discrete stochastic processes for which I contains equidistant sampling moments. The observations are made at discrete time moments $t_n = t_0 + (n-1)\Delta t$, where $n = 1, 2, \ldots, N$, Δt is the sampling interval, and t_0 is the initial time. The observed values $x_n \equiv x(t_n)$ are realizations of the corresponding random variables $X_n \equiv X(t_n)$. Although the number of observations is always finite, we assume that there is an infinite stochastic process $\{X_n, n = 0, \pm 1, \pm 2, \ldots\}$ whose realizations for $n < 1$ and $n > N$ have not

C. Vamoş and M. Crăciun, *Automatic Trend Estimation*, SpringerBriefs in Physics,
DOI: 10.1007/978-94-007-4825-5_1, © The Author(s) 2012

been observed. To distinguish between the infinite stochastic process which models the time evolution of the natural phenomenon and its measurements, we call *time series* the finite sequence of real numbers $\{x_n, n = 1, 2, ..., N\}$.

The joint cdf of the random variables $X_{n_1}, X_{n_2}, \ldots, X_{n_m}$ is the probability that their values are smaller than some given values

$$F_{\mathbf{n}}(\mathbf{x}) = P(X_{n_1} \leq x_1, \ldots, X_{n_m} \leq x_m),$$

where $\mathbf{n} = (n_1, \ldots, n_m)$ and $\mathbf{x} = (x_1, \ldots, x_m)$. For absolutely continuous random variables there exists the joint pdf $p_{\mathbf{n}}(\mathbf{x})$. A stochastic process is (strictly) *stationary* if for every index vector $\mathbf{n}$ and integer d we have $F_{\mathbf{n}+d\mathbf{1}} = F_{\mathbf{n}}$ or $p_{\mathbf{n}+d\mathbf{1}} = p_{\mathbf{n}}$, where $\mathbf{1} = (1, 1, \ldots, 1)$, i.e., its joint probabilities do not change under temporal translations. From this definition, for $m = 1$ it follows that all the components of a stationary process have the same probability distribution $p_n(x) = p(x)$ for all integers n. Such a stochastic process is called *identically distributed*.

The *autocovariance function* of a stochastic process with finite variance for all its components ($\sigma_n^2 < \infty$) is defined as

$$\gamma(n, m) = \langle (X_n - \langle X_n \rangle)(X_m - \langle X_m \rangle) \rangle. \tag{1.1}$$

Obviously $\gamma(n, n) = \sigma_n^2$. If the stochastic process is stationary, then

$$\gamma(n + d, m + d) = \gamma(n, m), \tag{1.2}$$

for all n, m, d integers and the autocovariance function depends only on the lag $h = n - m$ so that $\gamma(h) \equiv \gamma(h, 0)$. It is easy to show that $\gamma(0) \geq 0$, $\gamma(h) = \gamma(-h)$, and $|\gamma(h)| \leq \gamma(0)$ for any h. The *autocorrelation function* of a stationary stochastic process is defined as $\rho(h) = \gamma(h)/\gamma(0)$ and then $\rho(0) = 1$.

Usually the observed time series do not satisfy the condition imposed to strictly stationary stochastic processes. Furthermore, the analysis of time series is often reduced only to the statistical moments of second order. Therefore one defines a subclass of the stationary process more suitable for modeling of real phenomena. A stochastic process is *weak-stationary* if $\langle |X_n^2| \rangle < \infty$, $\langle X_n \rangle = \mu$ for all integers n and satisfies Eq. (1.2). A special weak-stationary process is the *white noise*, for which the components are uncorrelated $\gamma(h) = \sigma^2 \delta_{h0}$, where δ_{nm} is the Kronecker delta. Such a stochastic process is denoted by $X_n \sim WN(\mu, \sigma^2)$.

Another subclass of stationary processes contains the *independent and identically distributed* (i.i.d.) stochastic processes. The components of an i.i.d. process are mutually independent $p_{\mathbf{n}}(\mathbf{x}) = p_{n_1}(x_1) p_{n_2}(x_2) \ldots p_{n_m}(x_m)$. They are also identically distributed $p_{n_i}(x_i) = p(x_i)$ and then $p_{\mathbf{n}+h\mathbf{1}}(\mathbf{x}) = p(x_1) p(x_2) \ldots p(x_m) = p_{\mathbf{n}}(\mathbf{x})$ so that, if the stochastic process is infinite, the stationarity condition (1.2) is satisfied.

If the properties of the components of a stochastic process vary in time, then the stochastic process is *nonstationary*. As an example of nonstationary stochastic process we consider the *random walk* $\{X_n, n = 0, 1, 2, \ldots\}$ defined as

$$X_n = X_{n-1} + Z_n \quad \text{for} \quad n > 0, \tag{1.3}$$

where $\{Z_n\}$ is an i.i.d. stochastic process with zero mean, variance σ^2, and $X_0 = Z_0$. Obviously $\langle X_n \rangle = 0$ and for $n \leq m$ the autocovariance function given by Eq. (1.1) becomes

$$\gamma(n, m) = \langle X_n X_m \rangle = \langle X_n (X_n + Z_{n+1} + \cdots + Z_m) \rangle = \langle X_n^2 \rangle$$

because X_n depends only on $Z_0, Z_1, \ldots, Z_n$ which are independent of $Z_{n+1}, \ldots, Z_m$. Because

$$\langle X_n^2 \rangle = \sum_{k=0}^{n} \langle Z_k^2 \rangle + \sum_{k \neq l}^{n} \langle Z_k Z_l \rangle = (n+1)\sigma^2$$

we have

$$\gamma(n, m) = (1 + \min\{n, m\})\sigma^2. \tag{1.4}$$

Hence the autocovariance function of the random walk is not invariant to temporal translations and $\{X_n\}$ is a nonstationary process.

In practice we do not have access to random variables or stochastic processes but only to their realizations and we have to use the methods of the mathematical statistics in order to estimate the parameters of the observed phenomena. Let us consider a random variable X and one of its realizations $x^{(s)}$.[1] The set formed by S independent realizations $\{x^{(1)}, x^{(2)}, \ldots, x^{(S)}\}$ is called *sample of volume* S and it allows the estimation of the parameters of X. For instance the mean $\mu = \langle X \rangle$ is approximated by the *sample mean*

$$\mu^{\text{est}} \equiv \widehat{\mu} = \frac{1}{S} \sum_{s=1}^{S} x^{(s)}. \tag{1.5}$$

We make the convention that the quantities computed by means of a sample are denoted by a hat or with the superscript 'est'. By means of the law of large numbers one proves under rather general conditions that $\widehat{\mu}$ tends to μ when S tends to infinity. In the same way we define the *sample variance*

$$\widehat{\sigma}^2 = \frac{1}{S} \sum_{s=1}^{S} \left(x^{(s)} - \widehat{\mu} \right)^2. \tag{1.6}$$

Analogous relations can be used for a stationary time series $\{x_1, x_2, \ldots, x_N\}$. Instead of the sample mean (1.5) we define the *temporal mean*

[1] We have changed the usual notation x_s in order to avoid the confusion with the terms of a time series.

$$\overline{x} = \frac{1}{N}\sum_{n=1}^{N} x_n. \tag{1.7}$$

Since the associated stochastic process $\{X_n\}$ is stationary, all the terms in the sum have identical pdfs and $\overline{x}$ tends to its theoretical mean μ when N tends to infinity. A similar analogy can be made for the sample variance (1.6). The serial correlation of a time series is characterized by the *sample autocovariance function*

$$\widehat{\gamma}(h) = \frac{1}{N}\sum_{n=1}^{N-h}(x_{n+h} - \overline{x})(x_n - \overline{x}), \quad 0 \le h < N. \tag{1.8}$$

For $-N < h \le 0$, we have $\widehat{\gamma}(h) = \widehat{\gamma}(-h)$. If X_n is a linear combination of the components of an i.i.d. stochastic process with finite fourth order moment and the sum of the absolute values of the linear combination coefficients is finite, then the estimator (1.8) is biased, but its asymptotic distribution has the mean equal to the theoretical autocovariance function ([4], Chap. 7). The *sample autocorrelation function* is given by

$$\widehat{\rho}(h) = \widehat{\gamma}(h)/\widehat{\gamma}(0), \quad |h| < N. \tag{1.9}$$

1.2 Trend Definition and Estimation

Stochastic processes model the random phenomena as opposed to the deterministic phenomena which are modeled by numerical functions of time. There are many situations when different random and deterministic phenomena overlap. In the simplest case, a deterministic and a random phenomenon, mutually independent, are superposed (for example the instrumental noise affecting a measured physical quantity). The most frequently used model is the stochastic process

$$X_n = f_n + Z_n, \tag{1.10}$$

where $\{Z_n\}$ is a stationary stochastic process with zero mean $\langle Z_n \rangle = 0$ named *additive noise* and $f_n = f(t_n)$ are the values at the sampling moments of the deterministic function named *trend*.

We denote by $p_Z(z)$ the pdf of Z_n and by $p_X(x, n)$ that of X_n. Because $\{Z_n\}$ is stationary, p_Z does not depend on n. According to Eq. (1.10), p_X is equal to p_Z translated by f_n

$$p_X(x, n) = p_Z(x - f_n). \tag{1.11}$$

The explicit dependence of p_X on the time index n indicates that $\{X_n\}$ is a nonstationary process with the mean varying in time

$$\langle X_n \rangle = f_n + \langle Z_n \rangle = f_n.$$

Using this relation in Eq. (1.1) we obtain

$$\gamma_X(n, m) = \langle Z_n Z_m \rangle = \gamma_Z(n - m),$$

hence its autocovariance function is identical with that of the stationary noise, the nonstationarity of $\{X_n\}$ being restricted only to its mean.

We cannot apply the usual statistical methods to a single realization of a nonstationary stochastic process given by Eq. (1.10)

$$x_n = f_n + z_n. \tag{1.12}$$

If more time series are available $\{x_n^{(s)}\}$, $s = 1, 2, \ldots, S$, then for each term of the stochastic process a sample is available and we can evaluate the trend using the simple average given by Eq. (1.5)

$$f_n = \frac{1}{S} \sum_{s=1}^{S} x_n^{(s)}. \tag{1.13}$$

In practice more time series obtained under exactly the same conditions are rarely available. Usually we have to analyze a single time series and then the trend must be estimated and removed from the given time series.

The estimation of the trend $\{f_n\}$ from the time series $\{x_n\}$ is more accurate if the properties of the trend are distinct from those of the noise. Usually the trend is qualitatively characterized as being the component of a time series that is "slowly changing in time" [3]. It is implicitly assumed that the noise does not have this property. In order to formulate more precisely this observation, we denote by τ_Z the time scale of the stochastic process $\{Z_n\}$. It is defined as the time interval for which $\{z_n\}$ becomes uncorrelated. For example, a white noise has $\tau_Z = 1$. The condition that the trend has a slow variation in comparison with the noise means that the trend values remain strongly correlated over intervals of the order τ_Z. If we denote by τ_f the time scale for which the trend values are correlated, then the property of slow variation of the trend corresponds to the condition $\tau_Z < \tau_f$. Usually the trend has only a few monotonic components on the whole time series, so τ_f is of the magnitude order of the time series length. Then the condition $\tau_Z < \tau_f$ represents the requirement that the values of the noise become uncorrelated on time intervals smaller than the length of the analyzed time series.

Obviously, there are situations when $\tau_Z \sim \tau_f$, for example, if the noise is strongly correlated or if the time series is too short. In such cases the slowest fluctuations of the noise are confounded with the variations due to the trend, and their separation is very difficult. In such cases the false trend is called *stochastic trend*. It is only partially possible to avoid such errors when we have more information on the specific properties of noise or trend.

Confusion is often made between the trend definition and trend estimation. The definition is based on the different nature of the two terms in Eq. (1.10): the trend is the deterministic part of a time series with additive noise. Then it can be determined by means of Eq. (1.13) if we dispose of a sufficient number of time series (1.12) generated in the same conditions by the same phenomenon. Due to its random nature, the averaging in Eq. (1.13) eliminates the noise from the averaged time series. Difficulties occur when we have to estimate the trend from a single time series. Therefore a statement as "a rigorous and satisfactory definition of either the trend of nonlinear nonstationary data or the corresponding detrending operation still is lacking" [17] really refers to trend estimation, not to trend definition.

There are many methods of trend estimation [1, 13], but most of them belong to two major classes, parametric and nonparametric. A parametric method chooses for $f(t)$ a functional form depending on several parameters and then computes their values by regression. As a typical parametric method we analyze in Chap. 3 the *polynomial fitting*. The estimation of the trend obtained with this method becomes worse as the form of the function $f(t)$ becomes more complicated and as the noise serial correlation increases.

Among the nonparametric methods to estimate the trend we study in Chap. 4 the noise smoothing by *moving average* (MA). The *central moving average* (CMA) is defined as

$$\vartheta_x(n) = \frac{1}{2K+1} \sum_{k=-K}^{K} x_{n+k}, \tag{1.14}$$

where K is the semi-length of the averaging window given by a positive integer. Applying the CMA to the time series (1.12) we obtain

$$\vartheta_x(n) = \vartheta_f(n) + \vartheta_z(n). \tag{1.15}$$

Since the noise is more fluctuant than the trend, it is more damped by the CMA and then $\vartheta_x(n) \approx \vartheta_f(n)$. If the trend is not too much distorted by smoothing, then the averaged time series is a good estimator of the trend $f_n \approx \vartheta_x(n)$.

Every method for trend estimation can be transformed into an automatic algorithm if the optimum values of its parameters for the analyzed time series are automatically determined. The difficulty consists in the great variability of the possible time series. For instance, the ratio between the amplitude of the trend variations and the amplitude of noise fluctuations has a significant influence on the trend estimation quality. If the time series is dominated by noise, then the degree of the estimated polynomial trend must be small, such that the estimated trend should not follow the noise fluctuations. If the CMA is used, then K must take large values such that the noise should be strongly smoothed. Inversely, when the time series is dominated by trend, then the optimum degree of the polynomial trend increases and K decreases. The accuracy of the estimated trend also depends on the serial correlation of the noise. Therefore, before choosing the correct values of the parameters of the automatic algorithm,

we have to estimate the main properties of the time series containing an arbitrary nonmonotonic trend.

Sometimes the time series analysis is performed not in order to estimate the trend, but to study the noise. In this case we say that a *detrending* method is applied. *Differencing* is such an algorithm. We define the difference operator ∇_d for a positive integer d as

$$\nabla_d X_n = X_{n+d} - X_n, \tag{1.16}$$

where $\{X_n\}$ is a stochastic process. This definition also holds for deterministic functions. Since the operator ∇_d is linear, its action on the stochastic process (1.10) is separated into two terms

$$\nabla_d X_n = \nabla_d f_n + \nabla_d Z_n. \tag{1.17}$$

If the step d is small enough, then the term due to the trend is much smaller than the term due to the noise because according to its definition the trend has a slow variation. Hence we have the approximation

$$\nabla_d X_n \approx \nabla_d Z_n. \tag{1.18}$$

It is obvious that it is not always possible to find such a value for d, for example, if the noise fluctuations have an amplitude of the same magnitude order as the trend variation over a sample interval.

1.3 AR(1) Stochastic Process

The autoregressive-moving average (ARMA) processes are the fundamental models in time series theory. A detailed presentation can be found in [4]. In this book we use only a particular form of ARMA processes. The process $\{Z_n, n = 0, \pm 1, \pm 2, \ldots\}$ is an AR(1) process if it is stationary and for every n

$$Z_n = \phi Z_{n-1} + W_n. \tag{1.19}$$

where $\{W_n\} \sim WN(0, \sigma^2)$. One can specify more exactly the properties of such stochastic processes if the noise characteristics are better defined. For example, if the white noise is Gaussian, then the AR(1) process is also Gaussian.

By iteratively applying Eq. (1.19) h times we obtain

$$Z_n = W_n + \phi W_{n-1} + \cdots + \phi^{h-1} W_{n-h+1} + \phi^h Z_{n-h}. \tag{1.20}$$

Because the process $\{Z_n\}$ is stationary, the random variables Z_n and Z_{n-h} have the same norm and then the norm of the last term in Eq. (1.20) is ϕ^h times the norm of the left term. If $|\phi| < 1$, then for any positive number $\varepsilon < 1$, there exists $h > \ln \varepsilon / \ln |\phi| > 0$ such that the last term from the right side can be neglected

$$Z_n = W_n + \phi W_{n-1} + \cdots + \phi^{h-1} W_{n-h+1} + \mathcal{O}(\varepsilon) \tag{1.21}$$

for $\varepsilon \to 0$. From Eq. (1.21) it follows that the influence of the white noise reduces as h increases and Z_n depends only on the past values of the noise. Such a stochastic process obtained for $|\phi| < 1$ is called *causal*.

In the following we analyze the basic properties of the causal AR(1) process. If we take the mean of Eq. (1.19) we obtain $\langle Z_n \rangle = 0$. Its square is

$$Z_n^2 = \phi^2 Z_{n-1}^2 + 2\phi Z_{n-1} W_n + W_n^2.$$

In accordance with Eq. (1.21) the random variables Z_{n-1} and W_n are independent, and then $\langle Z_{n-1} W_n \rangle = 0$. If we take the mean of the last equation, we obtain a relation between the variances of successive random variables

$$\sigma_n^2 = \phi^2 \sigma_{n-1}^2 + \sigma^2.$$

Because the AR(1) process is stationary, it follows that for every n we have the same value for the variance $\sigma_n \equiv \sigma_s$ and then

$$\sigma_s^2 = \frac{\sigma^2}{1 - \phi^2}. \tag{1.22}$$

In order to compute the autocovariance function we multiply Eq. (1.19) with Z_{n-h} and we take the mean. Because the mean of Z_n vanishes and $\langle Z_{n-h} W_n \rangle = 0$, we obtain

$$\gamma(h) \equiv \langle Z_n Z_{n-h} \rangle = \phi \gamma(h-1)\,.$$

By applying successively this relation and taking into account that $\gamma(0) = \sigma_s^2$, we have

$$\gamma(h) = \sigma_s^2 \phi^h. \tag{1.23}$$

The spectral density is the Fourier transform of the autocovariance function

$$g(\nu) = \frac{1}{2\pi} \sum_{h=-\infty}^{\infty} \mathrm{e}^{-2\pi i h \nu} \gamma(h),$$

where the positive number ν is the frequency. By direct calculation with $\gamma(h)$ given by Eq. (1.23) we obtain

$$g(\nu) = \frac{\sigma^2}{2\pi} \frac{1}{1 - 2\phi \cos 2\pi\nu + \phi^2}. \tag{1.24}$$

We will use the causal AR(1) process for the most of the stationary noise models. One of its advantages is that its serial correlation depends on the single parameter ϕ.

In addition, it has been used in many models of natural phenomena: the wind speed fluctuations [7], radar signals [5], climatic phenomena variability [11], electroencephalographic activity [10], heart interbeat time series [6], the daily temperature fluctuations [8], X-ray emission from the active galactic nuclei [14], sunspots variability [12], etc. As a result of their simple mathematical properties and their direct physical interpretation, the realizations of AR(1) processes have been used as artificial time series to analyze some numerical algorithms for monotonic trend removal [15], some surrogate data test for nonlinearity [9] or for renormalization group analysis [2].

In the following we briefly discuss the solution of Eq. (1.19) for other values of the parameter ϕ. In the nonstationary case $|\phi| = 1$, all the terms in Eq. (1.20) have unit coefficients and for any delay h none of them can be neglected. Every random variable Z_n is an infinite sum of terms with the same norm, hence its norm is infinite and Eq. (1.19) has no stationary solution. For $\phi = 1$ the AR(1) process has the same recursive formula as the random walk (1.3), but while the former has no initial term, the latter begins with Z_0.

When $|\phi| > 1$, Eq. (1.21) is not true any more because in this case ϕ^h increases when h increases. However, the same reasoning as in the case $|\phi| < 1$ can be repeated if we write Eq. (1.19) in the form

$$Z_{n-1} = \phi^{-1} Z_n - \phi^{-1} W_n.$$

In this way we obtain a stationary causal AR(1) process because $|\phi^{-1}| < 1$, but in the reverse temporal direction and with the variance of the noise equal to $\phi^{-1}\sigma$. It means that the formulas can be obtained from those for the causal ones by replacing ϕ with ϕ^{-1} and σ with $\phi^{-1}\sigma$. Such a stochastic process is called *acausal*.

The values of the parameter ϕ can be both positive and negative. In order to qualitatively characterize the difference between the two situations we use the fact that for $\phi = 0$, the AR(1) process reduces to a white noise with uncorrelated terms. When $\phi > 0$, from Eq. (1.19) it follows that the fluctuations due to the white noise are superposed over the term ϕZ_{n-1} which memorizes a part of the previous value of the time series. Hence, for large ϕ the successive values of the time series are close to each other and the fluctuations due to the white noise are small. Therefore, in comparison with a realization of a white noise, for $\phi > 0$ the graphical representation of an AR(1) process is less fluctuant and resembles to a deterministic trajectory disturbed by a random fluctuation (Fig. 1.1c).

When $\phi < 0$ the white noise is superposed over the term $-|\phi| X_{n-1}$ which has an opposite sign to the previous term of the time series. Consequently, the white noise fluctuations are enhanced and the series values fluctuate more rapidly than the white noise, as shown in Fig. 1.1a. The successive values of the autocovariance function (1.23) are of opposite signs and the time series is called anticorrelated.

In practice the time series have a finite length and usually they are considered realizations of a finite part of a stochastic process of infinite length. For an AR(1) process, the first term of the time series is correlated with the preceding term which has not been recorded. But the first term of a numerically generated time series

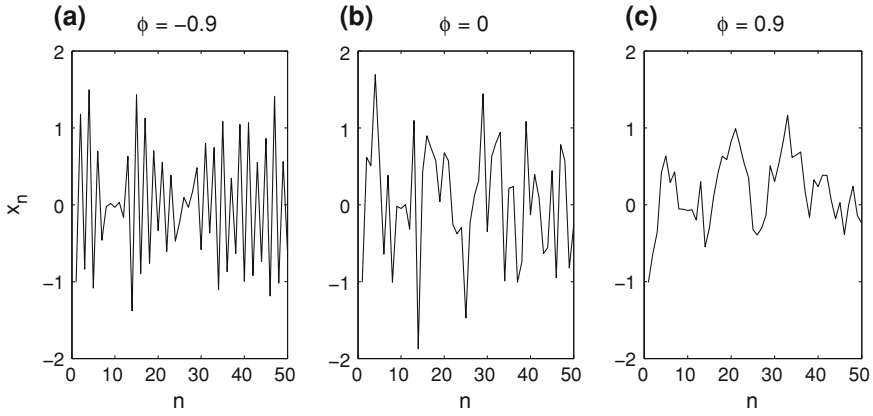

Fig. 1.1 Realizations of an anticorrelated (**a**), uncorrelated (**b**), and correlated (**c**) AR(1) process

cannot be related to realizations of other preceding random variables. Therefore, a numerically generated time series is never strictly a realization of a finite part of an ideal stationary stochastic process of infinite length. Since Eq. (1.19) defining the AR(1) process is recursive, the first term must be defined by an additional relation. As we show in the following, the manner in which this additional relation is chosen can essentially modify the properties of the stochastic process.

We call *finite AR(1) process* a stochastic process of finite length satisfying the recursive relation in Eq. (1.19). We consider only the case when the white noise $\{W_n\}$ is Gaussian, otherwise one can obtain only asymptotic distributions of the AR(1) process. Let us denote the finite AR(1) process by $\{\widehat{Z}_n\}$, $n = 1, 2, \ldots, N$. Because $\widehat{Z}_n$ satisfies Eq. (1.19) for $n > 1$, by successive applications of this relation we can express the terms of the stochastic process as a finite sum

$$\widehat{Z}_n = W_n + \phi W_{n-1} + \cdots + \phi^{n-2} W_2 + \phi^{n-1} \widehat{Z}_1. \tag{1.25}$$

In the following we consider only the causal AR(1) processes with $|\phi| < 1$. As shown above, the acausal process is equivalent with a causal one generated in reverse temporal order.

If $\widehat{Z}_1$ is a Gaussian random variable with variance $\widehat{\sigma}_1$ and zero mean, then from Eq. (1.25) it follows that $\widehat{Z}_n$ is the sum of n Gaussian random variables, hence it has also a Gaussian distribution with variance

$$\widehat{\sigma}_n^2 = \sigma^2 \left(1 + \phi^2 + \phi^4 + \cdots + \phi^{2(n-2)}\right) + \widehat{\sigma}_1^2 \phi^{2(n-1)}.$$

Applying the formula for the sum of a geometric series we have

$$\widehat{\sigma}_n^2 = \sigma_s^2 + \left(\widehat{\sigma}_1^2 - \sigma_s^2\right) \phi^{2(n-1)}, \tag{1.26}$$

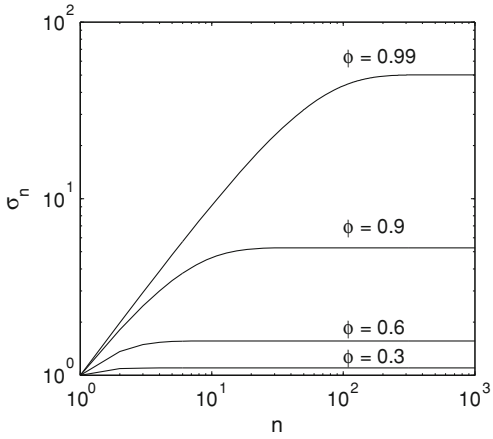

Fig. 1.2 The standard deviation of a finite AR(1) process for different values of ϕ when the first term coincides with the white noise $\widehat{Z}_1 = W_1$

where we have used Eq. (1.22). The variance of the finite AR(1) process has a constant term σ_s^2 equal with the variance of the infinite AR(1) process and a variable term which tends asymptotically to zero because $|\phi| < 1$. Hence the finite AR(1) process is nonstationary presenting transient effects, i.e., its variance approximates the theoretical one $\widehat{\sigma}_n \simeq \sigma_s$ only after a time interval t_0 for which ϕ^{2t_0} can be neglected.

For $\widehat{\sigma}_1 = \sigma_s$ the variable term in Eq. (1.26) vanishes and $\widehat{\sigma}_n = \sigma_s$ for all $n \leq N$. Hence, if for a finite AR(1) process we choose $\widehat{Z}_1 = (\sigma_s/\sigma)W_1$, then all the terms have the same variance. This choice is natural because it is more reasonable to take the first term of the finite AR(1) process similar to the stationary infinite AR(1) process and not to the white noise. In the following we show that for $\widehat{\sigma}_1 = \sigma_s$ the properties of the finite AR(1) process are identical to those of a finite sample of a stationary infinite AR(1) process.

The autocovariance function $\widehat{\gamma}(h, n) = \langle \widehat{Z}_n \widehat{Z}_{n-h} \rangle$ can be calculated only if $|h| < N$ and $h < n \leq N + h$. Unlike the autocovariance function (1.23) the quantity $\widehat{\gamma}(h, n)$ depends also on n since it exists only for certain values of n. Therefore $\{\widehat{Z}_n\}$ is not a stationary stochastic process in a strict mathematical meaning. However, when it exists, we can show with the same method as for Eq. (1.23) that $\langle \widehat{Z}_n \widehat{Z}_{n-h} \rangle = \phi \langle \widehat{Z}_{n-1} \widehat{Z}_{n-h} \rangle$. Then instead of Eq. (1.23) we obtain

$$\widehat{\gamma}(h, n) = \phi^h \widehat{\sigma}_{n-h}^2. \tag{1.27}$$

If we choose $\widehat{\sigma}_1 = \sigma_s$, then $\widehat{\sigma}_{n-h} = \sigma_s$ is constant and when $\widehat{\gamma}(h, n)$ exists it is identical to the covariance function $\gamma(h)$ in Eq. (1.23). Hence, if we want to numerically model a stationary infinite AR(1) process, then we have to use a finite AR(1) process $\{\widehat{Z}_n\}$ with $\widehat{\sigma}_1 = \sigma_s$. A more detailed analysis of the properties of the finite AR(1) process is available in [16].

The finite AR(1) process satisfying Eq. (1.19) for $\phi = 1$ and $\widehat{Z}_1 = W_1$ becomes the Gaussian random walk (see Sect. 1.1). From Eq. (1.4) the sample variance is equal to

$$\widehat{\sigma}_n^2 = \gamma(n, n) = n\sigma^2.$$

The relation between the Gaussian random walk and the quasistationary finite AR(1) process can be clarified if in Eq. (1.26) we take $\widehat{\sigma}_1 = \sigma$ corresponding to the choice of the first term for the Gaussian random walk $\widehat{Z}_1 = W_1$

$$\widehat{\sigma}_n^2 = \sigma_s^2 \left(1 - \phi^{2n}\right).$$

Figure 1.2 shows the variation of $\widehat{\sigma}_n^2$ for $\sigma = 1$ and different values of ϕ. For a given ϕ, at the beginning there is a nonstationary transient period before the stationary state of the AR(1) process is reached. As ϕ tends to 1, the transient region is expanded and at the limit it becomes infinite, such that for $\phi = 1$ an entirely nonstationary process is obtained. Hence the Gaussian random walk corresponds to the transient region of the finite AR(1) process extended to the infinity, whereas the stationary infinite AR(1) process corresponds to the stationary part of the graph. Therefore, to obtain a quasistationary finite AR(1) process for ϕ close to 1, the only possibility is to choose $\widehat{\sigma}_1 = \sigma_s$ completely eliminating in this way the transient region.

References

1. Alexandrov, T., Bianconcini, S., Dagum, E.B., Maass, P., McElroy, T. S.: A review of some modern approaches to the problem of trend extraction. Research Report Series US Census Bureau, Statistics 3. http://www.census.gov/srd/papers/pdf/rrs2008-03.pdf (2008)
2. Blender, R.: Renormalisation group analysis of autoregressive processes and fractional noise. Phys. Rev. E **64**, 067101 (2001)
3. Brockwell, P.J., Davies, R.A.: Introduction to Time Series and Forecasting. Springer, New York (1996)
4. Brockwell, P.J., Davies, R.A.: Time Series:Theory and Methods. Springer, New York (1996)
5. Gao, J., Hu, J., Tung, W., Cao, Y., Sarshar, N., Roychowdhury, V.P.: Assessment of long range correlation in time series: How to avoid pitfalls. Phys. Rev. E **73**, 016117 (2006)
6. Guzman-Vargas, L., Angulo-Brown, F.: Simple model of the aging effect in heart interbeat time series. Phys. Rev. E **67**, 052901 (2003)
7. Hallerberg, S., Altmann, E.G., Holstein, D., Kantz, H.: Precursors of extreme increments. Phys. Rev. E **75**, 016706 (2007)
8. Király, A., Jánosi, I.M.: Stochastic modeling of daily temperature fluctuations. Phys. Rev. E **65**, 051102 (2002)
9. Kugiumtzis, D.: Statically transformed autoregressive process and surrogate data test for nonlinearity. Phys. Rev. E **66**, 025201 (2002)
10. Liley, D.T., Cadusch, P.J., Gray, M., Nathan, P.J.: Drug-induced modification of the system properties associated with spontaneous human electroencephalographic activity. Phys. Rev. E **68**, 051906 (2003)
11. Maraun, D., Rust, H.W., Timmer, J.: Tempting long-memory—on the interpretation of DFA results. Nonlinear Proc. Geoph. **11**, 495–503 (2004)

12. Palus, M., Novotna, D.: Sunspot cycle: a driven nonlinear oscillator? Phys. Rev. Lett. **83**, 3406–3409 (1999)
13. Stephen, D., Pollock, G.: Statistical signal extraction and filtering: a partial survey. In: Belsley, D.A., Kontoghiorghes, E. (eds.) Handbook of Computational Econometrics, pp. 321–376. Wiley, New York (2009).
14. Timmer, J., Schwarz, U., Voss, H.U., Wardinski, I., Belloni, T., Hasinger, G., van der Klis, M., Kurths, J.: Linear and nonlinear time series analysis of the black hole candidate Cygnus X-1. Phys. Rev. E **61**, 1342–1352 (2000)
15. Vamoş, C.: Automatic algorithm for monotone trend removal. Phys. Rev. E **75**, 036705 (2007)
16. Vamoş, C., Şoltuz, Ş.M., Crăciun, M.: Order 1 autoregressive process of finite length. Rev. Anal. Numer. Theor. **36**, 199–214 (2007)
17. Wu, Z., Huang, N.E., Long, S.R., Peng, C.K.: On the trend, detrending, and variability of nonlinear and nonstationary time series. PNAS **18**, 14889–14894 (2007)
18. Wentzell, A.D.: A Course in the Theory of Stochastic Processes. McGraw-Hill, New York (1981)

Chapter 2
Monte Carlo Experiments

In this chapter we design a numerical algorithm to generate nonmonotonic trends with a diversity of shapes comparable to those encountered in practice. This original algorithm is essential for all the rest of the book because it provides the numerical trends on which the estimation methods are tested. Over these trends finite AR(1) noises (see Sect. 1.3) are superposed so that the resulting artificial time series depend on five independent parameters. In the case of the trend estimation algorithms the complexity of the problem is reduced because the accuracy of the estimated trend significantly depends only on three parameters: the time series length, the noise serial correlation, and the ratio between the amplitudes of the trend variations and noise fluctuations. Using Monte Carlo experiments we derive the accuracy of a simple method to estimate the serial correlation of an AR(1) noise.

2.1 Monte Carlo Statistical Ensembles

The Monte Carlo method does not have a rigorous and exhaustive definition. Even in the original paper where it was firstly presented as a general numerical method, instead of a definition, Metropolis and Ulam gave only a few examples [9]. In terms of statistical physics, a Monte Carlo experiment provides an approximation of the statistical ensemble associated to the phenomenon under investigation. For example, the *statistical ensemble* associated to a macroscopic state of a thermodynamic system is the set of all microscopic states which are compatible with the considered macroscopic state and form a multidimensional volume in phase space. By Monte Carlo simulations one intends to obtain a finite number of microscopic states randomly distributed in phase space which describe the volume associated to the statistical ensemble as accurately as possible.

From the point of view of the theoretical statistics, the Monte Carlo methods are computational algorithms that perform repeated random sampling. Generally by sampling we mean the selection of individual observations of a sample intended to

C. Vamoş and M. Crăciun, *Automatic Trend Estimation*, SpringerBriefs in Physics,
DOI: 10.1007/978-94-007-4825-5_2, © The Author(s) 2012

yield some knowledge about a statistical population. In the case of the thermodynamic system discussed above the population from which the selection is made is the statistical ensemble and the individuals are the microscopic states. Usually each individual is chosen randomly such that each individual has the same probability of being chosen at any stage during the sampling process and each subset of k individuals has the same probability of being chosen for the sample as any other subset of k individuals. This technique is known as simple random sampling and it is only one of the probability sampling schemes.

First a *Monte Carlo experiment* defines a statistical ensemble (population in statistical language). Then a numerical method is used to generate several microscopic states (individual representatives) from the statistical ensemble (population). Finally the results of the individual computations are aggregated into the final result by a statistical analysis procedure. Usually the Monte Carlo statistical ensemble is defined not theoretically, but by an effective numerical generation algorithm of the microscopic states. By repeating the generation algorithm several times, we obtain a sample on which we do statistical analysis.

As an example, we evaluate the accuracy of a simple method to estimate the correlation parameter ϕ of a finite AR(1) noise $\{z_n\}$. From Eqs. (1.27) and (1.9) we obtain for the sample autocorrelation function the expression $\widehat{\rho}(h) = \phi^h$. It follows that for $h = 1$ the sample autocorrelation function is a statistical estimator $\widehat{\phi}_\rho = \widehat{\rho}(1)$ for the correlation parameter of a finite AR(1) noise. Using a Monte Carlo experiment we can determine the properties of this estimator without deriving them by a complicated theoretical analysis of its asymptotic properties.

We consider a statistical ensemble composed of S finite AR(1) time series $\{z_n^{(s)}\}$, $1 \leq s \leq S$, numerically generated by means of the algorithm described in Sect. 1.3. Figure 2.1a shows the histogram of the values obtained for the estimated values $\widehat{\phi}_\rho^{(s)}$ using $S = 2000$ time series with length $N = 10000$ and the correlation parameter $\phi = 0.7$. The estimated values are symmetrically distributed around the real value and for a given $\alpha < 1$ we characterize their spread by the interquantile interval $I_\alpha(\widehat{\phi}_\rho) = (Q_{\alpha/2}(\widehat{\phi}_\rho), Q_{1-\alpha/2}(\widehat{\phi}_\rho))$. The probability that $\widehat{\phi}_\rho^{(s)} \in I_\alpha(\widehat{\phi}_\rho)$ is equal with $1 - \alpha$.

In Fig. 2.1b we present for different values of ϕ and N the limits of the interquantile intervals with $\alpha = 0.05$ from which the real value of ϕ is subtracted. These quantities characterize the *sampling errors* of the estimation method. The errors decrease when ϕ increases, i.e., the higher the serial correlation, the more accurate its estimation is. It can be numerically checked that the standard deviation of the errors is inversely proportional to $\sqrt{N}$, i.e., the longer the series, the smaller the errors are. For short time series ($N = 100$) and large values of ϕ the interquantile interval becomes asymmetric, its upper limit being closer to the real value of ϕ. This behavior occurs because always $\widehat{\phi}_\rho < 1$ and when ϕ is close to 1, the maximum upper error is limited by the value $1 - \phi$. For long time series ($N \geq 1000$) the sampling errors are less than 0.1 for all values of ϕ and for short times series ($N = 100$) it is two times larger.

In applications we do not know the value of the parameter that we estimate and we cannot use the sampling error to characterize the accuracy of the estimation. In this

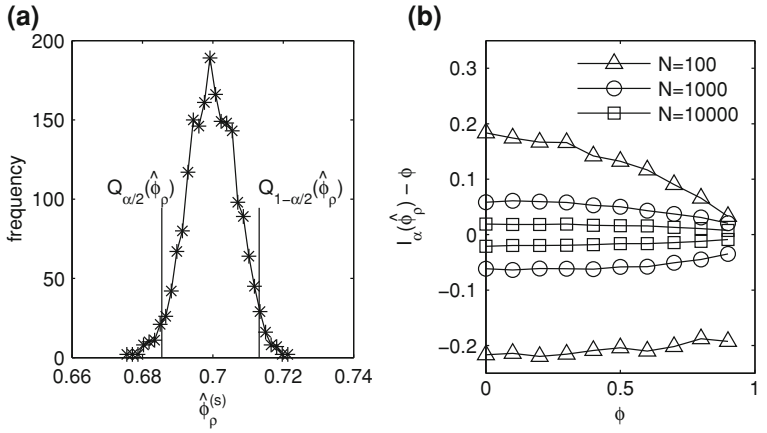

Fig. 2.1 **a** The histogram of the estimated values $\widehat{\phi}_\rho^{(s)}$ for finite AR(1) noises with $\phi = 0.7$ and $N = 10000$ and the interquantile interval for $\alpha = 0.05$; **b** the limits of the interquantile intervals with $\alpha = 0.05$ from which the real value of ϕ is subtracted

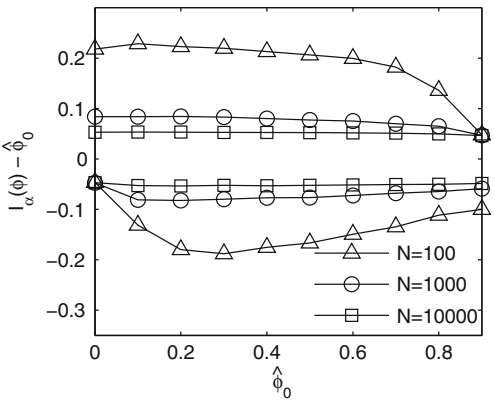

Fig. 2.2 The limits of the 95 % confidence intervals of the estimation $\widehat{\phi}_\rho$ from which the center $\widehat{\phi}_0$ of the variation range of the real values $\phi^{(s)}$ is subtracted

case we use the spreading of the possible values of ϕ for which we obtain a certain estimated value $\widehat{\phi}_\rho$. We generate a statistical ensemble of finite AR(1) time series with the values of real $\phi^{(s)}$, $1 \le s \le S$, uniformly distributed in a given interval $[\phi_{\min}, \phi_{\max}]$. For each time series we compute the estimated value $\widehat{\phi}_\rho^{(s)}$ and then we consider the time series for which $\widehat{\phi}_\rho^{(s)}$ belongs to an interval $[\widehat{\phi}_0 - \Delta\widehat{\phi}, \widehat{\phi}_0 + \Delta\widehat{\phi}]$ with given $\widehat{\phi}_0$ and $\Delta\widehat{\phi}$. The interquantile interval $I_\alpha(\phi)$ of the distribution of $\phi^{(s)}$ is called *confidence interval* and the number $1 - \alpha$ (or as a percentage $100\,\%(1 - \alpha)$) is the *confidence level*.

The 95 % confidence interval of the same estimator as that in Fig. 2.1 is presented in Fig. 2.2. The statistical ensemble contains 100000 finite AR(1) time series with $\phi^{(s)} \in [-0.05, 0.95]$. The estimated values $\widehat{\phi}_\rho^{(s)}$ are distributed in bins with the centers at $\widehat{\phi}_0 = 0, 0.1, \ldots, 0.9$ and with $\Delta\widehat{\phi} = 0.05$. In comparison with Fig. 2.1b

the independent variable is not the real value ϕ, but $\widehat{\phi}_0$, the center of the intervals containing the estimated values $\widehat{\phi}_\rho$. For long time series ($N = 10000$) the confidence intervals coincide with the bins $[\widehat{\phi}_0 - 0.05, \widehat{\phi}_0 + 0.05]$, hence if $\Delta\phi$ would be decreased, then the confidence intervals would become smaller. When $N = 1000$, the confidence intervals are slightly larger than the bins. For $N = 100$ the confidence intervals double and become slightly asymmetric because the values of $\phi^{(s)}$ are limited by $\phi_{\min} = -0.05$ and $\phi_{\max} = 0.95$.

2.2 Numerical Generation of Trends

Usually the trend estimation methods are evaluated by Monte Carlo experiments only for time series with monotonic trends (linear, power–law, exponential, and logarithmic) or periodic ones (sinusoidal). Such trends have a fixed functional form depending on several parameters. In the following chapters we test the trend estimation methods on numerically generated time series and we need artificial trends with shapes having a variability comparable with the trends of the real time series.

The numerical method to generate artificial trends should ensure effective control over their main features, as for example, the number and length of the monotonic segments of the artificial trend. Another important property of a trend is the amplitude of its variations over the monotonic segments, i.e., the difference between successive local extrema. The statistical ensemble of the generated trends should be homogeneous, the average distribution of the trends features should not vary with the location. If we need a nonhomogeneous ensemble of trends, then the trends can be easily rescaled to meet such a requirement.

The apparent choice of a complex nonmonotonic trend with significant variability is a high order polynomial trend with a large number of coefficients. If we choose the coefficients by means of a random algorithm, then the form of the generated trend is difficult to be controlled. Usually the generated trend has only a few parts with significant monotonic variation.

The number and length of monotonic segments of a polynomial trend can be manipulated by means of the position of its roots. Let us consider that we want to generate a polynomial trend with P monotonic parts. We select P distinct roots a_p, $p = 1, 2, \ldots, P$, with a uniform distribution within the definition interval $t \in (0, 1)$. The trend given by the polynomial

$$f(t) = \prod_{p=1}^{P} (t - a_p)$$

has $P - 1$ distinct local extrema. We have to choose the roots a_p not too close to each other. Suppose that we want to generate a time series with N values. Then we exclude the possibility that the interval between two successive roots has the length comparable to the time step $\Delta t = 1/N$, otherwise it would contain only a few

time steps. Therefore we impose the condition that each monotonic segment should contain more than a given number $\Delta N_{\min}$ of time steps.

In the following we describe a numerical algorithm that divides the interval (0, 1) into P subintervals satisfying the conditions previously imposed to the roots of the polynomial trend. The length of the subinterval p is chosen as a random number d_p uniformly distributed within the interval $[d_{\min}, 1]$, where $d_{\min}$ is a parameter that will be determined later. The union of all these subintervals is an interval with length $d = \sum_{p=1}^{P} d_p$ and each subinterval is divided by d such that the entire interval is rescaled to (0, 1). The interval (0, 1) is divided into $N - 1$ equal bins corresponding to the N values of the time series. Hence the superior limit of the subinterval p is equal to

$$N_p = 1 + \left[(N-1) d^{-1} \sum_{i=1}^{p} d_i \right], \tag{2.1}$$

where $[\cdot]$ is the integer part function. The subinterval p contains the time steps satisfying the condition $N_{p-1} < n \leq N_p$, where $N_0 = 0$. The number of time steps of any subinterval is equal with $\Delta N_p = N_p - N_{p-1}$.

We have to set the value of the parameter $d_{\min}$. If in Eq. (2.1) we do not take into account the integer part function, then the number of time steps of the segment p is approximately equal to $\Delta N_p \simeq d_p (N-1)/d$. The minimum of this quantity is obtained when $d_p = d_{\min}$ and $d_{p'} = 1$ for $p' \neq p$ and is equal to

$$\min(\Delta N_p) = \frac{d_{\min}(N-1)}{P - 1 + d_{\min}}.$$

Imposing the condition that $\min(\Delta N_p)$ equals the given minimum length of the monotonic segments $\Delta N_{\min}$, it follows

$$d_{\min} = \frac{(P-1)\Delta N_{\min}}{N - 1 - \Delta N_{\min}}.$$

Figure 2.3 contains some polynomial trends with the roots generated by the algorithm described above for $P = 10$ monotonic parts, $N = 1000$ values and $\Delta N_{\min} = 10$. The nonuniform distribution of the trend values is obvious. The variability of the trends near the boundaries is with almost two orders of magnitude greater than in the middle of the definition interval. The same nonuniform distribution is obtained if we choose the position of the local extrema of the polynomial trend, not of its roots. In general, all algorithms to generate polynomial trends suffer from this drawback which originates in the so-called "Runge phenomenon" [4].

It is possible to correct this problem if we introduce a mechanism to control the amplitude of monotonic variations. Such a solution is based on constructing monotonic semi-periods of sinusoid with random amplitude [11]. They are joined together such that the trend is continuous. The amplitude of the sinusoid associated

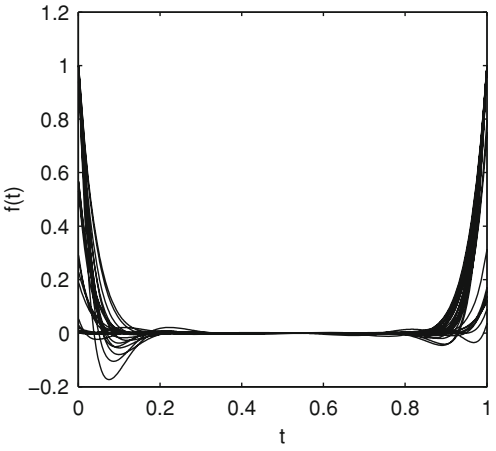

Fig. 2.3 Numerically generated polynomial trends with the roots uniformly distributed in the definition interval

with the subinterval p is a random number $A_p \in [0, 1]$ with a uniform probability distribution.

The sinusoid semi-periods are located in the definition interval [0, 1] by means of the numbers N_p generated by the algorithm previously used to set the polynomial roots. The value of the trend at a point n of the segment p, $N_p < n \leq N_{p+1}$, is given by the recurrence relation

$$f_n = f_{N_p} + c_p A_p \left[1 - \sin \frac{\pi}{2} \left(1 + 2\frac{n - N_p}{\Delta N_p}\right)\right] \tag{2.2}$$

where the term f_{N_p} is the last value of the previous segment $p - 1$ and for $p = 1$ we choose $f_0 = 0$. The coefficients c_p describe how the successive sinusoids are connected together. When $c_p = (-1)^p$, the first sinusoidal part is decreasing and the monotony of the other parts alternates such that the trend looks like a distorted sinusoid. Examples of such trends are given in Figs. 2.5, 3.2, and 4.5. If the values $c_p = \pm 1$ are randomly assigned, then the trend shape variability increases because it is possible to have several successive sinusoids with variation of the same sign (see Figs. 2.6, 6.2, 6.4, 7.5). Finally the mean of the series $\{f_n\}$ is removed from it.

The algorithm described above is characterized by three parameters: the length of the time series N, the number of monotonic segments P, and the minimum number of points in a monotonic segment $\Delta N_{\min}$. Figure 2.4 shows some numerically generated trends with $P = 10$, $N = 1000$, $\Delta N_{\min} = 10$ and the coefficient c_p alternately chosen. Unlike polynomial trends in Fig. 2.3, these trends have a large variability homogeneously distributed throughout the entire definition interval.

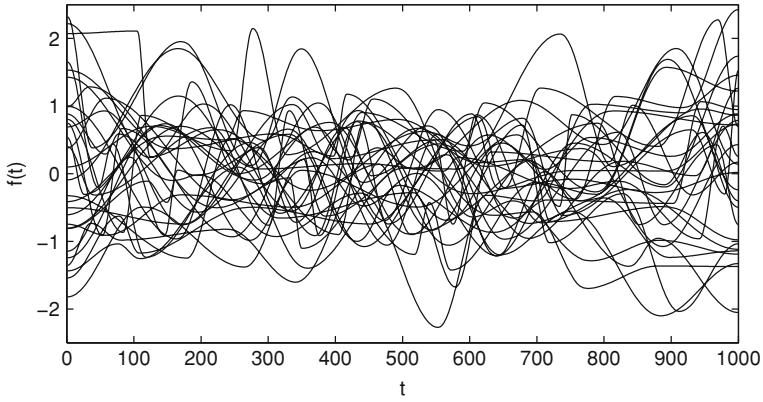

Fig. 2.4 Numerically generated trends as an alternate succession of random sinusoidal semi-periods

2.3 Numerical Generation of Noisy Time Series

In this section we present a numerical algorithm to generate a time series $\{x_n\}$ by superposing an AR(1) noise $\{z_n\}$ over a nonmonotonic trend $\{f_n\}$ generated by the algorithm described in the previous section. This algorithm is almost identical to that presented in [11] with the only difference that here we rescale the noise, not the trend. As discussed above, the trend $\{f_n\}$ is characterized by three parameters: N, $\Delta N_{\min}$, and P. The AR(1) noise is obtained by means of the numerical method described in Sect. 1.3 and has two parameters: ϕ for the serial correlation and σ_s^2 for its variance. We need an additional parameter r to characterize the ratio between the amplitudes of the trend variations and of the noise fluctuations defined as

$$r = \frac{\max\{f_n\} - \min\{f_n\}}{\max\{z_n\} - \min\{z_n\}}, \tag{2.3}$$

which is a more sensitive measurement of the classical signal to noise ratio. Because the noise variance depends on the new parameter, the noisy time series $\{x_n\}$ has five independent parameters: N, $\Delta N_{\min}$, P, ϕ, and r.

To obtain a time series characterized by these five parameters we generate a trend $\{f_n\}$ characterized by N, $\Delta N_{\min}$, and P and an AR(1) noise $\{z'_n\}$ with $\sigma_s = 1$ and a given ϕ. Then we rescale the noise according to the relation

$$z_n = \frac{z'_n}{r} \frac{\max\{f_n\} - \min\{f_n\}}{\max\{z'_n\} - \min\{z'_n\}}.$$

When $r > 1$ the time series $x_n = f_n + z_n$ is dominated by trend and when $r < 1$ by noise. Depending on the aim of the numerical experiment we generate time series

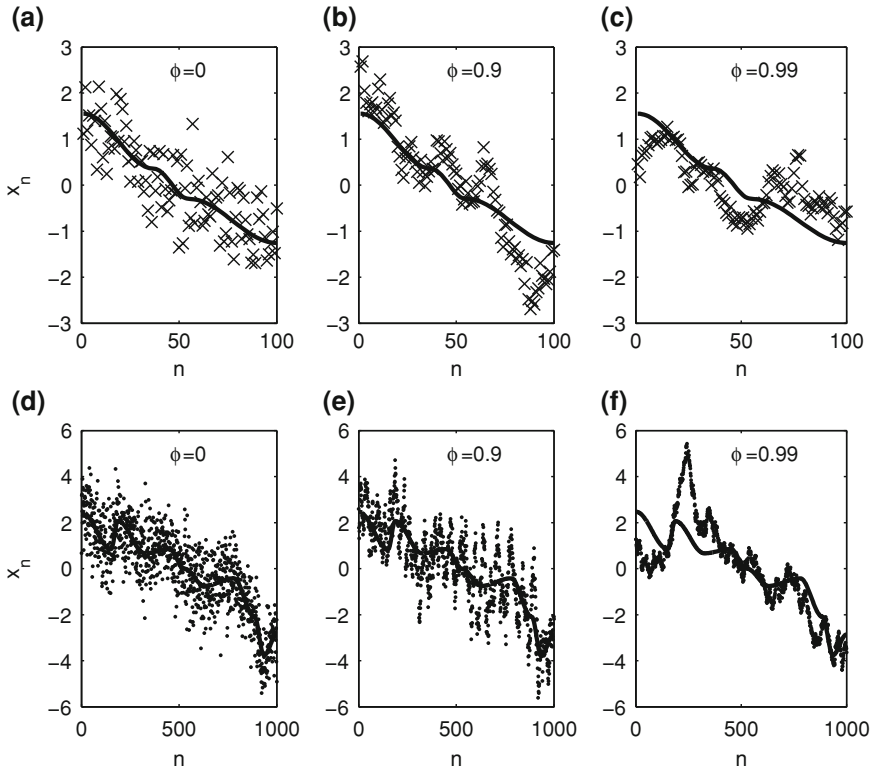

Fig. 2.5 Numerically generated time series for different values of N and ϕ with $r = 1$ and $\Delta N_{\min} = 10$

choosing different values for the five parameters. In the following we give some indication concerning the typical values of these parameters.

In order to obtain a description of each monotonic segment with an acceptable resolution we choose $\Delta N_{\min} > 10$, i.e., an order of magnitude greater than the time step. The value of $\Delta N_{\min}$ is related to the length of the time series and the number of monotonic parts through the relation $\Delta N_{\min} < N/P$. This condition also limits the values of N and P because for a given length N there is a maximum number of monotonic parts given by $P < N/\Delta N_{\min}$.

We impose the superior limit of ϕ to 0.9 because the AR(1) process with ϕ closer to unit has a special behavior similar to the Gaussian random walk (see Sects. 1.1 and 1.3), which must be analyzed with special methods [5]. In addition we consider for ϕ only positive values because few of the phenomena of interest are characterized by an anticorrelated noise. Hence the maximum range for the serial correlation parameter is $\phi \in [0, 0.9]$. When we choose the value of ϕ we have to take also into account the length of the time series. In Fig. 2.5 we present several noisy time series with $N = 100$ and $N = 1000$, different values of ϕ, and $r = 1$. For $\phi = 0.9$ and $N = 100$

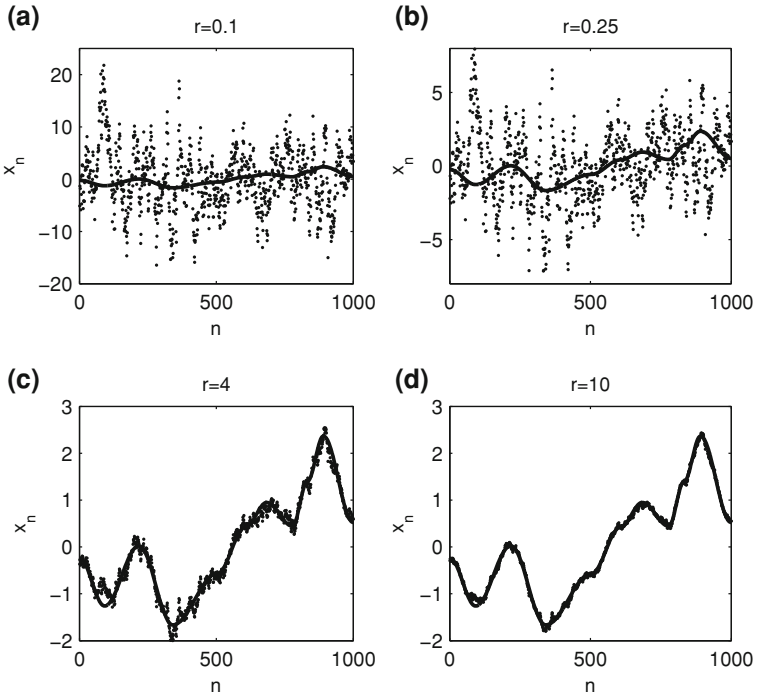

Fig. 2.6 Time series generated by superposing an AR(1) noise with $\phi = 0.9$ over the same trend but with different ratios r

a stochastic trend appears (Fig. 2.5b), i.e., the strongly correlated noise has variations that cannot be distinguished from those of the trend (see Sect. 1.1). For $N = 1000$ the time series with the same value $\phi = 0.9$ contains many noise fluctuations and the noise does not confound with the trend (Fig. 2.5e). But if we increase the serial correlation to $\phi = 0.99$, the noise variations are enhanced and the stochastic trend occurs even for $N = 1000$ (Fig. 2.5f).

Since for $r = 0$ the trend vanishes, we eliminate the small values of r. We choose the minimum value for r equal to 0.25 because, as shown in Fig. 2.6b, in this situation the shape of the total signal allows us to assume the presence of a trend even for strong serial correlation ($\phi = 0.9$). For smaller values of r the shape of trend is not observable (Fig. 2.6a). The maximum value of r that we use in our numerical simulations is 4 and corresponds to the time series in which the noise still has a large enough amplitude to allow the estimation of its parameters (Fig. 2.6c). For higher values of r the noise becomes negligible as in Fig. 2.6d. Hence the interval chosen for the variation of r is $r \in [0.25, 4]$.

Besides the previous criteria, we have chosen the variation ranges of the parameters of the artificial time series such that they also include the extreme situations for which the numerical algorithms fail. The most difficult problems occur for the short

time series and, in general, we have limited their length to $N = 100$ values. If such a short time series is characterized by the inferior limit of the variation range of the ratio $r = r_{\min} = 0.25$, i.e., it is strongly dominated by noise, then the trend estimated by any algorithm has large errors. Reversely, if the short time series is strongly dominated by trend ($r = r_{\max} = 4$), then the noise parameters are estimated with very large errors. If, in addition, the noise is strongly correlated ($\phi = \phi_{\min} = 0.9$), then the stochastic trend prevents any possibility to correctly separate the trend from the superposed noise.

For most of the numerical simulations in the next chapters we limit ourselves to AR(1) processes. Obviously, it is possible to choose other types of noises but it would mean an increase of the number of parameters describing the statistical ensemble. In order to preserve the presentation clarity we use noise models as simple as possible, retaining in the same time their essential characteristic, namely, the serial correlation. As shown in Fig. 2.5, the parameter ϕ controls the occurrence of the stochastic trend which is a major difficulty in an accurate trend estimation. The analysis of the noise models with more parameters is similar to that of the AR(1) noise, although more elaborate.

As a simple application of the algorithm presented above we analyze a method to estimate the parameter ϕ of a finite AR(1) noise. Unlike the method discussed in Sect. 2.1, this method can be applied to time series containing a trend. It allows a first guess of the serial correlation parameter without effectively detrending the time series. In some cases this estimation is not very accurate, but it can indicate to which category the time series belongs, for example, if it is strongly or weakly correlated. Such an approximate classification allows an adjustment of a more complex numerical algorithm to the characteristics of a particular time series.

If the approximation (1.18) is valid, then we can obtain information on the noise properties analyzing the differenced stochastic process $\{\nabla_d X_n\}$. From Eqs. (1.16) and (1.1), for a stationary noise with zero mean we have

$$\langle(\nabla_d Z_n)^2\rangle = \langle Z_{n+d}^2\rangle - 2\langle Z_{n+d} Z_n\rangle + \langle Z_n^2\rangle = 2\langle Z_n^2\rangle - 2\gamma(d).$$

Substituting Eq. (1.23) in the previous relation, we obtain for an AR(1) process the relation

$$\langle(\nabla_d Z_n)^2\rangle = 2\sigma_s^2(1 - \phi^d).$$

Because $\{Z_n\}$ is stationary, this relation holds for all indexes n. Applying it for $d = 1$ and $d = 2$ it results a system of equations from which the parameter ϕ can be computed

$$\phi = \frac{\langle(\nabla_2 Z_n)^2\rangle}{\langle(\nabla_1 Z_n)^2\rangle} - 1. \tag{2.4}$$

Since $\{Z_n\}$ is a stationary process, the ensemble average of one of its terms $\langle Z_n\rangle$ can be estimated by the temporal mean (1.7) of one of its realization $\{z_n\}$. For any n, the term $\langle(\nabla_d Z_n)^2\rangle$ is replaced by $(N - d)^{-1} \|\nabla_d z_n\|^2$ where

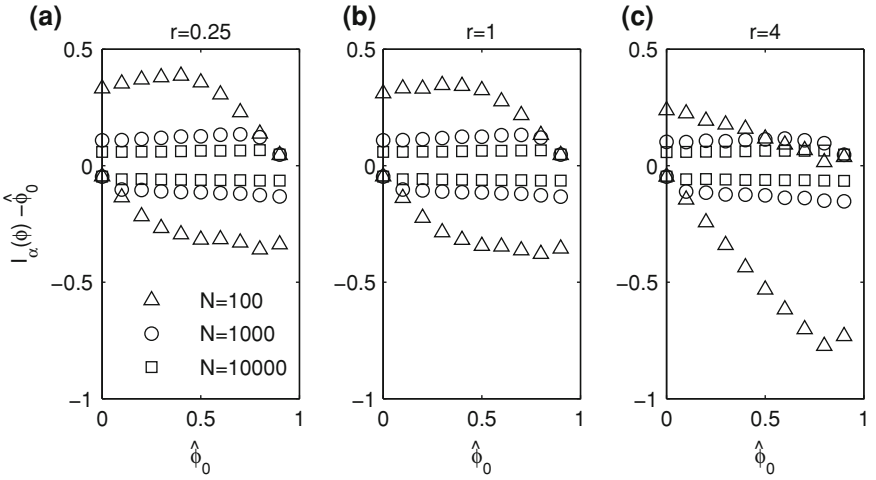

Fig. 2.7 The limits of the 95 % confidence intervals of the estimation $\widehat{\phi}_x$ from which the center $\widehat{\phi}_0$ of the variation range of the real values $\phi^{(s)}$ is subtracted

$$\|\nabla_d z_n\|^2 = \sum_{n=1}^{N-d} (\nabla_d z_n)^2 \tag{2.5}$$

is the usual quadratic norm of finite sequences. We use the approximation (1.18)

$$\nabla_d x_n \approx \nabla_d z_n,$$

written for a realization of the stochastic process $\{X_n\}$ which for small d is sufficiently accurate. Then the estimation (2.4) of ϕ using only the values of the time series $\{x_n\}$ is

$$\widehat{\phi}_x = \frac{N-1}{N-2} \frac{\|\nabla_2 x_n\|^2}{\|\nabla_1 x_n\|^2} - 1. \tag{2.6}$$

From this equation it is possible to result $\widehat{\phi}_x > 1$ corresponding to an acausal AR(1) noise (see Sect. 1.3). Such situations generally occur when ϕ is near 1 and then we correct the estimation of the serial correlation parameter by the maximum value of the parameter used to generate the time series $\widehat{\phi}_x = \phi_{\max} = 0.95$. Another correction is made when from the estimation we obtain $\widehat{\phi}_x < 0$, then we make the substitution $\widehat{\phi}_x = 0$.

In order to determine the confidence intervals of the estimation (2.6) we use Monte Carlo experiments. The results presented in Fig. 2.7 are obtained for the same AR(1) noises as those in Fig. 2.2, but superposed over numerically generated trends. Whatever the value of the ratio r, for long time series ($N \geq 1000$), the accuracy of the estimation $\widehat{\phi}_x$ is almost the same, showing that for such time series differencing allows the removal of the trend influence. The confidence intervals are symmetric with

respect to estimated values and approximately twice larger than those determined for the estimation $\widehat{\phi}_\rho$ (see Fig. 2.2). Hence the presence of the trend worsens the estimation quality, but the errors remain smaller than ± 0.1. The correction $\widehat{\phi}_x = 0.95$ when $\widehat{\phi}_x > 1$ causes the decrease of the superior limit of the confidence interval for $\widehat{\phi}_0 > 0.8$.

For the short time series ($N = 100$) the confidence intervals significantly increase and become asymmetric due to the influence of the corrections made at the boundaries of the variation range of $\phi^{(s)}$. For the time series dominated by trend ($r = 4$) the errors become unacceptably large. As discussed before in this section, this behavior indicates that for an efficient numerical processing, the nonstationary time series must have at least several hundreds of values.

2.4 Statistical Hypothesis Testing

Before processing a time series it is important to estimate the probability that it contains a significant deterministic trend on the basis of a statistical hypothesis test. First, we formulate the null hypothesis, in this case that there is no deterministic trend in the considered time series. Then we define a statistical quantity T, named statistic, which has the ability to detect the existence of a deterministic trend. We choose the value α of the *significance level* equal with the probability of incorrectly rejecting the null hypothesis. If we know the pdf of T, then we can compute the *critical value* t_α which delimits the values of T for which one can consider that the null hypothesis is true. The observed value t of T is compared with t_α and we can decide to reject or not the null hypothesis.

In the following we describe the Mann–Kendall test designed to detect a monotonic trend in a time series $\{x_n\}$. It is used especially in hydrology, other environmental sciences, and econometrics for short time series, even containing only several dozens of values [7]. It is based on the quantity $S_{MK} = P_{MK} - M_{MK}$, where P_{MK} is the number of the pairs $x_n > x_m$ for $n > m$ and M_{MK} is the number of the pairs $x_n < x_m$ for $n > m$. If $\{x_n\}$ are independent observations, then for $N > 10$ the random variable

$$T_{MK} = \begin{cases} (S_{MK} - 1)/\sigma_{MK} & \text{if} \quad S_{MK} > 0 \\ 0 & \text{if} \quad S_{MK} = 0 \\ (S_{MK} + 1)/\sigma_{MK} & \text{if} \quad S_{MK} < 0 \end{cases}$$

with $\sigma_{MK}^2 = N(N-1)(2N+5)/18$ follows a standard normal distribution. The null hypothesis that there is no monotonic trend is rejected when the computed value t_{MK} of T_{MK} is greater in absolute value than the critical value $t_{\alpha/2}$, where α is the chosen significance level. Hence, if $-t_{\alpha/2} < t_{MK} < t_{\alpha/2}$, then the probability that the time series $\{x_n\}$ does contain a monotonic trend is equal with α.

In general, the derivation of the probability distribution of a statistic T is difficult and usually it can be made only asymptotically, for infinite time series. Using Monte

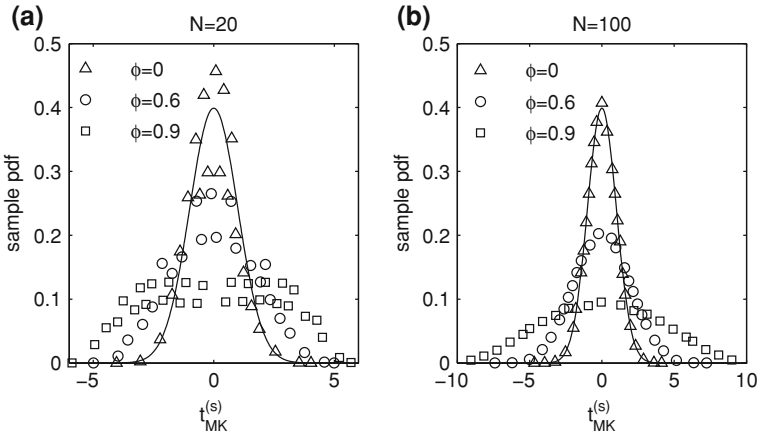

Fig. 2.8 The sample pdfs of the values $t_{MK}^{(s)}$ of the Mann–Kendall statistic for finite AR(1) noises without trend compared with the theoretical pdf for Gaussian white noise (*continuous line*)

Carlo experiments one can determine the sample pdf for any length of the time series and for any value of its parameters. For example in Fig. 2.8 we present the sample pdfs of $t_{MK}^{(s)}$ for $S = 100000$ finite AR(1) noises with $\sigma_s = 1$ and different values for N and ϕ. For $N = 100$ and $\phi = 0$ the sample pdf is indistinguishable from the theoretical pdf (Fig. 2.8b). As ϕ increases, the sample pdf is moving away from the theoretical pdf for uncorrelated noise. The pdf for correlated noise can be derived theoretically [6], but here we limit ourselves to the simplest case.

The asymptotic pdf becomes inadequate when the time series length decreases. For short uncorrelated time series ($N = 20$ and $\phi = 0$) the asymptotic pdf is correct only for $t_{MK} > 1$ (Fig. 2.8a). In the neighborhood of $t_{MK} = 0$ a significant fluctuation of the sample pdf occurs about the asymptotic pdf. The amplitude of these oscillations decreases when ϕ increases, but the oscillations extend at higher values of t_{MK}. Such a behavior is very difficult to be deduced theoretically, proving the usefulness of Monte Carlo experiments in hypothesis testing.

For a given significance level α we can determine the critical value $t_{\alpha/2}$ from the sample pdfs obtained by Monte Carlo experiments. Figure 2.9 shows $t_{\alpha/2}$ for $\alpha = 0.05$ and several values of N and ϕ. The dashed line represents $t_{\alpha/2}$ obtained from the asymptotic pdf of the uncorrelated noise and it coincides with that numerically determined for $\phi = 0$. When the serial correlation increases, the difference between the real and the asymptotic critical values becomes larger. For long time series we obtain a higher critical value, because for large N the probability distribution has a support that stretches to higher values of $t_{MK}^{(s)}$ (see Fig. 2.8).

Sometimes it is also important to analyze the possibility that the null hypothesis is not true and we erroneously decide that it is true. Therefore we have to determine the probability to accept the null hypothesis when it is false. The alternative hypothesis of the Mann–Kendall test is not clearly defined. If the null hypothesis (signal without

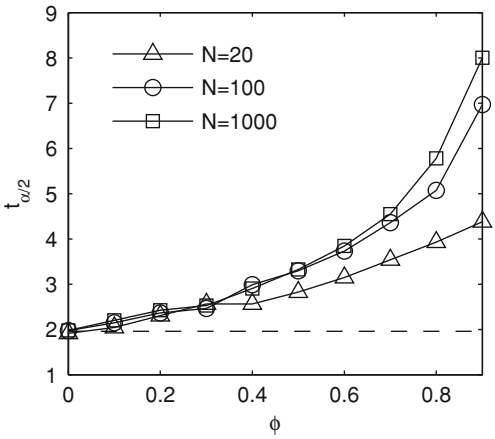

Fig. 2.9 The critical value $t_{\alpha/2}$ for a significance level $\alpha = 0.05$ of the Mann–Kendall test for AR(1) noises without trend in terms of the correlation parameter ϕ.The *dotted line* represents $t_{\alpha/2}$ obtained from the asymptotic pdf for Gaussian white noise

monotonic trend) is not true, then the signal contains a monotonic trend, but we have to mention what kind of trend from the infinity of possible functional forms is assumed. In this book we do not use this type of statistical test.

We are interested in statistical tests for the i.i.d. time series defined in Sect. 1.1. There are tests that check if the correlations between the terms of time series are negligible by means of the sample autocorrelation function (1.9). They are based on the fact that for i.i.d. time series the values of the sample autocorrelation function asymptotically form a normal i.i.d time series with mean zero and variance N^{-1} ([3], p. 223). This property is obtained as a special case of Bartlett's theorem that is valid for an arbitrary ARMA process ([3], p. 221). It is possible that, as in the case of financial time series (see Sect. 4.3), the time series is uncorrelated, but the amplitudes of its fluctuations are strongly correlated and then it is necessary to check if its absolute values are also i.i.d. It can be shown that in this case Bartlett's theorem remains valid and one can practically apply the same tests as for the initial time series.

In accordance with Bartlett's theorem, $\widehat{\rho}(h)$ has a normal distribution with mean zero and variance N^{-1}. A simple test that the time series is i.i.d. asks that less than 5 % of the values $\widehat{\rho}(h)$ should be outside the bounds $\pm 1.96 N^{-1/2}$. A more complex test is the Box–Pierce test [2] for ARMA residuals which is based on the statistic

$$Q = N \sum_{j=1}^{h} \widehat{\rho}^2(j)$$

that is approximately chi-squared distributed. These tests and their versions impose only the condition that the sample autocorrelation function should not have too large values, but in numerical processing it is possible that the sample autocorrelation function becomes too small. To avoid this possibility we compare the distribution of the sample autocorrelation function with the theoretical normal distribution.

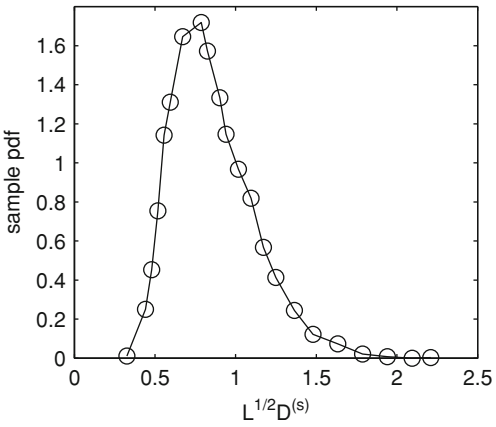

Fig. 2.10 The sample pdf of the Kolmogorov–Smirnov statistic $\sqrt{L}D^{(s)}$ for $S = 1000$ normal distributions

The Kolmogorov–Smirnov (KS) test establishes the resemblance between two distributions ([10], Sect. 14.3). If $\widehat{F}_L(x)$ is an empirical cdf with L values, then the KS statistic with respect to a theoretical cdf $F(x)$ is

$$D = \max_x \left| \widehat{F}_L(x) - F(x) \right| . \tag{2.7}$$

According to the Kolmogorov theorem the asymptotic pdf of this quantity is given by

$$\lim_{L\to\infty} P\{\sqrt{L}D < \alpha\} = \sum_{k=-\infty}^{\infty} (-1)^k e^{-k^2\alpha^2}$$

and it does not depend on the form of the distribution $F(x)$. The sum in this formula can be numerically calculated [8], but it can be also determined by a Monte Carlo experiment. In Fig. 2.10 we present the sample pdf of $\sqrt{L}D^{(s)}$ for $S = 10000$ normal distributions with zero mean, unit variance, and $L = 1000$ values. The mean of the quantity $\sqrt{L}D^{(s)}$ is 0.859 and its standard deviation is 0.259. Hence, even if $\widehat{F}_L(x)$ is obtained from a realization of a random variable with the cdf identical with $F(x)$, D is nonvanishing. Only asymptotically for $L \to \infty$ it becomes zero when $\lim_{L\to\infty} \widehat{F}_L = F$.

According to Bartlett's theorem, the sample autocorrelation function of a Gaussian i.i.d. time series forms a normal i.i.d. series with mean zero and variance N^{-1}. In order to test whether the time series is i.i.d. we calculate the KS statistic (2.7) for the first $N/4$ values of the sample autocorrelation function as recommended in [1]

$$\varepsilon = \max_{1\le h\le N/4} \left| \widehat{F}(\widehat{\rho}(h)) - G(\widehat{\rho}(h)) \right| , \tag{2.8}$$

where $G(x) = \Phi(x)/\sqrt{N}$ and $\Phi(x)$ is the normal cdf. Figure 2.11 shows the average $\langle\varepsilon\rangle$ over statistical ensembles with $S = 100$ for the autocorrelation function of the

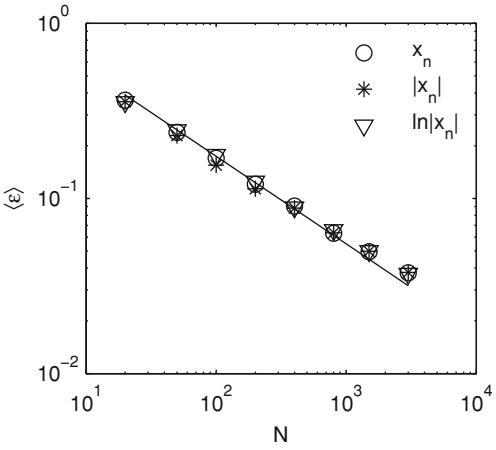

Fig. 2.11 The average of Kolmogorov–Smirnov statistic of the first $N/4$ values of the autocorrelation function of Gaussian i.i.d. white noise (o), of its absolute values (*), and of the logarithm of its absolute values (∇)

Gaussian white noises $\{x_n\}$, for their absolute values $\{|x_n|\}$, and for the logarithm of their absolute values $\{\ln|x_n|\}$. As mentioned before the three quantities are practically identical.

References

1. Box, G., Jenkins, G., Reinsel, G.: Time Series Analysis: Forecasting and Control, 3rd edn. Prentice-Hall, Upper Saddle River (1994)
2. Box, G.E.P., Pierce, D.A.: Distribution of the autocorrelations in autoregressive moving average time series models. J. Am. Stat. Assoc. **65**, 1509–1526 (1970)
3. Brockwell, P.J., Davies, R.A.: Time Series: Theory and Methods, 2nd edn. Springer, New York (1996)
4. Fornberg, B.: A Practical Guide to Pseudospectral Methods. Cambridge University Press, Cambridge (1998)
5. Hamilton, J.D.: Time Series Analysis. Princeton University Press, Princeton (1994)
6. Hirsch, R.M., Slack, J.R.: A nonparametric trend test for seasonal data with serial dependence. Water Resour. Res. **20**, 727 (1984)
7. Kendall, M.G.: Rank Correlation Methods. Griffin, London (1975)
8. Marsaglia, G., Tsang, W., Wang, J.: Evaluating kolmogorov's distribution. J. Stat. Softw. **8**(18), 1–4 (2003)
9. Metropolis, N., Ulam, S.: The Monte Carlo method. J. Am. Stat. Assoc. **44**, 335–341 (1949)
10. Press, W.H., Teukolsky, S.A., Vetterling, W.T., Flannery, B.P.: Numerical Recipes in C. The Art of Scientific Computing, 2nd edn. Cambridge University Press, Cambridge (1992)
11. Vamoş, C., Crăciun, M.: Serial correlation of detrended time series. Phys. Rev. E **78**, 036707 (2008)

Chapter 3
Polynomial Fitting

In this chapter we analyze the well-known polynomial fitting method by means of the Monte Carlo experiments with artificial time series generated by the algorithm presented in the previous chapter. Unlike the theoretical results obtained in mathematical statistics, our conclusions are valid for arbitrary trends, not only for polynomial trends. The accuracy of the estimated polynomial trend depends mainly on the ratio r between amplitudes of the trend variations and noise fluctuations. When the noise is small ($r > 1$), the estimated trend has a strong resemblance with the real trend and the noise serial correlation has negligible influence on it. Conversely, when the time series is dominated by noise ($r < 1$), the accuracy significantly decreases and it becomes even worse for noises with strong serial correlation. We conclude that the polynomial fitting is recommended for time series with small noise and simple trend with small number of local extrema. The example from astrophysics shows that the optimum degree of the polynomial trend can be determined by searching the most suited stochastic model for the noise contained in the time series.

3.1 Polynomial Fitting

Regression is a numerical method to analyze the relationship between a dependent variable and one or more independent variables. *Linear regression* refers to a linear dependence on the unknown parameters to be estimated from data. If the dependent variable is given by observed data $\{x_1, x_2, \ldots, x_N\}$ and the independent variables by $K < N$ vectors $\{a_{k1}, a_{k2}, \ldots, a_{kN}\}$, $k = 1, 2, \ldots, K$, then the linear regression model

$$x_n = \sum_{k=1}^{K} \beta_k a_{kn} + z_n \ . \tag{3.1}$$

contains the real unknown parameters β_k and the disturbances $\{z_1, z_2, \ldots, z_N\}$.

C. Vamoş and M. Crăciun, *Automatic Trend Estimation*, SpringerBriefs in Physics,
DOI: 10.1007/978-94-007-4825-5_3, © The Author(s) 2012

The linear regression is used to formulate the parametric methods of trend estimation from a noisy time series. Suppose that the trend in the stochastic process described by Eq. (1.10) is a linear superposition of several elementary trends $f^{(k)}(t)$ with known shapes

$$f(t) = \sum_{k=1}^{K} \beta_k f^{(k)}(t) \ . \tag{3.2}$$

The functions $f^{(k)}(t)$ can have any functional form (linear, polynomial, sinusoidal, exponential, etc.). Denoting by $a_{kn} = f^{(k)}(t_n)$ the values of the elementary trends at the sampling moments, Eq. (1.10) becomes the linear regression model (3.1). The trend estimation reduces to the computation of the unknown parameters β_k.

One can determine the values of β_k in Eq. (3.1) by means of the ordinary least square (OLS) method [3]. The OLS estimates of β_k (denoted b_k) are obtained by minimizing the sum of squared residuals

$$S = \sum_{n=1}^{N} \left(x_n - \sum_{k=1}^{K} \beta_k a_{kn} \right)^2 \ .$$

By differentiating with respect to β_l and equating the derivatives with zero we obtain the following system of equations for b_k

$$\sum_{k=1}^{K} b_k \left(\sum_{n=1}^{N} a_{ln} a_{kn} \right) = \sum_{n=1}^{N} x_n a_{ln} \ , \quad l = 1, 2, \ldots, K \ . \tag{3.3}$$

The analysis of this system of equations is presented in Appendix A.

If the noise $\{z_n\}$ is Gaussian i.i.d. with variance σ^2, then the estimated values b_k are normally distributed about the real value β_k with the probability distribution (A.6) derived in Appendix A. Under the same conditions, the variance σ^2 is approximated by s^2 given by Eq. (A.7) and has a $\chi^2(N-K)$ distribution. These statistical estimates are made under the assumption that the elementary trends $f^{(k)}(t)$ in Eq. (3.2) are known. But in practice these assumptions are rarely satisfied and the errors can be estimated only by Monte Carlo experiments.

In the following we consider that the trend is a polynomial of degree $K-1$

$$f(t) = \sum_{k=1}^{K} \beta_k t^{k-1} \ . \tag{3.4}$$

Then the system of equations (3.3) allows the estimation of the coefficients of the polynomial. The results presented in Appendix A hold only if the real trend is indeed polynomial and if it has the same degree as the estimated polynomial trend. Here we analyze what happens when the second hypothesis is not true. In the next section we consider the general case when the real trend is not polynomial.

As an example we consider that the trend $f(t), t \in [0, 1]$, is a polynomial of degree $q_0 = 4$ having the roots $x_0 \in \{0.1, 0.4, 0.5, 0.8\}$. Over this trend we superpose Gaussian i.i.d. noises with ratio $r \in \{0.25, 0.5, 1, 2, 4\}$ between the amplitude of the trend variations and that of the noise fluctuations given by Eq. (2.3). Because we consider that we do not know the trend degree, we apply the polynomial fitting with different number of coefficients $K = 2, 3, \ldots, 11$. The degree of the estimated polynomial trend is $q = K - 1$.

We quantify the global resemblance between the estimated trend

$$\widetilde{f}(t) = \sum_{k=1}^{K} b_k t^{k-1} \tag{3.5}$$

and the real trend $f(t)$ by the index

$$\eta = \frac{\left\| \widetilde{f}_n - f_n \right\|}{\left\| f_n - \overline{f} \right\|} , \tag{3.6}$$

where $\|\cdot\|$ is the usual quadratic norm. The temporal mean of the trend $\overline{f}$ (see Eq. 1.7) is subtracted from the denominator because we want the index η to depend only on the shape of the trends. Otherwise it would depend on a constant added to the trend. We compute the resemblance indexes $\eta^{(s)}$ for a statistical ensemble of $S = 1000$ artificial time series with $N = 1000$ values.

In Fig. 3.1a we present the variation of the average index $\langle \eta \rangle$ in function of the degree q of the estimated polynomial trend. The maximum resemblance is obtained when q coincides with the degree of the real trend $q_0 = 4$. If the degree of the estimated trend is smaller than q_0 ($q < q_0$), the errors are large because the polynomial has not enough coefficients to follow the shape of the real trend. When the degree of the estimated trend increases ($q > q_0$), the estimated trend is influenced by the noise fluctuations and η increases. In this case the variation of the index η is strongly influenced by the ratio r. When the time series is dominated by noise ($r = 0.25$), the index η is several times greater than when the time series is dominated by trend ($r > 1$).

The coefficients b_k of the estimated trend (3.5) have a similar behavior in function of q. Figure 3.1b shows the mean coefficient b_2 of the linear term in the estimated trend and the continuous line represents the coefficient β_2 of the real trend. For $q < q_0$ the variations of b_2 are very large, behavior related to the fact that the estimated polynomial trend cannot describe the shape of the real trend with a higher degree. Due to the noise contribution, for $q > 7$ the coefficient b_2 also depends on the value of the ratio r.

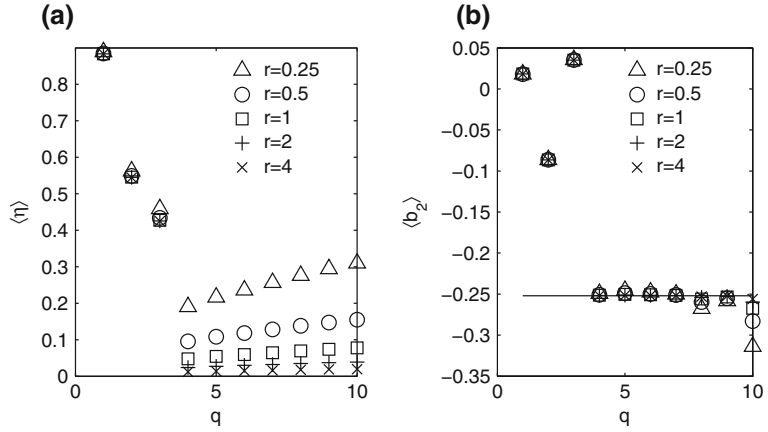

Fig. 3.1 **a** The mean resemblance index $\langle\eta\rangle$ between the real and the estimated polynomial trends. **b** The mean coefficient b_2 of the linear term of the estimated polynomial trend in comparison with the coefficient β_2 of the real trend (*continuous line*)

3.2 Polynomial Fitting of Artificial Time Series

In the previous section we assumed that the real trend was polynomial. Now we use time series obtained by means of the algorithm described in Sect. 2.3. In Fig. 3.2 we present the results of polynomial fitting for a single numerically generated time series with $N = 1000$ values. The trend has $P = 5$ monotonic segments and is plotted with a bold dashed line. A Gaussian white noise (not shown in the figure) with the amplitude smaller than the amplitude of the trend variations ($r = 4$) is superposed over the trend. The thin curves in Fig. 3.2 are the estimated polynomial trends with $q \leq 15$. As the polynomial degree q increases, the estimated trend approaches to the real one in jumps. This discontinuous variations are also visible in the dependence of the index η on the degree q (the $\times$ markers in Fig. 3.3a).

For the estimated linear trend ($q = 1$) the resemblance index is $\eta \approx 1$ (the $\times$ marker in Fig. 3.3a). The estimated parabolic trend ($q = 2$) describes better the overall shape of the real trend (Fig. 3.2a) and the resemblance index drops to $\eta \approx 0.7$ (Fig. 3.3a). By increasing the polynomial degree by one ($q = 3$) the estimated trend changes very little and η remains at the same value. When $q = 4$ and $q = 5$, the estimated trend has four monotonic segments but with the variations amplitude smaller than those of the real trend (Fig. 3.2b). A significant decrease of η occurs for $q = 6$ (Fig. 3.3a) when the shape of the estimated trend becomes almost identical to the real one for $t > 0.3$ (Fig. 3.2c). Until $q = 9$ the shape of the trend and the value of η do not vary significantly. For $q = 10$ another jump of the estimated trend towards the real one occurs due to a better approximation of the real trend in the region with the steepest slope (Fig. 3.2d). Subsequently η continues to decrease to the minimum obtained for $q_0 = 23$ (Fig. 3.3a). For higher degrees the influence of the noise on the estimated trend becomes dominant and η slightly increases.

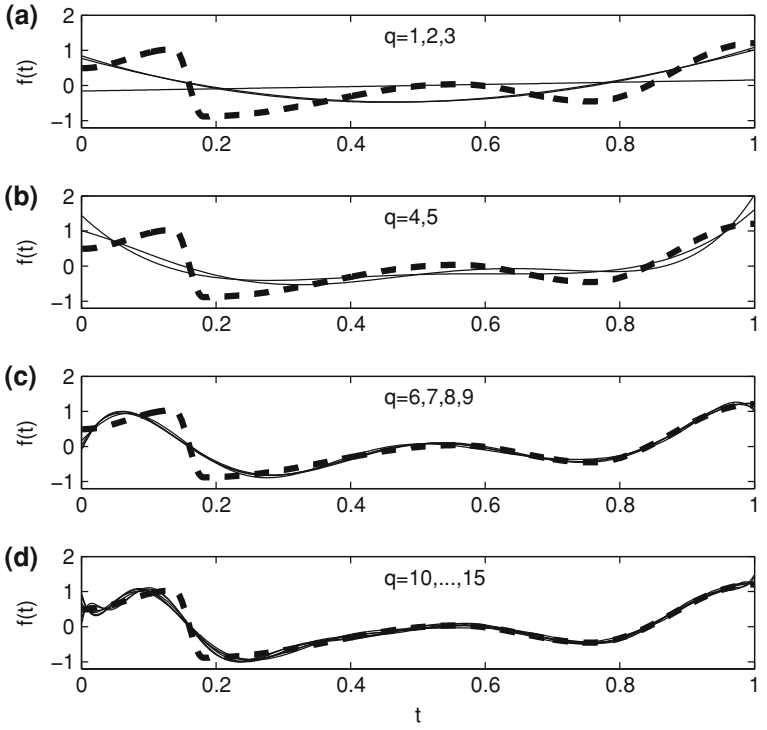

Fig. 3.2 The estimated polynomial trends (*thin lines*) of degrees $q \leq 15$ for a trend (*bold dashed line*) over which a white noise is superposed with $r = 4$ (not shown in figure)

This behavior is different from that observed in the previous section in case of polynomial trends. If the real trend is polynomial, then the maximum resemblance is obtained when the degree of the estimated polynomial trend coincides with the real one (Fig. 3.1). When the real trend is not polynomial, then it is gradually approximated by the estimated polynomial trend, its shape adjusting itself by successive jumps. When q becomes too large, the estimated trend is more influenced by the noise and η begins to increase. The degree q_0 for which the maximum resemblance is obtained is smaller ($q_0 = 11$) when the time series is dominated by the noise (the plot for $r = 0.25$ in Fig. 3.3a).

The behavior of the coefficients of the estimated polynomial trend is also different from the case when the real trend is polynomial. As one can see in Fig. 3.3b, the coefficients increase exponentially with q until the optimum degree q_0 is reached. Their values are greater than those for a polynomial trend by several orders of magnitude. Figure 3.3b shows the coefficients b_k only for $r = 4$, but they have the same behavior for the other values of the ratio r.

From the results presented in Fig. 3.3a we obtain information concerning the interpretation of the values of the resemblance index η. If we do not succeed to estimate

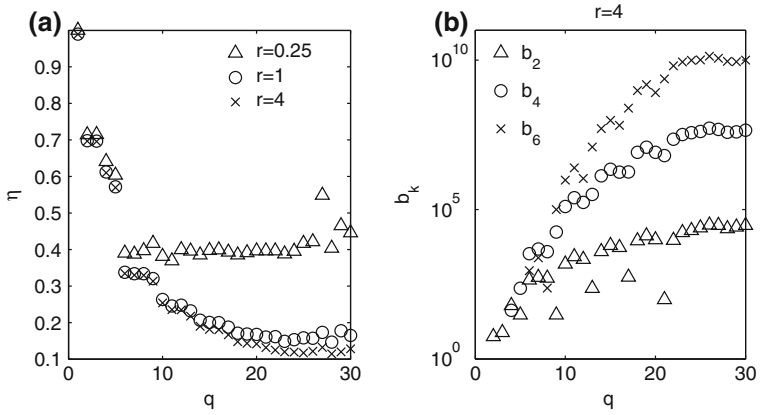

Fig. 3.3 **a** The resemblance index η of the estimated polynomial trends in function of their degree q for several values of the ratio r. **b** Three coefficients of the estimated polynomial trend in terms of the degree q for $r = 4$

the trend, then in Eq. (3.6) we consider $\widetilde{f}_n \equiv \overline{f}$ and $\eta = 1$. Hence $0.5 < \eta < 1$ indicates that there is a slight resemblance between the real and the estimated trend. When $0.2 < \eta < 0.5$ the two trends have the same global shape with approximately the same number of monotonic segments. A good resemblance of the trends is indicated by a small value of the index $\eta < 0.2$. Finally, $\eta > 1$ means that the estimated trend differs from the real one more than a constant trend and the estimated trend removal would cause a distortion of the information contained by the time series. In such cases the trend removal is not recommended [1].

In the following we use a Monte Carlo experiment to evaluate the accuracy of polynomial fitting. The results in Fig. 3.4 are obtained for statistical ensembles of $S = 1000$ numerically generated time series with the method presented in Sect. 2.3 with $N = 1000$, $P = 5$ and different values for ϕ and r. First we analyze the case of the white noise (Fig. 3.4a) that continues the previous discussion for a single time series. The average $\langle\eta\rangle$ over the statistical ensemble of the resemblance index has the same shape as in Fig. 3.3a except that the variation with respect to q is smooth, without jumps. For $r = 1$ the minimum $\langle\eta\rangle_{\min}$ separating the estimated polynomial trends which are dominated by the shape of the real trend from those dominated by noise occurs for $\langle q\rangle_{\min} = 21$. The average $\langle\eta_{\min}\rangle$ of the minimum with respect to q of the index η for each time series (marked with a star in figure) is slightly smaller than $\langle\eta\rangle_{\min}$. Also the value $\langle q\rangle_{\min}$ is close to the average of the degrees at which the minimum occurs for each time series $\langle q_{\min}\rangle = 19.25$ showing that the variability of the individual time series due to the noise is not significant. If the time series are dominated by noise ($r < 1$), then the minimum of the average resemblance index $\langle\eta\rangle_{\min}$ increases and occurs at lower values of the degree $\langle q\rangle_{\min}$ of the estimated trend. In addition, the noise influence on the estimated trend increases for higher values of $q > \langle q\rangle_{\min}$, making the minimum of $\langle\eta\rangle_{\min}$ to be more clearly defined.

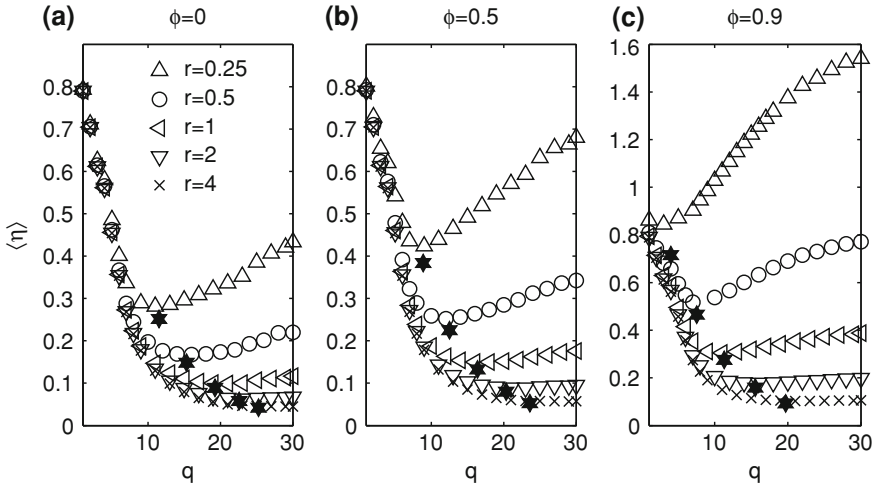

Fig. 3.4 The average index η for estimated polynomial trends. The average $\langle \eta_{\min} \rangle$ of the minimum with respect to q of the index η for each time series is marked with a star

Conversely, for $r > 1$ at large values of $q > \langle q \rangle_{\min}$ the average index $\langle \eta \rangle$ does not vary significantly because the influence of noise on the estimated trend is small.

If the noise has a significant serial correlation (Fig. 3.4b, c), then its influence on the estimated trend grows and the minima of $\langle \eta \rangle_{\min}$ increase and occur at smaller degrees of the estimated polynomial trend. This behavior is due to the correlation between successive values of noise which can be confounded with the variation of the trend (stochastic trend). For time series dominated by noise ($r = 0.25$) with strong serial correlation ($\phi = 0.9$) even for small degrees q of the estimated trend the noise influence is dominant and $\langle q \rangle_{\min} = 3$. In the case of time series with low noise ($r > 1$) the influence of serial correlation is very small.

Finally we analyze the influence of the time series resolution and the number of the monotonic parts on the estimated polynomial trend. The minimum $\langle \eta \rangle_{\min}$ plotted in Fig. 3.5 is determined on statistical ensembles of 100 numerically generated time series with $\phi = 0$ and different values for N, P, and r. For $P = 5$ monotonic segments, from Fig. 3.5a it follows that by increasing the number N of the time series values, a better resemblance with the real trend is obtained, at least for time series dominated by noise ($r \leq 1$). When the noise is small ($r = 4$) the increase of N does not influence the quality of the estimated polynomial trend.

In Fig. 3.5b, for $N = 1000$ fixed, we present the variation of $\langle \eta \rangle_{\min}$ with respect to the number P of monotonic segments. The resemblance between the actual and estimated trend becomes smaller when P increases for all values of r. Hence the accuracy of the trend obtained by polynomial fitting significantly depends on the complexity of the real trend, i.e., the number of its monotonic segments.

From Fig. 3.5a, b it follows that the parameters N and P have opposite influences on the resemblance index η. In order to find which of the two parameters has a stronger

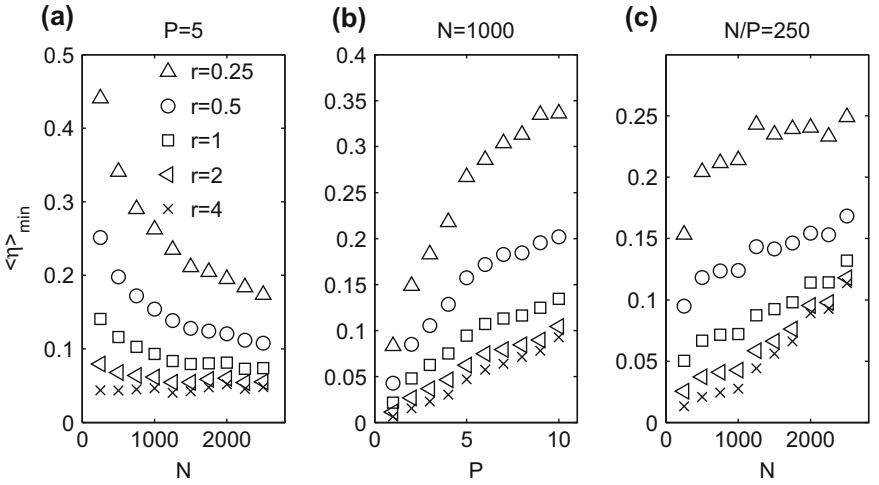

Fig. 3.5 The minimum with respect to the degree q of the average resemblance index $\langle\eta\rangle$ for estimated polynomial trends

influence, in Fig. 3.5c both N and P are varied such that the average resolution on a monotonic segment remains constant $N/P = 250$. When N varies from 250 to 2500, the number of monotonic segments varies from 1 to 10. For all values of r, $\langle\eta\rangle_{\min}$ increases when N and P increase in the same way as for N fixed and P variable. Hence the accuracy of the polynomial fitting is mainly controlled by the number of the monotonic segments.

In conclusion, the polynomial fitting is recommended for time series with simple trends. If the trend has many local extrema, i.e., many monotonic segments, then the degree of the estimated polynomial trend rapidly increases and the order of magnitude of its coefficients becomes very large. The accuracy also worsens if the trend has very steep segments where the resolution of the time series is small.

3.3 An Astrophysical Example

In this section we estimate the polynomial trend of a time series describing the variability of the X-ray flux from a Seyfert galaxy. The data are extracted from Heasarc Exosat ME archive for the Seyfert galaxy NGC5506 and are represented in Fig. 3.6a. This time series was modeled by a first order autoregressive process with the parameters $\phi = 0.994$ and $\sigma = 0.722$ [6]. We use this result as a benchmark for our model presented in [4].

We perform polynomial fittings with degrees from 0 to 15. When $q = 0$ the trend is a constant equal to the average of time series. As remarked in Sect. 3.2, when the degree of the polynomial trend increases, the shape of the trend does not

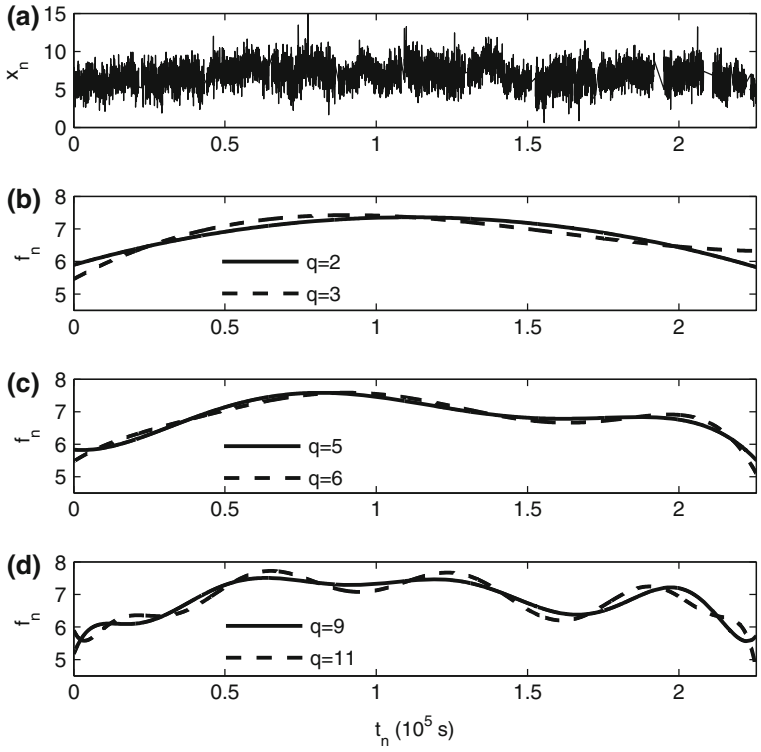

Fig. 3.6 **a** The X-ray flux from the Seyfert galaxy NGC5506. **b–d** The estimated polynomial trends with the greatest values of the resemblance index $\delta\eta$ between two estimated polynomial trends having successive degrees

change monotonically. For certain polynomial degrees, the trend has more significant changes, while for other degrees the shape remains practically unchanged. Now we do not know the real trend and we cannot compute the resemblance index (3.6). Instead we define an index that measures the difference between two estimated polynomial trends having successive degrees

$$\delta\eta(q) = \left\| \widetilde{f}_n(q) - \widetilde{f}_n(q-1) \right\| \ .$$

Its values with respect to the degree q are plotted in Fig. 3.7. Higher values of $\delta\eta(q)$ are related to the degrees q for which the shape of the estimated trend significantly varies.

In Fig. 3.6b–d we plot six of the estimated polynomial trends with the greatest values of the index $\delta\eta$. The most significant variation of the estimated trend occurs for $q = 2$ when the polynomial trend describes the overall shape of the signal. For $q = 3$ and $q = 4$ the parabolic form of the estimated trend is preserved, the higher value of $\delta\eta(3)$ in Fig. 3.7 being related to a slight shift of the parabola. Only for

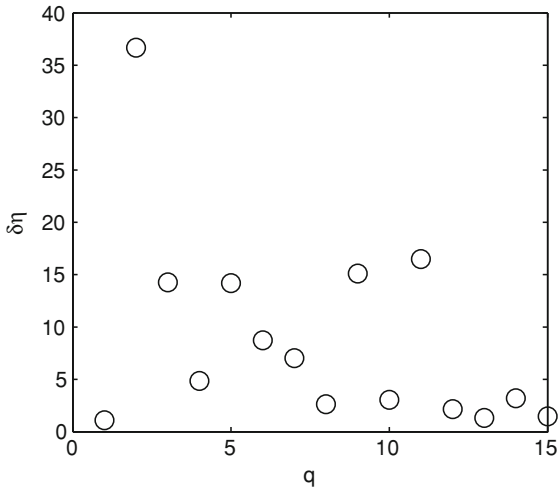

Fig. 3.7 The resemblance index $\delta\eta$ between two estimated polynomial trends having successive degrees

$q = 5$ the polynomial trend describes the different shapes of the first and second half of the time series. For higher values of polynomial degree ($q = 6, 7, 8$) the trend shape slightly changes by the enhancement of the second local maximum. The polynomial trend can better follow the details of the time series shape when $q = 9$ and the estimated trend has more local extrema. The next significant variation of $\delta\eta$ occurs at $q = 11$ when the amplitudes of the local extrema of the trend increase. For greater values of q the estimated trend does not significantly vary anymore.

We determine the model of the random component of the data using the periodogram of the time series [2]. Because 584 out of $N = 7532$ values of the time series are missing, the mean of the time series is assigned to these missing data and we obtain an equidistant time series with a sampling interval $\Delta t = 30s$. The periodogram $I(\lambda_k)$ of the time series from which its mean value was subtracted is plotted in Fig. 3.8a. Here the Fourier frequencies are equal with the numbers $\lambda_k = 2\pi k/N$ with k integers which fall in the interval $(-\pi, \pi]$. Due to the large statistical variability of the periodogram, it has to be averaged [5]. The averaged periodogram $I_K(\lambda_k)$ by a central moving average (1.14) with $K = 10$ is plotted in Fig. 3.8b. The averaging procedure results in loosing data at low frequencies.

At higher Fourier frequencies than $\lambda_0 = 0.02 \cdot 2\pi$ the averaged periodogram is parallel to the abscissa. Therefore we model the random component of this time series as a superposition of a white noise over a correlated noise whose spectrum appears in periodogram only at low frequencies $\lambda < \lambda_0$. Over this part of the spectrum we fit the theoretical spectrum of an AR(1) process given by Eq. (1.24) and represented by the continuous line in Fig. 3.8. The power spectrum is plotted in double logarithmic scale and we measure the similarity between the theoretical spectrum $g(\lambda)$ and the periodogram by the quantity

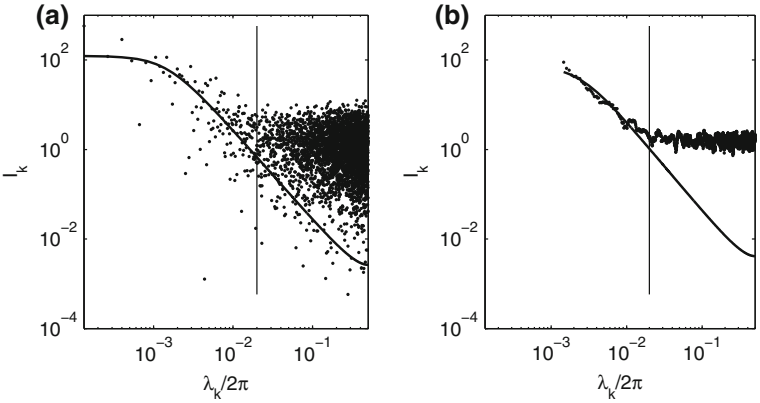

Fig. 3.8 The initial **a** and the averaged **b** periodogram compared with the power spectrum of the AR(1) model of the noise (*continuous line*). The *vertical thin lines* marks the frequency λ_0 separating the white noise from the AR(1) noise

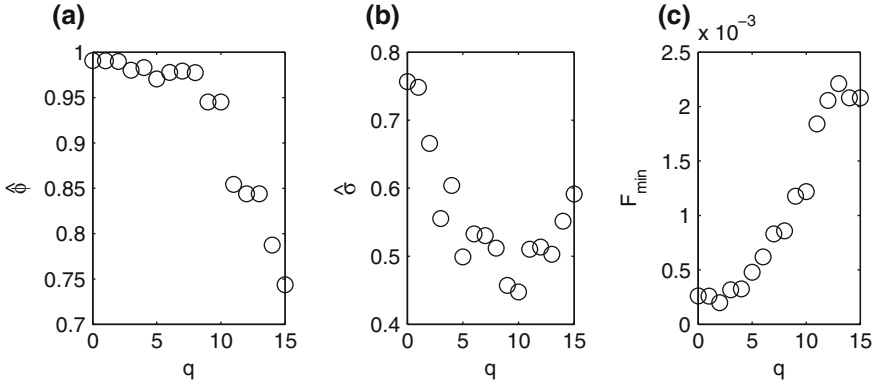

Fig. 3.9 The values of the estimated AR(1) parameters and the minima of $F(\phi, \sigma)$

$$F(\phi, \sigma) = \sum_{k>0} p_k \left(\log_{10} g(\lambda_k) - \log_{10} I_K(\lambda_k)\right)^2 , \tag{3.7}$$

where p_k are weights taking into account the fact that in logarithmic scale the points of the periodogram are nonuniformly distributed, the distance between them being much smaller at high frequencies. For $k > 1$ we take the weight p_k equal to half the distance between the Fourier frequencies adjacent to λ_k

$$p_k = \frac{1}{2}(\log_{10} \lambda_{k+1} - \log_{10} \lambda_{k-1}) = \frac{1}{2} \log_{10} \frac{k+1}{k-1} .$$

For $k = 1$ we use the distance between the first two Fourier frequencies $p_1 = \log_{10}(\lambda_2/\lambda_1) = \log_{10} 2$. We denote with $\widehat{\phi}$ and $\widehat{\sigma}$ the spectral estimators obtained by minimizing the function (3.7).

The estimated AR(1) parameters after detrending the time series with polynomials of various degrees are presented in Fig. 3.9a, b. The minimum of the function $F(\phi, \sigma)$ with respect to the degree q is plotted in Fig. 3.9c. The smallest value of $F(\phi, \sigma)$ is obtained for $q = 2$ and corresponds to the polynomial trend that provides the most suited noise model for the observed data. The estimated AR(1) parameters for $q = 2$ are $\widehat{\phi} = 0.990$ and $\widehat{\sigma} = 0.666$ which are close to those reported in [6].

We have to address the problem of the statistical significance of our estimates and of the optimal polynomial trend. The minimum value of $F(\phi, \sigma)$ for $q = 2$ is only slightly smaller than that for $q = 0$ and $q = 1$. Then are we entitled to detrend the time series by a parabolic trend or it is preferable to let it remain undetrended as in [6]? The results in the previous section are not applicable because in this case the noise is a mixture of a white noise and an AR(1) strongly correlated noise.

In order to establish the statistical significance of the estimated trend we resort to a Monte Carlo experiment. The statistical ensemble contains time series with $N = 7532$ values obtained by superposing over the estimated parabolic trend a white noise and an AR(1) noise with the parameters previously determined $\widehat{\phi} = 0.990$ and $\widehat{\sigma} = 0.666$. From the average value of the horizontal part of the periodogram in Fig. 3.8b we determine the standard deviation of the white noise as $\sigma_{\text{white}} = 1.262$. For each of the 10000 artificial time series of the statistical ensemble we estimate by polynomial fitting with $q = 2$ a perturbed trend and we compute its resemblance index with respect to the parabolic trend in Fig. 3.6b. We obtain the average index $\langle \eta \rangle = 0.317$ with the standard deviation $\sigma_\eta = 0.166$ indicating that if the initial time series contains such a parabolic trend, then it can be estimated with good accuracy despite of the existence of a complex noise with large amplitude. Hence we conclude that the time series in Fig. 3.6a contains a significant trend that can be estimated by polynomial fitting.

References

1. Andreas, E.L., Trevino, G.: Using wavelets to detect trends. J. Atmos. Ocean. Technol. **14**, 554–564 (1997)
2. Brockwell, P.J., Davies, R.A.: Time Series: Theory and Methods. Springer, New York (1996)
3. Hamilton, J.D.: Time Series Analysis. Princeton University Press, Princeton (1994)
4. Morariu, V.V., Vamoş, C., Pop, A., Şoltuz, Ş.M., Buimaga-Iarinca, L.: Autoregressive modeling of the variability of an active galaxy. Rom. J. Phys. **55**, 676–686 (2010)
5. Stoica, P., Moses, R.L.: Introduction to Spectral Analysis. Prentice Hall, New Jersey (1997)
6. Timmer, J., Schwarz, U., Voss, H.U., Wardinski, I., Belloni, T., Hasinger, G., van der Klis, M., Kurths, J.: Linear and nonlinear time series analysis of the black hole candidate Cygnus X-1. Phys. Rev. E **61**, 1342–1352 (2000)

Chapter 4
Noise Smoothing

The central moving average (CMA) is one of the simplest and most used method to filter out the noise fluctuations from a time series and it depends on a single parameter, the semi-length K of the averaging window. We introduce the repeated central moving average (RCMA) which depends on an additional parameter (the number i of averagings) and allows a gradual smoothing of the time series. There are many more sophisticated nonparametric methods for trend estimation [4], but we concentrate our analysis on a detailed analysis of an algorithm depending on a small number of parameters. Using Monte Carlo experiments we analyze the properties of the RCMA with boundary conditions obtained by a normalized form of padding with zero of the time series. We show that roughly the same smoothing is obtained either by repeating many times an averaging with a small K or by repeating fewer times an averaging with a large K. We also prove that any form of moving average introduces a spurious serial correlation in the smoothed time series. The accuracy of the trend estimated by the RCMA depends on the ratio r between the amplitudes of the trend variations and noise fluctuations and on the noise serial correlation in the same way as the accuracy of the estimated trend by polynomial fitting. The RCMA trend does not mainly depend on the number of the monotonic segments of the trend, but on the average resolution of the monotonic segments. The real time series of the returns of the S&P500 index is processed by the CMA in order to determine the financial volatility time series. The optimum semilength of the averaging window is found using the condition that the estimated noise should be uncorrelated.

4.1 Repeated Central Moving Average

The *moving average* (MA) of an infinite time series $\{x_n, n = 0, \pm 1, \pm 2, \ldots\}$ is defined as a normalized linear combination

$$\vartheta_x(n) = \sum_{k=K_-}^{K_+} w_k x_{n+k} \tag{4.1}$$

C. Vamoş and M. Crăciun, *Automatic Trend Estimation*, SpringerBriefs in Physics,
DOI: 10.1007/978-94-007-4825-5_4, © The Author(s) 2012

with the weighting coefficients $w_k \geq 0$

$$\sum_{k=K_-}^{K_+} w_k = 1.$$

The length of the *averaging window* $[n + K_-, n + K_+]$ is $T = K_+ - K_- + 1$. We implicitly consider that $w_{K_-} > 0$ and $w_{K_+} > 0$, otherwise new limits $K_\pm$ can be found such that the coefficients w_{K_-} and w_{K_+} should be nonvanishing. The summation limits can also be infinite or zero. If $K_- \geq 0$ ($K_+ \leq 0$), then only future (past) values of the time series are included in the MA. If $K = -K_- = K_+ > 0$ and $w_{-k} = w_k$ for all k, then the MA is symmetric and $T = 2K + 1$.

By choosing different values of the coefficients w_k we obtain different types of MAs. The simplest one is the arithmetic mean of the values within a finite averaging window with all the coefficients equal to a constant $w_k = T^{-1}$. In the symmetric case $w_k = (2K + 1)^{-1}$ we obtain the so-called *central moving average* (CMA)

$$\vartheta_x(n) = \frac{1}{2K+1} \sum_{k=-K}^{K} x_{n+k} \tag{4.2}$$

identical with Eq. (1.14). Another commonly used MA is the exponential weighted MA defined by the recursive formula

$$\vartheta_x(n) = \alpha x_n + (1 - \alpha)\vartheta_x(n - 1),$$

where $0 < \alpha < 1$ is a real parameter. The weighting coefficients in this case are $w_k = \alpha(1 - \alpha)^{-k}$ for $k \leq K_+ = 0$ and $K_- = -\infty$, hence only the past values of the time series are averaged.

In fact all the time series encountered in practice are finite and when the averaging window contains a time series boundary we cannot apply formulas like Eq. (4.1) with the same coefficients for all the moments t_n. For a finite time series $\{x_n, n = 1, 2, \ldots, N\}$ the MA is defined as

$$\vartheta_x(n) = \sum_{m=1}^{N} c_{n,m} x_m,$$

where the weighting coefficient $c_{n,m}$ describes the weight with which the term m contributes at the average value at n. The coefficients $c_{n,m} \geq 0$ satisfy the normality condition

$$\sum_{m=1}^{N} c_{n,m} = 1$$

for each n. If $n + K_- \geq 1$ and $n + K_+ \leq N$, then the averaging window contains neither of the two boundaries and the coefficients can be expressed by means of the coefficients w_k

$$c_{n,m} = w_{m-n} \quad \text{for} \quad n + K_- \leq m \leq n + K_+$$

and zero in the rest. For other values of n, $c_{n,m}$ has to be determined from some boundary conditions supplementing the information contained in the coefficients w_k.

A method to design boundary conditions is to extend the finite time series to an infinite one. The most usual ways to extend the time series are the following: the time series is assumed to be zero outside its support; it is assumed to be periodic; it is extended by repeating the first and the last values respectively; it is extended by its mirror image, whereby it becomes symmetric with respect to the boundaries [5].

In the following we shall use extensively a normalized form of the padding with zeros of CMA. The explicit expression of the averaging near the boundaries is

$$\vartheta_x(n) = \frac{1}{n + K} \sum_{k=-n+1}^{K} x_{n+k} \tag{4.3}$$

when $n \leq K$, i.e., $c_{n,m} = (n + K)^{-1}$ for $m \leq n + K$ and $c_{n,m} = 0$ for $m > n + K$, and

$$\vartheta_x(n) = \frac{1}{N - n + K + 1} \sum_{k=-K}^{N-n} x_{n+k} \tag{4.4}$$

when $n \geq N - K + 1$, i.e., $c_{n,m} = (N - n + K + 1)^{-1}$ for $m \geq n - K$ and $c_{n,m} = 0$ for $m < n - K$. Hence near boundaries we use a CMA with smaller averaging window.

In practice we do not know from the beginning the optimum degree of smoothing for a particular time series. Therefore, it is useful to apply gradual averagings and by means of an optimization criterion to decide when to stop the smoothing. Depending on the phenomenon under consideration and the mathematical model, one can find specific optimization criteria. For example, in Sect. 4.3 we use the condition that the estimated noise should be as close as possible to an i.i.d. time series. A numerical criterion is introduced in Chap. 7 in order to design an automatic form of the CMA.

A simple method to vary the degree of smoothing is to repeat several times an averaging with a lower degree of smoothing. By applying i times a MA characterized by coefficients $c_{n,m}$ we obtain a different MA with the coefficients denoted by $c^{(i)}_{n,m}$. We call this smoothing method the *repeated moving average* (RMA) algorithm. If $\{\vartheta_x(n; i)\}$ is the time series obtained after i successive MA, then the result obtained by applying other j additional MA can be described by the linear transformation

$$\vartheta_x(n; i + j) = \sum_{m=1}^{N} c^{(j)}_{n,m} \vartheta_x(m; i). \tag{4.5}$$

Because the first MA is applied to the initial time series, we have $\vartheta_x(n;0) = x_n$. By expressing the averages $\vartheta_x(n;i+j)$ and $\vartheta_x(n;i)$ in Eq. (4.5) in function of $\vartheta_x(n;0)$ it results the transformation law

$$c_{n,s}^{(i+j)} = \sum_{m=1}^{N} c_{n,m}^{(j)} c_{m,s}^{(i)}. \tag{4.6}$$

A relation similar to Eq. (4.5) holds for the weighting coefficients $w_k^{(i)}$ for a RMA with i repetitions. Applying Eq. (4.1) iteratively to an infinite time series the averaging window is progressively extended and new coefficients are obtained

$$\vartheta_x(n;i) = \sum_{k=iK_-}^{iK_+} w_k^{(i)} x_{n+k}. \tag{4.7}$$

This formula resembles to Eq. (4.5), but the coefficients $w_k^{(i)}$ are not related to each other by a simple relation as Eq. (4.6). Nevertheless, the relation $c_{n,m}^{(i)} = w_{m-n}^{(i)}$ is preserved when the interval (iK_-, iK_+) does not contain the boundaries of the finite time series.

We numerically compute the coefficients $c_{n,m}^{(i)}$ for the CMA by applying Eq. (4.6) iteratively for $j = 1$ and successive values of i. Because the values of $c_{n,m}^{(1)}$ are known (see Eqs. (4.2), (4.3), (4.4)) we can recursively obtain all the coefficients. Figure 4.1 shows the values of $c_{150,m}^{(i)}$ obtained for the normalized form of the padding with zeros of the *repeated central moving average* (RCMA) with $K = 10$ and $i \leq 10$ for a time series with $N = 300$ values. The interval where $c_{150,m}^{(i)} = w_{m-150}^{(i)}$ is nonzero increases by $2K$ with each new averaging. After i repetitions the length of this interval is $2iK + 1$. The short horizontal segment in Fig. 4.1 represents the initial coefficients $c_{150,m}^{(1)} = (2K+1)^{-1}$ which have a constant value. In Fig. 4.1b the y-axis has a logarithmic scale so that one can see that for large m after several averagings the coefficients become very small and they do not significantly influence the result of the RCMA.

If the averaging window does not reach the boundaries of the time series, the coefficients remain symmetric as in Fig. 4.1 and equal to $w_{m-n}^{(i)}$. Near the boundaries the coefficients $c_{n,m}^{(i)}$ become asymmetric (Fig. 4.2). The asymmetry always occurs at $m = K + 1$ and the maximum values of the coefficients always occur inside the boundaries of the time series. Therefore the values near the boundaries are influenced with a greater weight by the values located inside the time series.

An important feature of any MA method is that it introduces a spurious serial correlation in the smoothed time series. In Appendix B we compute the autocovariance function of an i.i.d. noise averaged by a MA and of the noise estimated from a time series containing a trend. Hence, if we are interested in the statistical properties of the noise, then we have to take into account that the statistics of the estimated noise is changed by MA.

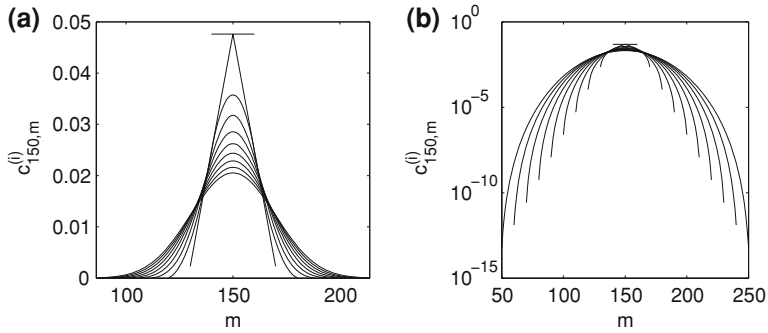

Fig. 4.1 The coefficients $c^{(i)}_{150,m}$ of the normalized form of the padding with zeros of the RCMA for $K = 10$ and $i \leq 10$ if the averaging window does not reach the boundaries of the time series. The y-axis has a linear scale in panel **(a)** and a logarithmic scale in panel **(b)**

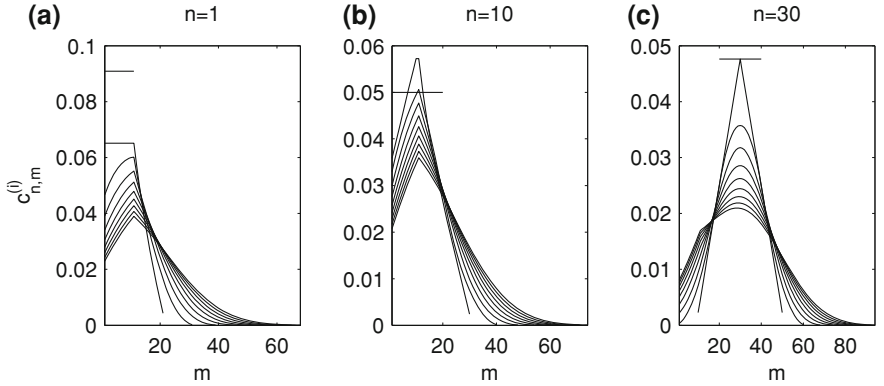

Fig. 4.2 The coefficients $c^{(i)}_{n,m}$ of the normalized form of the padding with zeros of the RCMA for $K = 10$ and $i \leq 10$ for three different locations near the boundary of the time series

Another property of the RCMA is that roughly the same smoothing is obtained either by repeating many times an averaging with a small K or by repeating fewer times an averaging with a large K. To illustrate this feature we consider two RCMAs with $K_1 < K_2$ in the simple case when the averaging windows do not contain the boundaries of the time series. For i_1 repetitions of the first averaging and i_2 repetitions of the second averaging we measure the similarity between the two RCMAs by the index

$$\varepsilon_{2,1} = \sum_n \left| w_n^{(i_1)}(K_1) - w_n^{(i_2)}(K_2) \right|,$$

where $w_n^{(i)}$ is defined in Eq. (4.7). By exhaustive search we determine the number of repetitions $i_{2,1}$ of the second averaging giving the maximum similarity of the coefficients, i.e., the minimum of $\varepsilon_{2,1}$. Figure 4.3a presents the results of this analysis

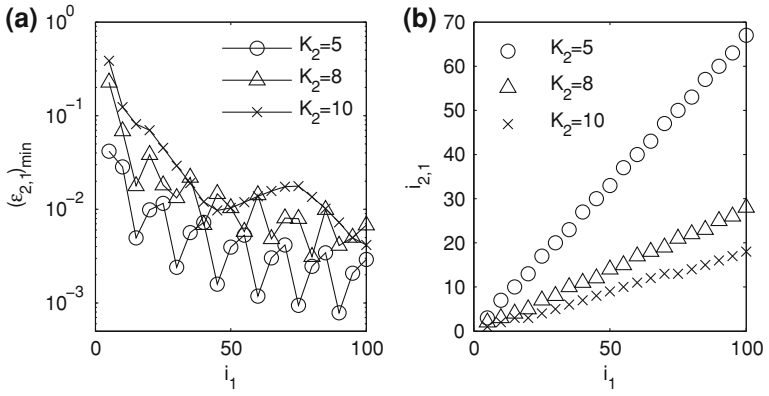

Fig. 4.3 **a** The minimum of the similarity index between the RCMA with averaging window semi-length K_2 and the RCMA with $K_1 = 4$ repeated i_1 times. **b** The number of repetitions $i_{2,1}$ for which the two RCMAs have the greatest similarity

for $K_1 = 4$ and three different values of K_2. Although the minimum error $(\varepsilon_{2,1})_{\min}$ has a decreasing global trend, it has oscillations due to discontinuous variations of parameters on which $\varepsilon_{2,1}$ depends. For a small number of repetitions $(\varepsilon_{2,1})_{\min}$ is greater than 0.1, but for $i_1 > 40$ it is always smaller than 0.01.

Hence the same smoothing can be obtained for different values of K and i. Figure 4.3b shows that there is a linear relationship between $i_{2,1}$ and i_1. In order to study the influence of this property on the trend estimation accuracy we use statistical ensembles of numerically generated time series. They have the length N =1000, the number of monotonic parts of the trend is $P = 5$, the serial correlation parameter is $\phi = 0$ and the ratio of the amplitude variations of trend and noise $r \in \{0.25, 0.5, 1, 2, 4\}$. The other parameter of the artificial time series is $\Delta N_{\min} = 20$.

For each artificial time series $s \leq S$ and each given value of K we determine the minimum of the resemblance index $\eta^{(s)}$ given by Eq. (3.6) by exhaustive search over the number of repetitions $i \leq 50$. We use statistical ensembles of $S = 100$ artificial time series and we compute the average value of the minimum error $\langle \eta_{\min} \rangle$ and the average number of repetitions $\langle i_{\min} \rangle$ for which the minimum is obtained. The results are presented in Fig. 4.4. The graphs of $\langle \eta_{\min} \rangle$ have extended plateaus with almost constant values because approximately the same smoothing is obtained for different values of the parameters K and i. The plateaus are limited due to two types of causes.

First, in the case of the time series dominated by noise ($r < 1$) at small values of K the average minimum index $\langle \eta_{\min} \rangle$ is greater than the value of the plateau (Fig. 4.4a). These situations occur when the number of repetitions at which the minimum of η is obtained is equal to the limit $i \leq 50$ imposed to the number of iterations. As one can notice in Fig. 4.3b, for small K we need a larger number of iterations to obtain maximum resemblance than for higher values of K. When the required number of iterations is greater than the limit $i \leq 50$, the required degree of smoothing cannot

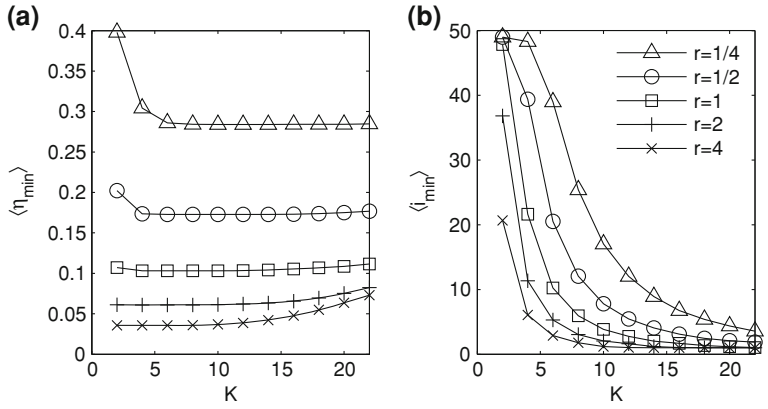

Fig. 4.4 The average value of the minimum resemblance index $\langle \eta_{\min} \rangle$ **(a)** and the average number of repetitions $\langle i_{\min} \rangle$ **(b)** for which the minimum is obtained for trends estimated by the RCMA

be obtained. If we increased the number of repetitions up to $i \leq 100$, then η would be reduced to the value of the plateau.

Secondly, for the time series dominated by trend ($r > 1$) the plateaus are limited at greater values of K (Fig. 4.4a). In these cases, even for a single averaging the smoothing of the time series is too strong and the minimum of η is obtained for $i = 1$ and then $\langle i_{\min} \rangle = 1$ in Fig. 4.4b. By repeating the averaging the deformation of the trend grows and the error η increases.

In conclusion, if the value of K belongs to the plateau of the graph $\langle \eta_{\min} \rangle$, then there exists a number of repetitions i supplying the best approximation of the real trend. In the following we choose for K a value belonging to the plateau and determine the value of $i_{\min}$ by exhaustive search. This condition can be always satisfied if we decrease enough the value of K, but this leads to the increase of the value $i_{\min}$ and of the running time of the algorithm.

4.2 Smoothing of Artificial Time Series

In this section we analyze the errors of the trend estimated by the RCMA and compare them with the errors for polynomial fitting analyzed in Sect. 3.2. We measure the error by means of the resemblance index η defined by Eq. (3.6). First we analyze the variation of the index η for a single time series, the same series as that examined by polynomial fitting in Fig. 3.3. The real trend is represented by thick dashed line in Fig. 4.5a, the random component is a white noise (not shown in the figure) and the ratio of the trend and noise variations is $r = 1$. For $K = 10$ the index η is represented in Fig. 4.5b with circle markers in function of the number of repetitions i. Unlike the variation of η for the polynomial fitting in Fig. 3.3a, for the RCMA the variation of η is smooth, without jumps. The thin lines in Fig. 4.5a represent the

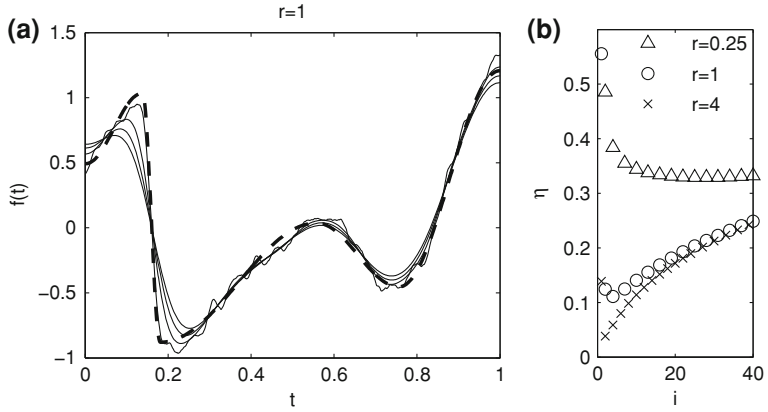

Fig. 4.5 **a** The trends estimated by the RCMA with $K = 10$ for $i = 2, 20, 40, 60$ iterations (*thin lines*) of a trend (*bold dashed line*) over which a white noise (not shown in figure) is superposed with $r = 1$. **b** The resemblance index η of the trends estimated by the RCMA in function of the number of iterations for different ratios r

trends estimated by the RCMA for $i = 2, 20, 40, 60$. By increasing the number of averagings, the estimated trend becomes more dissimilar to the real one. The minimum of η is obtained for $i = 4$ when the estimated trend contains partially damped noise fluctuations.

If the time series is dominated by noise ($r = 0.25$), then the minimum of η is obtained for $i_{\min} = 26$ and its value is much greater than for $r = 1$ (see Fig. 4.5b). In this case the noise fluctuations should be strongly damped and then the trend is also strongly distorted. Conversely, when the time series is dominated by trend ($r = 4$), the minimum index is obtained for two averagings that sufficiently attenuate the noise fluctuations without a significant distortion of the trend. For $i > 20$ the indexes η for the time series with $r = 1$ and $r = 4$ coincide, showing that the noise is completely attenuated and the error is due only to the trend distortion.

By comparing the minima of η in Fig. 4.5b with those obtained for the same time series using polynomial fitting presented in Fig. 3.3a, it follows that the trend estimated by the RCMA approximates better the real trend. But, as remarked above, the trend estimated by smoothing preserves noise fluctuations, while that obtained by polynomial fitting does not. The values of η for the estimated trends in Fig. 4.5 confirm the interpretation given in Sect. 3.2 to the values of the index η, i.e., $\eta < 0.2$ indicates a good resemblance between the estimated and real trends, whereas $0.2 < \eta < 0.5$ an acceptable resemblance.

The results in Fig. 4.6 are obtained by processing statistical ensembles of 100 numerically generated time series with $N = 1000$, $P = 5$, and various values for ϕ and r. We analyze first the case of white noise (Fig. 4.6a) which continues the previous discussion for a single time series. The average index $\langle\eta\rangle$ on the statistical ensemble has the same shape as that in Fig. 4.5b for a single time series. The minimum $\langle\eta\rangle_{\min}$ of the average resemblance index is close to the average of the minima for each

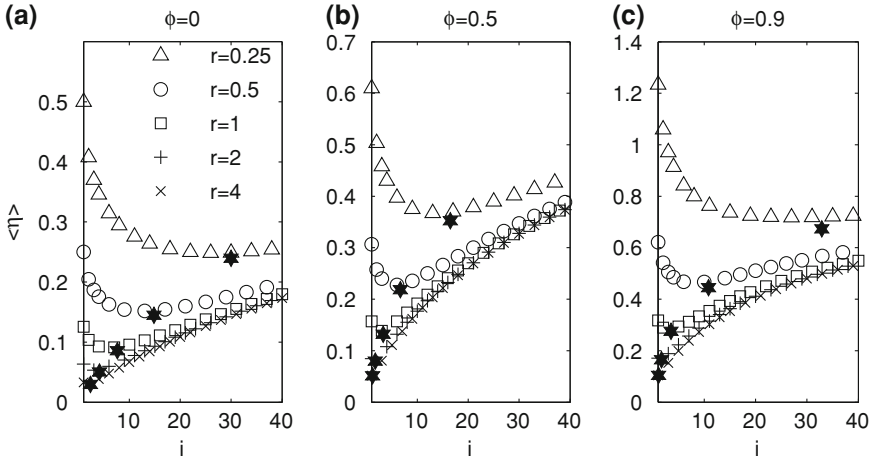

Fig. 4.6 The average index η for trends estimated by the RCMA. The average of the minima of the index η for each time series is marked with a *star*. The semi-lengths of the RCMA are $K = 10, 20, 30$ corresponding to the increasing values of the serial correlation parameter ϕ

time series $\langle \eta_{\min} \rangle$ (marked with a star in the figure) showing that the variability of individual time series due to the noise is not significant. If the time series is dominated by noise ($r < 1$), then $\langle \eta \rangle_{\min}$ increases and occurs at a large number of averaging repetitions. Conversely, for $r > 1$ the minimum $\langle \eta \rangle_{\min}$ decreases and is obtained for a small number of repetitions.

If the noise has a significant serial correlation (Figs. 4.6b, c), then the minima of $\langle \eta \rangle$ increase. This behavior is due to the fact that the correlation between successive values of noise is confounded with the variation of trend values (stochastic trend). For time series strongly dominated by trend ($r = 4$) $\langle \eta \rangle_{\min}$ remains below 0.1 and for $\phi = 0$ it is below 0.05. If the amplitude of the trend variations equals that of noise fluctuations ($r = 1$), then the error increases from $\langle \eta_{\min} \rangle = 0.1$ to 0.3 for strongly correlated noise. For the time series dominated by noise the minimum average index is greater than 0.5 for $\phi = 0.9$ when the stochastic trend is much larger than the deterministic one. Hence the errors of the estimated trend become very large when the series is dominated by noise and become intolerable if the noise is also strongly correlated.

If we compare these results with those obtained for polynomial fitting (see Fig. 3.4) we observe that many of the properties presented above are the same. The accuracy of the trend estimated by the RCMA mainly depends on the ratio r. Its accuracy is high when the time series is dominated by trend ($r > 1$) and it is only slightly influenced by the noise serial correlation. When the time series is dominated by noise ($r < 1$), the RCMA accuracy decreases and it becomes strongly dependent on the noise serial correlation. The observation previously made for a single time series remains valid in general: the minimum of the resemblance index η is slightly smaller for RCMA than for polynomial fitting, but the estimated trend contains residual noise fluctuations.

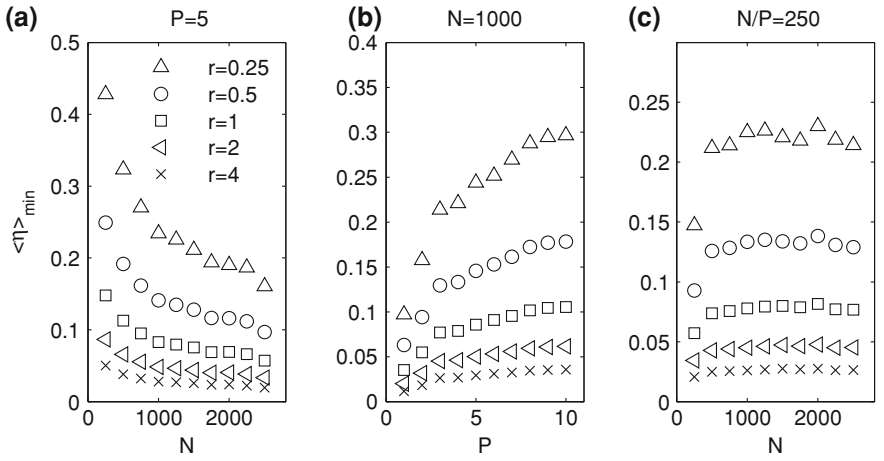

Fig. 4.7 The minimum with respect to the number i of the averaging repetitions of the average resemblance index $\langle\eta\rangle$. The smoothing is obtained by the RCMA with $K = 20, 12, 10, 8, 6$ corresponding to the increasing values of the ratio r

Finally we analyze the influence of the time series resolution and the number of the monotonic parts on the trend estimated by the RCMA. The minimum $\langle\eta\rangle_{\min}$ with respect to the averagings number i of the average index $\langle\eta\rangle$ on statistical ensembles of 100 numerically generated time series with $\phi = 0$ and different values for N, P, and r is plotted in Fig. 4.7. For $P = 5$ fixed, by increasing the number N of the time series values, a better resemblance with the real trend is obtained (Fig. 4.7a), especially for time series dominated by noise ($r \leq 1$).

In Fig. 4.7b for $N = 1000$ fixed, we present the variation of $\langle\eta\rangle_{\min}$ with respect to P. The resemblance between the actual and estimated trend becomes smaller when P increases, especially for small values of r. In order to determine which of the opposite influences of N and P is more important, in Fig. 4.7c both the number of values N and the number of monotonic parts P are varied such that the average resolution on a monotonic part remains constant $N/P = 250$. One observes that $\langle\eta\rangle_{\min}$ is almost constant proving that the accuracy of estimated trend depends primarily on the ratio N/P.

Compared with the results for the polynomial fitting presented in Fig. 3.5, we note that the variation of $\langle\eta\rangle_{\min}$ for fixed N or P is similar with that of the RCMA. But in the case of the RCMA the opposite influences of N and P balance each other, so that the RCMA accuracy is almost constant for fixed average resolution N/P of the monotonic segments of trend. Hence if the number of monotonic parts of the trend has the most important influence for polynomial fitting, for the RCMA it is more important the average resolution on the monotonic parts. Therefore the trend estimation by polynomial fitting is recommended when the real trend has a small number of local extrema and the smoothness of the estimated trend is especially demanded.

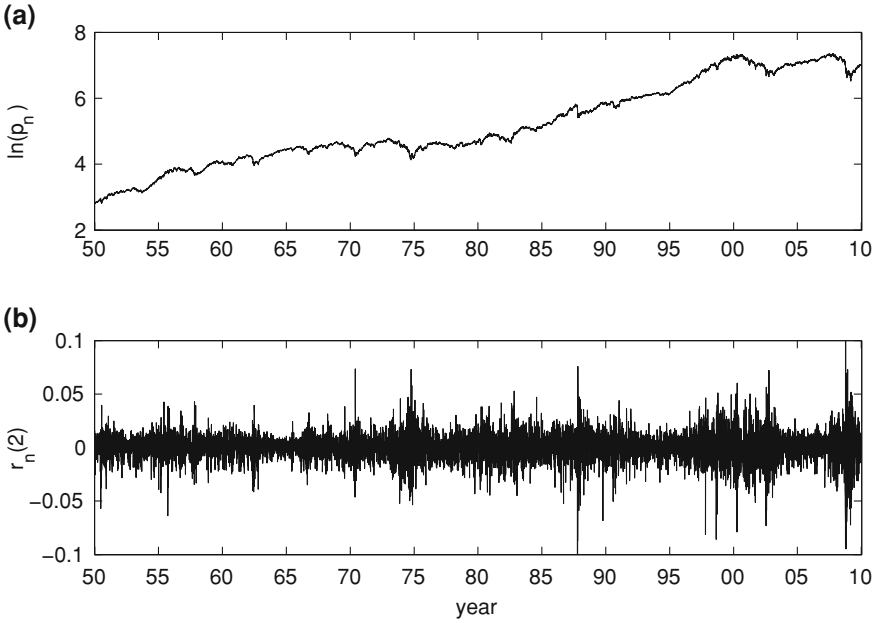

Fig. 4.8 The logarithm of the index S&P500 **(a)** and the logreturns for $\Delta n = 2$ days **(b)**

The RCMA is recommended for time series having a large enough resolution on each monotonic part of the trend.

4.3 A Financial Example

In the case of a real time series the trend is unknown and the optimal smoothing can be chosen only if we use additional assumptions on the noise properties. We illustrate this procedure with the financial time series of the S&P500 index in the interval between January 1st, 1950 and December 31st, 2009 containing $N = 15097$ trading days. If we denote by $\{p_n, n = 1, 2, \ldots, N\}$ the daily closing values of the S&P500 index, then the logreturn over the temporal interval $[n, n + \Delta n]$ is defined as

$$r_n(\Delta n) = \ln p_{n+\Delta n} - \ln p_n.$$

For $\Delta n > 1$ the logreturns are computed only for $n = k\Delta n$, otherwise the temporal intervals would overlap each other and spurious correlations would occur between the logreturns. The mean is extracted from the final time series $\{r_n(\Delta n)\}$. In Fig. 4.8 we present the logarithm of the index S&P500 and the logreturns for $\Delta n = 2$.

In quantitative finance the logreturns time series $\{r_n(\Delta n), n = 1, 2, \ldots, N\}$ is modeled as a finite sample of an infinite discrete heteroskedastic process $\{R_n(\Delta n), n = 0, \pm 1, \pm 2, \ldots\}$. This means that the stochastic process is equal to the product of two terms

$$R_n(\Delta n) = V_n(\Delta n)\ Z_n(\Delta n). \tag{4.8}$$

For Δn fixed $\{Z_n(\Delta n)\}$ is a Gaussian i.i.d. stochastic process with unit variance and the variance of its absolute values is equal to a constant $\sigma^2_{|Z|}$

$$\begin{aligned} \langle Z_n(\Delta n)\ Z_s(\Delta n)\rangle &= \delta_{ns}, \\ \langle |Z_n(\Delta n)|\ \ |Z_s(\Delta n)|\rangle &= \sigma^2_{|Z|}\delta_{ns}. \end{aligned} \tag{4.9}$$

The unpredictable random nature of the process $\{Z_n(\Delta n)\}$ models the efficiency property of financial markets, i.e., the logreturns cannot be forecast based on their previous values. The first factor $\{V_n(\Delta n)\}$ in Eq. (4.8) models the *volatility* which measures the degree of instability of financial markets and it is strictly positive.

The financial time series have specific properties, called "stylized facts" [6]. In Fig. 4.8b one can see the *volatility clustering*, i.e., the small and large volatility values are not homogeneously distributed, but they separately cluster together in well delimited time periods. As a consequence, the sample autocorrelation function of the absolute values of the logreturns $\widehat{\rho}_{|r|}(h)$ has significant nonvanishing values for large time lags. For example, when $\Delta n = 2$, the serial correlations are significant up to hundreds of days (see Fig. 4.9b). Using 1000 random permutations of the time series $\{r_n(2)\}$ we have computed the boundaries of the 95 % confidence interval that the sample autocorrelation function of $\{|r_n(2)|\}$ has vanishing values. Only for $h > 300$ the values of $\widehat{\rho}_{|r|}(h)$ are situated within these boundaries.

As shown in Fig. 4.9a the sample autocorrelation function is negligible and the logreturns are uncorrelated. Only a few of the values of $\widehat{\rho}_r(h)$ are outside the boundaries of the 95 % confidence interval so we can conclude that the sample autocorrelation function of $\{r_n(2)\}$ has vanishing values. This property results from the randomness of the sign variations of the white noise $\{Z_n\}$ in Eq. (4.8), which is preserved by multiplication with the strictly positive values of the volatility $\{V_n\}$.

By taking the logarithm of the absolute value of Eq. (4.8) we obtain a superposition of an i.i.d noise on a trend related to the volatility

$$\ln|R_n(\Delta n)| = \ln V_n(\Delta n) + \ln|Z_n(\Delta n)|. \tag{4.10}$$

The volatility of a time series generated by such a process can be estimated by the RCMA discussed in previous sections. The pdf of the stochastic process $\{\ln|Z_n(\Delta n)|\}$ is no longer Gaussian and its average $\mu_{\ln|Z|} = \langle \ln|Z_n(\Delta n)|\rangle = -0.6352$ is nonvanishing. Another method to estimate the volatility is to apply the CMA directly to the absolute values of the returns [8].

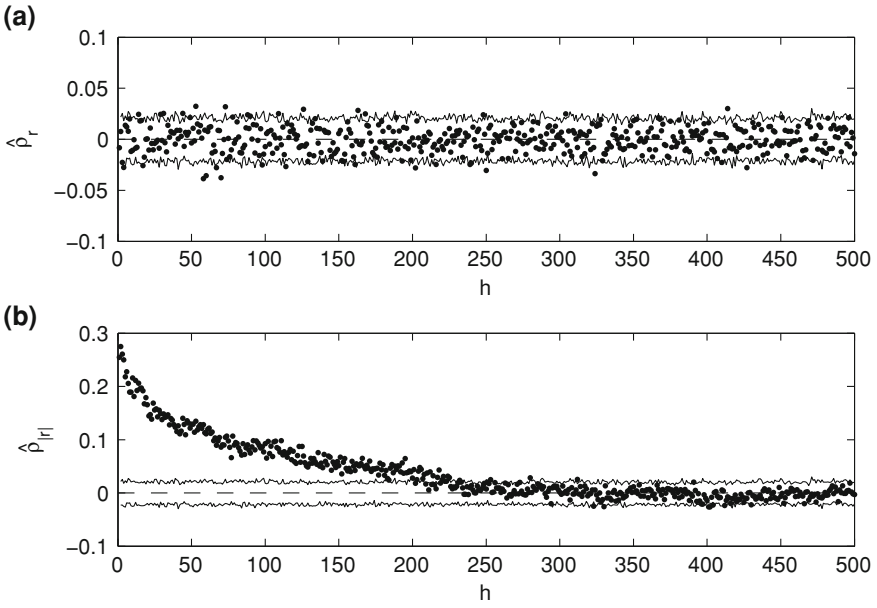

Fig. 4.9 The sample autocorrelation function of the logreturns **(a)** and of the absolute values of the logreturns **(b)** for S&P500

In order to keep the discussion simple, we use for smoothing a single averaging CMA without repeating it. Then the only parameter is the semi-length of the averaging window K. We apply the CMA given by Eq. (4.2) to a realization of the stochastic process (4.10) and introducing the notation $x_n(\Delta n) = \ln|r_n(\Delta n)|$ we obtain

$$\vartheta_x(n;\, \Delta n,\, K) = \vartheta_{\ln v}(n;\, \Delta n,\, K) + \vartheta_{\ln|z|}(n;\, \Delta n,\, K), \tag{4.11}$$

where $\{v_n\}$ and $\{z_n\}$ are realizations of the stochastic processes $\{V_n\}$ and $\{Z_n\}$, respectively. The CMA of the noise term $\vartheta_{\ln|z|}(n;\, \Delta n,\, K)$ is not zero and it can be approximated by $\mu_{\ln|Z|}$. Then from Eq. (4.11) it follows that we can try to estimate the volatility by the quantity $\exp\left\{\vartheta_x(n;\, \Delta n,\, K) - \mu_{\ln|Z|}\right\}$ and according to Eq. (4.8) the estimator of the white noise is

$$\zeta_n(\Delta n,\, K) = r_n(\Delta n)/\exp\left\{\vartheta_x(n;\, \Delta n,\, K) - \mu_{\ln|Z|}\right\}$$

which depends on the semi-length K of the averaging window. We have to find the optimum value K_0 by imposing the condition that the time series $\{|\zeta_n(\Delta n,\, K)|\}$ should be uncorrelated in accordance with the properties of the noise $\{Z_n\}$ given by Eq. (4.9).

According to Eq. (2.8) we measure the deviation from normality of the sample autocorrelation function of the time series $\{|\zeta_n(\Delta n,\, K)|\}$, denoted $\widehat{\rho}_{|\zeta|}$, by the

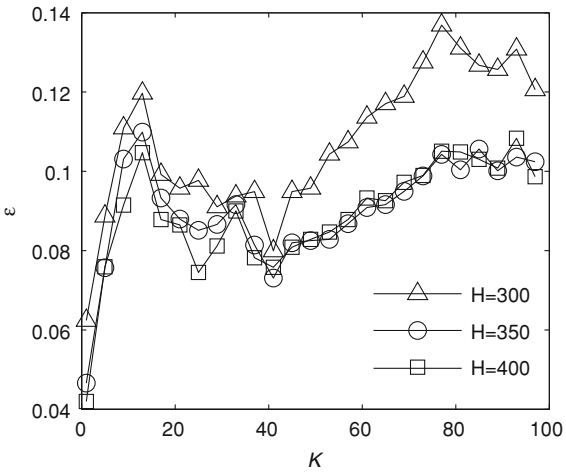

Fig. 4.10 The variation of the index ε for $\Delta n = 1$ with respect to the semi-length K of the averaging window for several values of H

Kolmogorov–Smirnov statistic

$$\varepsilon(K, H) = \max_{1 \le h \le H} \left| \widehat{F}(\widehat{\rho}_{|\zeta|}(h)) - G(\widehat{\rho}_{|\zeta|}(h)) \right|, \tag{4.12}$$

where $\widehat{F}$ is the sample cdf of $\widehat{\rho}_{|\zeta|}$ and G is the theoretical cdf of the normal distribution with Bartlett's parameters. Unlike Eq. (2.8) the maximum in Eq. (4.12) is taken over a variable number H of values of the sample autocorrelation function. By numerical experiments we have checked that the autocorrelation function of $\{|Z_n(\Delta n)|\}$ verifies Bartlett's formula if $\{Z_n(\Delta n)\}$ is normally distributed (see Sect. 2.4). We take into account the spurious correlation introduced by averaging given by Eq.(B.3) in Appendix B and we extract it from the sample autocorrelation function. The quantity ε is an index measuring the nonnormality of $\widehat{\rho}_{|\zeta|}$, i.e., the serial correlation of $\{|\zeta_n(\Delta n, K)|\}$. If K_0 is the value for which ε is minimum, then the volatility estimator is

$$\widehat{v}_n(\Delta n) = \exp\left\{\vartheta_x(n; \Delta n, K_0) - \mu_{\ln|Z|}\right\}$$

and from Eq. (4.8) it follows that the estimator of the white noise is $\widehat{z}_n(\Delta n) = r_n(\Delta n)/\widehat{v}_n(\Delta n)$.

The index ε defined by Eq. (4.12) depends on two parameters (K and H). The choice of the number H of sample autocorrelation function values is a difficult problem with no simple solution [3]. As a general rule H should be smaller than a quarter of the time series length [1]. However, the sample autocorrelation function $\widehat{\rho}_{|\zeta|}$ is different from that of a white noise especially for the small values of h. Moreover, if the serial correlation for small h is reduced, then the entire autocorrelation function becomes negligible. As a consequence, we limit the value of H in Eq. (4.12) only to several hundreds.

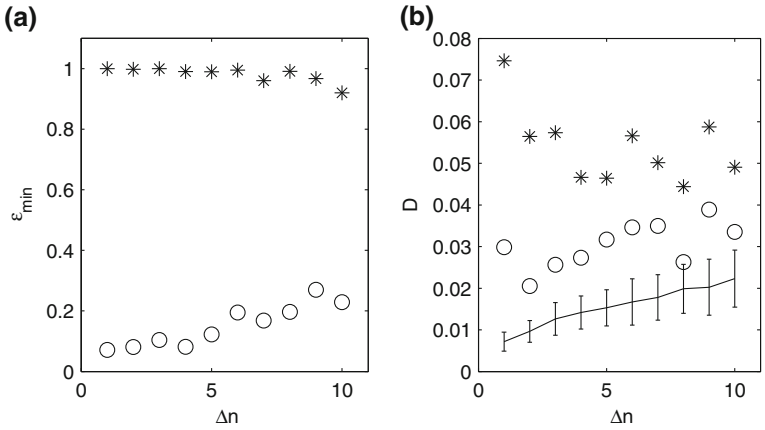

Fig. 4.11 **a** The minimum of ε (o) compared with the index ε_0 of the initial logreturns series ($*$). **b** The Kolmogorov–Smirnov statistic D of the estimated white noises $\{\widehat{z}_n(\Delta n)\}$ (o) and of the logreturns $\{r_n(\Delta n)\}$ ($*$) for different time scales Δn

In Fig. 4.10 we present the variation of the index ε for $\Delta n = 1$ with respect to the semi-length K of the averaging window for several values of H. Each curve has two minima. The first minimum corresponds to an averaging with $K = 1$ which does not significantly damp the fluctuations of the logreturns, so that the average $\vartheta_x(n; \Delta n, K)$ preserves these fluctuations. We are interested in the second minimum occurring for larger K for which the average $\vartheta_x(n; \Delta n, K)$ is slowly varying. For other values of Δn the first minimum is greater than the second one or even does not exist and then the minimum supplying the estimated volatility coincides with the global minimum.

The minimum values $\varepsilon_{\min}$ of the index ε for which the white noise is estimated are compared in Fig. 4.11a with ε_0 which is the index ε computed for the initial logreturns time series. The minimum of ε is obtained by exhaustive search over $K \leq 150$ and $H < 360$. For all time scales Δn, ε_0 is close to 1, indicating that all the values of the sample autocorrelation function implied in the computation of ε_0 lay outside the variation range of the autocorrelation function of a white noise (see Fig. 4.9b). The values of $\varepsilon_{\min}$ are significantly smaller than the corresponding ε_0 showing that the estimated white noise is much closer to an uncorrelated time series than the initial logreturns.

In order to verify the normality of the estimated white noise, we compute the Kolmogorov–Smirnov statistic D defined by Eq. (2.7) for $\{\widehat{z}_n\}$. Figure 4.11 shows the values of D for $\Delta n \leq 10$ compared with those for the normalized initial logreturns D_0. The error bars represent the mean and the standard deviation of D for statistical ensembles of 1000 numerically generated i.i.d. Gaussian time series having the same length as $r_n(\Delta n)$. The results show that the estimated white noise $\{\widehat{z}_n(\Delta n)\}$ has a probability distribution closer to a Gaussian than that of the initial logreturns.

Figure 4.12 shows the logreturns, the estimated volatility, and the estimated white noise after the year 2000 for $\Delta n = 2$. The increased variability of the logreturns

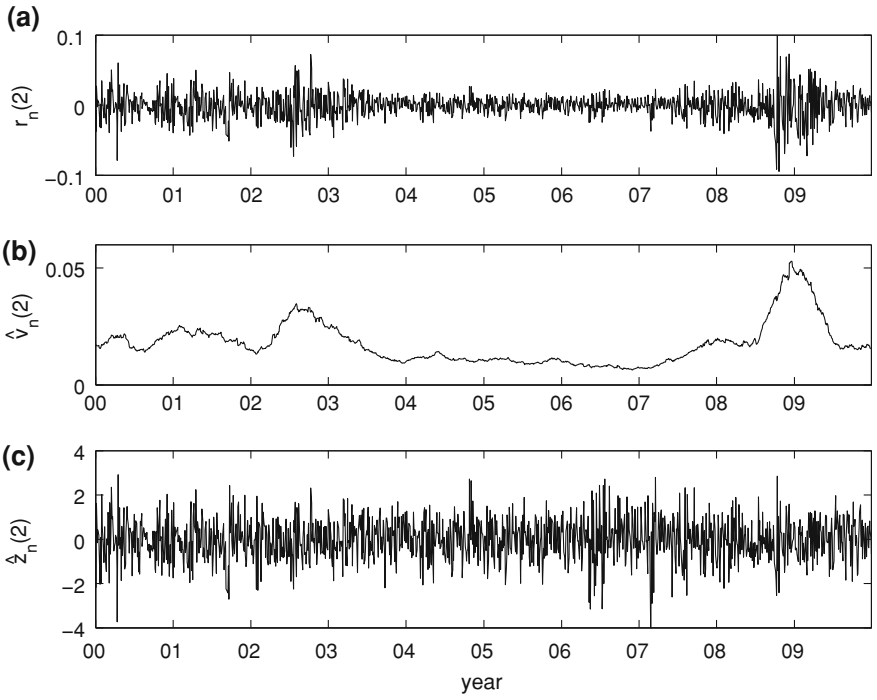

Fig. 4.12 The logreturns **(a)**, the estimated volatility **(b)**, and the estimated white noise **(c)** for the daily S&P500 index after the year 2000 for $\Delta n = 2$

amplitude (volatility clustering) is no more present in the estimated white noise. Even the most volatile period of the financial crisis which started in 2008 is correctly described by the estimated volatility.

References

1. Box, G., Jenkins, G., Reinsel, G.: Time Series Analysis: Forecasting and Control, 3rd edn. Prentice-Hall, Upper Saddle River (1994)
2. Brockwell, P.J., Davies, R.A.: Time Series: Theory and Methods, 2nd edn. Springer, New York (1996)
3. Escanciano, J.C., Lobato, I.N.: An automatic portmanteau test for serial correlation. J. Econ. **151**, 140–149 (2009)
4. Goodall, C.: A survey of smoothing techniques. In: Fox, J., Long, J.S. (eds.) Modern Methods of Data Analysis, p. 176. Sage Publications, Newbury Park (1990)
5. Karlsson, G., Vetterli, M.: Extension of finite length signals for sub-band coding. Signal Process. **17**, 161–168 (1989)
6. Poon, S.H.: A Practical Guide to Forecasting Financial Market Volatility. Wiley, Chichester (2005)

7. Press, W.H., Teukolsky, S.A., Vetterling, W.T., Flannery, B.P.: Numerical Recipes in C. The Art of Scientific Computing, 2nd edn. Cambridge University Press, New York (1992)
8. Vamoş, C., Crăciun, M.: Separation of components from a scale mixture of Gaussian white noises. Phys. Rev. E **81**, 051125 (2010)

Chapter 5
Automatic Estimation of Monotonic Trends

In this chapter we design an automatic algorithm to estimate monotonic trends over which an arbitrary stationary noise is superposed. It approximates the trend by a piecewise linear curve obtained by dividing into subintervals the time series values, instead of the time domain. The slope of each linear segment of the estimated trend is proportional to the average one-step displacement of the time series values included into the corresponding subinterval, therefore the method is referred to as average conditional displacement (ACD). This algorithm was first presented in [5]. Using Monte Carlo experiments we show that for AR(1) noises the accuracy of the ACD algorithm is comparable with that of the polynomial fitting and moving average but it has the advantage to be automatic. For time series with nonmonotonic trends the ACD algorithm determines one of the possible monotonic components which can be associated to the trend. As an illustration we apply the ACD algorithm to a paleoclimatic time series to determine the periods with a significant monotonic temperature variation.

5.1 Average Conditional Displacement (ACD) Algorithm

As a rule, the trend estimation methods do not take into account whether the estimated trend is or is not monotonic. The exceptions are the methods in which the trend is explicitly looked for as a monotonic function, usually a linear, exponential, or logarithmic function. The weakness of such an approach is the limited number of available monotonic functional forms. If an enrichment of the functional forms is attempted, for example, using polynomials of order greater than one, then the monotony property is lost. From this point of view the advantage of the ACD method is that it can describe a much richer set of monotonic trends as piecewise linear functions.

In practice there are situations when the trend monotony is especially required, for example the ascertainment of the time periods with monotonic variation of the global atmospheric temperature (see Sect. 5.4). Another important application of the ACD

C. Vamoş and M. Crăciun, *Automatic Trend Estimation*, SpringerBriefs in Physics,
DOI: 10.1007/978-94-007-4825-5_5, © The Author(s) 2012

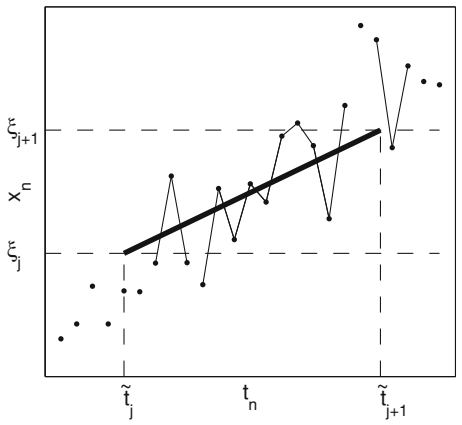

Fig. 5.1 The one-step variations of the time series $\{x_n\}$ entering into the computation of the average conditional displacement $\widetilde{g}_j$ are plotted with *thin straight segments*. The *thick segment* is the ACD approximation of the monotonic variation of the trend of the time series in the interval $(\xi_j, \xi_{j+1}]$

algorithm is the interpolation of the sample cdf. By definition the cdf is monotonic increasing and the usual interpolation methods (spline functions for example) do not always preserve the monotony.

Let us consider a time series $\{x_n\}$. We define the disjoint intervals $I_j = (\xi_j, \xi_{j+1}]$, $j = 1, 2, \ldots, J$, such that any value x_n is contained in an interval I_j. We denote by N_j the number of values x_n lying within I_j. The simplest solution to build them is to use homogeneous intervals, i.e., the values of N_j should differ from each other by a unit at the most. Then the total number of intervals J is the single parameter describing the distribution of the series values. The numerical algorithm to generate the intervals I_j is presented in Sect. 5.2.

The one-step variation of the time series is $\delta x_n = x_{n+1} - x_n$. For a given interval I_j, we compute the sample average of δx_n under the condition that the initial or final values should be included in this interval

$$\widetilde{g}_j = \frac{1}{2N_j}\left(\sum_{x_n \in I_j} \delta x_n + \sum_{x_{n+1} \in I_j} \delta x_n\right). \tag{5.1}$$

This quantity characterizes the average variation of the time series within I_j. In Fig. 5.1 the straight segments correspond to the one-step variations δx_n entering in the computation of the average conditional displacement $\widetilde{g}_j$. If all the values $\widetilde{g}_j$ have the same sign, then we can use them to determine a numerical approximation of the monotonic trend by a piecewise linear curve denoted $\widetilde{f}(t)$.

We construct the estimated monotonic trend $\widetilde{f}(t)$ by iteratively computing the straight segments of the estimated trend. Let us assume that the estimated trend is increasing and all the quantities $\widetilde{g}_j$ are positive. The j-th straight segment is delimited by the points $(\widetilde{t}_j, \xi_j)$ and $(\widetilde{t}_{j+1}, \xi_{j+1})$ and it has the slope equal with $\widetilde{g}_j$. The time moments corresponding to the boundaries of the straight segments are given by the recursive relation

$$\widetilde{t}_{j+1} = \widetilde{t}_j + \frac{\xi_{j+1} - \xi_j}{\widetilde{g}_j} \ ,$$

where $\widetilde{t}_1 = t_1$. Then for $t \in (\widetilde{t}_j, \widetilde{t}_{j+1})$, the estimated trend is

$$\widetilde{f}(t) = \xi_j + (t - \widetilde{t}_j)\widetilde{g}_j \ . \tag{5.2}$$

In Fig. 5.1 the straight segment of the estimated trend is plotted with thick continuous line.

When the estimated trend is decreasing and all the quantities $\widetilde{g}_j$ are negative, the order in which the values ξ_j are assigned to the straight segments is reversed. Now the j-th straight segment is delimited by the points $(\widetilde{t}_j, \xi_{J-j+2})$ and $(\widetilde{t}_{j+1}, \xi_{J-j+1})$ and it has the slope equal with $\widetilde{g}_{J-j+2}$. The moments $\widetilde{t}_j$ are computed by the recursive relation

$$\widetilde{t}_{j+1} = \widetilde{t}_j + \frac{\widetilde{\xi}_{J-j+1} - \widetilde{\xi}_{J-j+2}}{\widetilde{g}_{J-j+2}} \ ,$$

where $\widetilde{t}_1 = t_1$. Instead of Eq. (5.2) we have

$$\widetilde{f}(t) = \xi_j + (t - \widetilde{t}_j)\widetilde{g}_{J-j+2} \ .$$

In Appendix C we show that if the trend is monotonic and the noise stationary, then the continuous analogue of the numerical method described above supplies an approximation of the trend contained in a noisy time series. The approximation improves if the trend is linear and the noise is symmetric. However, we want to apply the ACD algorithm to any time series $\{x_n\}$, even if it does not contain a monotonic trend (see Sect. 5.5) and then we introduce several modifications which assure the correct functioning of the algorithm in all cases.

The estimated trend $\widetilde{f}(t)$ is a piecewise linear function defined on $t \in [t_1, \widetilde{t}_{J+1}]$ and we denote by $\widetilde{T} = \widetilde{t}_{J+1} - t_1$ the length of the time interval resulting from the construction. We have to determine the values taken by $\widetilde{f}(t)$ at the sampling moments $t_n = t_1 + n\Delta t$. The problem is that the estimated length $\widetilde{T}$ is in general different from the length $T = t_N - t_1$ of the initial time series. We find the optimum correspondence using the translation or the scaling of the estimated trend $\widetilde{f}(t)$ with respect to the time series $\{x_n\}$.

If the estimated trend covers a longer period of time than the time series $\widetilde{T} > T$, then we translate the sampling moments through the distance $\tau \in [0, \widetilde{T} - T]$, so that only a portion of length T of the estimated trend is discretized $\widetilde{f}_n = \widetilde{f}(t_n + \tau)$. When $\widetilde{T} < T$ we translate the estimated trend through the distance $\tau \in [0, T - \widetilde{T}]$ and then for $t_n \in [\tau, \tau + \widetilde{T}]$ the estimated trend is discretized in the usual way $\widetilde{f}_n = \widetilde{f}(t_n - \tau)$. Outside this interval we extend the estimated trend assuming that it has a constant value: $\widetilde{f}_n = \widetilde{f}(0)$ for $t_n < \tau$ and $\widetilde{f}_n = \widetilde{f}(\widetilde{T})$ for $t_n > \tau + \widetilde{T}$. The optimum value τ_0 of the translation distance τ is obtained by minimizing the squared norm of the difference $\{x_n - \widetilde{f}_n\}$.

Another method is to expand or contract the interval $[t_1, \widetilde{t}_{J+1}]$ such that it is superposed over the interval $[t_1, t_N]$. Then the scaled discretized estimated trend is given by the equation

$$\widetilde{f}_n = \widetilde{f}\left[t_1 + \frac{n-1}{N-1}(\widetilde{T} - t_1)\right].$$

Between the translation or the scaling we choose the one which provides the minimum squared norm of the difference $\{x_n - \widetilde{f}_n\}$.

A numerical problem of the ACD algorithm described above is the possibility that some values $\widetilde{g}_j$ could be very close to zero and then the linear segments of the estimated trend in the corresponding intervals I_j would be almost parallel to the time axis. In such a case the estimated trend would be artificially deformed and its length would be much longer than the length of the initial time series ($\widetilde{T} \gg T$). In order to eliminate this possibility we impose the additional condition that the absolute value of the slope $\widetilde{g}_j$ should have an inferior bound. We denote by $n_j^{\min}$ ($n_j^{\max}$) the time step when the time series takes for the first time (the last time) a value in the interval I_j. The difference $n_j^{\max} - n_j^{\min}$ is the number of time steps during which the time series takes all the values contained in I_j. Then we choose the minimum value of the estimated trend slope as $|\widetilde{g}_j| \geq (\xi_{j+1} - \xi_j)/(n_j^{\max} - n_j^{\min})$.

If the quantities $\widetilde{g}_j$ have different signs, then the monotonic estimated trend $\widetilde{f}_n$ cannot be determined and the time series fluctuations are smoothed by means of the normalized form of the padding with zeros of the RCMA described in Sect. 4.1. Therefore the numerical algorithm of trend estimation using the ACD method consists in a succession of trend extractions and RCMA smoothings. We denote by $\{x_n^{(i)}\}$ the time series obtained after a succession of i extractions and smoothings. Initially $x_n^{(0)} = x_n$. If $\widetilde{g}_j$ have the same sign, then we estimate the component $\widetilde{f}_n^{(i+1)}$ of the estimated monotonic trend and we remove it from the previous residuals

$$x_n^{(i+1)} = x_n^{(i)} - \widetilde{f}_n^{(i+1)} \,. \tag{5.3}$$

If $\widetilde{g}_j$ do not have the same sign, then we consider that $\widetilde{f}_n^{(i+1)} = 0$ for all n and $\{x_n^{(i+1)}\}$ is computed by the RCMA. At the end, the estimated trend is the sum of the removed monotonic components

$$\widetilde{f}_n = \sum_{k=1}^{i} \widetilde{f}_n^{(k)}$$

and the estimated noise is

$$\widetilde{z}_n = x_n - \widetilde{f}_n \,.$$

The time series processing is interrupted if one of the following conditions is satisfied:
1. the standard deviation of the residuals $\{x_n^{(i)}\}$ is ρ times smaller than the noise standard deviation estimated from the initial time series using the method presented

in Appendix D (ρ is a given positive real number); 2. the standard deviation of $\{x_n^{(i+1)}\}$ is larger than that of the previous step $\{x_n^{(i)}\}$; 3. by adding the component $\{\widetilde{f}_n^{(i+1)}\}$ the estimated trend becomes nonmonotonic.

The initial CMAs are performed on small averaging windows such that the trend should be deformed as little as possible. For the first smoothing we use $K = 1$. If the quantities $\widetilde{g}_j$ do not acquire the same sign, then K is gradually increased by a unit for each new smoothing up to a maximum value K_f and then the next smoothings are computed keeping this value for K. In this way a compromise is made between the computing efficiency and the requirement that the trend should not be distorted when the noise is small.

5.2 Automatic ACD Algorithm

The ACD algorithm performance is analyzed using statistical ensembles of time series with as various as possible characteristics. We generate the time series according to Eq. (1.12) superposing realizations of an AR(1) process over a monotonic trend. The method is identical with that described in Sect. 2.3, but this time the trend is monotonic.

Since in Sect. 5.3 the ACD algorithm is compared with the polynomial fitting, we do not use polynomial trends. If the trend were of the same functional form as the function used in the polynomial fitting, then the ACD method would be disadvantaged. We choose the trend as the rational function

$$f(t) = t/(a - t), \tag{5.4}$$

$t \in [0, 1]$ with $a > 1$ because its slope $f'(t) = a/(a - t)^2$ has a nonhomogeneous distribution with its extreme values at the boundaries of the definition domain: the minimum at $t = 0$ and the maximum at $t = 1$. It is more difficult to estimate such a trend than one with the maximum slope in the interior of the definition domain because numerical algorithms are more difficult to be implemented near the boundaries of the time series. In our case the ratio between the maximum and the minimum slope is $f'(1)/f'(0) = (1 - 1/a)^{-2}$.

An artificial time series with a monotonic trend (5.4) and an AR(1) noise is characterized by four parameters: the time series length N, the trend parameter a, the correlation parameter of the noise ϕ, and the ratio r between the amplitudes of the trend variations and the noise fluctuations defined by Eq. (2.3). The resolution of the time series is varied with one order of magnitude choosing the time series lengths in the interval $N \in [100, 1000]$. The parameter a takes values in the interval $a \in [1.1, 2.0]$, so that the ratio of the extreme values of the slope varies between 121 and 4. As we have justified in Sect. 2.3, for the other two parameters we chose as variation range the intervals $\phi \in [0, 0.9]$ and $r \in [0.25, 4]$. The time series in Fig. 5.2 correspond to the extreme values of a for $\phi = 0.9$ and $r = 1$. The shape

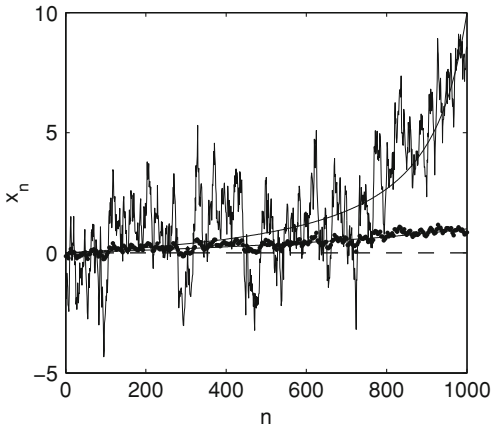

Fig. 5.2 Two artificial time series obtained by superposing an AR(1) noise with $\phi = 0.9$ over two monotonic trends with $a = 1.1$ and $a = 2.0$ such that $r = 1$ for both of them

of the two time series is different because the amplitude of the noise fluctuations is much larger for $a = 1.1$. In this case the trend variation is ten times greater than for $a = 2$ and, since the ratio r has the same value, it follows that in the first case the noise standard deviation is ten times larger.

In order to apply the ACD method as an automatic algorithm there are two things to be done: 1. to distribute the time series values into disjoint intervals I_j; 2. to determine the values of two parameters (the maximum value ρ of the ratio between the noise standard deviation estimated from the initial time series and the final residual standard deviation and the maximum value K_f of the semi-length of the averaging interval). In this section we show how these tasks can be automatically accomplished regardless of the statistical properties of the noise. We determine the values of the ACD parameters using numerical experiments with AR(1) noises. For noises of other types it is possible that other values of the parameters could be optimal.

When we distribute the time series values into disjoint intervals I_j, we have to take into account two opposite requests. If the noise fluctuations are much smaller than the trend variation, then it is recommended to use a large number of intervals I_j in order to describe the trend shape as accurately as possible. On the contrary, if the noise fluctuations are much larger, then it is better to use a small number of intervals I_j, each of them containing more time series values such that the noise fluctuations are smoothed as much as possible by the average in Eq. (5.1). Therefore we distribute the time series values into a number of disjoint intervals inversely proportional with the noise standard deviation σ_Z^{est} estimated by Eq. (D.6) derived in Appendix D which uses only the values of the initial time series

$$J = \left[(x_{\max} - x_{\min})/\sigma_Z^{\text{est}}\right], \tag{5.5}$$

where $x_{\max}$ and $x_{\min}$ are, respectively, the maximum and the minimum value of the time series $\{x_n\}$ and $[\cdot]$ is the integer number of the expression between brackets.

The value J obtained in this way could be too large or too small. We impose a minimum value to this parameter $J_{\min} = 2$ such that the trend estimated by the ACD algorithm should contain at least two linear segments. In order to obtain a maximum limit for J we use the minimal length $N_{\min} = 14$ determined in Appendix D for which, in average, a white noise is not confounded with a time series containing a monotonic trend. Hence if the value obtained from Eq. (5.5) is smaller than $J_{\min}$, then we impose $J = J_{\min}$ and if it is larger than $J_{\max} = [N/N_{\min}]$, then $J = J_{\max}$.

With the determined number of intervals J we can compute the number of the values x_n contained in the interval I_j

$$N_j = [jN/J] - [(j-1)N/J],$$

where $[\cdot]$ is the integer part function. The values of N_j differ from each other at most with a unit. The limits of the intervals I_j are equal with $\xi_j = (x_{N_j+1} + x_{N_j})/2$ for $j = 2, 3, \ldots, J$. The first and the last boundaries of the intervals are $\xi_1 = \min\{x_n\}$ and $\xi_{J+1} = \max\{x_n\}$, respectively.

After the intervals I_j have been determined, we analyze the influence of the parameters ρ and K_f on the ACD performance. First we consider the influence of ρ. Since no criterion to choose the values of the parameter K_f has been established, its values are randomly distributed over its variation range according to a homogeneous probability. The parameter K_f controls the speed of the damping of the noise fluctuations by averaging. The semi-length K_f of the maximum averaging window takes values in the interval $K_f/N \in [0.01, 0.1]$.

For the iteration i we introduce the ratio $\rho^{(i)} = \sigma_Z^{\text{est}}/\sigma_x^{(i)}$ where σ_Z^{est} is the estimation (D.6) of the noise standard deviation and $\sigma_x^{(i)}$ is the standard deviation of the residuals $\{x_n^{(i)}\}$ given by Eq. (5.3). We have to establish the threshold ρ such that when $\rho^{(i)} < \rho$ we stop the iterations of the ACD algorithm considering that the noise fluctuations have been sufficiently damped. It is known that the sample mean of a random variable with a Gaussian distribution has the standard deviation $\sqrt{S}$ times smaller than that of the random variable, where S is the number of realizations in the sample. Considering that the reduction by $\sqrt{S}$ times of the standard deviation is a measure of the maximum effectiveness that a statistical method can have, we assume that in the case of the ACD method the standard deviation of the final residuals can be at best $\sigma_Z/\sqrt{N}$ as well. Then we assume that the optimum value of ρ is $\rho_{\text{opt}} = \sqrt{N}$.

Figure 5.3 shows the dependence of the average accuracy of the ACD method on the parameter ρ for random values of the parameter K_f. We evaluate the accuracy of the trend estimation by means of the resemblance index η defined by Eq. (3.6) averaged on statistical ensembles of 100 time series with length $N = 1000$ and with different values of the parameters a and r. When $\rho \leq \rho_{\text{opt}}$, the average index $\langle\eta\rangle$ reaches a stationary value in all cases. This kind of behavior shows that the value $\rho_{\text{opt}} = \sqrt{N}$ assures the retrieval of all available information from the original time series. For $\rho > \rho_{\text{opt}}$, $\langle\eta\rangle$ increases for increasing ρ, showing the importance of a right choice of ρ, because if ρ has a too high value, then the trend is incompletely estimated.

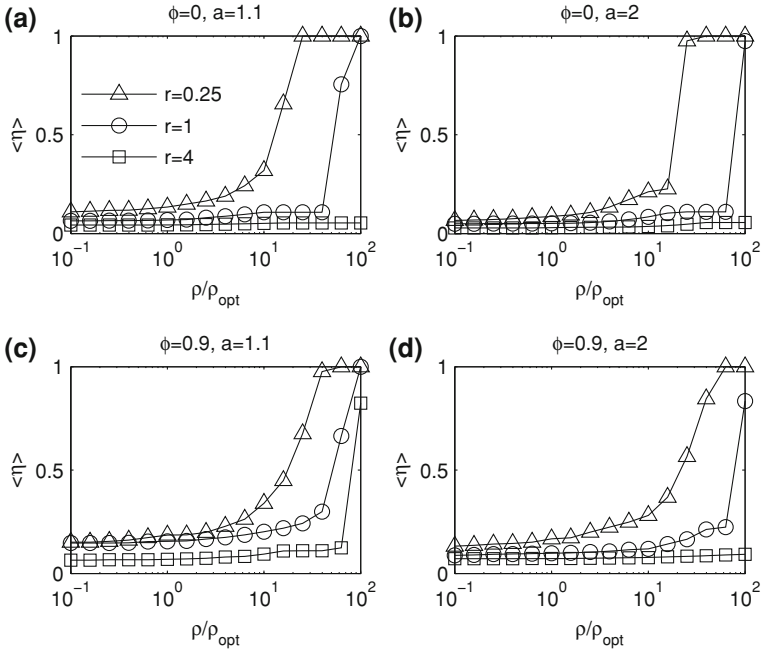

Fig. 5.3 The dependence of the accuracy of the ACD method on the parameter ρ for random values of the parameter K_f for time series with $N = 1000$ values

The average evaluation index $\langle\eta\rangle$ depends in the first place on the ratio r: $\langle\eta\rangle$ is small when the time series is dominated by trend ($r > 1$). The value of $\langle\eta\rangle$ depends on the parameters a and ϕ too, but to a smaller extent. When the serial correlation increases (Figs. 5.3c, d) or the trend slope increases (Figs. 5.3a, c) then the average index $\langle\eta\rangle$ increases too. When the time series length decreases, the accuracy of the ACD method decreases. For example, when $N = 100$, $\langle\eta\rangle$ increases roughly twice than for $N = 1000$ (results not shown in the figure). However, $\rho_{\text{opt}} = \sqrt{N}$ remains in the stationary region of the graphs. These results show that the optimum value of the parameter $\rho_{\text{opt}} = \sqrt{N}$ is correct.

With the optimum value of the first parameter established, we continue with the analysis of the parameter K_f. The larger K_f, the larger the averaging windows and the more strongly damped the noise fluctuations are, but at the same time the more distorted the trend is. Figure 5.4 contains the results for time series with $a = 1.1$ and $\phi = 0.9$ for which the worst estimations of the trend have been obtained in Fig. 5.3c. Over the entire variation range of K_f the average resemblance index $\langle\eta\rangle$ has approximately the same value. Also for the other values of a and ϕ the behavior of $\langle\eta\rangle$ is similar, therefore we can choose for K_f an arbitrary value in its variation range. To save computing time we use its maximum value $K_f = 0.1N$.

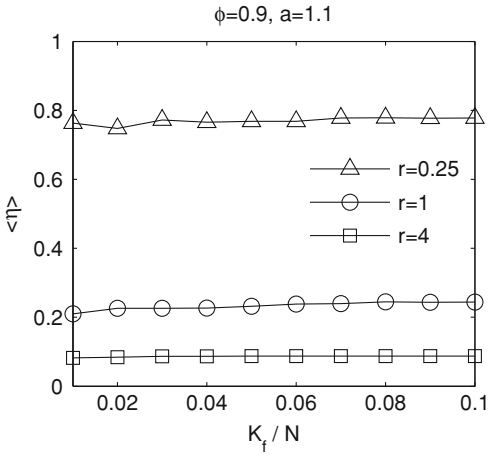

Fig. 5.4 The dependence of the accuracy of the ACD method on the parameter K_f for $\rho = \sqrt{N}$

The application of the automatic ACD algorithm to a given time series $\{x_n\}$ consists in several steps.[1] First σ_Z^{est} is computed from Eq. (D.6) and using it, the values x_n are distributed in the intervals I_j. Then the iterative process described in Sect. 5.1 is performed and the monotonic components $\{\widetilde{f}_n^{(i)}\}$ of the estimated trend are obtained. The process is stopped using the previously determined parameters $\rho_{\text{opt}} = \sqrt{N}$ and $K_f = 0.1N$. This automatic form of the ACD method for monotonic trend estimation was substantiated for time series with stationary noise, but it can be useful as well if the noise is nonstationary [5].

We exemplify how the ACD works for a time series with $N = 1000$ values containing a white noise ($\phi = 0$) and with the ratio $r = 3$ (Fig. 5.5a). After three averagings the first monotonic component of the estimated ACD trend is extracted (the dashed line in Fig. 5.5b). The residuals $\{x_n^{(4)} = x_n^{(3)} - \widetilde{f}_n^{(4)}\}$ obtained by subtracting this component from the averaged time series (Fig. 5.5c) still contain a monotonic variation extracted after other 8 additional averagings (Fig. 5.5d). The subtraction of the second component of the ACD trend causes the resemblance index to drop from $\eta^{(3)} = 0.099$ to $\eta^{(12)} = 0.075$. The variation of the second component is significantly smaller than that of the first component. The last component of the estimated trend extracted at the step $i = 64$ is decreasing and its global variation is ten times smaller than that of the previous component (Fig. 5.5f). As a result of its subtraction the resemblance index has a slight decrease reaching $\eta^{(64)} = 0.069$ and the algorithm stops because the residuals obtained after the subtraction of $\{\widetilde{f}_n^{(65)}\}$ satisfy the condition $\rho^{(65)} < \rho_{\text{opt}}$.

Figure 5.6 shows the average resemblance index for the final automatic form of the ACD algorithm applied to statistical ensembles with $S = 100$ artificial time series with the monotonic trend (5.4). In all cases, when N increases, the accuracy of the

[1] The automatic ACD algorithm is implemented by the MATLAB program `trendacd` freely accessible on web.

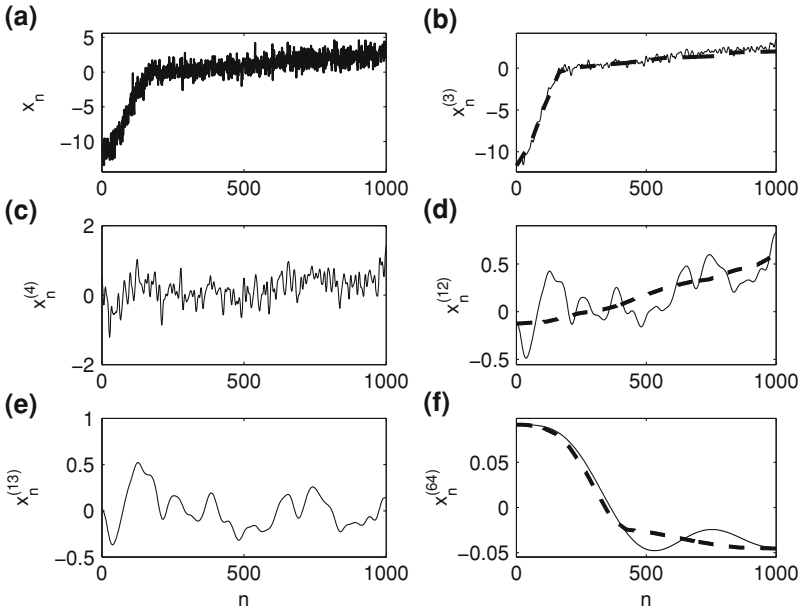

Fig. 5.5 The time series **a** and the residuals **b–f** resulted during the application of the ACD algorithm. The three monotonic components of an estimated trend are plotted with *dashed lines* in panels **b**, **d**, and **f**

trend estimation is improved. The strongest influence on the accuracy of the ACD method is that of the ratio r between the amplitudes of the trend variations and the noise fluctuations. When the time series is dominated by trend ($r = 4$), the average index $\langle \eta \rangle < 0.2$ indicates a very good resemblance between the estimated and the real trend for all the values of the other parameters. In the case of the time series with equal weight of the trend and the noise ($r = 1$), although $\langle \eta \rangle$ remains smaller than 0.5, however, it begins to depend significantly on the serial correlation and the trend slope. When the noise dominates the time series ($r = 0.25$), the index η increases significantly. If in addition the noise is strongly correlated ($\phi = 0.9$) and the trend slope is large ($a = 1.1$), we have $\langle \eta \rangle > 1$ indicating that the estimated trend does not contain any useful information on the real trend (Fig. 5.6c).

5.3 Evaluation of the ACD Algorithm

The evaluation of the automatic ACD algorithm is performed by comparison with the two classical methods of trend estimation presented in Chaps. 3 and 4. The trend is estimated for artificial time series with monotonic trends of the same type as those described in Sect. 5.2. The resemblance index η_{ACD} of the automatic ACD algorithm

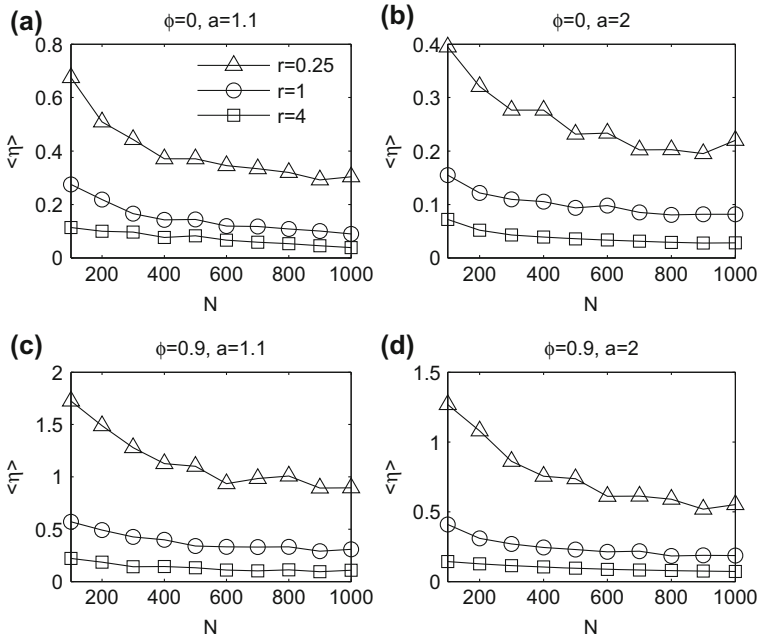

Fig. 5.6 The average resemblance index for the automatic ACD algorithm

is considered as a reference and is subtracted from the resemblance index of the other two algorithms.

Figure 5.7 shows the results of the comparison with the polynomial fitting of degrees from 1 to 10 for time series with $N = 1000$ values. We notice that the polynomial fitting does not dispose of a parameter equivalent to ρ of the ACD algorithm, therefore the polynomial trend is estimated in a single step and there is no possibility to control the accuracy of the final residuals. Since η_{ACD} does not depend on the polynomial degree, the curves plotted in Fig. 5.7 describe the dependence of the average $\langle \eta_{\text{poly}} \rangle$ on the polynomial degree, but translated with a constant $-\langle \eta_{\text{ACD}} \rangle$, different for each type of time series. The positive values of the quantity $\langle \eta_{\text{poly}} - \eta_{\text{ACD}} \rangle$ indicate that on average the accuracy of the ACD algorithm is better than that of the polynomial fitting.

For $q > 4$ the two methods have roughly equal accuracies, with small fluctuations depending on the values of the three parameters characterizing the time series (ϕ, r, and a). At small values of q, the ACD algorithm has a significantly larger accuracy than the polynomial fitting. The exceptions occur only for $\phi = 0.9$ and $r = 0.25$ when the estimated polynomial trend worsens as the polynomial degree increases (Figs. 5.7c, d). In these cases the time series are dominated by a strong correlated noise, so that the increase of q causes the estimated polynomial trend to be more influenced by the noise fluctuations. Hence the accuracy of the ACD algorithm is slightly better than that of the polynomial fitting.

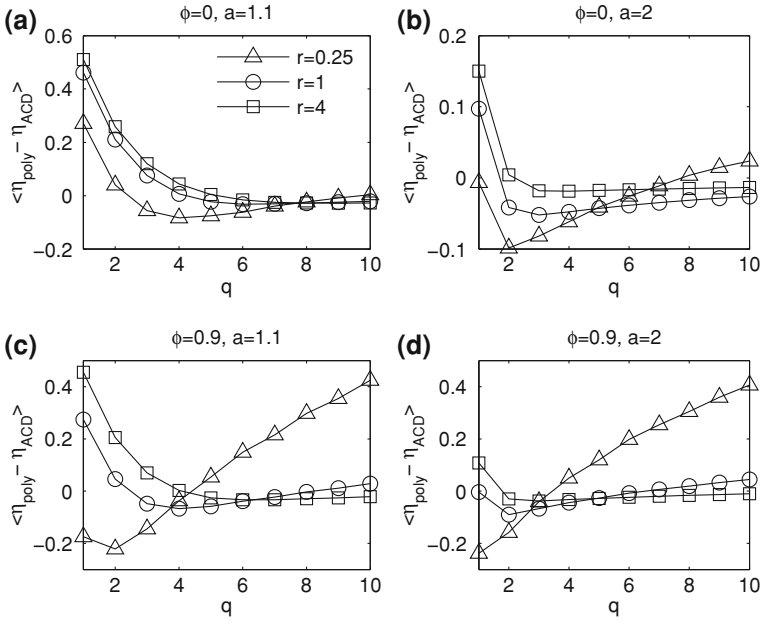

Fig. 5.7 Comparison of the accuracy of the automatic ACD algorithm with that of the polynomial fitting for time series with $N = 1000$

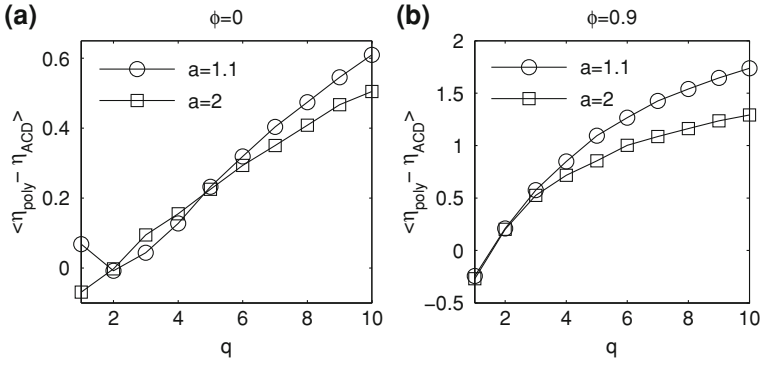

Fig. 5.8 Comparison of the accuracy of the automatic ACD algorithm with that of the polynomial fitting for $N = 100$ and $r = 0.25$

The accuracy of the estimated trend also depends on the time series resolution. For $N = 100$ the ACD algorithm has a significantly better accuracy than the polynomial fitting. Especially for time series dominated by noise ($r = 0.25$) the estimated polynomial trend is more dissimilar to the real trend, even for white noise (Fig. 5.8a). When the noise is strongly correlated ($\phi = 0.9$) the accuracy is several times smaller than for $N = 1000$ (Fig. 5.8b).

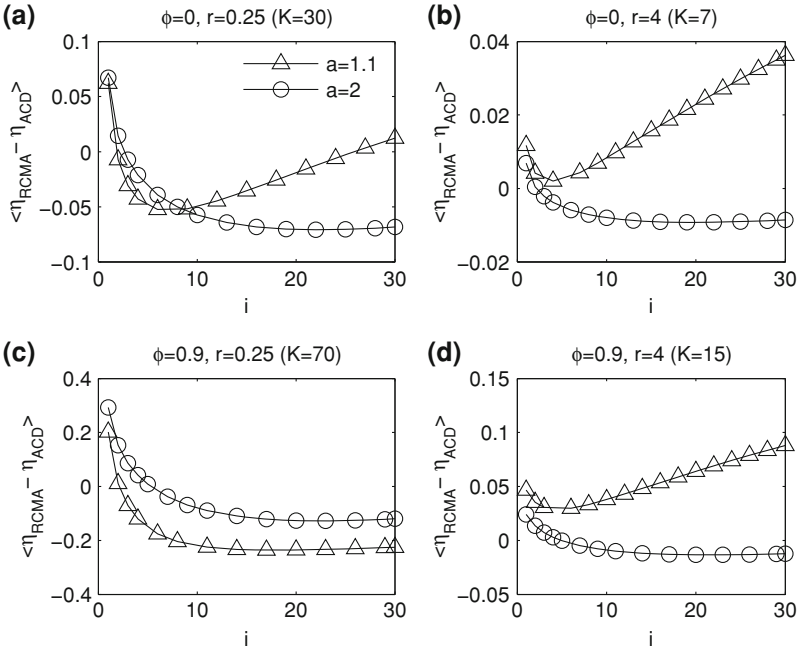

Fig. 5.9 Comparison of the accuracy of the automatic ACD algorithm with that of the RCMA for time series with $N = 1000$

The second comparison method is the RCMA analyzed in Chap. 4. Figure 5.9 contains the results of the comparison of the ACD algorithm with the RCMA for different values of the parameters ϕ, r, and a. We have chosen the values of K within the plateaus described in Sect. 4.1 such that the optimum value of the averagings number should be included in the same interval $i \in [1, 30]$. The numerical simulations are grouped differently than for polynomial fitting in Fig. 5.7.

For the time series with small trend slope ($a = 2$) the average resemblance index $\langle \eta_{\mathrm{RCMA}} \rangle$ is almost stationary for i around its optimum value and it is smaller than $\langle \eta_{\mathrm{ACD}} \rangle$. If the slope is steeper ($a = 1.1$), then the minimum of $\langle \eta_{\mathrm{RCMA}} \rangle$ is well-defined and most of its values are larger than $\langle \eta_{\mathrm{ACD}} \rangle$. The exception is the time series dominated by strongly correlated noise ($\phi = 0.9$ and $r = 0.25$) for which the average index $\langle \eta_{\mathrm{RCMA}} \rangle$ has an almost identical variation as that for $a = 2$ (Fig. 5.9c). We remark that, as expected, if the averagings number i is small and the time series is not smoothed enough, then the ACD trend has a better accuracy than that obtained by RCMA. These results indicate that the RCMA has a better accuracy in general, but it is necessary to take into account that the RCMA trend preserves many fluctuations from the smoothed noise.

From the comparisons presented above, we conclude that, on average, the ACD algorithm estimates monotonic trends with an accuracy comparable with that of the other two analyzed classical methods. But, unlike the ACD algorithm, the two clas-

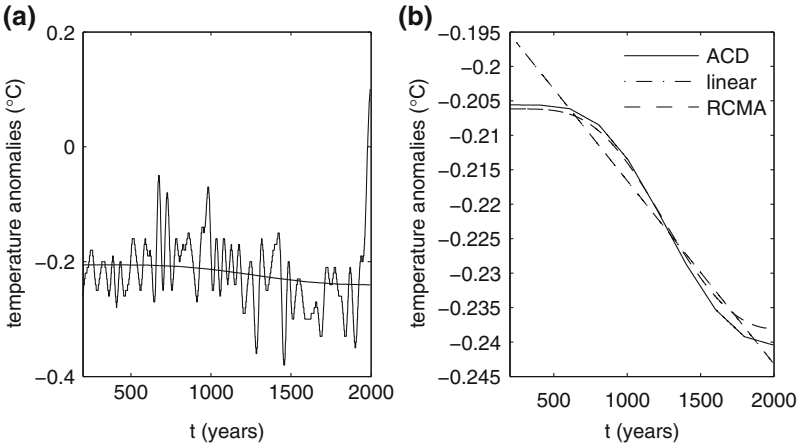

Fig. 5.10 Global mean annual temperature anomalies and the trend estimated by means of the ACD algorithm (*continuous lines*). In panel **b** the trends estimated by means of the linear fitting (*dash-dot line*) and RCMA (*dashed line*) are also presented

sical methods are not monotonic and can neither of them guarantee the monotony of the estimated trend. Analogue results were obtained by comparing the ACD algorithm with the jump process [6] which is a nonparametric estimation method of the same type as the RCMA [5].

5.4 A Climatological Example

The ACD algorithm removes only the monotonic component of a trend. To illustrate this special characteristic we estimate by means of the three algorithms compared with each other in the previous section the monotonic component of a real time series from paleoclimatology. Figure 5.10a shows the global mean annual temperature anomalies during the period A.D. 200–1995 with respect to the Northern Hemisphere mean annual temperature over 1856–1980 discussed in [2] and freely accessible from [3].

The temperature anomalies series contains $N = 1796$ values, many of them repeating themselves. In fact there are only 48 distinct values and the distribution of the time series values into disjoint intervals used in the ACD algorithm demands that all the values should be distinct. To satisfy this request without distorting the initial signal, we superpose on the original values quantities randomly generated with a homogeneous probability distribution on a range 1000 times smaller than the minimum difference between two distinct time series values.

The ACD algorithm has estimated the trend of the paleoclimatologic time series by a single component after 128 averagings. The time series values are distributed

into $J = 9$ intervals. In Fig. 5.10b the ACD estimated trend is compared with those estimated by the other two methods. Most of the trends estimated by nonlinear polynomial fitting are nonmonotonic. The linear trend, the only polynomial trend which is assuredly monotonic, exaggerates the global variation of the time series and does not provide any information on the slope variation over the analyzed time interval. In the case of the RCMA, the time series is averaged until all nonmonotonic variations of the estimated trend are eliminated. For $K = 10$, $i = 7096$ averagings are needed and the computing time becomes prohibitive.

We analyze the influence of noise fluctuations on the estimated trend by means of the surrogate time series obtained by superposing over the estimated trend $\{\widetilde{f}_n\}$ noises $\{\zeta_n\}$ as similar as possible with the real one. For each surrogate time series $\{\widetilde{f}_n + \zeta_n\}$ we estimate by means of the ACD algorithm a new trend $\{\widetilde{F}_n\}$ which differs from the initially estimated trend $\{\widetilde{f}_n\}$. We compare the two trends by their type of monotony. The fraction of the trends $\{\widetilde{F}_n\}$ having the same monotony as the initial trend $\{\widetilde{f}_n\}$ is a measure of the probability that the real trend is indeed monotonic.

In order to generate the surrogate noises $\{\zeta_n\}$ we have to choose a definite stochastic model for the estimated noise $\{\widetilde{z}_n = x_n - \widetilde{f}_n\}$. We use an AR(1) stochastic process with the parameters equal to the sample standard deviation $\widehat{\sigma}_Z$ of the estimated noise given by Eq. (1.6) and the serial correlation parameter computed by means of the sample autocorrelation function $\widehat{\phi}_\rho = \widehat{\rho}(1)$ discussed in Sect. 2.1. For the paleoclimatological time series in Fig. 5.10a we obtain $\widehat{\sigma}_Z = 0.0605\,^\circ\mathrm{C}$ and $\widehat{\phi}_\rho = 0.986$ indicating a strong serial correlation. From the estimated trends $\{\widetilde{F}_n\}$ for 400 surrogate series, only 169 are decreasing showing that the association of a monotonic trend with the initial climatological series is hazardous.

Climatologists are interested in the time periods with a monotonic temperature variation associated with geophysical processes of global scale, as for example, the global warming in the last century. We have applied the surrogate time series method described above to the global temperature anomalies over time intervals measured in centuries, i.e., intervals $[t_1, t_2]$ with $t_1 = 200, 300, \ldots, 1900$, $t_2 = 300, 400, \ldots, 1900, 1995$ and $t_2 > t_1$. We have used 400 surrogates for each time interval. The results are presented in Fig. 5.11. There are only five time periods (marked in the figure with bigger algebraic signs) to which we can associate a monotonic trend with a probability greater than 0.95. The three adjacent big plus signs correspond to three periods with lengths 100, 200, and 300 years, all ending in 1995. Their significant monotonic increase is due to the global warming in the twentieth century. Another period with significant temperature increase is [400, 1000]. The big minus sign indicates the temperature decrease in the fifteenth century. For 7 time periods to which no sign is attached in the figure, the ACD algorithm did not succeed to associate a monotonic trend to the temperature variation. The rest of the estimated trends cannot be considered "significant", conclusion which coincides with that in [2].

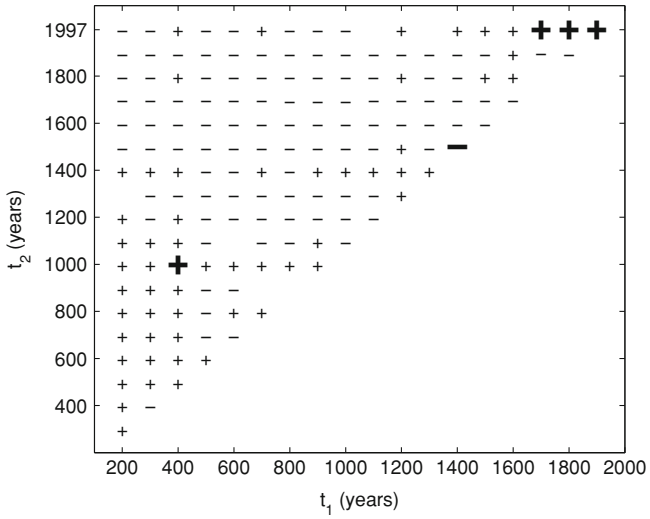

Fig. 5.11 Time periods $[t_1, t_2]$ over which the estimated variation of the global average temperature has an increasing ($+$*sign*) or decreasing ($-$*sign*) trend. The *big signs* indicate the periods with a significant monotonic trend component

5.5 Monotonic Components of Nonmonotonic Trends

So far in this chapter we have analyzed only time series with monotonic trend. In this section we consider time series without trend or with nonmonotonic trend. In these cases, as well, the trend evaluated by ACD is monotonic and it may or may not correspond to a monotonic component of the actual trend. We can quantitatively evaluate this property if we compare the results obtained on the same statistical ensemble by the ACD algorithm with the results of the Mann–Kendall test for monotony detection discussed in Sect. 2.4.

First we analyze time series without trends, containing only a stationary noise. In order to obtain a complete analysis of the ACD algorithm performance, we process statistical ensembles of $S = 1000$ time series of lengths down to the inferior limit of validity of the Mann–Kendall test $N = 10$. Figure 5.12a shows the probability α_{MK} of the Mann–Kendall test with 5 % significance level (see Sect. 2.4) applied to an AR(1) noise without trend, in terms of the correlation parameter ϕ, for different lengths of the time series. This is the probability to erroneously consider that an AR(1) noise is in fact a monotonic time series.

The ACD algorithm is not a statistical test and the analogue of the probability α_{MK} is the probability α_{ACD} that the ACD algorithm estimates a monotonic component for an AR(1) noise. For the same time series as those used for the Mann–Kendall test, the ACD algorithm estimates a monotonic trend much more frequently than it could occur as a random event (Fig. 5.12a). In fact for all time series lengths and

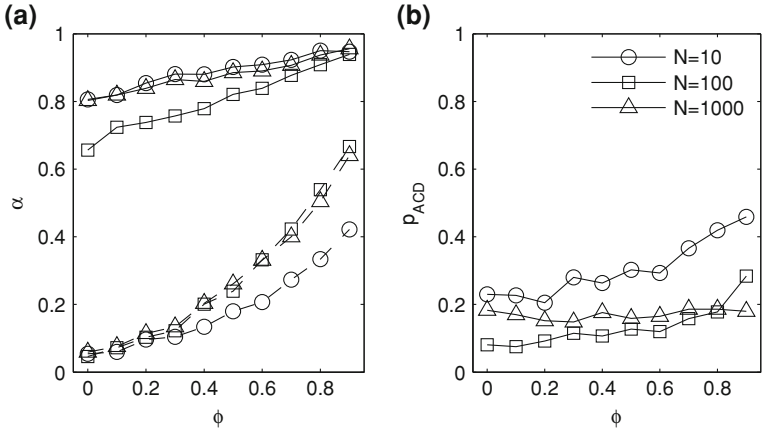

Fig. 5.12 **a** The probability α_{MK} of the Mann–Kendall test with 5 % significance level applied to finite AR(1) noises without trend (*dashed line*) and the probability α_{ACD} that the ACD algorithm estimates a monotonic component for the same time series (*continuous line*). **b** The probability p_{ACD} that more than 95 % of surrogate time series have the same monotony as the ACD trend

serial correlations, a monotonic component is attributed to more than 70 % of the AR(1) noises. But the majority of the estimated monotonic trends have negligible slope being almost constant. Unlike α_{MK}, the probability α_{ACD} increases when the time series length increases only for $N > 40$. For shorter time series this dependence is reversed because the ACD algorithm imposes that the time series values should be distributed in only two intervals ($J = 2$).

For AR(1) noise the Mann–Kendall test can be compared with the surrogate method to assess the statistical significance of the ACD trend presented in Sect. 5.4. In Fig. 5.12b we plot the probability p_{ACD} that more than 95 % of surrogate time series have the same monotony as the ACD trend estimated for the AR(1) noise. The probability p_{ACD} is computed for the same statistical ensembles as the Mann–Kendall statistic and it is much smaller than the probability α_{ACD}, i.e., only a small fraction of the estimated monotonic components are significant. The Mann–Kendall test is designed to test the monotony of a time series, not the existence of a monotonic component, but for a noise without trend the two hypotheses are identical. The comparable values of α_{MK} and p_{ACD} in Fig. 5.12 show that the surrogate method can be used as a test for the existence of a monotonic component in a time series.

To analyze the behavior of the ACD algorithm for nonmonotonic trends, we introduce a trend composed of a monotonic linear trend and a sinusoid

$$f(t) = pt + c \sin 2\pi t, \tag{5.6}$$

where $t \in [0, 1]$ and p and c are positive real parameters. This formula allows us to make a continuous transition from a monotonic to a nonmonotonic trend. When $c = c_0 = p/(2\pi) \simeq 0.159$, the function $f(t)$ has an inflection point at $t = 0.5$ with

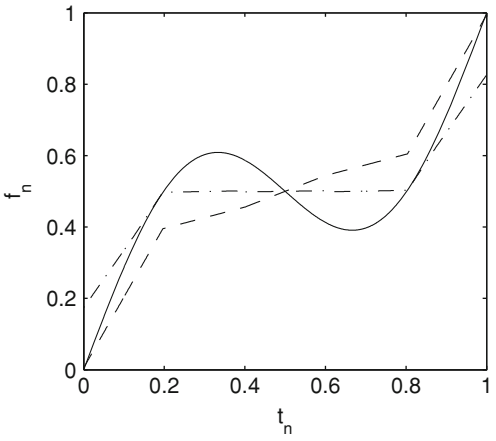

Fig. 5.13 The monotonic trend estimated by the ACD algorithm (*dashed line*) from a nonmonotonic trend (*continuous line*). The *dash–dot line* is the rejected estimated trend obtained after $i = 25$ steps of the ACD algorithm

the tangent parallel to the Ox axis. Therefore, if $c \leq c_0$ the trend is monotonic, and if $c > c_0$ it is nonmonotonic. The parameter p controls the amplitude of the global variation of the trend.

In Fig. 5.13 we give an example of monotonic component estimated by the ACD algorithm for a nonmonotonic trend without superposed noise. We use a time series with $N = 1000$ values obtained by discretizing the nonmonotonic trend (5.6) with $p = 1$, $c = 2c_0$ and without superposed noise. It contains the monotonic linear trend $f_{\text{mon}}(t) = t$ and the sinusoid that is an oscillation without any monotonic component. The trend estimated by the ACD algorithm (the dashed line in Fig. 5.13) is significantly different from the monotonic component $f_{\text{mon}}(t)$ of the trend, which is a straight line.

In Sect. 5.2 we have designed the automatic ACD algorithm to process time series containing a stationary noise. The time series in Fig. 5.13 contains only a deterministic trend and we have to discuss the behavior of the ACD algorithm in this case. Although there is no noise, from the condition (D.5) we obtain $d_0 = 276$ which is the temporal scale for which the variation of the differenced trend $\{\nabla_d f_n\}$ is dominated by the single oscillation present in the trend. That is why the value of d_0 is approximately equal to the distance between the abscissas of the two local extrema. From Eq. (D.6) we obtain $\sigma_Z^{\text{est}} = 0.182$ which measures the order of magnitude of the oscillation of the trend $f(t)$ and Eq. (5.5) gives the number of the intervals $J = 5$ dividing the time series values.

Without any averaging, from the initial time series, the monotonic component plotted by the dashed line in Fig. 5.13 is obtained. It replaces the oscillation of the trend $f(t)$ by a part of the estimated trend with a smaller slope than the rest of the estimated trend. After 23 averagings, another monotonic component is extracted, this time a decreasing one. The sum of the two monotonic components is plotted in Fig. 5.13 with the dash–dot line. This new estimated trend replaces the oscillation of the initial trend $f(t)$ by an almost horizontal plateau. But because the plateau

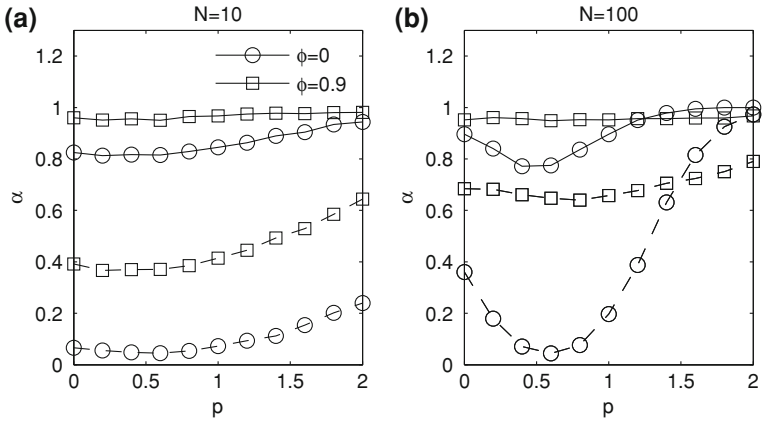

Fig. 5.14 The probability α_{MK} of the Mann–Kendall test with 5 % significance level applied to nonmonotonic trends with AR(1) noises (*dashed line*) and the probability α_{ACD} of the ACD algorithm for the same time series (*continuous line*)

contains several very small fluctuations, the new estimated trend is not monotonic and the ACD algorithm rejects it and stops the iterations. However, the second estimated trend could be easily transformed in a monotonic one. Hence both the two estimated trends could be considered as possible monotonic components of the initial trend.

This example shows that the ACD algorithm also works for deterministic time series, not only for noisy ones. The noise fluctuations are replaced by the oscillations of the deterministic trend and d_0 is proportional with the average time scale of these oscillations and σ_Z^{est} with their amplitude. We also remark that for a nonmonotonic trend one cannot unequivocally define a monotonic component and the ACD monotonic estimated trend is only one of the infinity of the existing possibilities.

In Fig. 5.14 we extend the analysis in Fig. 5.12 for time series with nonmonotonic trends. AR(1) noises with $\sigma_Z = 1$ are superposed over the nonmonotonic trends described by Eq. (5.6) with $c = 2c_0$ and different values of the parameter $p \in [0, 2]$. When $p = 2$, the trend becomes monotonic with an inflection point at $t = 0.5$. Like in Fig. 5.12, the probability α_{ACD} that the ACD algorithm estimates a monotonic component is much greater than the Mann–Kendall probability α_{MK}, reinforcing the previous conclusion that the ACD algorithm estimates the monotonic component of the trend.

References

1. Andreas, E.L., Trevino, G.: Using wavelets to detect trends. J. Atmos. Ocean. Technol. **14**, 554–564 (1997)
2. Jones, P.D., Mann, M.E.: Climate over past millennia. Rev. Geophys. **42**, 1–42 (2004)
3. Jones, P.D., Mann, M.E.: Climate over past millenia. Data contribution series #2004-085. In: NOAA/NGDC Paleoclimatology Program, http://www.ncdc.noaa.gov/paleo/pubs/jones2004/ (2004)
4. Kendall, M.G.: Rank Correlation Methods. Griffin, London (1975)
5. Vamoş, C.: Automatic algorithm for monotone trend removal. Phys. Rev. E **75**, 036705 (2007)
6. Zhao, S., Wei, G.W.: Jump process for the trend estimation of time series. Comput. Stat. Data Anal. **42**, 219–241 (2003)

Chapter 6
Estimation of Monotonic Trend Segments from a Noisy Time Series

An arbitrary nonmonotonic trend is composed by a succession of monotonic segments limited by its local extrema. A superposed noise breaks up the trend monotonic variations into many small fluctuations, but the global shape of the trend is recognizable because the trend local extrema have a larger time scale than those induced by noise. By rigorously defining the time scale of a local extremum we design an automatic algorithm to estimate the trend local extrema from a noisy time series. The estimation accuracy is improved if the noisy time series is first smoothed such that the noise fluctuations are damped. Using the ACD algorithm for monotonic trend estimation presented in the previous chapter we evaluate the significance of the estimated local extrema. As an example we analyze a biophysical time series for which we estimate the large scale monotonic segments of the trend.

6.1 Time Scale of Local Extrema

Representation of the time series by means of the distribution of their local extrema or monotonic parts is used in data mining to preprocess the time series. The result is a segmentation of the time series in a sequence of intervals over which a certain property is constant [2]. For example, the time series can be represented as a set of straight line segments [4], solution which, for monotonic trends, resembles with the ACD algorithm presented in Chap. 5. Segments with a dominant monotonic variation, not necessarily linear, can be obtained using some of the local extrema of the time series. But the importance of different local extrema for the time series shape significantly varies and we have to quantify it by introducing the notion of scale. For simple quasiperiodic signals the local scale is the distance between two successive local extrema [3]. For more complex time series it is possible to introduce a parameter of scale by smoothing the signal with a mask of variable size and to establish the local extrema which are not excluded by smoothing [8]. By continuously varying the scale parameter one can track local extrema as they change with the scale and identify the

C. Vamoş and M. Crăciun, *Automatic Trend Estimation*, SpringerBriefs in Physics,
DOI: 10.1007/978-94-007-4825-5_6, © The Author(s) 2012

significant local extrema for a given time scale [5, 8]. In this chapter we define and analyze the time scale of a local extremum in a similar way as in [1].

Let us consider a time series $\{x(n), n \leq N\}$.[1] The value $x(n)$ in the interior of the time series ($1 < n < N$) is a local maximum if $x(n) > x(n-1)$ and $x(n) > x(n+1)$. For the time series boundaries the condition is reduced to a single inequality $x(1) > x(2)$ or $x(N) > x(N-1)$, respectively. The inverse inequalities define a local minimum. The number of the local extrema of a noisy time series can be quite large and their importance for the time series shape can be very different. In order to characterize a local extremum $x(n)$ we introduce its time scale as the maximum distance over which $x(n)$ is also a global extremum.

We say that Δn is the *time scale* of the local maximum $x(n)$ if

$$\Delta n = \max \left\{ d \leq N \mid x(n) > x(m) \text{ for } m \neq n \text{ and } d_{\inf} \leq m \leq d_{\sup} \right\} , \qquad (6.1)$$

where $d_{\inf} = \min\{1, n-d\}$ and $d_{\sup} = \max\{n+d, N\}$. The time scale of a local minimum is defined similarly but by the reverse inequality $x(n) < x(m)$. Because the definition is given by means of strict inequalities, it holds only for time series with distinct values. The case of time series with repeating values is discussed at the end of this section.

We assume that the time series $\{x(n), n \leq N\}$ has J local extrema located at n_j, $j = 1, 2, \ldots, J$. They are ordered in ascending order of their positions $n_j < n_{j+1}$ for all j. We denote by Δn_j the time scale of the local extremum located at n_j. The importance of a local extremum for the shape of the time series is given by the magnitude of its time scale. The global extrema have the maximum time scale equal with N. All the other local extrema have smaller time scales. Then we can obtain a description characterized by a given time scale Δn of the time series by the set of the local extrema with $\Delta n_j > \Delta n$. We want to construct a *partition of scale* Δn as a succession of local maxima and minima such that the segments between them approximate the parts of the time series with a monotonic variation. But generally the local extrema with $\Delta n_j > \Delta n$ do not form an alternation of maxima and minima, therefore we have to add to the partition some local extrema with $\Delta n_j < \Delta n$.

It is possible that two local extrema of the same type with time scales larger than Δn succeed one to another as the minima at n_j and n_{j+2} in Fig. 6.1a. The line segment of length $2\Delta n$ centered at the local maximum $x(n_{j+1})$ intersects the time series graph so that $\Delta n_{j+1} < \Delta n$. In such a situation, the local maximum $x(n_{j+1})$ is inserted between the minima $x(n_j)$ and $x(n_{j+2})$ in the partition of scale Δn.

It is also possible that a local minimum is larger than the neighboring local maximum as in Fig. 6.1b. In this case between the two local extrema at n_j and n_{j+3} with time scales greater than Δn there are two additional local extrema at n_{j+1} and n_{j+2} with time scales smaller than Δn. The additional minimum $x(n_{j+2})$ is smaller than the local maximum $x(n_{j+3})$ and the additional maximum $x(n_{j+1})$ is larger than the local minimum $x(n_j)$, so that the variation of the time series in the interval $[n_j, n_{j+3}]$

[1] In this chapter we use the notation $x(n)$ instead of x_n.

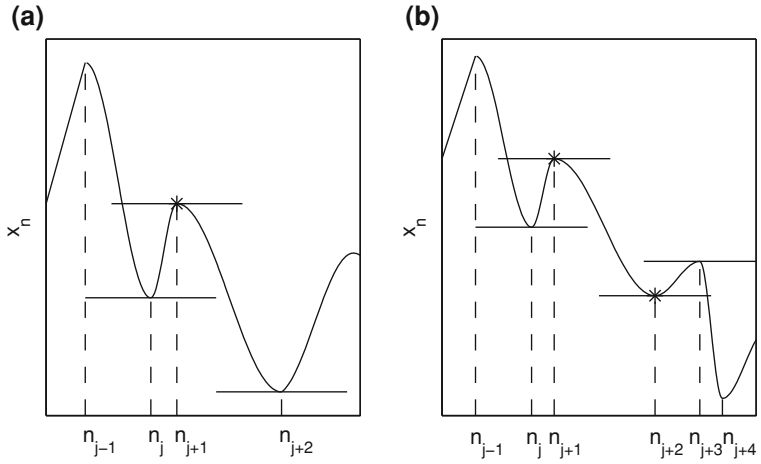

Fig. 6.1 Local extrema with $\Delta n_j < \Delta n$ are marked with an *asterisk* and are added to a partition of scale Δn

cannot be considered monotonic. In order to correct such a situation we include the two intermediate additional extrema in the partition of scale Δn.

Once the additional local extrema are added to the partition of scale Δn, it is necessary to verify if the boundaries of the time series have the time scale greater than Δn. If either of them does not, we add it as a local extremum. In this way the entire time series is partitioned by the succession of local maxima and minima into segments which can be approximated by monotonic curves. We denote by $\widetilde{J}_{\Delta n}$ the number of the local extrema belonging to the partition of scale Δn and by $\widetilde{n}_j(\Delta n)$ their positions. If $\Delta n = 1$, then all the local extrema are included in the partition and $\widetilde{J}_1 = J$. In Appendix E we present an algorithm to iteratively identify the local extrema of a partition with a given Δn.[2]

In Fig. 6.2 we give an example of partitions of a trend without noise for a succession of time scales. The trend is numerically generated by the method described in Sect. 2.2 for $N = 1000$, $P = 50$, and $\Delta N_{\min} = 10$. Because we want to obtain a large variety of shapes of the generated trends, we randomly choose the coefficient $c_p = \pm 1$ so that the trends have a number of monotonic segments smaller than P. In the example in Fig. 6.2a the trend has 22 monotonic segments and $J = 23$ local extrema. The positions of the maxima in Fig. 6.2a are indicated by continuous vertical lines and the positions of the minima by dashed vertical lines. We have partitioned this time series for time scales distributed according to the powers of 2, $\Delta n \in \{1, 2, 2^2, \ldots, 2^9, 1000\}$. In Fig. 6.2b we identify by circular markers the positions of the local extrema of time scales greater than Δn and by asterisks the positions of the additional local extrema used to obtain an alternation of maxima and minima.

[2] The automatic algorithm to build a partition of time scale Δn is implemented by the MATLAB program `local_extrema` freely accessible on web.

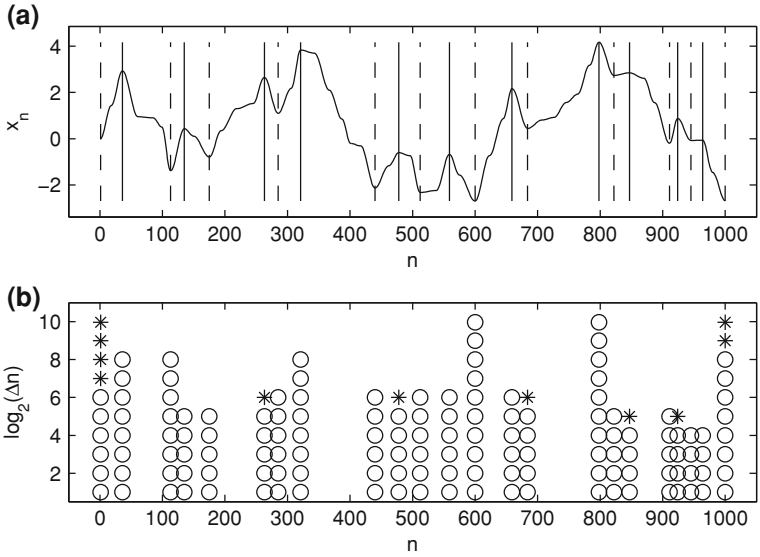

Fig. 6.2 The partitions at different time scales Δn of a nonmonotonic trend without a superposed noise

If $\Delta n \leq 2^4 = 16$, then the partition contains all the local extrema of the trend. When $\Delta n = 2^5 = 32$ the three local extrema between the minimum at $n_{19} = 911$ and the end of the trend have the time scales smaller than $\Delta n = 32$. For example the local maximum at $n_{20} = 924$ is not of scale $\Delta n = 32$ because there are greater values $x(n)$ for $n > n_{20} - \Delta n = 892$. Between the two local maxima of scale $\Delta n = 32$ situated at $n = 911$ and $n = 1000$ the additional minimum marked with an asterisk is inserted at $n = 924$.

Increasing further the time scale Δn, the less significant variations of the trend are gradually eliminated. For $\Delta n = 2^6 = 64$ all the local extrema between the local maximum at $n_{16} = 798$ and the end of the time series are ignored and this segment of the trend may be considered monotonic at this time scale. The local extrema at $n_4 = 135$ and $n_5 = 175$ are also considered irrelevant at this time scale. For the next two values of Δn only 7 local extrema are left and they approximate even at a larger scale the monotonic parts of the trend. For the last two values of the time scale ($\Delta n = 2^9 = 512$ and $\Delta n = N = 1000$) only the global extrema of the time series and its boundaries remain.

We apply the partitioning algorithm to a statistical ensemble of 1000 time series composed of trends without noise with the same parameters as those of the trend in Fig. 6.2. Figure 6.3 shows the average ratio of the number $\widetilde{J}_{\Delta n}$ of estimated local extrema of a partition with time scale Δn and the real number J of local extrema. The average $d_{\min}$ on the statistical ensemble of the minimum distance between two successive local extrema is represented by a dashed vertical line. For Δn smaller than $d_{\min}$ the partition contains all the local extrema of the time series. For larger values

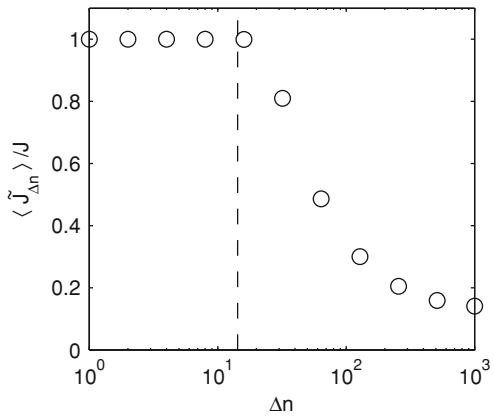

Fig. 6.3 The average relative number of the estimated local extrema as a function of the time scale Δn

of Δn the number of the estimated extrema $\widetilde{J}_{\Delta n}$ decreases reaching the minimum when the time scale equals the length of the time series N.

Now we extend the definition (6.1) for time series containing repeating values. An example of such a time series is the paleoclimatological data analyzed in Sect. 5.4. Let us consider that $x(n_1) = x(n_1 + 1) = \cdots = x(n_2) = x_0$, with $x(n_1 - 1) \neq x_0$ and $x(n_2 + 1) \neq x_0$. The constant value x_0 is a maximum of scale Δn if

$$\Delta n = \max \left\{ d \leq N \,|\, x_0 > x(m) \text{ for } m \neq n \text{ and } d_{\text{inf}} \leq m \leq d_{\text{sup}} \right\},$$

where $d_{\text{inf}} = \min\{1, n_1 - d\}$ and $d_{\text{sup}} = \max\{n_2 + d, N\}$. The position of this local extremum is considered the middle of the interval $[n_1, n_2]$.

6.2 Local Extrema of Noisy Time Series

Due to the noise fluctuations, the local extrema $\{x(n_j)\}$, $j \leq J$, of a noisy time series are much more numerous than those of the trend contained in it. To distinguish them we denote the positions of the trend local extrema by $\{n_j^{(0)}\}$, $j \leq J^{(0)}$. But the time scales of the local extrema related to noise are significantly smaller than the time scales of the trend local extrema. Then a partition of a scale $\Delta n > 1$ filters out the local extrema with time scales smaller than Δn and its estimated local extrema located at $\{\widetilde{n}_j(\Delta n)\}$, $j \leq \widetilde{J}_{\Delta n}$, approximate the local extrema of the trend located at $\{n_j^{(0)}\}$. In this section we determine the time scale for which the best approximation is obtained. Analogue tasks are performed, for instance, by the min–max filter [7] or the multi-scale method to determine the significant local extrema of a time series [8].

We analyze the partitioning of a single noisy time series obtained by superposing a Gaussian white noise on a trend generated according to the method described in Sect. 2.3 (Fig. 6.4a). The trend is characterized by the parameters $N = 1000$, $P = 30$,

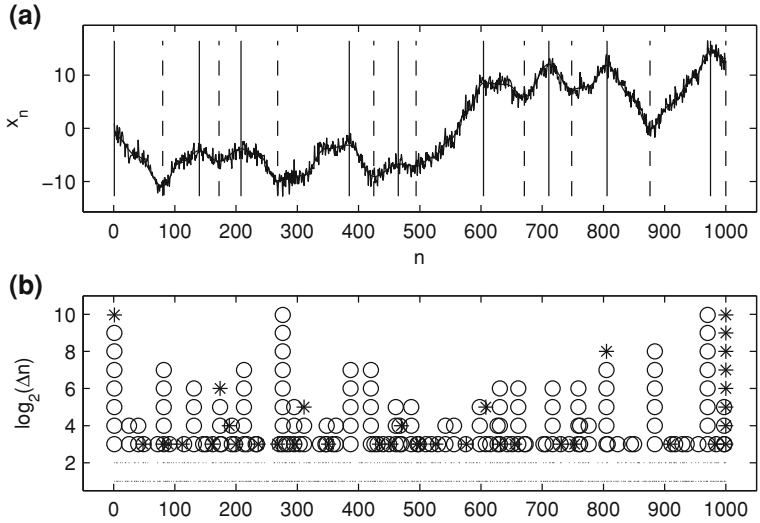

Fig. 6.4 The partition in monotonic segments of a nonmonotonic trend on which a white noise is superposed

and $\Delta N_{\min} = 20$. The time series is dominated by trend ($r = 4$) so that the influence of the noise is small, the trend local extrema are not significantly distorted and our presentation does not become too intricate. The positions of the trend maxima are indicated by continuous vertical lines and the minima by dashed vertical lines. The partitions for different time scales are presented in Fig. 6.4b.

For $\Delta n = 1$ we obtain all $\widetilde{J}_1 = J = 663$ local extrema of the noisy time series graphically marked by points, not by circles. Although reduced, the number of the estimated extrema for the next two time scales remains very large. For $\Delta n = 2^3 = 8$ we obtain $\widetilde{J}_8 = 106$, much larger than the number of the trend local extrema $J^{(0)} = 18$. The closest number of the estimated extrema is obtained for $\Delta n = 2^6 = 64$, $\widetilde{J}_{64} = 16$. For $\Delta n \geq 2^9 = 512$ only the global extrema of the time series and its boundaries are included in the partition. Hence the number of the estimated local extrema does not coincide with that of the trend local extrema for any of the time scales.

Not only the number of the estimated local extrema is affected by the noise presence, but their positions too. For example, the global maximum of the trend is located at $n_{17}^{(0)} = 975$, whereas for all Δn the estimated one at $n = 970$. It is difficult to evaluate the importance of these errors in comparison with the errors of the number of the estimated extrema. For instance, for $\Delta n = 2^5 = 32$ we obtain five local extrema ($\widetilde{J}_{32} = 23$) more than the number of the trend local extrema ($J^{(0)} = 18$). We have to establish if the influence of those five additional extrema is more important in the total approximation error of the trend local extrema than the error due to the two neglected local extrema for $\Delta n = 2^6 = 64$. The partitioning errors related to the position of the estimated local extrema considerably increase

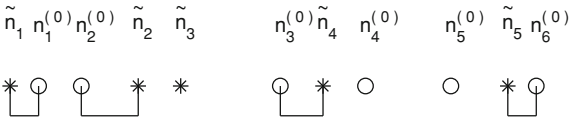

Fig. 6.5 Coupling between the trend maxima (*circles*) and estimated maxima (*asterisks*) for a partition of scale Δn

with the noise amplitude, therefore it is necessary to define a quantitative method to evaluate the approximation of the trend local extrema in which both types of errors should be included.

First we establish the estimated local extrema which correspond to the local extrema of the trend. We have to compare only the local extrema of the same type (the estimated local maxima with the trend local maxima and the minima with each other). In the following we assume that the maxima and the minima have been separated and we discuss only the case of the maxima. As we have previously stated, we denote by $\{n_j^{(0)},\, j = 1, 2, \ldots, J^{(0)}\}$ the positions of the local maxima of the trend and by $\{\widetilde{n}_j(\Delta n),\, j = 1, 2, \ldots, \widetilde{J}_{\Delta n}\}$ the positions of the estimated local maxima for a given time scale Δn. We have to decide for a local maximum $x(n_j^{(0)})$ of the trend if there exists an estimated local maximum $x(\widetilde{n}_k)$ which approximates it. The criterion is that there should be no other estimated maximum closer to $x(n_j^{(0)})$ than $x(\widetilde{n}_k)$ and no other trend maximum closer to $x(\widetilde{n}_k)$ than $x(n_j^{(0)})$. In the following we describe an iterative algorithm to determine the couples of the estimated and trend maxima.

We denote by $\{m_j,\, j = 1, 2, \ldots, J^{(0)} + \widetilde{J}_{\Delta n}\}$ the sequence obtained by joining together the two sequences of the maxima positions $\{n_j^{(0)}\}$ and $\{\widetilde{n}_j(\Delta n)\}$ and arranging it in increasing order. The distances between the successive local maxima, no matter whether they belong to the trend or to the estimated one, are given by the differenced sequence $\{\nabla_1 m_j = m_{j+1} - m_j\}$. First we analyze the pair of local maxima for which $\nabla_1 m_j$ is minimum. If one of the two local maxima is estimated and the other characterizes the trend, then we couple them in the sense that the maximum of the partition approximates the maximum of the trend. If the two maxima belong both of them to the partition or to the trend, then we do not couple them. We continue to couple the local maxima in increasing order of the terms $\nabla_1 m_j$ with the difference that if one of the two maxima has been previously included in a pair, then we do not couple it again. We do the same for the sequence of local minima.

For instance, for the repartition of trend and estimated local maxima in Fig. 6.5 the coupled maxima are the pairs located at $(n_1^{(0)}, \widetilde{n}_1)$, $(n_2^{(0)}, \widetilde{n}_2)$, $(n_3^{(0)}, \widetilde{n}_4)$, and $(\widetilde{n}_5, n_6^{(0)})$. After the maxima pairing is realized, a certain number of trend and estimated local maxima remain not included in any pair. In our example, $n_4^{(0)}$ and $n_5^{(0)}$ are not estimated and $\widetilde{n}_3$ approximates no trend maximum.

We define the partitioning error proportional to the sum of the distances between all the trend and estimated local extrema. First we sum the distances between the trend and estimated local extrema contained in all the pairs previously formed. In our example

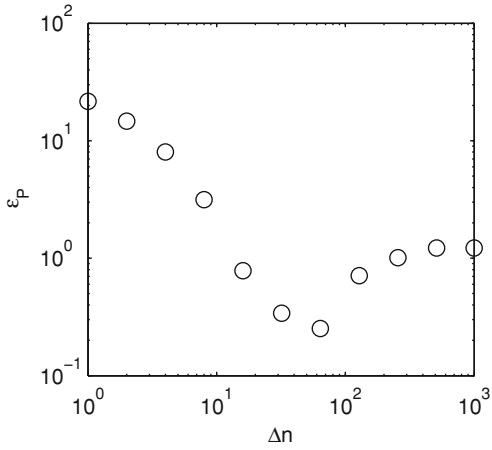

Fig. 6.6 The partitioning error ε_P for a single noisy time series with respect to the time scale Δn

$$d_1 = |n_1^{(0)} - \widetilde{n}_1| + |n_2^{(0)} - \widetilde{n}_2| + |n_3^{(0)} - \widetilde{n}_4| + |\widetilde{n}_5 - n_6^{(0)}| \, .$$

Then we sum the distances between the estimated local extrema which are not included in a pair and the nearest trend local extremum. For the example in Fig. 6.5

$$d_2 = |n_2^{(0)} - \widetilde{n}_3| \, .$$

Finally we sum the distances between the trend local extrema which are not included in a pair and the nearest trend local extremum

$$d_3 = |n_3^{(0)} - n_4^{(0)}| + |n_5^{(0)} - n_6^{(0)}| \, .$$

In this way we take into account that it is difficult to separate by partitioning two close trend local extrema. The partitioning error $\varepsilon_P(\Delta n)$ is obtained by dividing the sum of all these distances to the time series length $\varepsilon_P = (d_1 + d_2 + d_3)/N$ and it is a function of the scale Δn.

Figure 6.6 shows the variation of the partitioning error ε_P in function of the time scale Δn in the case of the time series in Fig. 6.4. The minimum of ε_P occurs for $\Delta n = 2^6 = 64$, but the error for $\Delta n = 2^5 = 32$ is comparable. When the time scale Δn is close to the time series length, the error ε_P is approximately equal to 1 because in this situation the number of the estimated local extrema is small and the main contribution to ε_P is given by the trend local extrema which have not been coupled with an estimated local extremum and the sum of the distances between them is roughly equal to the time series length ($d_1/N \ll 1, d_2/N = 0, d_3/N \approx 1$). Reversely, for small time scale Δn the main contribution to ε_P is given by the estimated local extrema which approximate none of the trend local extrema and their large number causes the error ε_P to take values much larger than 1 ($d_1/N < 1$, $d_2/N \gg 1, d_3/N = 0$).

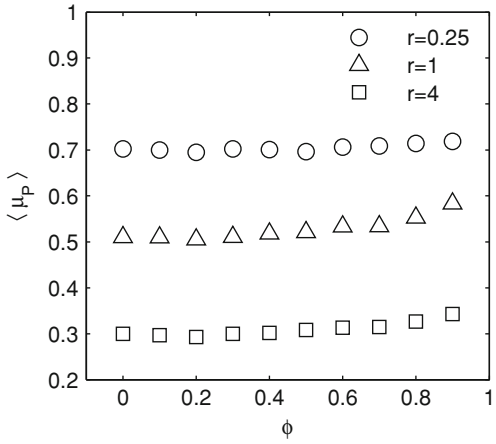

Fig. 6.7 The average $\langle \mu_P \rangle$ of the minimum in terms of the scale Δn of the partitioning error ε_P

The minimum value $\varepsilon_P = 0.252$ of the partitioning error in Fig. 6.6 corresponds to an accurate approximation of the local extrema of the trend (see Fig. 6.4 and the comments on it). When $\varepsilon_P \approx 1$, as for example for $\Delta n = N$, only the global shape of the time series is captured by the estimated local extrema. When $\varepsilon_P > 1$, as for example when $\Delta n = 1$, the number of the estimated local extrema is much larger than the number of those of the trend. Hence we may say that if $\varepsilon_P < 1$, then there exists useful information in the time series partition and if $\varepsilon_P < 0.3$, then there exists a good resemblance between the estimated and trend local extrema.

In order to analyze the influence of the noise amplitude and its serial correlation on the partitioning accuracy we apply the partitioning algorithm to statistical ensembles of 1000 time series with given values of the parameters r and ϕ of the AR(1) noise. The trend is characterized by the parameters $N = 1000$, $P = 30$, and $\Delta N_{\min} = 20$. Each time series is partitioned using time scales $\Delta n \in \{10, 20, \ldots, 200\}$ and the minimum $\mu_P = \min_{\Delta n}\{\varepsilon_P(\Delta n)\}$ of the partitioning error ε_P with respect to Δn is determined. The averages on the statistical ensembles of the minimum μ_P are plotted in Fig. 6.7. The serial correlation has little influence on the average partitioning error, i.e., it slightly increases with ϕ. The influence of the noise amplitude is much more significant. While for $r = 4$ the average partitioning error is about 0.3, for $r = 1$ it increases to 0.5, and when the noise is dominant ($r = 0.25$) the average partitioning error becomes 0.7.

6.3 Local Extrema of RCMA Trends

The approximation of the local extrema of the trend by the estimated local extrema can be improved if before looking for local extrema the time series is smoothed by means of the RCMA described in Sect. 4.1. Then not the initial time series is

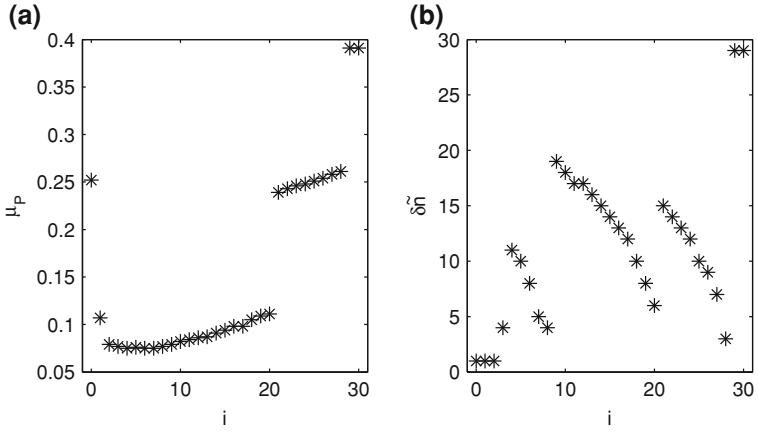

Fig. 6.8 **a** The minimum $\mu_P(i)$ of the partitioning error $\varepsilon_P(\Delta n, i)$ with respect to the time scale Δn. **b** The minimum $\delta\widetilde{n}(i)$ of the intervals between two successive local extrema

partitioned, but the trend estimated by means of the RCMA. The noise fluctuations are damped by averaging more strongly than the variations of the slowly varying trend. Then the shape of the averaged time series is closer to the trend shape than the shape of the unaveraged time series. However, we have to take into account that the trend is distorted by smoothing and several of its small local extrema may disappear.

The RCMA introduces two additional parameters: the length of the averaging window (K) and the number of repeated averagings (i). We apply the RCMA to the time series in Fig. 6.4 using $K = 5$ fixed so that only i varies. Then the partitioning error also depends on this parameter and we explicitly introduce this dependence by the notation $\varepsilon_P(\Delta n, i)$. The minimum $\mu_P(i) = \min_{\Delta n}\{\varepsilon_P(\Delta n, i)\}$ with respect to the time scales $\Delta n \in \{1, 2, 2^2, \ldots, 2^9, 1000\}$ is plotted in Fig. 6.8a. In comparison with the minimum partitioning error $\varepsilon_P = 0.252$ obtained without any averaging (see Fig. 6.6), for many values of i the averaging provides a much smaller error. For some values of the averagings number i the partitioning error has large discontinuous variations.

Both the positions $\widetilde{n}_j(\Delta n, i)$ of the estimated local extrema of the time series averaged i times and their number $\widetilde{J}_{\Delta n}(i)$ also depend on both the scale Δn and the averagings number i. Then $\widetilde{n}_j(1, i)$ are the positions of all the local extrema of the time series averaged i times. The minimum distance between two successive local extrema for a given i is $\delta\widetilde{n}(i) = \min_j\{\widetilde{n}_{j+1}(1, i) - \widetilde{n}_j(1, i)\}$ and is plotted in Fig. 6.8b. When $i \leq 2$ the noise fluctuations are not eliminated by averaging and there are successive local extrema separated by a single time step, $\delta\widetilde{n}(i) = 1$. For more averagings the variation of $\delta\widetilde{n}(i)$ is more complicated.

By averaging, not only the noise fluctuations are damped, but also the shape of the trend is altered, its local extrema are displaced and even they can be eliminated. For example the two local extrema of the trend initially situated at $n_9^{(0)} = 467$

and $n_{10}^{(0)} = 486$ (see Fig. 6.4) are eliminated from the averaged time series after $i = 21$ averagings. But before that, after $i = 20$ averagings, the corresponding local extrema of the averaged time series have the minimum distance $\delta\widetilde{n}(20) = 6$ between two successive local extrema. After $i = 21$ averagings the closest successive local extrema in the averaged time series correspond to the local extrema of the trend initially situated at $n_{17}^{(0)} = 975$ and $n_{18}^{(0)} = 1000$ and now they are separated by $\delta\widetilde{n}(21) = 15$ time steps. By increasing the averagings number the distance between these two local extrema of the averaged time series decreases and for $i = 29$ they are eliminated from the averaged time series.

When the local extrema disappear from the averaged time series the minimum partitioning error $\mu_P(i)$ suddenly increases for the same values of i (see Fig. 6.8a). There are other discontinuous variations of $\delta\widetilde{n}(i)$ to which no discontinuous variations of $\mu_P(i)$ correspond because they are caused by the elimination of estimated local extrema due to the noise and not to the trend. For example after $i = 4$ averagings two local extrema of the averaged time series situated at $n = 300$ and $n = 304$ are eliminated and they do not belong to the local extrema of the trend.

The previous analysis is performed for a single time series. Now we extend it to the statistical ensembles used in the previous section to compute the average minimum partitioning error. There the time series have not been averaged and in the notation of this section those results correspond to the minimum $\mu_P(0)$ with respect to the time scales $\Delta n \in \{10, 20, \ldots, 200\}$ of the partitioning error $\varepsilon_P(\Delta n, 0)$. In Fig. 6.9a we have copied the results presented in Fig. 6.7. We apply to the time series in the statistical ensembles the RCMA with $K = 10$ and limit the averagings number to $i \leq 50$. In Fig. 6.9b we present the average of the minimum of the partitioning error $\varepsilon_P(\Delta n, i)$ with respect to both the time scale Δn and the averagings number i, i.e., $\langle \min_i \{\mu_P(i)\} \rangle$. One can see that, by averaging the time series, the estimation of the trend local extrema is considerably improved especially when the noise has a small serial correlation.

In the case of observational time series the trend is unknown and therefore we cannot compare the estimated local extrema with the real ones in order to compute the partitioning error and to determine the optimum number of averagings. That is why we need a special criterion to choose the averagings number. We denote by i_0 the number of averagings for which all successive local extrema of the averaged time series are separated by more than one time step, i.e., $\delta\widetilde{n}(i_0) > 1$. When this condition is satisfied, then the noise fluctuations with the highest frequency are removed by the RCMA. Figure 6.9c shows the average $\mu_P(i_0)$, i.e., the minimum with respect to the time scale Δn of the partitioning error $\varepsilon_P(\Delta n, i_0)$ computed for i_0 averagings. The accuracy of the estimation of the trend local extrema is slightly worse than the exact minimum in Fig. 6.9b. We conclude that by averaging the time series until the condition $\delta\widetilde{n}(i_0) > 1$ is satisfied, the accuracy of the partitioning algorithm is significantly improved for all tested values of the parameters ϕ and r.

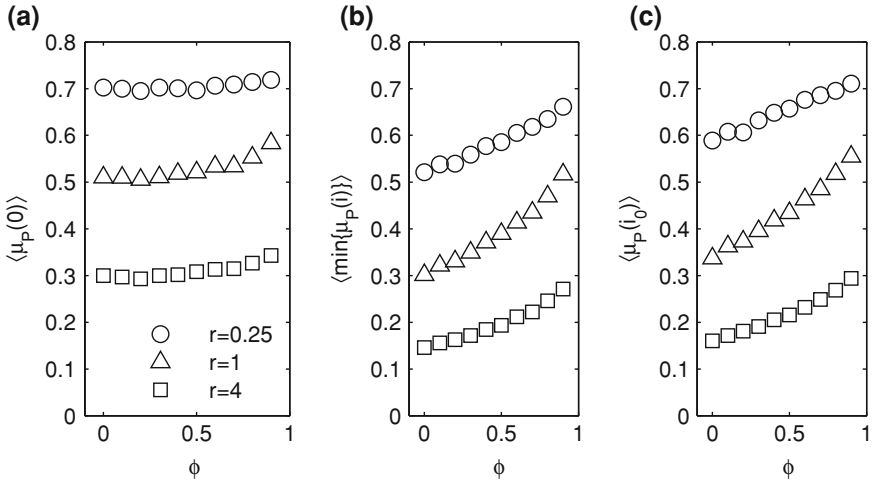

Fig. 6.9 The average of the minimum partitioning errors ε_P for unaveraged time series (**a**), the averaged ones (**b**), and those averaged i_0 times (**c**)

6.4 Significant Local Extrema of a Real Time Series

Until now, in this chapter we have analyzed only artificial time series with known trend which allow an accurate computation of the partitioning error ε_P of the estimated local extrema. The partitioning of a real time series with an unknown trend requires a criterion to select the trend local extrema from the estimated ones. In this section we partition the time series of cell membrane undulations analyzed in [6] and plotted in Fig. 6.10. We concluded that the minimum flickering frequency $\nu_{\min} = 0.3\,\text{Hz}$ implies a maximum semi-period of 16 time steps for the flickering fluctuations. This means that the variations of the cell surface with larger time scale should not be attributed to cell flickering. They should be estimated as a trend and then removed from the time series in order to keep only the fluctuations due to the phenomenon of interest. Therefore we may smooth the time series until all the distances between the successive local extrema of the averaged time series become larger than 16 time steps. Using the notations introduced in the previous section this means that, for a given K, we have to determine the averagings number i_0 for which the inequality $\delta\widetilde{n}(i_0) \geq 16$ is satisfied.

The minimum of the distances between successive local extrema with respect to the averagings number is plotted in Fig. 6.11 for three different values of K. For $K = 5$ even after 20 averagings the minimum of the distances between successive local extrema is smaller than 16 time steps. When $K = 10$ we obtain $i_0 = 6$ for which $\delta\widetilde{n} = 20$. In accordance with the RCMA properties discussed in Sect. 4.1, for $K = 20$ the necessary smoothing is obtained after only $i_0 = 3$ averagings. We also notice that for some values of i the minimum distance $\delta\widetilde{n}$ has sudden variations. As

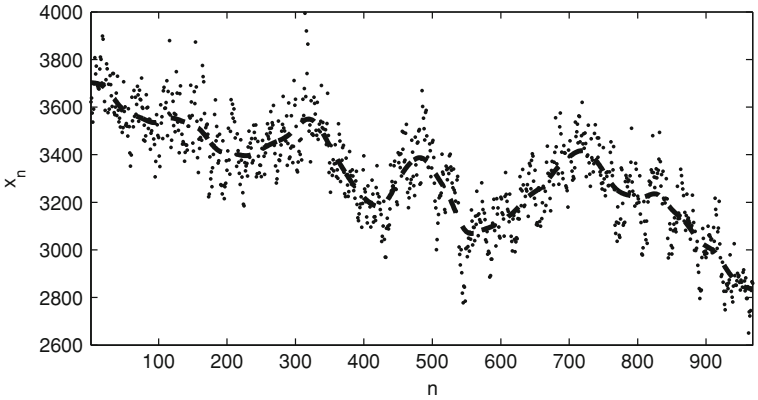

Fig. 6.10 The fluctuations of the relative area of a human red blood cell freely floating in a fluid and the trend estimated by the RCMA (*dashed line*)

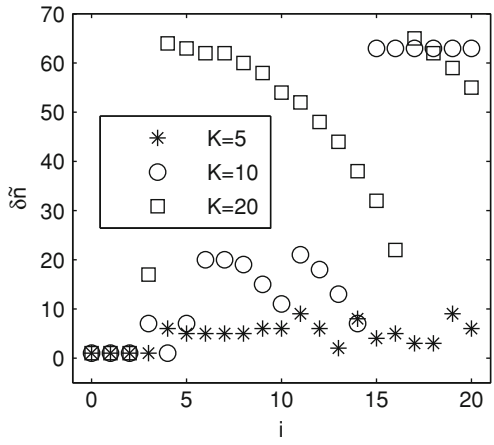

Fig. 6.11 The minimum of the distances between successive local extrema with respect to the averagings number for the cell flickering time series

shown in the previous section they correspond to the elimination by smoothing of a pair of local extrema from the averaged time series.

To ease the notation, in this section we denote the estimated RCMA trend by $f_n(i_0)$ with the dependence on the averagings number explicitly shown. No confusion can occur because in this section the real trend is not known. For $K = 10$ and $i_0 = 6$ the RCMA trend has $\widetilde{J}_1(6) = 12$ local extrema and it is plotted in Figs. 6.10 and 6.12a. In Fig. 6.12a the positions of local maxima are marked by continuous straight lines and of the local minima by dashed straight lines. Figure 6.12b shows the positions of the estimated local extrema of $f_n(i_0)$ for different time scales Δn. Since we are interested only in the variations for which the distance between successive local extrema is larger than 16 time steps, it follows that the distance between successive

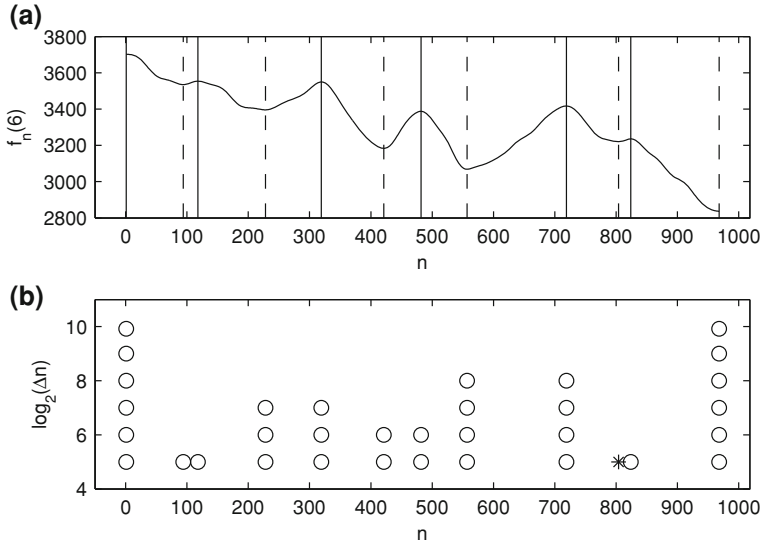

Fig. 6.12 The RCMA trend of the cell flickering time series (**a**) and the estimated local extrema for different values of the time scale Δn (**b**)

local extrema of the same type (maxima with respect to maxima and minima with respect to minima) is double. That is why the minimum time scale used to partition the RCMA trend is $\Delta n = 2^5 = 32$. The other time scales are powers of 2 up to 512 and they also include the length of the time series $\Delta n \in \{2^5, 2^6, \ldots, 2^9, 968\}$.

When $\Delta n = 2^9 = 512$ only the two global extrema which coincide with the boundaries of the averaged time series are found. The first pair of local extrema at $n = 557$ and $n = 719$ appears for $\Delta n = 2^8 = 256$. Afterwards the local extrema located at $n = 228$ and $n = 319$ and those at $n = 421$ and $n = 482$ are successively obtained. For $\Delta n = 2^5 = 32$ the partition contains all the local extrema of the RCMA trend. Now we have to find a criterion to determine which of these local extrema could be attributed to the real unknown trend.

The criterion that we propose is related to the observation that between two successive local extrema of different type the real trend has a monotonic variation. According to Eq. (1.15) the RCMA trend is composed by two parts

$$f_n(i) = \vartheta_f(n; i) + \vartheta_z(n; i) \; . \tag{6.2}$$

The monotonic variations of the smoothed trend $\vartheta_f(n; i)$ in Eq. (6.2) approximate those of the real trend. The estimated trend $f_n(i_0)$ is nonmonotonic over these intervals especially due to the fluctuations of the smoothed noise $\vartheta_z(n; i)$. In order to identify these local extrema related to the noise we compare the deviations from monotony of the RCMA trend with the fluctuations of the term $\vartheta_z(n; i)$ estimated by means of surrogate noises.

For a fixed time scale Δn, on each segment of the RCMA trend limited by two successive local extrema we apply the ACD algorithm described in Chap. 5. We denote by $g_n(\Delta n)$ the trends obtained in this way, plotted in Fig. 6.13 with dashed line and by

$$\delta f_n = f_n(i_0) - g_n(\Delta n)$$

the difference between the RCMA and ACD trends. We quantitatively characterize the deviations from monotony of the RCMA trend over the j segment of a partition of scale Δn by the quantity

$$A_j(\Delta n) = \max\{\delta f_n\} - \min\{\delta f_n\}, \quad n \in [\widetilde{n}_j, \widetilde{n}_{j+1}] \, .$$

In order to estimate the amplitude of the fluctuations of the term $\vartheta_z(n; i)$ in Eq. (6.2) we use the surrogate time series method presented in Sect. 5.4. We model the estimated noise $z_n = x_n - f_n(i_0)$ by an AR(1) stochastic process with the parameters equal to the sample standard deviation $\widehat{\sigma}$ of the estimated noise and the serial correlation parameter computed by means of the sample autocorrelation function, $\widehat{\phi} = \widehat{\rho}(1)$ (see Sect. 2.1). For the flickering time series in Fig. 6.10 we obtain $\widehat{\sigma} = 104$ and $\widehat{\phi} = 0.595$. We apply to the surrogate noise $\{\zeta_n\}$ the same RCMA as for the flickering time series ($K = 10$ and $i_0 = 6$) and we compute its fluctuations magnitude

$$A_\zeta = \max\{\vartheta_\zeta(n; i_0)\} - \min\{\vartheta_\zeta(n; i_0)\}, \quad n \in [1, N] \, .$$

We reduce the fluctuations of A_ζ due to different surrogates by computing the average $\langle A_\zeta \rangle$ over a statistical ensemble of $S = 30$ surrogates. For the flickering time series we have obtained $\langle A_\zeta \rangle = 146$.

We introduce the index

$$\chi_j(\Delta n) = A_j(\Delta n)/\langle A_\zeta \rangle \tag{6.3}$$

characterizing the magnitude of the nonmonotonic variations of the RCMA trend over the segment j. If $\chi_j < 1$, then the local extrema of the RCMA trend within the segment j could be attributed to the noise fluctuations and we cannot be sure that they correspond to actual local extrema of the trend. We have computed the index χ_j for all segments of the partitions presented in Fig. 6.12. The values of χ_j for four time scales are given in Fig. 6.13 together with the monotonic ACD trends for each segment. For the entire RCMA trend ($\Delta n = 512$) we obtain $\chi_1(512) = 2.80$ indicating that it contains significant nonmonotonic variations (Fig. 6.13a).

By reducing the time scale to $\Delta n = 256$ the RCMA trend is divided into three segments, each of them with an index χ smaller than the value 2.80 obtained for $\Delta n = 516$ (Fig. 6.13b). The ACD trend of the middle segment is almost identical with the RCMA trend and the value of the index is $\chi_2(256) = 0.03$. The segment at the end of the time series preserves a pair of local extrema and has a higher value $\chi_3(256) = 0.23$. The segment from the beginning of the time series has

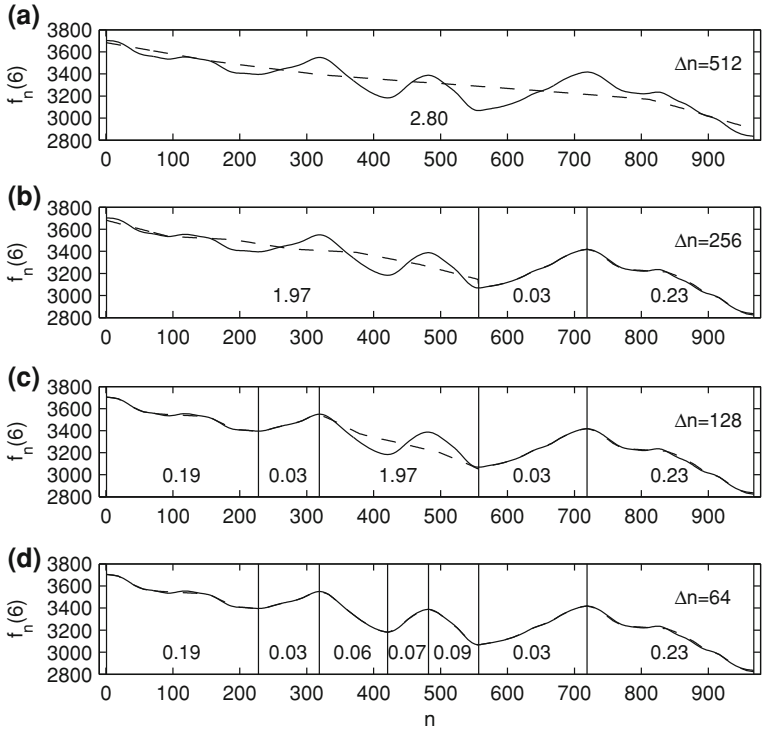

Fig. 6.13 The partitioning of the RCMA trend of the flickering time series with different time scales Δn and the values of the index χ measuring the nonmonotonic variations of the RCMA trend segments

important nonmonotonic variations as seen from the large value of $\chi_1(256) = 1.97$. By halving the time scale this segment is divided into three smaller segments, one of them preserving a high value of the index $\chi_3(128) = 1.97$ (Fig. 6.13c).

By a new halving of the time scale seven segments are obtained, all of them with an index smaller than 1 (Fig. 6.13d). Hence for $\Delta n = 64$ all significant nonmonotonic variations of the RCMA trend have been identified. This assessment is corroborated with the coincidence between the seven segments obtained for $\Delta n = 64$ and those visually identified in Fig. 6.10. The other four extrema of the RCMA trend have too small amplitudes to be associated with the real trend. They are identified by reducing the time scale to $\Delta n = 32$ so that the two segments from the beginning and the end of the time series are also partitioned into three segments each (see Fig. 6.12a).

References

1. Fink, E., Gandhi, H.S.: Compression of time series by extracting major extrema. J. Exp. Theor. Artif. In. **23**, 255–270 (2011)
2. Höppner, F.: Time series abstraction methods—a survey. In: Schubert, S., Reusch, B., Jesse, N. (eds.) Proceedings GI Jahrestagung Informatik, Workshop on Knowledge Discovery in Databases, Dortmund, September 2002, pp. 777–786. Bonner Köllen Verlag, Bonn (2002)
3. Huang, N.E., Shen, Z., Long, S.R., Wu, M.C., Shih, H.H., Zheng, Q., Yen, N.C., Tung, C.C., Liu, H.H.: The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis. Proc. R. Soc. Lond. A **454**, 903–995 (1998)
4. Keogh, E., Chu, S., Hart, D., Pazzani, M.: An online algorithm for segmenting time series. In: First IEEE International Conference on Data Mining (ICDM'01), pp. 289–296. IEEE Computer Society Press, Los Alamitos (2001)
5. Lindeberg, T.: Effective scale: a natural unit for measuring scale-space lifetime. IEEE T. Pattern Anal. **15**, 1068–1074 (1993)
6. Vamoş, C., Crăciun, M.: Serial correlation of detrended time series. Phys. Rev. E **78**, 036707 (2008)
7. Verbeek, P.W., Vrooman, H.A., van Vliet, L.J.: Low-level image processing by max–min filters. Signal Process. **15**, 249–258 (1988)
8. Witkin, A.P.: Scale space filtering. In: Bundy, A. (ed.) Proceedings of the 8th International Joint Conference on Artificial Intelligence, Karlsruhe, FRG, August 1983, pp. 1019–1022. William Kaufmann Inc., Los Alto (1983)

Chapter 7
Automatic Estimation of Arbitrary Trends

In this final chapter we transform the RCMA algorithm presented in Chap. 4 in an automatic algorithm. Instead of the two parameters controlling the RCMA we introduce a single parameter equal to the minimum distance $\delta\widetilde{n}$ between two successive local extrema of the smoothed time series. Its optimum value is determined as a function of the estimated serial correlation of the noise and of the estimated ratio between the amplitudes of the trend variations and noise fluctuations. The accuracy of the automatic RCMA is measured by Monte Carlo experiments and it is only slightly smaller than the maximum accuracy obtained by exhaustive search of all the RCMA trends. As an illustration we use the automatic RCMA to estimate the trend from a financial time series and by means of the partitioning algorithm presented in Chap. 6 we evaluate the significance of the local extrema of the estimated trend.

7.1 Automatic RCMA (AutRCMA)

An automatic algorithm for trend estimation has to adjust its parameters to the features of the analyzed time series. Both the polynomial fitting (Chap. 3) and the RCMA (Chap. 4) may be automated using numerical methods to estimate the time series parameters. We have chosen the RCMA for the following reasons:

- The RCMA accuracy does not depend on the length N of the time series, but on the resolution of trend monotonic segments (Sect. 4.2). If we use the polynomial fitting, then the time series with a large number of monotonic parts P of the trend have to be first partitioned since the accuracy of the polynomial fitting decreases when P increases (Sect. 3.2). In addition, the number of the trend monotonic parts is difficult to be estimated for arbitrary real time series.
- As shown in Sect. 4.2, the accuracy of the RCMA is higher than that of the polynomial fitting.
- The RCMA has two parameters to which we have to assign values (the semi-length K of the averaging window and the averagings number i) whereas the polynomial

C. Vamoş and M. Crăciun, *Automatic Trend Estimation*, SpringerBriefs in Physics,
DOI: 10.1007/978-94-007-4825-5_7, © The Author(s) 2012

fitting depends only on the polynomial degree. However, in the following we replace the two parameters by a single one.

- The RCMA introduces spurious autocorrelations in the estimated noise (see Appendix B). But the automatic trend estimation imposes high smoothings such that the introduced autocorrelation is usually negligible.

In Sect. 6.4 we have chosen the smoothing by imposing the condition that the minimum length $\delta\widetilde{n}$ of the monotonic segments of the averaged time series should be greater than a threshold. In this way the time series smoothing is controlled by a single parameter $\delta\widetilde{n}$, not by two parameters. The threshold was chosen using information about the studied phenomena, the cell flickering. For an arbitrary time series such information is not available and therefore $\delta\widetilde{n}$ has to be chosen taking into account the statistical properties of the time series.

We resume the evaluation of the accuracy of the trend estimated by the RCMA plotted in Fig. 4.6 replacing the dependence on the averagings number i by the dependence on the minimum $\delta\widetilde{n}$ of the monotonic variations of the trend estimated by RCMA defined in Sect. 6.3. Figure 7.1 shows the results obtained using statistical ensembles containing 100 artificial time series with $N = 1000$, $P = 10$, $\Delta N_{\min} = 50$ and the constant c_p in Eq. (2.2) randomly taking the values +1 and −1. In all cases $K = 10$ and the correlation parameter ϕ takes six distinct values. We have plotted with dashed line the average of the minimum with respect to the averagings number i of the resemblance index $\langle \eta_{\min} \rangle$ which in Fig. 4.6 is plotted by asterisks.

The graphs in Fig. 7.1 resemble to those in Fig. 4.6. The only difference is the disappearance of the segment corresponding to the small values of the averagings number i, explained by the fact that the first point represented in Fig. 7.1 has $\delta\widetilde{n} = 5$, which already implies a large number of averagings. The average resemblance index $\langle \eta \rangle$ mainly depends on the ratio r between the amplitudes of the trend variations and of the noise fluctuations. For the time series dominated by trend ($r > 1$) the minimum of $\langle \eta \rangle$ occurs at small values of $\delta\widetilde{n}$, i.e., for a weak smoothing. When the time series are dominated by noise ($r < 1$), the minimum of $\langle \eta \rangle$ is displaced to large values of $\delta\widetilde{n}$.

The accuracy of the estimated trend is also influenced by the correlation parameter of the noise ϕ, mainly for the time series dominated by noise ($r < 1$). For instance, when $r = 0.25$, the minimum of $\langle \eta \rangle$ for the white noise ($\phi = 0$) occurs for $\delta\widetilde{n} \approx 45$, while for the strongly correlated noise ($\phi = 0.9$) the minimum occurs for $\delta\widetilde{n} > 100$. The minimum of $\langle \eta \rangle$ for the time series dominated by trend ($r > 1$) is always situated at $\delta\widetilde{n} = 5$ and is independent on ϕ. What varies with respect to ϕ is the slope of the graph of $\langle \eta \rangle$ which, for small ϕ, rapidly increases with $\delta\widetilde{n}$.

From Fig. 7.1 we can determine for each value of r and ϕ the value of $\delta\widetilde{n}$ for which the minimum of $\langle \eta \rangle$ is obtained and when in average the estimated RCMA trend is the most similar to the real one. However, for the real time series we do not know the values of r and ϕ and we have to use estimated values r^{est} and ϕ^{est} which can significantly differ from the real ones. An acceptable accuracy of the estimated trend can be obtained introducing an approximation of the precise values of $\delta\widetilde{n}$ for which $\langle \eta \rangle$ is minimum.

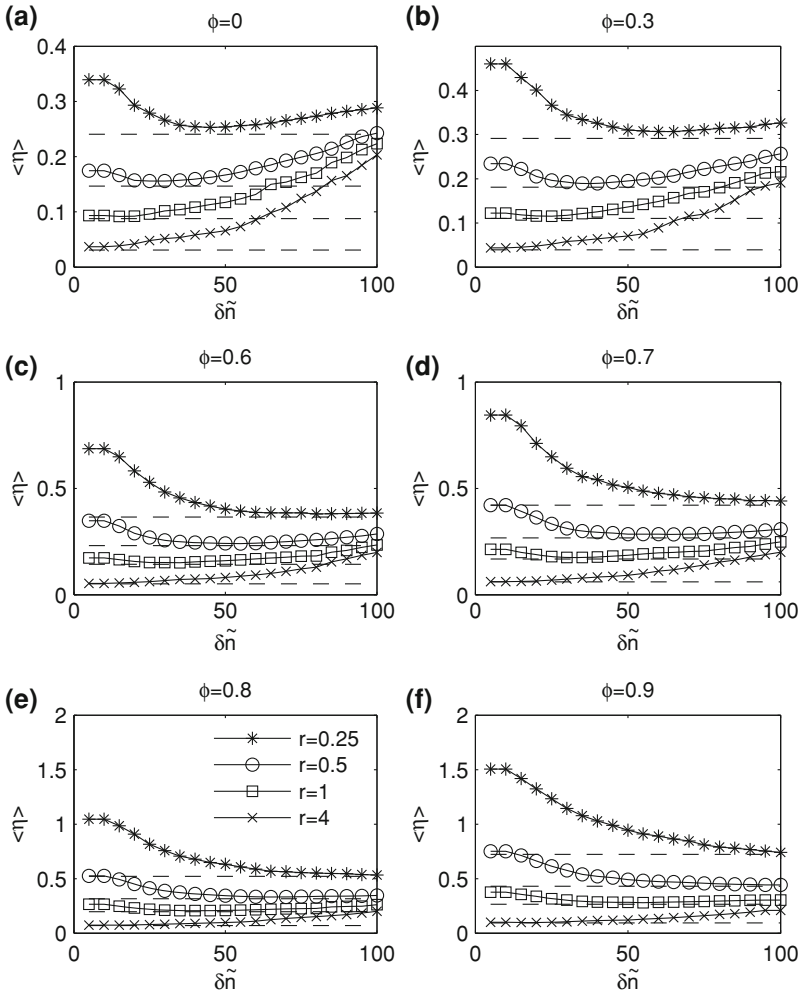

Fig. 7.1 The average resemblance index η for the trends estimated by the RCMA as a function of the minimum length $\delta\tilde{n}$ of the monotonic segments of the estimated trend. The average of the minimum of the index η for each time series is marked by an *horizontal dashed line*

We divide the time series into four classes: the time series dominated by noise ($r < 1$), those dominated by trend ($r > 1$), the time series with weakly correlated noise ($\phi < 0.6$), and those with strongly correlated noise ($\phi > 0.6$). For each class we choose from Fig. 7.1 a value of $\delta\tilde{n}$ corresponding approximately to the minimum of $\langle\eta\rangle$ for the whole class considered. These values are given in Table 7.1. The larger $\delta\tilde{n}$, the stronger the smoothing and in order to increase the processing speed we choose larger values of K.

Table 7.1 The values of the smoothing parameters

Classes of time series	$\delta\widetilde{n}$	K
$r > 1$ and $\phi < 0.6$	20	10
$r > 1$ and $\phi > 0.6$	40	20
$r < 1$ and $\phi < 0.6$	60	20
$r < 1$ and $\phi > 0.6$	80	30

For a real time series we do not know the values of the parameters ϕ and r and we have to estimate them. The noise correlation parameter can be estimated by means of Eq. (2.6) and the associated confidence level is plotted in Fig. 2.7. For long enough time series the accuracy of this estimation is high. But in order to deduce it we have used the restrictive hypothesis that the noise is of AR(1) type. For more complex noises one can make a similar analysis.

The ratio r is estimated by means of Eq. (F.1) in Appendix F. This algorithm does not make any assumption on the noise type, it is only tested on artificial time series with AR(1) noise. In average, when $\phi = 0$, if $r^{\text{est}} < 1.5$, then with a probability greater than 0.95 the time series is dominated by noise $r < 1$ (Fig. F.2). For strongly correlated time series ($\phi = 0.9$) the condition $r^{\text{est}} < 1.5$ is not enough restrictive, but for such time series there are many other sources of errors and we preserve this condition in all cases. Hence we consider that in average the time series dominated by noise ($r < 1$) are characterized by the condition $r^{\text{est}} < 1.5$ and reversely for those dominated by trend.

To conclude, the automatic algorithm based on RCMA consists of two steps. First we estimate the parameters ϕ^{est} and r^{est} and using them we classify the time series into one of the four classes presented above. Using the values of $\delta\widetilde{n}$ and K obtained from Table 7.1 we estimate the trend by means of RCMA. In Fig. 7.2 we present the average resemblance index $\langle\eta\rangle$ obtained by means of this automatic algorithm for statistical ensembles of 100 artificial time series with $N = 1000$ and $P = 10$. The dashed line represents the average of the minimum with respect to the number of averagings $i \leq 100$ of the resemblance index $\langle\eta_{\min}\rangle$. The small difference between the two average resemblance indices in the two cases shows that the quality of the automatically estimated trend does not significantly decrease. Similar results are obtained for $\Delta N_{\min} = 20$ and $\Delta N_{\min} = 50$. The accuracy decreases when the average resolution over the monotonic segments (N/P) decreases (see Sect. 4.2).

We apply this automatic algorithm to the logarithm of the S&P500 index analyzed in Sect. 4.3. We consider the index values from 1st January 1996 to 31st December 2010, i.e., N=3526 values (Fig. 7.3a). The trend estimated by AutRCMA is plotted by a continuous line and the vertical continuous lines mark the position of the estimated local extrema. Using the algorithm presented in Sect. 6.4 we identify the significant monotonic segments of the estimated RCMA trend by means of the nonmonotony index χ_j defined by Eq. (6.3). The local extrema limiting these segments are marked in Fig. 7.3a by thick vertical lines.

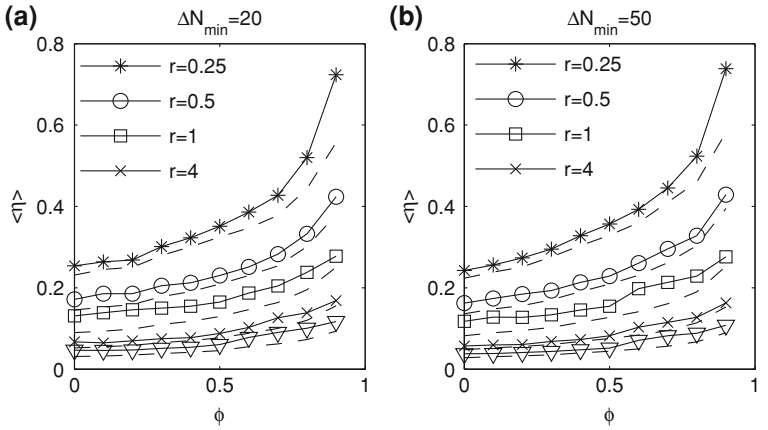

Fig. 7.2 The average resemblance index $\langle \eta \rangle$ of the trends estimated by the automatic RCMA (*continuous line*) and the average of the minimum of the resemblance index $\langle \eta_{\min} \rangle$ (*dashed line*)

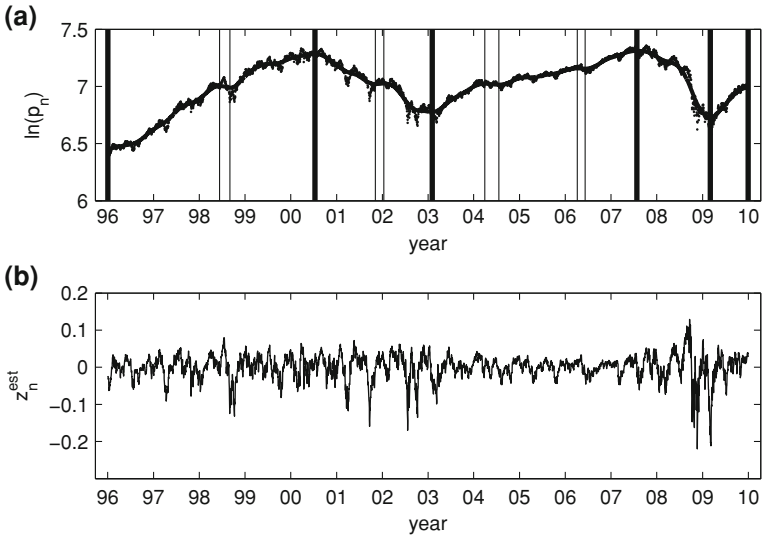

Fig. 7.3 **a** The logarithm of the S&P500 index from the 1st January 1996 to the 31 December 2010 and the estimated AutRCMA trend (*continuous line*). **b** The estimated noise obtained after removing the trend

If we remove the AutRCMA trend from the time series of the S&P500 index we obtain the estimated noise $\left\{z_n^{\text{est}}\right\}$ represented in Fig. 7.3b. In comparison with the logreturns in Fig. 4.8b, the estimated noise has different properties. Figure 7.4 shows the ACF of the estimated noise and of its absolute value. While the logreturns are uncorrelated (Fig. 4.9a), the estimated noise has a strong correlation (Fig. 7.4a). By differencing any correlation between the logreturns is eliminated, while by removing

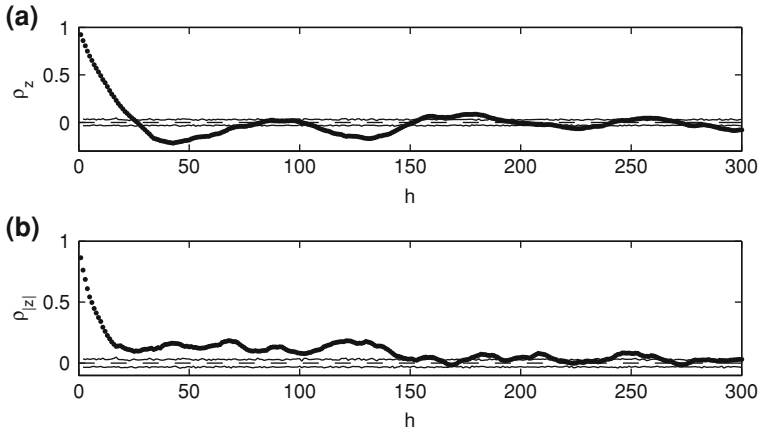

Fig. 7.4 ACF of the estimated noise (**a**) and of its absolute values (**b**)

the estimated trend a much larger correlation between the values of the estimated noise is kept.

The estimated noise has also several features similar to the logreturns. They both are heteroskedastic, i.e., the fluctuations amplitude has slow variations of large time scale, as one can see by comparing the autocorrelation functions of their absolute values in Figs. 7.4b and 4.9b. The estimated noise has an asymmetric stable distribution with the negative values larger than the positive ones similar to the financial returns [3].

Although the estimated noise is not an AR(1) time series and not even Gaussian, the automatic RCMA succeeded to estimate the trend because the time series is dominated by trend ($r > 1$). As expected, when the noise has small amplitude, its properties do not significantly influence the accuracy of the estimated trend. For AR(1) noise this behavior was discussed in relation with the results in Fig. 7.1.

As we have shown in Sect. 4.2, the accuracy of the RCMA mainly depends on the average resolution N/P of the trend monotonic segments. The values of the parameters $\delta\widetilde{n}$ and K in Table 7.1 are determined for an average resolution $N/P = 100$. Hence for time series with an average resolution of the monotonic segments significantly smaller than 100, new values of $\delta\widetilde{n}$ and K have to be established. Obviously, the smaller the average resolution, the smaller the values of $\delta\widetilde{n}$ and K are.

7.2 Statistical Significance of the Local Extrema of the AutRCMA Trend

In this section we continue the analysis of the trend of the financial time series estimated in the previous section. We compute the probability that the local extrema of the estimated trend correspond to local extrema of the real trend. We expect that this probability be larger for the local extrema which dominate the trend shape and it decreases when their importance decreases. In the case of the financial time series the periods with monotonic variations between successive local extrema correspond to some lasting macroeconomic and financial conditions and help to identify the economic cycles [2].

We analyze the statistical properties of the trend estimated by AutRCMA by means of surrogate time series. We generate them with the parameters similar to those estimated for the financial time series: $N = 3500, \phi = 0.86, r = 2.7, P = 27$, and $\Delta N_{\min} = 20$. In Fig. 7.5a we plot one of these artificial time series. The real trend and the trend estimated by AutRCMA are difficult to be distinguished from each other, as indicated also by the resemblance index $\eta = 0.125$. The real and the estimated local extrema can be compared in Fig. 7.5a, where the positions of the real extrema are represented by vertical lines in the lower half of the figure, and the positions of the estimated ones in the upper half of the figure. The local maxima are marked by continuous vertical lines and the minima by dashed vertical lines. Both the real and the estimated trend have $J = \widetilde{J} = 14$ local extrema each. But the real local extrema at $n_{12} = 3130$ and $n_{13} = 3176$ have no correspondents among the estimated ones and the estimated trend contains two local extrema at $\widetilde{n}_5 = 1205$ and $\widetilde{n}_6 = 1292$, both due to the noise, that do not correspond to real local extrema.

First we calculate the nonmonotony index χ_j given by Eq. (6.3) for the whole trend estimated by AutRCMA for the artificial time series in Fig. 7.5a and we obtain $\chi_1 = 14.1$. The value of χ_1 much larger than the unit, shows that the AutRCMA trend contains important nonmonotonic variations as one can see from its shape. Therefore we are looking for the most important local extrema by means of which the AutRCMA trend can be split into shorter segments with a smaller nonmonotony index.

In order to do that, we apply to the AutRCMA trend the partitioning algorithm described in Sect. 6.1 for the maximum temporal scale $\Delta n = N$. If, for this time scale, we cannot obtain more than two local extrema and the AutRCMA trend cannot be partitioned, then we reduce the time scale to half and repeat the halving until at least three local extrema are obtained or the scale becomes smaller than a fixed threshold. By decreasing step by step the time scale, the first local extrema that are found are the most important for the shape of the analyzed time series. As minimum threshold of the time scale we choose the minimum length $\delta\widetilde{n}$ of a monotonic part of the trend.

In the case of AutRCMA trend in Fig. 7.5 we obtain for $\Delta n = N$ the global maximum at $\widetilde{n}_{13} = 2898$, the global minimum at $\widetilde{n}_{14} = 3500$, and the additional minimum at $\widetilde{n}_1 = 1$ (see Fig. 7.5b). These local extrema are the important ones for

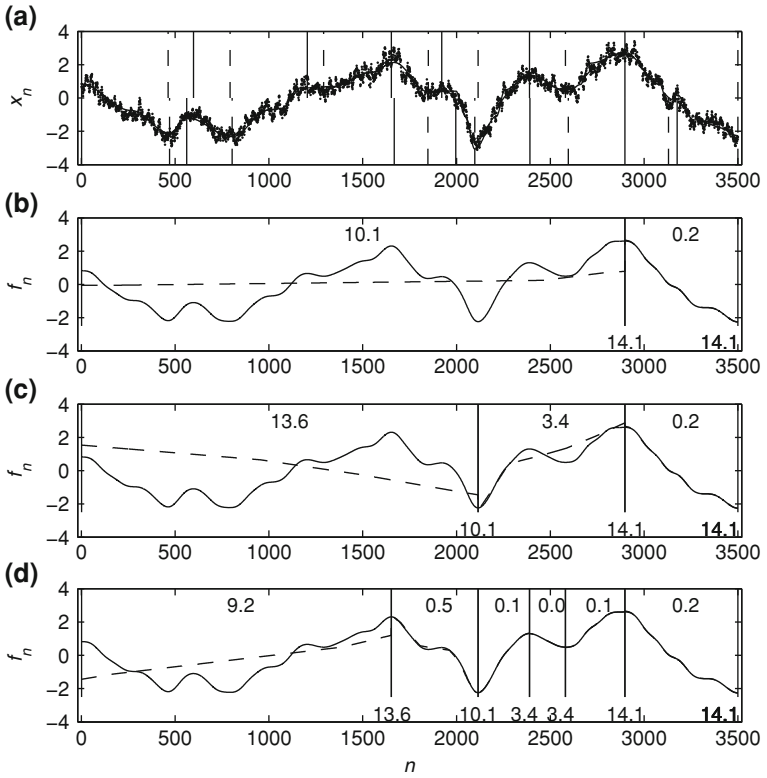

Fig. 7.5 **a** The real and estimated local extrema for an artificial time series; **b–d** successive steps of the partitioning of the estimated trend into monotonic segments

the shape of the AutRCMA trend and they explain the nonmonotony index $\chi_1 = 14.1$ previously obtained and characterizing the entire estimated trend. That is why we quantitatively measure the importance of these local extrema in the AutRCMA trend shape by the nonmonotony index itself $\theta_{13} = \chi_1 = 14.1$ and $\theta_{14} = \chi_1 = 14.1$. Hence we denote by θ_j the nonmonotony index related to the jth local extremum. These values are marked in Fig. 7.5b at the basis of the vertical continuous lines pointing the local extrema positions identified in this step. We give no value θ_1 to additional extremum at $\widetilde{n}_1 = 1$ because it is not one of the main local extrema satisfying the condition imposed by Eq. (6.1).

As a result of the first partitioning step we have split the AutRCMA trend into two segments (see Fig. 7.5b) which, in the second step, we process separately in the same way as the whole estimated trend. For $1 \leq n \leq \widetilde{n}_{13} = 2898$ we obtain a nonmonotony index $\chi_1 = 10.1$, which shows that this segment contains important nonmonotonic variations of the estimated trend. The interval $\widetilde{n}_{13} = 2898 \leq n \leq N = 3500$ does not contain any local extremum, as can be seen from the value of the index $\chi_2 = 0.2 < 1$. The values of χ_j are marked in the upper half of Fig. 7.5b at the middle

of the corresponding segment of the estimated trend. By partitioning the first segment we obtain the local minimum at $\widetilde{n}_{10} = 2115$, the global maximum at $\widetilde{n}_{13} = 2898$, and the additional minimum at $\widetilde{n}_1 = 1$ (see Fig. 7.5c). Hence $\theta_{10} = \chi_1 = 10.1$. The value $\theta_{13} = 14.1$ from the previous step has been assigned to the global maximum $\widetilde{n}_{13}$ such that now we do not assign to it the new value of χ.

In the third step we first analyze the part $1 \leq n \leq \widetilde{n}_{10} = 2115$ from the AutRCMA trend for which we obtain $\chi_1 = 13.6$ (Fig. 7.5c). This value is assigned to the local extremum $\widetilde{n}_7 = 1653$ (Fig. 7.5d). In the interval $\widetilde{n}_{10} = 2115 \leq n \leq \widetilde{n}_{13} = 2898$ we assign the value $\chi_2 = 3.4$ to the local extrema $\widetilde{n}_{11} = 2389$ and $\widetilde{n}_{12} = 2581$. This procedure goes on until values θ_j are assigned to all the local extrema of the AutRCMA trend. We note that the estimated local extrema $\widetilde{n}_5 = 1205$ and $\widetilde{n}_6 = 1292$ which do not correspond to any real local extrema have the value $\theta_5 = \theta_6 = 0.6$ showing that the nonmonotonic variation to which they are related is unimportant. The same is true for the extrema $n_8 = 1850$ and $n_9 = 1921$ for which $\theta_8 = \theta_9 = 0.5 < 1$, even if they correspond to two local extrema of the real trend.

The algorithm described above can be applied to any time series and it places the local extrema with respect to their contributions to the nonmonotonic variations of the trend.[1] For example it can be applied to the real trend and then we can make a correlation between the values of θ_j for real trend and for the estimated trend. To distinguish between them we denote the latter by θ_j^{est}. Using this information, in the following we provide a statistical analysis of the way in which the AutRCMA trend reproduces the shape of the real trend. The statistical ensemble is composed by 1000 artificial time series of the same type as the time series analyzed above with $N = 3500$, $\phi = 0.86$, $r = 2.7$, and $\Delta N_{\min} = 20$. The number of semi-periods of sinusoid is randomly chosen from the values $P = 14, 15, \ldots, 50$, such that the numerically generated trend should have a shape of a similar complexity with that of real financial time series.

The complexity of the time series shape is given by the number of the local extrema of the real trend or the number J of monotonic parts. In the case of the real time series we do not know the real trend, therefore we analyze the statistical properties of the statistical ensemble with respect to the number J^{est} of monotonic parts of the AutRCMA trend. In Fig. 7.6 we plot the average value and the standard deviation of the resemblance index η and index ε_P which measures the correspondence between the estimated and real local extrema. The minimum value of $J^{\text{est}} = 5$ has been obtained only for a single time series, which does not provide a sufficient statistics to calculate the statistical properties. Therefore we have plotted only the values $J^{\text{est}} \geq 9$, and in the corresponding statistics for $J^{\text{est}} = 9$ the time series for which $J^{\text{est}} < 9$ are introduced as well. For the same reasons we have plotted only the values for $J^{\text{est}} \leq 25$, although the maximum value of J^{est} is 32. From the results obtained it follows that the AutRCMA trend approximates very well the real trend ($\langle \eta \rangle \approx 0.10$)

[1] The automatic RCMA algorithm supplemented with the identification of significant monotonic segments of the estimated trend is implemented by the MATLAB program `trendrcma` freely accessible on web.

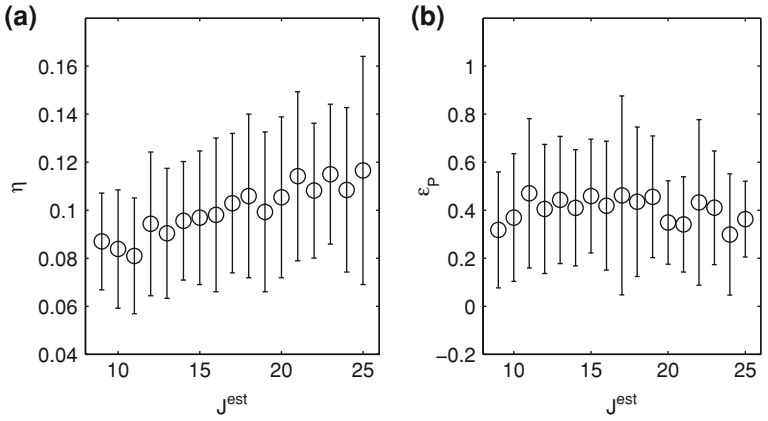

Fig. 7.6 The average value and standard deviation of the resemblance index η (**a**) and of the index ε_P (**b**) for the trends estimated with AutRCMA

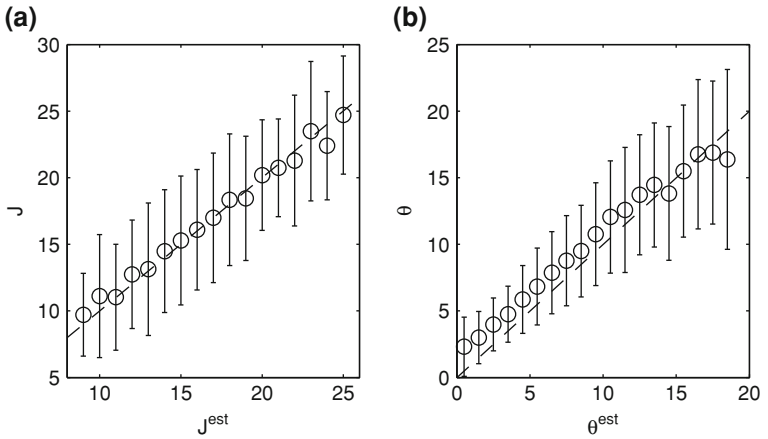

Fig. 7.7 **a** The average and standard deviation of the number of the real local extrema with respect to the estimated local extrema number. **b** The average and standard deviation of the nonmonotony index θ assigned to the real local extrema with respect to θ^{est} of the estimated local extrema

and that its accuracy slowly decreases with J^{est}. The correspondence between the real and estimated local extrema is weak ($\langle \varepsilon_P \rangle \approx 0.41$) and does not vary with J^{est}.

In order to analyze the correspondence between the real and estimated local extrema, in Fig. 7.7a the average and standard deviation of the number J of the real local extrema are plotted with respect to J^{est} . From the manner in which the average values of J follow the main diagonal (dotted line) it follows that in average the AutRCMA trend has the same number of local extrema as the real trend. However, the large values of the standard deviation show that the difference between the number of real and estimated local extrema can be rather large.

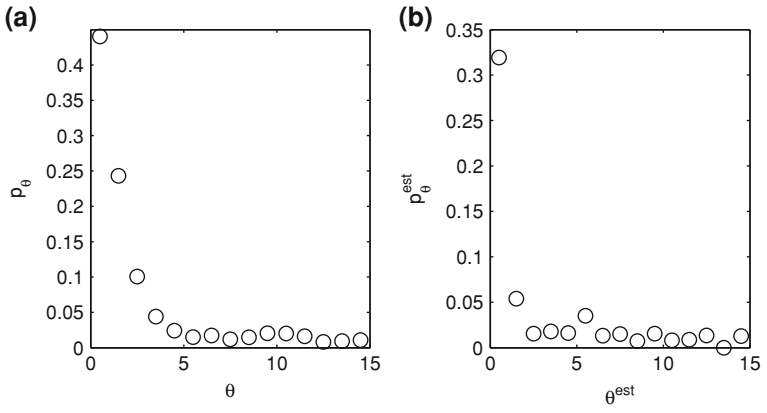

Fig. 7.8 **a** The probability that a real local extremum with a nonmonotony index equal to θ is not identified by AutRCMA trend. **b** The probability that an estimated local extremum with a nonmonotony index equal to θ^{est} has not a correspondent in the real trend

Figure 7.7b shows the average and standard deviation of the nonmonotony index θ_j of the real local extrema with respect to θ_j^{est} of the estimated local extrema. Only the real local extrema correctly identified by the AutRCMA trend are taken into account. For $\theta^{\text{est}} < 15$, the real value of the nonmonotony index is in average a little larger, while for the large values of θ^{est} it is smaller. Even if the standard deviation of θ becomes two times larger at large values of θ^{est}, there is an important correlation between the nonmonotony index of the estimated and real local extrema.

The results presented in Fig. 7.7b concern only the real local extrema correctly identified by AutRCMA trend. For the other local extrema another analysis is needed. We define p_θ as the probability that a real local extremum with a nonmonotony index equal to θ is not identified by AutRCMA trend and p_θ^{est} the probability that an estimated local extremum with a nonmonotony index equal to θ^{est} has not a correspondent in the real trend. The values of these probabilities are represented in Fig. 7.8. In both cases the probability that these errors occur rapidly decreases, for $\theta > 5$ and $\theta^{\text{est}} > 5$ being smaller than 0.03. Only for $\theta < 3$ the probability p_θ that the real local extrema are not identified is larger than 0.1, for $\theta < 1$ increasing to 0.43. The probability p_θ^{est} that the estimated local extrema do not correspond to some real local extrema is larger than 0.1 only for $\theta^{\text{est}} < 1$, when $p_\theta^{\text{est}} = 0.32$.

The significant local extrema with $\theta_j > 10$ of the financial time series in Fig. 7.3a are those at 12.07.2000, 04.02.2003, and 26.07.2007 and they are related to the major events of the global economy. For the crisis 2008–2009 we have the local minimum on the 03.03.2009 with $\theta_j = 6.2$. The other local extrema have small values of θ_j (less than 1), which shows that they are related to small fluctuations of the estimated trend, which may be caused by the noise superposed over the trend. According to the previous analysis, the probability that the estimated local extrema correspond to real local extrema is larger than 95 % (see Fig. 7.8b). Also, the probability that real local

extrema with $\theta > 6$ are not identified is smaller than 3 % (see Fig. 7.8a). Hence, with high probability, the trend estimated by AutRCMA contains the main characteristics of the real trend.

These results show that the variation of S&P500 index can be modelled as a superposition of a noise over a deterministic trend which describes the evolution of the global economy. Another possible modelling is by means of a nonstationary noise generated by "a unit root process" ([1], Chap. 15).

References

1. Hamilton, J.D.: Time series analysis. Princeton University Press, Princeton (1994)
2. Schumpeter, J.A.: Business cycles. A theoretical, historical and statistical analysis of the capitalist process. McGraw-Hill, New York (1939)
3. Voit, J.: The statistical mechanics of financial markets, 3rd edn. Springer, Berlin (2005)

Appendix A
Statistical Properties of the Linear Regression

In order to analyze the system of equations (3.3) we use the matrix notation

$$\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_N \end{bmatrix}, \quad \mathbf{z} = \begin{bmatrix} z_1 \\ z_2 \\ \vdots \\ z_N \end{bmatrix}, \quad \beta = \begin{bmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_K \end{bmatrix}, \quad \mathbf{b} = \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_K \end{bmatrix}, \quad \mathbf{a}_n = \begin{bmatrix} a_{1n} \\ a_{2n} \\ \vdots \\ a_{Kn} \end{bmatrix},$$

$$\mathbf{A} = \begin{bmatrix} \mathbf{a}_1^T \\ \mathbf{a}_2^T \\ \vdots \\ \mathbf{a}_N^T \end{bmatrix} = \begin{bmatrix} a_{11} & a_{21} & \cdots & a_{K1} \\ a_{12} & a_{22} & \cdots & a_{K2} \\ \vdots & \vdots & \cdots & \vdots \\ a_{1N} & a_{2N} & \cdots & a_{KN} \end{bmatrix}.$$

Because we define $\mathbf{a}_n$ as column vectors, the indexes of matrix $\mathbf{A}$ with dimensions of $(N \times K)$ are reversed with respect to the usual notation. In matrix notation Eq. (3.1) becomes

$$\mathbf{x} = \mathbf{A}\beta + \mathbf{z} \tag{A.1}$$

and the system (3.3) has the form

$$(\mathbf{A}^T\mathbf{A})\mathbf{b} = \mathbf{A}^T\mathbf{x}.$$

If $\mathbf{A}^T\mathbf{A}$ is nonsingular with dimensions of $(K \times K)$, then the estimated parameters are

$$\mathbf{b} = (\mathbf{A}^T\mathbf{A})^{-1}\mathbf{A}^T\mathbf{x}. \tag{A.2}$$

The quantity $\mathbf{Ab}$ is an estimation of the term $\mathbf{A}\beta$ and from Eq. (A.1) it follows that the sample estimate of the white noise $\mathbf{z}$ is

$$\widehat{\mathbf{z}} = \mathbf{x} - \mathbf{Ab} = \mathbf{Mx}, \tag{A.3}$$

where $\mathbf{M}$ is the $(N \times N)$ matrix

C. Vamoş and M. Crăciun, *Automatic Trend Estimation*, SpringerBriefs in Physics,
DOI: 10.1007/978-94-007-4825-5, © The Author(s) 2012

$$\mathbf{M} = \mathbf{I}_N - \mathbf{A}(\mathbf{A}^T\mathbf{A})^{-1}\mathbf{A}^T .$$

It is easy to show that $\mathbf{M}$ is symmetric ($\mathbf{M} = \mathbf{M}^T$) and idempotent ($\mathbf{M}^2 = \mathbf{M}$). Replacing Eq. (A.1) in Eq. (A.3) and taking into account that $\mathbf{MA} = \mathbf{0}$ we obtain

$$\widehat{\mathbf{z}} = \mathbf{M}(\mathbf{A}\beta + \mathbf{z}) = \mathbf{M}\mathbf{z}. \tag{A.4}$$

Now we can study the statistical properties of the estimated parameters. Replacing Eq. (A.1) in Eq. (A.2) it results that

$$\mathbf{b} = \beta + (\mathbf{A}^T\mathbf{A})^{-1}\mathbf{A}^T\mathbf{z}. \tag{A.5}$$

The same relation holds between the stochastic processes $\mathbf{B}$ and $\mathbf{Z}$ whose realizations are $\mathbf{b}$ and $\mathbf{z}$, respectively. If the noise has zero mean $\langle \mathbf{Z} \rangle = \mathbf{0}$, then the above equation implies that the estimation of the parameters is unbiased $\langle \mathbf{B} \rangle = \beta$. For the variance–covariance matrix we have

$$(\mathbf{B} - \beta)(\mathbf{B} - \beta)^T = (\mathbf{A}^T\mathbf{A})^{-1}\mathbf{A}^T\mathbf{Z}\mathbf{Z}^T\mathbf{A}(\mathbf{A}^T\mathbf{A})^{-1}.$$

If the noise is i.i.d. with variance σ^2, $\langle \mathbf{Z}\mathbf{Z}^T \rangle = \sigma^2\mathbf{I}_N$, then from the previous equation it follows that the variance of $\mathbf{B}$ is $\sigma^2(\mathbf{A}^T\mathbf{A})^{-1}$. When $\mathbf{Z}$ is Gaussian, because of the linearity of Eq. (A.5), $\mathbf{B}$ is also Gaussian

$$\mathbf{B} \sim N(\beta, \sigma^2(\mathbf{A}^\mathbf{T}\mathbf{A})^{-1}). \tag{A.6}$$

This relation synthesizes the statistical properties of the parameters estimated by linear regression if the noise is Gaussian i.i.d.

The derivation of the statistical properties of the noise estimated by Eq. (A.4) is more difficult. If the noise is i.i.d. it can be shown that

$$s^2 = \frac{1}{N-K}\widehat{\mathbf{Z}}^T\widehat{\mathbf{Z}} \tag{A.7}$$

is an unbiased estimator of the noise variance σ^2 ([1], Sect. 8.1). When the noise is Gaussian, the random variable $\widehat{\mathbf{Z}}^T\widehat{\mathbf{Z}}/\sigma^2$ is distributed $\chi^2(N-K)$. If the noise is non-Gaussian, then only asymptotic results can be obtained for the OLS estimators ([1], Sect. 8.2).

When the trend is the polynomial (3.4) the matrix $\mathbf{A}$ becomes a Vandermonde matrix with elements $a_{kn} = t_n^{k-1}$ where t_n are the sampling moments. A numerical method to compute $\mathbf{b}$ from Eq. (A.2) uses the QR factorization of matrix $\mathbf{A}$.

Reference

1. Hamilton, J.D.: Time Series Analysis. Princeton University Press, Princeton (1994)

Appendix B
Spurious Serial Correlation Induced by MA

Let us consider an i.i.d. stochastic process $\{Z_n\}$ and denote by

$$\Theta_n = \sum_{k=K_-}^{K_+} w_k Z_{n+k} \tag{B.1}$$

the stochastic process corresponding to a MA given by Eq. (4.1). The autocovariance function (1.1) of this stochastic process is equal to

$$\gamma_\Theta(h) = \langle \Theta_n \Theta_{n+h} \rangle = \sum_{j,k=K_-}^{K_+} w_j w_k \langle Z_{n+j} Z_{n+h+k} \rangle .$$

Taking into account that $\langle Z_{n+j} Z_{n+h+k} \rangle = \sigma_Z^2 \delta_{j,h+k}$, we obtain for $h \leq K_+ - K_- + 1$

$$\gamma_\Theta(h) = \sigma_Z^2 \sum_{k=K_-}^{K_+} w_k w_{k+h} \tag{B.2}$$

and $\gamma_\Theta(h) = 0$ for $h > K_+ - K_- + 1$. In conclusion, by averaging the i.i.d. process becomes correlated over the length of the averaging window.

This spurious correlation introduced by the MA has important consequences for the noise estimated from a time series $x_n = f_n + z_n$. According to Eq. (1.15), the trend estimated by MA is

$$\widetilde{f}_n = \vartheta_x(n) = \vartheta_f(n) + \vartheta_z(n) \ .$$

The estimated noise is then equal to

$$\widetilde{z}_n = x_n - \widetilde{f}_n = f_n - \vartheta_f(n) + z_n - \vartheta_z(n).$$

We denote by

$$\widetilde{Z}_n = f_n - \vartheta_f(n) + Z_n - \Theta_n$$

C. Vamoş and M. Crăciun, *Automatic Trend Estimation*, SpringerBriefs in Physics,
DOI: 10.1007/978-94-007-4825-5, © The Author(s) 2012

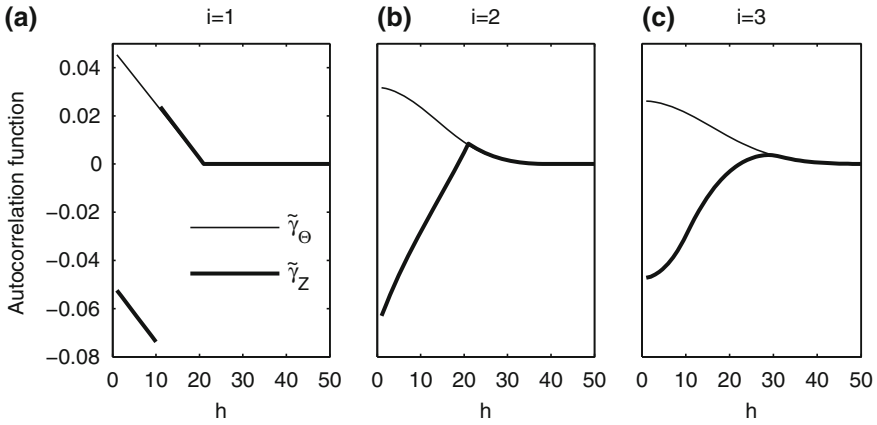

Fig. B.1 The autocovariance function of an i.i.d. noise averaged by the RCMA with $K = 10$ (*thin line*) and of the noise estimated by means of the RCMA (*thick line*) for three different number of repeated averagings

the stochastic process associated to the estimated noise and its autocovariance function is equal to

$$\begin{aligned}\widetilde{\gamma}_Z(h) &= \left\langle (\widetilde{Z}_n - \langle \widetilde{Z}_n \rangle)(\widetilde{Z}_{n+h} - \langle \widetilde{Z}_n \rangle) \right\rangle \\ &= \langle (Z_n - \Theta_n)(Z_{n+h} - \Theta_{n+h}) \rangle.\end{aligned}$$

Substituting Eq. (B.1) and using again the relation $\langle Z_n Z_{n+h} \rangle = \sigma_Z^2 \delta_{h,0}$, the previous equation becomes

$$\widetilde{\gamma}_Z(h) = \sigma_Z^2 \delta_{h,0} - (w_h + w_{-h})\sigma_Z^2 + \gamma_\Theta(h). \tag{B.3}$$

Besides the autocorrelation (B.2), the autocovariance function of the estimated noise has two additional terms due to the correlation between the noise $\{Z_n\}$ and the averaged noise $\{\Theta_n\}$.

In the case of the CMA, explicit expressions can be written because the weighting coefficients $w_k = (2K+1)^{-1}$ are constant. Then the autocovariance function (B.2) has a linear variation for $h \leq 2K+1$

$$\gamma_\Theta(h) = \frac{2K+1-h}{(2K+1)^2}\sigma_Z^2$$

and zero for $h > 2K+1$. It is represented in Fig. B.1a for $K = 10$ and $\sigma_Z = 1$ by a thin line. The autocovariance function of the estimated noise $\widetilde{\gamma}_Z(h)$ differs from $\gamma(h)$ only for $h \leq K$ by a constant term and it is represented by a thick line. We remark the anticorrelation of the estimated noise for small lags h.

Equations (B.2) and (B.3) hold also for RMA if w_n is replaced by $w_n^{(i)}$. When i increases, the variation of the autocovariance function is described by polynomials of increasing degree (Figs. B.1b and c). In addition, by increasing the number of repetitions the interval within which $\widetilde{\gamma}_Z$ is nonzero extends and its value decreases. In all cases, the estimated noise is anticorrelated for small lags h.

Appendix C
Continuous Analogue of the ACD Algorithm

We consider a time series $\{x_n\}$ of the type (1.12) as a realization of the stochastic process $\{X_n\}$ given by Eq. (1.10) with a monotonic trend. The average number of values of the time series $\{x_n\}$ about a given value ξ is proportional with the average density $\langle N_\xi \rangle$ of the number of the time series values in the neighborhood of ξ. In order to compute $\langle N_\xi \rangle$ we have to sum the probabilities that at time t_n the value x_n is within the vicinity of ξ. According to Eq. (1.11), the pdfs of Z_n and of X_n are identical with the exception of a translation by f_n, i.e., $p_X(x, n) = p_Z(x - f_n)$ and then

$$\langle N_\xi \rangle = \sum_{n=1}^{N} p_Z(\xi - f_n). \tag{C.1}$$

For an infinitesimal interval $I_\xi = (\xi - \delta\xi/2, \xi + \delta\xi/2)$ the number of values x_n lying within I_ξ is equal with $\langle N_\xi \rangle \delta\xi$. The discrete analogues of these quantities defined in Sect. 5.1 are the interval I_j and the number of values N_j lying within the interval I_j.

We introduce a quantity characterizing the average variation of the time series about the fixed value ξ which is the continuous analogue of that defined by Eq. (5.1). We define the average one-step displacement with the initial value in the neighborhood of ξ

$$g(\xi)\delta\xi = \frac{1}{\langle N_\xi \rangle} \sum_{n=1}^{N-1} p_Z(\xi - f_n) \langle \delta X_n | X_n \in I_\xi \rangle, \tag{C.2}$$

where $\delta X_n = X_{n+1} - X_n$. A similar relation is obtained if the final value is included in I_ξ and similar formulas are obtained. Using the conditional probability density for successive values of the stationary stochastic process $\{Z_n\}$ denoted by $p_Z(z''|z')$, we write

C. Vamoş and M. Crăciun, *Automatic Trend Estimation*, SpringerBriefs in Physics,
DOI: 10.1007/978-94-007-4825-5, © The Author(s) 2012

$$\langle \delta X_n | X_n \in I_\xi \rangle = \delta\xi \int_{-\infty}^{+\infty} (x-\xi)\, p_Z(x - f_{n+1} | \xi - f_n)\ \mathrm{d}x,$$

where the relation $p_X(x, n+1|\xi, n) = p_Z(x - f_{n+1}|\xi - f_n)$ is used. From the simple change of variables $z = x - f_{n+1}$, the definition of the conditional probability, and the consistency condition for the joint probability density $p_Z(z'', z')$

$$p_Z(\xi - f_n) = \int_{-\infty}^{+\infty} p_Z(z, \xi - f_n)\ \mathrm{d}z,$$

it follows that Eq. (C.2) becomes

$$g(\xi) = \frac{1}{\langle N_\xi \rangle} \sum_{n=1}^{N-1} (f_{n+1} - f_n)\, p_Z(\xi - f_n) + \langle \varepsilon_\xi \rangle. \tag{C.3}$$

If the trend is linear, then $f_{n+1} - f_n = a\,\delta t$ is constant and from Eq. (C.1) it follows that the first term in Eq. (C.3) is equal with the one-step trend variation $a\,\delta t$. In the general case of the nonlinear trends, it is equal with the average trend variation in the neighborhood of the value ξ of the time series. The second term in Eq. (C.3) is due to the noise and in the following we investigate the conditions when it is zero because then $g(\xi)$ is given only by the trend variation.

The second term in Eq. (C.3) is equal with

$$\langle \varepsilon_\xi \rangle = \frac{1}{\langle N_\xi \rangle} \sum_{n=1}^{N-1} \int_{-\infty}^{+\infty} (z + f_n - \xi) p_Z(z, \xi - f_n)\ \mathrm{d}z.$$

If the trend variation at one time step is much smaller than that of the noise ($|f'(t_n)\delta t| \ll \sigma_Z$ for all n), then the sum can be approximated by an integral. If, in addition, the trend is linear $f(t) = at + b$ and we make the change of variables $\theta = \xi - f(t)$, then we obtain

$$\langle \varepsilon_\xi \rangle = \frac{a}{\langle N_\xi \rangle \delta t} \int_{\xi - f_1}^{\xi - f_N} d\theta \int_{-\infty}^{+\infty} (z - \theta) p_Z(z, \theta)\ \mathrm{d}z.$$

Consider that the noise Z_n has the symmetry properties $p_Z(z) = p_Z(-z)$ and $p_Z(z', z'') = p_Z(-z', -z'')$. Then it follows that for $\xi = (f_1 + f_N)/2$ the term $\langle \varepsilon_\xi \rangle$ vanishes and its value increases when ξ approaches the extreme values of the time series. If the noise is moderately asymmetric, then $\langle \varepsilon_\xi \rangle$ vanishes for a different ξ, but close to $(f_1 + f_N)/2$.

When $\langle \varepsilon_\xi \rangle$ is negligible, then $g(\xi)/\delta t$ is an approximation of the trend slope $f'(t)$, but the two quantities have different arguments. We can express the trend slope in terms of ξ as $f'(f^{-1}(\xi))$. The trend estimated by the ACD method denoted

by $\mathscr{F}(t)$ is determined from the requirement that its derivative expressed with respect to the function values (not to the time argument t) be proportional to the average one-step displacement

$$g(\xi) = \mathscr{F}'(\mathscr{F}^{-1}(\xi))\,\delta t\ .$$

This relation holds only if $\mathscr{F}(t)$ is invertible, i.e., if $g(\xi)$ preserves the same sign over all its domain of definition. This condition is the analogue of the condition in Sect. 5.1 that all the quantities $\widetilde{g}_j$ should be positive.

Appendix D
Standard Deviation of a Noise Superposed over a Monotonic Trend

In this Appendix we propose and analyze a method to estimate the standard deviation of an arbitrary stationary noise superposed over a monotonic trend. We apply the quadratic norm defined by Eq. (2.5) to a realization $\{\nabla_d x_n\}$ of the differenced stochastic process (1.17) and we obtain

$$\|\nabla_d x_n\|^2 = \|\nabla_d f_n\|^2 + \|\nabla_d z_n\|^2 + 2\sum_{n=1}^{N-d}(\nabla_d f_n)(\nabla_d z_n). \tag{D.1}$$

The last term in this equation is negligible if the time series is long enough. We prove this property by decomposing it into the sum

$$\sum_{n=1}^{N-d}(\nabla_d f_n)(\nabla_d z_n) = \sum_{n=1}^{N-d} f_n z_n - \sum_{n=1}^{N-d} f_{n+d} z_n - \sum_{n=1}^{N-d} f_n z_{n+d} + \sum_{n=1}^{N-d} f_{n+d} z_{n+d}\ .$$

Since $\langle Z_n \rangle = 0$ and the noise is not correlated with the trend, all the four terms in the right hand side are negligible.

The noise term in Eq. (D.1) is equal with

$$\|\nabla_d z_n\|^2 = \sum_{n=1}^{N-d} z_{n+d}^2 + \sum_{n=1}^{N-d} z_n^2 - 2\sum_{n=1}^{N-d} z_n z_{n+d}. \tag{D.2}$$

If the noise is stationary and uncorrelated, then for any $d > 0$ the last term is negligible. Because the noise is stationary, the first two sums in Eq. (D.2) can be expressed by the sample variance (1.6), so that

$$\| \nabla_d z_n \|^2 \approx 2(N-d)\widehat{\sigma}_Z^2.$$

For small d the first term in Eq. (D.1) is negligible because it measures the variation during a few time steps of the slowly varying trend and Eq. (D.1) becomes

$$\| \nabla_d x_n \|^2 \approx 2(N-d)\widehat{\sigma}_Z^2. \tag{D.3}$$

C. Vamoş and M. Crăciun, *Automatic Trend Estimation*, SpringerBriefs in Physics,
DOI: 10.1007/978-94-007-4825-5, © The Author(s) 2012

Hence for an uncorrelated stationary noise and a monotonic trend we obtain the best estimation of the noise standard deviation for $d = 1$

$$\sigma_Z^{\text{est}} = \frac{1}{\sqrt{2(N-1)}} \parallel \nabla_1 x_n \parallel . \tag{D.4}$$

If the noise is correlated, a larger value of d is needed in order to keep z_n and z_{n+d} in Eq. (D.2) uncorrelated and then the first term in Eq. (D.1) due to the trend increases. As shown in the following, if the trend is monotonic, there is a criterion to establish the optimum value of d. In this case the term $\|\nabla_d f_n\|^2$ in Eq. (D.1) is monotonically increasing with respect to d for $d \le D = [(N+1)/2]$, where $[\cdot]$ is the integer part function. It follows that when $d \le D$ increases, the left hand term in Eq. (D.1) can decrease only due to the term $\|\nabla_d z_n\|^2$. We denote by d_0 the smallest integer number $d \le D$ for which

$$\parallel \nabla_{d_0+1} x_n \parallel^2 \; < \; \parallel \nabla_{d_0} x_n \parallel^2, \tag{D.5}$$

i.e., the number of time steps for which the first decrease of $\parallel \nabla_d x_n \parallel^2$ takes place. Hence d_0 is the value of d for which the variations of $\{\nabla_d x_n\}$ are dominated by noise. If the time series does not contain any noise, then $\parallel \nabla_{d_0} x_n \parallel = \parallel \nabla_{d_0} f_n \parallel$ is monotonically increasing for any monotonic trend and there is no d_0 satisfying Eq. (D.5). Therefore the existence of d_0 indicates that the time series contains significant nonmonotonic variations which can be due to both noise or nonmonotonic trend.

Since we intend to estimate only the magnitude order of σ_Z, we neglect in Eq. (D.1) the term due to the trend and Eq. (D.3) for $d_0 > 1$ becomes

$$\sigma_Z^{\text{est}} = \frac{1}{\sqrt{2(N-d_0)}} \parallel \nabla_{d_0} x_n \parallel \tag{D.6}$$

which is the generalization of Eq. (D.4). When the noise standard deviation is very small (r is large), it is possible that there is no value d_0 for which $\|\nabla_d x_n\|^2$ decreases. Since we need an estimated value for σ_Z in such situations as well, we apply Eq. (D.5) to the differenced time series $\{\nabla_1 x_n\}$, i.e., we analyze the variation with respect to d of the quantity $\|\nabla_d(\nabla_1 x_n)\|^2$. In this way the real trend $\{f_n\}$ is replaced in Eq. (D.1) with $\{\nabla_1 f_n\}$ which although nonmonotonic, has a variation amplitude much smaller than the initial trend. With the value of d_0 established, the noise standard deviation is estimated with Eq. (D.6).

Let us prove that, as assumed above, the differenced trend $\{\nabla_d f_n, n = 1, 2, \ldots, N-d\}$ has the square norm

$$\|\nabla_d f_n\|^2 = \sum_{n=1}^{N-d} (f_{n+d} - f_n)^2$$

monotonically increasing with respect to d if the trend is monotonic. We have used this property to define d_0. Because $\nabla_1 f_n = f_{n+1} - f_n$ we have

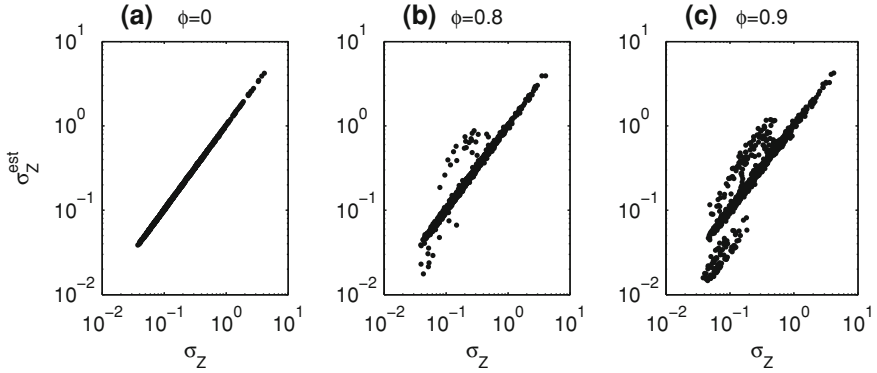

Fig. D.1 Correlation between the standard deviation σ_Z of a finite AR(1) noise superposed on a monotonic trend and the estimated standard deviation of the noise σ_Z^{est}

$$\nabla_d f_n = \nabla_1 f_n + \nabla_1 f_{n+1} + \cdots + \nabla_1 f_{n+d-1}.$$

By straight calculation we have for $d < D$

$$\|\nabla_{d+1} f_n\|^2 - \|\nabla_d f_n\|^2 = \sum_{n=1}^{N-2d-1} \left[\nabla_1 f_{n+d} \left(\nabla_1 f_{n+d} + 2 \sum_{i=1}^{d} \nabla_1 f_{n+i-1} \right) \right] + 2 \sum_{n=1}^{d} \left(\nabla_1 f_{N-n} \sum_{i=1}^{n} \nabla_1 f_{N-i-d} \right).$$

When the trend is monotonic, all the products $\nabla_1 f_n \nabla_1 f_m$ are positive and the right hand term of the above equation is positive. Hence $\|\nabla_d f_n\|^2$ is monotonic with respect to d for $d \leq D = [(N+1)/2]$.

The accuracy of the estimation (D.6) is tested on statistical ensembles of $S = 1000$ artificial time series as those described in Sect. 5.2. The time series have $N = 1000$ values, three different values of the parameter ϕ, and random values for a and r within their maximal range of variation. In Fig. D.1 a good correlation between the actual value σ_Z and the estimated one from Eq. (D.6) is noticeable. Only for $\phi \geq 0.8$ there are estimated values which significantly differ from the real values of ϕ (Fig. D.1b and c). Some of them are several times larger, others several times smaller than σ_Z. They appear when the time series are dominated by trend (r close to 4 and σ_Z small). Only 0.9 % of the time series are affected by this problem for $\phi = 0.8$ and 10 % for $\phi = 0.9$. The causes of this behavior are explained in the following.

The estimated value σ_Z^{est} is used to divide the time series values into the intervals I_j (see Sect. 5.2). That is why the estimation need not be very accurate since the ACD algorithm can compensate large variations of the number of the intervals I_j. It is more important that in all cases the order of magnitude of the estimation is correct. From the results presented in Fig. D.1 one can see that for the

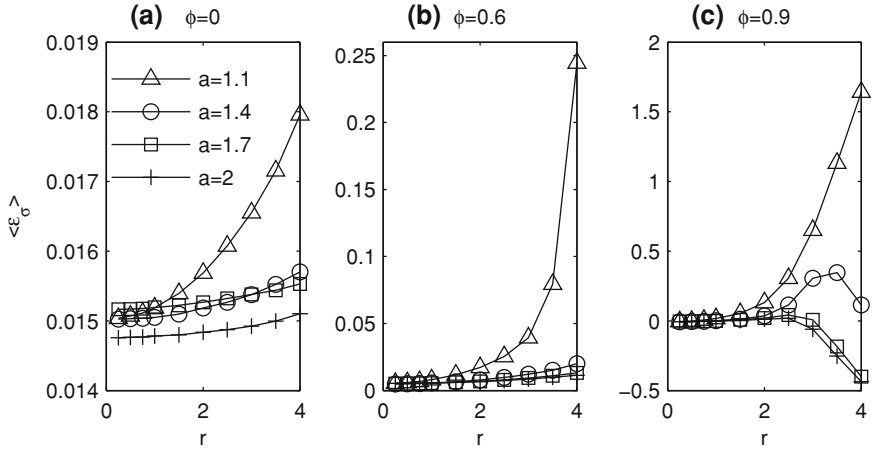

Fig. D.2 The average relative errors of the estimated noise standard deviation $\langle \varepsilon_\sigma \rangle$

chosen values of the parameters of the artificial time series the estimated values satisfy this condition. If the parameters take extreme values, then it is possible that the estimation deteriorates. For example, if $\phi = 0.99$, then σ_Z^{est} is one order of magnitude smaller than the real value. The performed tests show that for ϕ even closer to 1 the estimation quality remains almost constant. The estimation quality is the same for other functional forms of the monotonic trend (polynomial, exponential, logarithmic, etc.)

In order to study the factors that influence the estimation accuracy we use statistical ensembles of $S = 10000$ artificial time series with different values of the parameters ϕ, r, and a. The average relative error $\langle \varepsilon_\sigma \rangle$ of σ_Z^{est} is plotted in Fig. D.2. In the case of the white noise the average relative error is smaller than 2 % in all cases (Fig. D.2a). For larger serial correlation ($\phi = 0.6$) it reaches 20 % for $a = 1.1$ (steep slope trends) and $r = 4$ (time series dominated by trend) (Fig. D.2b). The average time lag $\langle d_0 \rangle$ for the same statistical ensembles presented in Figs. D.3a and b indicate that the cause of this behavior is the increase of d_0 given by Eq. (D.6) and used to compute σ_Z^{est}. When ϕ increases, the successive values of the noise are more correlated, so that in Eq. (D.1) the contribution of the term $\|\nabla_d z_n\|^2$ is smaller. If in addition the noise amplitude is also small, then the noise fluctuations are dominant in Eq. (D.1) only for large lags d. Hence when d_0 is large and the slope is also large ($a = 1.1$), then the term $\|\nabla_{d_0} f_n\|^2$ due to the trend is not negligible and the estimation error grows significantly.

The average relative error $\langle \varepsilon_\sigma \rangle$ is even larger for $\phi = 0.9$ (Fig. D.2c), but in this case there are many situations when d_0 cannot be determined from the initial time series. The fluctuations of the term $\|\nabla_d z_n\|^2$ in Eq. (D.1) are so small that they cannot change the monotony of the trend term $\|\nabla_d f_n\|^2$ for the small values of d. For the intermediate values of d the trend term is rapidly increasing and its monotonic variations cannot be altered. Only for d close to $D = [(N+1)/2]$ the

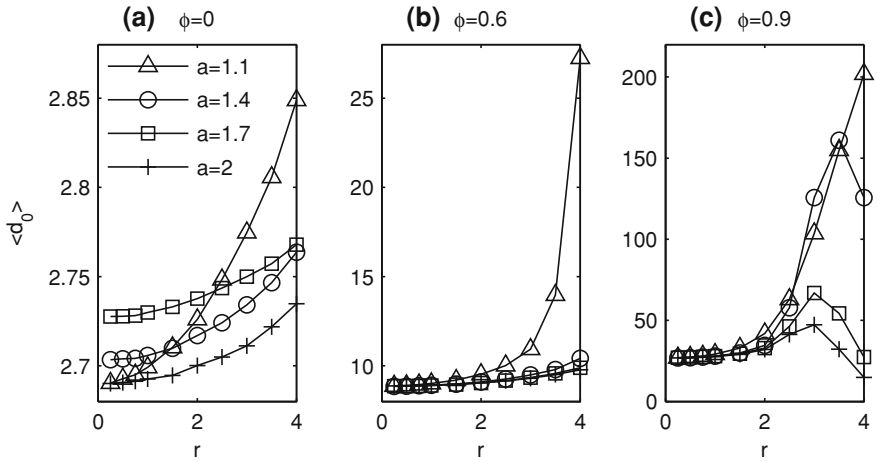

Fig. D.3 The average time lag d_0 used to estimate the noise standard deviation $\langle \varepsilon_\sigma \rangle$

combined contribution of the last two terms in Eq. (D.1) can change the monotony of $\|\nabla_d f_n\|^2$ and overestimated values of σ_Z^{est} are obtained, separated from the real values by a positive gap (see Fig. D.1c). Sometimes no d_0 is found and then we use the differenced time series $\{\nabla_1 x_n\}$ in which the trend is strongly damped and the noise is decorrelated, so that the value for d_0 is smaller than the correct one and σ_Z^{est} is underestimated. For this reason when a is large, the average relative error in Fig. D.2c becomes negative and in Fig. D.3c the average d_0 decreases. This explains why some values of σ_Z^{est} are several times smaller than σ_Z in Fig. D.1c.

The estimation previously described can be formally applied to time series with length $N \geq 4$, but the results obtained for very small N are not reliable. If the analyzed time series is a Gaussian white noise, then it needs to be long enough to allow us to obtain the lag d_0 without differencing it. When it is too short, it has only a few fluctuations and with a high probability they can be confounded with a monotonic trend. Then there is no d_0 satisfying condition (D.5) as if the time series contained negligible noise, although in reality it is only noise. We want to establish the minimum length $N_{\min}$ for which d_0 always exists for a Gaussian noise.

We make Monte Carlo experiments on statistical ensembles of $S = 1000$ realizations of a Gaussian white noise ($\phi = 0$) without trend with unit standard deviation. Beginning with $N = 12$ the percentage of the time series for which d_0 is obtained from the differenced time series $\{\nabla_1 x_n\}$ and not from the initial one $\{x_n\}$ is negligible (see Fig. D.4a). However we choose $N_{\min} = 14$ in order to be sure that the confusion between a white noise and a time series with a monotonic trend cannot occur in any case. Figure D.4a also contains the estimations for an AR(1) noise with $\phi = 0.9$. For large values of ϕ the stochastic trend becomes important enough to be interpreted in Eq. (D.6) as a deterministic trend. Then even for $N = 20$ in many cases d_0 cannot be determined from $\{x_n\}$. These results are

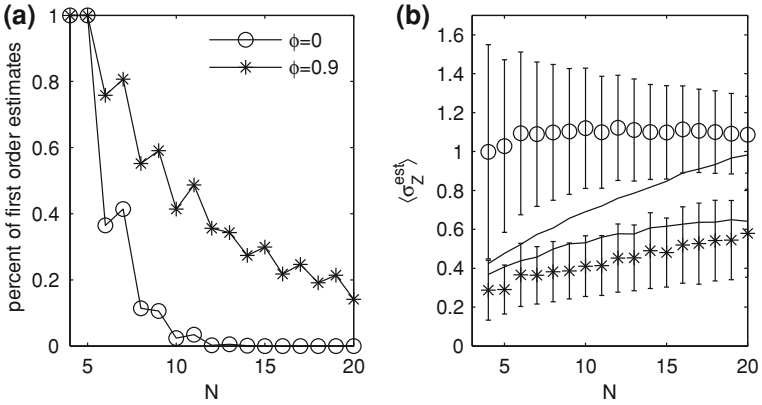

Fig. D.4 The percentage of the time series for which the estimation of d_0 is obtained from the differenced time series $\{\nabla_1 x_n\}$ and not from the initial one $\{x_n\}$ (**a**) and the average estimated standard deviation of the AR(1) noise (**b**)

consistent with the Mann–Kendall test for monotonic trends [1], which is applicable only if $N > 10$.

The average standard deviation estimated by means of Eq. (D.6) for the same AR(1) noises and its errors are plotted in Fig. D.4b. Even if σ_Z^{est} has a high variability, in all cases it has the same order of magnitude as the real value of the noise standard deviation $\sigma_Z = 1$. The quality of this estimation is comparable with that supplied by the sample standard deviation of the time series (the continuous line in Fig. D.4b). Hence the estimation (D.6) works acceptably even for such short time series.

Reference

1. Kendall, M.G.: Rank Correlation Methods. Griffin, London (1975)

Appendix E
Construction of a Partition of Scale Δn

In the following we describe a numerical iterative algorithm to build a partition of scale Δn defined in Sect. 6.1. We denote by n_c the time step up to which the time series analysis has been accomplished. The analysis continues by searching for the next local extremum of scale Δn in the interval $[n_c + 1, n_c + \Delta n]$. If such a local extremum is found at $\widetilde{n}_j$, then we make $n_c = \widetilde{n}_j$ and the analysis continues in the same way. If no local extremum is found, then we make n_c equal to $n_c + \Delta n$. The algorithm is initiated considering that $n_c = 0$ and it is stopped when $n_c \geq N$.

Let us assume that the positions $\widetilde{n}_j$, with $j < \widetilde{J}_n$, of the first local extrema contained in a partition of scale Δn have been determined and then $n_c = \widetilde{n}_j$. To facilitate the presentation we consider that $x(\widetilde{n}_j)$ is a local maximum of time scale greater than Δn and the time series is that presented in Fig. E.1. If we draw a line segment with the middle at the local maximum $\widetilde{n}_j$ and the ends at $\widetilde{n}_j - \Delta n$ and $\widetilde{n}_j + \Delta n$, it does not intersect the graph of the time series. Otherwise there would be values of the time series greater than $x(\widetilde{n}_j)$ at a smaller distance than n and the maximum would have a time scale smaller than Δn.

The thickened part of the graph of the time series in Fig. E.1 corresponds to the interval $[\widetilde{n}_j + 1, \widetilde{n}_j + \Delta n]$ where the next local extremum of the partition is searched for. A higher local maximum cannot exist in this interval because otherwise the condition (6.1) would not be satisfied by $x(\widetilde{n}_j)$. Hence the local extremum in the interval $[\widetilde{n}_j + 1, \widetilde{n}_j + \Delta n]$ may be only a minimum. We denote by n_- the position of the minimum of the time series values $x(n)$ with $n \in [\widetilde{n}_j + 1, \widetilde{n}_j + \Delta n]$. If in the interval $[n_- - \Delta n, n_- + \Delta n]$ there are no values x_n smaller than $x(n_-)$, then $x(n_-)$ has the time scale greater than Δn. Graphically, this means that the straight segment with length $2\Delta n$ and the middle at $x(n_-)$ does not intersect the time series graph. Then we can make the identification $\widetilde{n}_{j+1} = n_-$ and the search for the next local extremum (a maximum) continues in the same way for $n_c = \widetilde{n}_{j+1}$.

If a smaller minimum than $x(n_-)$ exists in the interval $[n_- - \Delta n, n_- + \Delta n]$, then it lays either in the subinterval $[n_- - \Delta n, \widetilde{n}_j]$ of the time series already analyzed, or in the subinterval $[\widetilde{n}_j + \Delta n, n_- + \Delta n]$ which has not been analyzed yet. Then $\Delta\widetilde{n}_j < \Delta n$ and we continue the analysis of the time series for $n_c = \widetilde{n}_j + \Delta n$. We

C. Vamoş and M. Crăciun, *Automatic Trend Estimation*, SpringerBriefs in Physics,
DOI: 10.1007/978-94-007-4825-5, © The Author(s) 2012

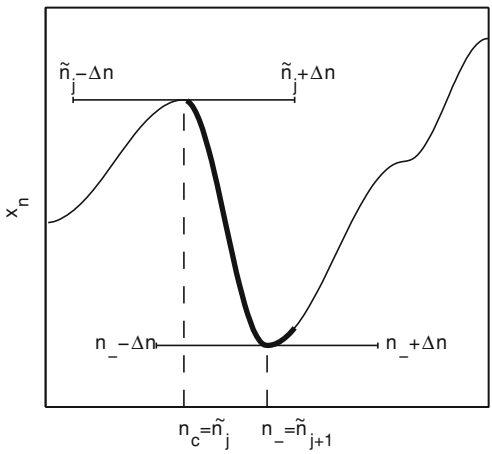

Fig. E.1 Search of a local extremum with the time scale larger than Δn

determine the position n_+ of the maximum and the position n_- of the minimum in the interval $[n_c + 1, n_c + \Delta n]$. By verifying the condition (6.1) on the intervals $[n_\pm - \Delta n, n_\pm + \Delta n]$ we determine which of the two extrema is a local extremum of scale Δn. The nearest of them to n_c is identified with the next local extremum of the partition of scale Δn from where the analysis of the time series continues. If neither of the two extrema satisfies the condition (6.1), the analysis continues from $n_c + \Delta n$.

For each new extremum added to the partition we check if it is of the same type as the preceding one (see Fig. 6.1a). If this is the case, then an additional local extremum is introduced such that a succession of maxima and minima is obtained. If the new extremum is a local minimum greater than the preceding local maximum, respectively if it is a local maximum smaller than the preceding local minimum (see Fig. 6.1b), then two additional local extrema are added to the partition.

Appendix F
Estimation of the Ratio Between the Trend and Noise Magnitudes

The automatic processing of the time series requires simple methods to estimate the properties of a time series. For example in Sect. 2.3 we have presented a method to estimate the serial correlation parameter ϕ based on the differenced time series. An equal important influence on the numerical results is due to the ratio r between the amplitudes of the trend variations and of the noise fluctuations defined by Eq. (2.3). For time series with a monotonic trend we have described in Appendix D a numerical method to estimate the noise standard deviation which is equivalent to the estimation of the ratio r. In this Appendix we show how a first guess of r can be obtained even if the trend is nonmonotonic.

As in Chap. 6 we denote by n_j, $j = 1, 2, \ldots, J$, the local extrema of the time series $\{x_n\}$. We define the maximum of the absolute value of the differences between successive local extrema separated at most by 10 time steps (an interval one order of magnitude larger than the time series resolution)

$$A_x = \max_j \left\{ |x(n_{j+1}) - x(n_j)| \ \text{ for } n_{j+1} - n_j \leq 10 \right\}.$$

Since the variation of the trend for the corresponding time interval is small, this quantity is an approximation of the noise amplitude $\max\{z_n\} - \min\{z_n\}$ in the definition of r given by Eq. (2.3). For the amplitude of the trend variations we have

$$\begin{aligned} \max\{f_n\} - \min\{f_n\} &= \max\{x_n - z_n\} - \min\{x_n - z_n\} \\ &\geq \max\{x_n\} - \max\{z_n\} - \min\{x_n\} + \min\{z_n\} \\ &\approx \max\{x_n\} - \min\{x_n\} - A_x \ . \end{aligned}$$

Then a first tentative estimation of the ratio r is the relation

$$r^{est} = \frac{\max\{x_n\} - \min\{x_n\}}{A_x} - 1 \ .$$

C. Vamoş and M. Crăciun, *Automatic Trend Estimation*, SpringerBriefs in Physics,
DOI: 10.1007/978-94-007-4825-5, © The Author(s) 2012

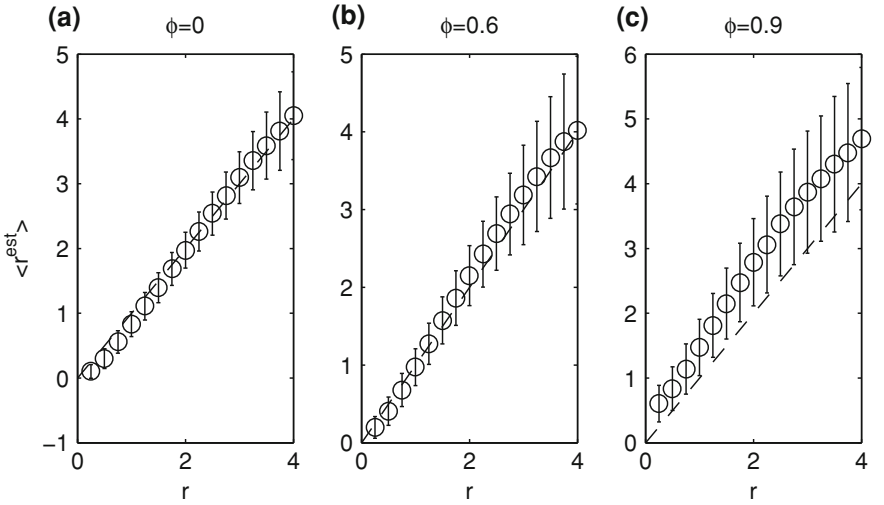

Fig. F.1 The average and the standard deviation of the estimated ratio r^{est} for different values of the parameters r and ϕ

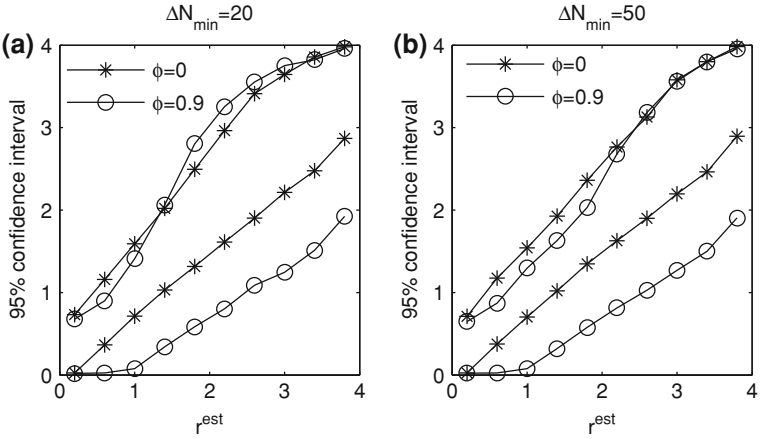

Fig. F.2 The 95 % confidence levels of the estimated ratio r^{est}

From the previous inequality it follows that in general $r^{est} > r$. We can correct this bias by diminishing the estimated value by the average of r^{est} for Gaussian white noises equal to 0.257. Then the proposed estimation of the ratio r is

$$r^{\text{est}} = \frac{\max\{x_n\} - \min\{x_n\}}{A_x} - 1.257. \tag{F.1}$$

The average r^{est} on statistical ensembles of 100 time series with $N = 1000$ and different values of r and ϕ are plotted in Fig. F.1. The trends of these time series

are formed by $P = 30$ sinusoid semi-periods longer than $\Delta N_{\min} = 20$ time steps (see Sect. 2.2). Excepting the time series dominated by noise and not strongly correlated, the averages on the statistical ensembles of the estimated ratio r^{est} are always larger than the real values of r. These errors increase when the noise is strongly correlated and when the trend dominate the time series (see Fig. F.1c). For $\phi = 0.9$ and $r > 3$ the errors begin to decrease but their values remain significant. The standard deviation of r^{est} increases when ϕ and r increase and when $\phi = 0.9$ and $r > 3$ it becomes greater than 1.

In order to establish the utility of this estimation we compute the 95 % confidence levels of the estimation r^{est} using the numerical method described in Sect. 2.1. The results are plotted in Fig. F.2 for two values of the minimum length $\Delta N_{\min}$ of the monotonic segments of the trend. The statistical ensemble has 10000 time series with $N = 1000$ values. The number of the monotonic segments of the trends is randomly chosen from the values $P = 1, 2, \ldots, 5$ and the ratio r has random values uniformly distributed in the range $r \in [0.01, 4]$. The confidence intervals do not depend on $\Delta N_{\min}$ and for strong serial correlation ($\phi = 0.9$) the confidence interval is two times larger than for a white noise.