In this chapter we shall discuss some basic topics of time series analysis, including the classical decomposition of a time series into deterministic trend and seasonal components and a random component, as well as spectral density estimation. Special topics include cases of missing observations, hidden additive components, and bivariate time series.

5.1 Estimation of Trend and Seasonal Components and Scale Function

A time series (process) is a set of pairs of observations \((X_1, Y_1), (X_2, Y_2), \ldots, (X_n, Y_n)\) where each response \(Y_i\) has been recorded at a specific time \(X_i\), and traditionally \(X_1 < X_2 < \cdots < X_n\). Then, the simplest classical decomposition model of a time series is

\[
Y_i := f(X_i) + S(X_i) + \sigma(X_i)\varepsilon_{X_i},
\]

where \(f(x)\) is a slowly changing function known as a trend component; \(S(x)\) is a periodic function with period \(T\) (that is, \(S(x + T) = S(x)\) for all \(x\)), known as a seasonal (cyclical) component (it is also customarily assumed that the integral or sum of the values of the seasonal component over the period is zero); \(\sigma(x)\) is called a scale function (it is also often referred to, especially in finance and econometrics literature, as a volatility); and \(\varepsilon_{X_i}\) are random components that may be dependent, and in this case the responses \(Y_i\) become dependent as well. Recall that the familiar phrase “a random
walk down Wall street” is motivated by this type of classical decomposition, and a primary argument in the literature is about the presence or absence of a deterministic part and about the type of a random walk.

A typical feature of a time series is that predictors \( X_l \) are equidistant integers. Thus, without loss of generality, we may set \( X_l = l \). Then a time series is completely described by the responses \( \{ Y_l, l = 1, 2, \ldots \} \), which may be treated as a sequence of regular observations in time, and this explains why such a sequence is called a time series. Of course, many practical examples are indeed sequences in time, but there are plenty of other examples; for instance, data may be collected in space. In the latter case the data are often referred to as spatial data, and there is even a special branch in statistics, known as geostatistics, that is primarily concerned with the analysis of such data. A particular example will be considered in Section 6.7. In this chapter, for the sake of clarity, we shall use only time series terminology and assume that data are collected sequentially in time.

Another typical feature of a time series is that the errors \( \{ \varepsilon_1, \varepsilon_2, \ldots \} \) in (5.1.1) may be dependent. Moreover, the case of dependent errors is the main topic in time series analysis. Thus, we begin our discussion with a short introduction to a class of ARMA processes, which are a good tool to model series of dependent random variables. Then we shall discuss methods of estimation of a trend, seasonal component, and scale function.

- **Causal ARMA Processes.** The main assumption about the class of time series (5.1.1) that we wish to consider is that the noise \( \varepsilon_l \) is a realization of a so-called second-order zero-mean stationary time series \( \{ \varepsilon_l \} = \{ \ldots, \varepsilon_{-1}, \varepsilon_0, \varepsilon_1, \ldots \} \) such that (i) \( E\{\varepsilon_l^2\} < \infty \) for all \( l \), that is, the second moment is finite; (ii) \( E\{\varepsilon_l\} = 0 \) for all \( l \), that is, the expectation is zero; (iii) the autocovariance function \( \gamma(l, s) := E\{\varepsilon_l \varepsilon_s\} \) satisfies the relation \( \gamma(l, s) = \gamma(l+h, s+h) \) for all \( l, s, \) and \( h \), that is, a translation in time does not affect the autocovariance function.

Note that property (iii) implies that \( \gamma(l, s) = \gamma(l-s) = \gamma(s-l) \). To see this just set \( h = -s \) and \( h = -l \). Thus a second-order zero-mean stationary time series is characterized by its autocovariance function \( \gamma(h) \) at the lag \( h \). Also note that no assumptions about higher moments or about distributions of the errors are made.

The simplest kind of second-order stationary error is one in which the random variables \( \{ \varepsilon_l \} \) are uncorrelated (that is, \( \gamma(h) = 0 \) for \( h \neq 0 \)), with mean 0 and variance 1. Let us denote such time series by \( \{ Z_l \} \) and call it a standard white noise. A classical example is a time series of iid standard Gaussian random variables, which is the white noise that we shall use in all the following simulations, and we call it a standard Gaussian white noise.

Then a wide variety of dependent second-order stationary processes can be generated by using a white noise and a set of linear difference equations. This leads us to the notion of an autoregressive moving average process of orders \( p \) and \( q \), an ARMA(\( p, q \)) process for short. By definition, the process \( \{ X_t, t = \ldots, -1, 0, 1, \ldots \} \) is said to be an ARMA(\( p, q \)) process if \( \{ X_t \} \) is
5.1 Estimation of Trend and Seasonal Components and Scale Function

Second-order stationary and for every \( t \),

\[
X_t - a_1 X_{t-1} - \cdots - a_p X_{t-p} = \sigma (Z_t + b_1 Z_{t-1} + \cdots + b_q Z_{t-q}), \tag{5.1.2}
\]

where \( \{Z_t\} \) is a standard white noise, \( \sigma > 0 \), the orders \( p \) and \( q \) are nonnegative integers, and \( a_1, \ldots, a_p, b_1, \ldots, b_p \) are real numbers. For the case of a Gaussian white noise we shall refer to the corresponding ARMA process as a Gaussian ARMA process.

Two particular classical examples of an ARMA process are a moving average MA(\( q \)) process, which is a moving average of \( q + 1 \) consecutive realizations of a white noise,

\[
X_t = \sigma (Z_t + b_1 Z_{t-1} + \cdots + b_q Z_{t-q}), \tag{5.1.3}
\]

and an autoregressive AR(\( p \)) process satisfying the difference equation

\[
X_t - a_1 X_{t-1} - \cdots - a_p X_{t-p} = \sigma Z_t. \tag{5.1.4}
\]

Each of these examples plays an important role in the analysis of time series. For instance, prediction of values \( \{X_t, t \geq n + 1\} \) in terms of \( \{X_1, \ldots, X_n\} \) is relatively simple and well understood for an autoregressive process; see Exercise 5.1.16. Also, for a given autocovariance function it is simpler to find an AR process with a similar autocovariance function. More precisely, if an autocovariance function \( \gamma(j) \) vanishes as \( j \to \infty \), then for any integer \( k \) one can easily find an AR(\( k \)) process with the autocovariance function equal to \( \gamma(j) \) for \( |j| \leq k \). The “negative” side of an AR process is that it is not a simple issue to find a stationary solution for (5.1.4), and moreover, it may not exist. For instance, the difference equation \( X_t - X_{t-1} = \sigma Z_t \) has no stationary solution, and consequently there is no AR(1) process with \( a_1 = 1 \). The discussion of such tricky things is beyond the scope of this book, and in what follows a range for the coefficients that “keeps us out of trouble” will always be specified.

The advantages of a moving average process are its simple simulation, the given expression for a second-order stationary solution, and that it is very close by its nature to white noise, namely, while realizations of a white noise are uncorrelated, realizations of an MA(\( q \)) process are uncorrelated whenever the lag is larger than \( q \). The “minus” of MA processes is that, surprisingly, they are not so easy for prediction and estimation as AR processes. Thus, among the two, typically AR processes are used for modeling and prediction. Also, AR processes are often used to approximate an ARMA process.

Now we are in a position to define a causal (future-independent) ARMA process (or more specifically, a causal process with respect to an underlying white noise \( \{Z_t\} \)). The idea is that it is quite natural to expect that an ARMA time series \( \{X_t\} \) depends only on current and previous (but not future!) realizations of the white noise. Thus, we say that an ARMA process \( \{X_t\} \) generated by a white noise \( \{Z_t\} \) is causal if \( X_t = \sum_{j=0}^{\infty} c_j Z_{t-j} \), where the coefficients \( c_j \) are absolutely summable. Clearly, MA(\( q \)) processes are
causal, but not all AR(p) processes are; for instance, a stationary process corresponding to the difference equation $X_t - 2X_{t-1} = Z_t$ is not causal. We shall not elaborate more on this issue and note only that below, we consider simulations of only Gaussian ARMA(1, 1) processes corresponding to the difference equation $X_t - aX_{t-1} = \sigma(Z_t + bZ_{t-1})$ with $|a| < 1$ and $-a \neq b$. It may be directly verified (Exercise 5.1.17) that for such a this equation has a stationary and causal solution $X_t = \sigma Z_t + \sigma(a + b) \sum_{j=1}^{\infty} a^{j-1} Z_{t-j}$.

This ends our brief discussion of ARMA processes.

The aim of the next subsections is to explain methods of estimation of the deterministic components $f(x)$, $S(x)$, and $\sigma(x)$ in (5.1.1) where $X_l = l$ and noise $\{\varepsilon_l\}$, $l = 1, \ldots, n$, is zero-mean and second-order stationary. A comprehensive example that combines all the steps is postponed until Section 5.3 because finding periods of seasonal components is based on estimation of the spectral density, which is discussed in Section 5.2.

- **Estimation of a Trend.** There is no surprise that time series analysis customarily uses methods of estimation of a trend that are also used by regression analysis, namely, methods such as parametric least-squares regression or smoothing by means of a moving average. On the other hand, the nonparametric orthogonal series approach, developed in Chapter 4, seems an attractive alternative to these classical methods. Indeed, if a time series has a deterministic term that is written as $\sum_{j=0}^{\infty} \theta_j \phi_j(x)$, then the low-frequency part of this series,

$$f(x) := \sum_{j=0}^{J_{\text{max}}} \theta_j \phi_j(x), \quad 0 \leq x \leq n,$$

(5.1.5)

can be referred to as a trend component (or simply trend). Here $\{\phi_j\}$ are elements of a basis in $L_2([0, n])$ and $\theta_j$ are the Fourier coefficients. The choice of $J_{\text{max}}$ is typically up to the practitioner, who defines the meaning of the trend and seasonal components in the frequency domain.

Then the data-driven universal estimator of Section 4.2 can be used to estimate the trend. (Recall that to use the universal estimator we always rescale data onto $[0, 1]$.) Moreover, the estimator is greatly simplified by the fact that its cutoff should be at most $J_{\text{max}}$. Then, all the examples considered in Chapter 4 can be viewed as some particular time series.

- **Estimation of a Scale Function.** The primary concern of the classical time series theory is that the stochastic term in (5.1.1) should be second-order stationary, that is, the scale function $\sigma(x)$ should be constant. Since this is typically not the case, the usually recommended approach is to transform a data set at hand in order to produce a new data set that can be successfully modeled as a stationary time series. In particular, to reduce the variability (volatility) of data, Box–Cox transformations are recommended when the original positive observations $Y_1, \ldots, Y_n$ are converted to $\psi_\lambda(Y_1), \ldots, \psi_\lambda(Y_n)$, where $\psi_\lambda(y) := (y^\lambda - 1)/\lambda$, $\lambda \neq 0$, and $\psi_\lambda(y) := \log(y)$, $\lambda = 0$. By a suitable choice of $\lambda$, the variability may be significantly reduced.
5.1 Estimation of Trend and Seasonal Components and Scale Function

Apparently, the nonparametric technique of Section 4.3 may be used as well. Firstly, we use the nonparametric estimator of a scale function suggested in Section 4.3. All the examples considered in Section 4.3 illustrate how the approach works. Then the original observations are divided by the estimate, and this should give us a new data set with a nearly constant variability of its stochastic term.

• Estimation of a Seasonal Component. Traditional time series analysis assumes that the period $T$ of an underlying seasonal component $S(x)$ is given. (We shall discuss in the next two sections how to find the period with the help of the spectral density; also note that in many practical examples, such as daily electricity demands or monthly average temperatures, periods of possible cyclical components are apparent.) By definition, $S(x + T) = S(x)$ for any $x$, and if a time series is defined at integer points, then $\sum_{l=1}^{T} S(l) = 0$ (a seasonal component should be zero-mean).

Using these two assumptions, classical time series theory recommends the following method of estimating a seasonal component. First, a given time series is detrended by the formula $\tilde{Y}_l = Y_l - \tilde{f}(l)$, where $\tilde{f}(l)$ is an estimated trend. Then, the natural procedure for estimating $S(j)$ is the sample mean estimate

$$\tilde{S}(j) := [(n-j)/T]^{-1} \sum_{r=0}^{[(n-j)/T]} \tilde{Y}_{j+rT}, \quad j = 1, 2, \ldots, T. \quad (5.1.6)$$

Recall that $[a]$ denotes the rounded-down $a$. Note that (5.1.6) is a nonparametric estimate because no parametric underlying model is assumed.

To understand how this conventional method performs, let us consider a simple example. Assume that $\tilde{Y}_l = S(l) + \sigma \varepsilon_l, l = 1, 2, \ldots, n$, where $n = kT$, $k$ is integer, and $\varepsilon_1, \varepsilon_2, \ldots$ are iid standard normal. Then

$$\tilde{S}(j) = S(j) + \sigma k^{-1} \sum_{r=1}^{k} \varepsilon_{j+rT} = S(j) + \sigma k^{-1/2} \eta_j, \quad j = 1, 2, \ldots, T, \quad (5.1.7)$$

where $\eta_j := k^{-1/2} \sum_{r=1}^{k} \varepsilon_{j+rT}$ are again iid standard normal. Thus, if $k$ is large enough (that is, if $n$ is large and $T$ is relatively small), then the conventional estimator should perform well.

On the other hand, if $k$ is small and, respectively, the period $T$ is large (and this is a rather typical case in many applications), then another approach may be used. It is apparent that (5.1.7) is an equidistant nonparametric regression model with $S(x)$ being the regression function and the period $T$ being the sample size. Thus, the universal nonparametric estimator of Section 4.2 (or Section 4.2) may be used straightforwardly to estimate $S(j)$ based on $T$ observations (5.1.7). Note that the nonparametric estimator smoothes the conventional estimate (5.1.6).
1. Uniform

2. Normal

3. Angle

FIGURE 5.1. Three simulated time series shown by squares connected by the dotted lines, with no trends and different seasonal components of period $T = 20$. The underlying seasonal components, shown by solid lines, are the Uniform, the Normal, and the Angle corner functions minus 1. Stochastic terms are generated by a Gaussian ARMA(1, 1) process $\varepsilon_t - a\varepsilon_{t-1} = \sigma(Z_t + bZ_{t-1})$, where $\{Z_t\}$ are iid standard normal, $a = 0.4$, $b = 0.3$, and $\sigma = 1$. The length $n$ of the realizations is controlled by the argument $n$. The period $T$ of seasonal components is controlled by the argument $Per$. The parameters $\sigma$, $a$, and $b$ of the ARMA(1,1) noise are controlled by the arguments $sigma$, $a$, and $b$. Use $|a| < 1$. The argument $set.seas$ allows one to choose any 3 corner functions as the underlying seasonal components. $[n=100, Per=20, sigma=1, a=.4, b=.3, set.seas=c(1,2,6)]$

Figure 5.1 shows 3 simulated time series (with no trends) of length $n = 100$ where seasonal components have period $T = 20$ and they are the corner functions Uniform, Normal, and Angle minus 1. Noise terms are generated by a Gaussian ARMA(1, 1) process $\varepsilon_t - 0.4\varepsilon_{t-1} = Z_t + 0.3Z_{t-1}$.

Let us consider these particular time series. It is known that the time series in Figure 5.1.1 has neither trend nor seasonal component, while the two others do have seasonal components, but is it apparent from the data? And what do we mean here by a trend and a seasonal component?

Let us begin the discussion with the second question. According to (5.1.5), the trend and seasonal components are separated in the frequency domain. Because it is easier to think about the frequency domain in terms of periods, let a deterministic periodic component with period less than $T_{\text{max}}$ be referred to as a seasonal component, and as a trend component otherwise. For instance, if we set $T_{\text{max}} = 40$, then no pronounced trend with this or larger period is visible in Figure 5.1.1, while a seasonal component with period between 10 and 20 is a likely bet. Note that such an illusion of the presence of a seasonal component is a trademark of ARMA processes. Moreover, long-memory processes, considered in Section 4.8, may create...
5.1 Estimation of Trend and Seasonal Components and Scale Function

even an illusion of a trend component. Now assume that \( T_{max} = 5 \). In this case apparently no seasonal component is present, but a slightly smoothed dotted line that connects the observations may be a possible bet on an underlying trend.

Now let us look at the second diagram with the underlying Normal seasonal component. The fact that this is very pronounced seasonal component makes it easier to conclude that a seasonal component does exist. On the other hand, it is not an easy task to estimate it; just look at the time between 70 and 90 where the shape of this component is completely lost due to the noise.

The third diagram illustrates another typical challenge caused by dependent observations. Look at the first half of the observations; here a majority of observations are above the seasonal component. The situation changes for the second part of the observations. This is what may cause great confusion in any estimate, and this is what the dependency means.

Now let us return to our discussion of the separation of seasonal component and trend. Here all depends on the choice of \( T_{max} \), or in other words, on what we mean by a slowly changing trend component. Fortunately, typically this is a clear-cut issue for practical applications. For instance, for a long-term money investor, \( T_{max} \) is about several years, while for an active stock trader it may be just several days or even hours.

If \( T_{max} \) is specified, then \( J_{max} \) in (5.1.5) is defined as the minimal integer such that \( \varphi_{J_{max}}(x + T_{max}) \approx \varphi_{J_{max}}(x) \) for all \( x \). For instance, for the cosine basis on \([0, n]\) with the elements \( \varphi_0(t) := n^{-1/2}, \varphi_j(t) := (n/2)^{-1/2} \cos(\pi j t/n), \; j = 1, 2, \ldots, 0 \leq t \leq n \), we get

\[
J_{max} = \left\lfloor \frac{2n}{T_{max}} \right\rfloor.
\] (5.1.8)

Now let us return to the first question, namely, can we visualize any trend or seasonal component in the particular realizations shown in Figure 5.1? Assume that \( T_{max} \) is defined approximately correctly, say \( T_{max} = 30 \). In other words, if we detect a deterministic cyclical component with period less than 30, then it is a seasonal component; otherwise it is a trend. Then, even in this case of correctly chosen \( T_{max} \), it is not easy (or even impossible) to correctly realize the underlying seasonal components. The issue, of course, is that the stochastic term is relatively large.

Now let us look at how the conventional estimator (5.1.6) and the nonparametric estimator perform for time series generated as in Figure 5.1.

Figure 5.2 exhibits estimated values of seasonal components for 8 different seasonal components, which are our familiar corner functions minus 1. What we see is an example of how to choose an optimal smoothing. Recall that the nonparametric estimator smoothes estimates calculated by the conventional estimator, and it performs well only if the period of a seasonal component is relatively large. In all the cases, with the apparent exception of the Delta (and maybe the Strata), the nonparametric estimator performs
188  5. Nonparametric Time Series Analysis for Small Samples

FIGURE 5.2. Seasonal components computed by the conventional method (5.1.6) (shown by triangles connected by dotted lines) and by the nonparametric universal estimator (dashed lines), which smooths the conventional estimates. The underlying seasonal components are shown by solid lines. Time series are generated similarly to ones shown in Figure 5.1. (The first 5 arguments allow one to change the simulated time series; they are explained in the caption of Figure 5.1. The rest of the arguments control the coefficients of the universal estimator explained in the caption of Figure 4.5.) \[n=100, Per=20, sigma=1, a=.4, b=.3, s0=.5, s1=.5, CJ0=4, CJ1=.5, CJM=6, cT=4, r=2, cB=2\]

well. In other words, it smoothes correctly. For the cases of the Strata and apparently the Delta, the conventional method is better. Exercise 5.1.14 is devoted to choosing optimal values of coefficients of the nonparametric estimator, and Exercise 5.1.15 to the cases where it is worthwhile to employ this estimator. In short, we should keep in mind that in the regression model (5.1.7) the period \(T\) of an estimated seasonal component (regression function) plays the role of the sample size. Thus the case of the period \(T = 20\) is a challenging problem for a nonparametric estimator.

5.2 Estimation of Spectral Density

There are two rather distinct approaches to the analysis of stationary time series: the spectral (frequency) domain approach and the time domain (dynamic) approach. The particular strength of the spectral approach is the simplicity of visualization of periodicities and separation long-term and short-term effects, whereas the time domain approach with its explicit equations for an underlying time series, and an important particular case of ARMA\((p, q)\) processes, is easy for predictions and describing the dynamics
of time series. This section is concerned with the first approach, namely, with nonparametric spectral density estimation, but with eyes open to the possibility that an underlying process is an ARMA($p, q$) time series.

Analysis of time series is customarily based on the assumption of second-order stationarity after removing the trend and seasonal components and (if necessary) rescaling the original data. This explains why both the study and estimation of second-order characteristics is the most important topic in the analysis of such time series. Let $X_t$, for $t = \ldots, -1, 0, 1, \ldots$, be a second-order stationary time series with mean 0 and autocovariance function $\gamma(j) := E\{X_{t+j}X_t\}$. Then the second-order properties of a time series are completely described by its autocovariance function, or, equivalently, under mild conditions (for instance, a sufficient condition is $\sum_{j=-\infty}^{\infty} |\gamma(j)| < \infty$), by its Fourier transform, which is called the spectral density function,

$$f(\lambda) := (2\pi)^{-1} \sum_{j=-\infty}^{\infty} \gamma(j) \cos(j\lambda) = (2\pi)^{-1} \gamma(0) + \pi^{-1} \sum_{j=1}^{\infty} \gamma(j) \cos(j\lambda), \quad -\pi < \lambda \leq \pi. \quad (5.2.2)$$

Here the frequency $\lambda$ is in units radians/time, and to get (5.2.2) we used the relation $\gamma(-j) = \gamma(j)$.

Because the autocovariance function is symmetric, the spectral density is also symmetric in $\lambda$ about 0, i.e., the spectral density is an even function. Thus, it is customary to consider a spectral density on the interval $[0, \pi]$. The spectral density is also a nonnegative function (like the probability density), and this explains why it is called a density.

Formula (5.2.1) shows why the spectral density is such a good tool for searching for periodicities; indeed, a peak in $f(\lambda)$ at frequency $\lambda = \lambda^*$ indicates a possible periodic phenomenon with period

$$T^* = \frac{2\pi}{\lambda^*}. \quad (5.2.3)$$

This formula explains why spectral domain analysis is the main tool in searching after periods of seasonal components. The next section gives us an example of how to use this formula.

Now let us explain how to estimate the spectral density. Let a finite realization $X_1, \ldots, X_n$ of a second-order stationary time series (recall that we always assume that its mean is zero) be given. The classical sample autocovariance estimator is defined as

$$\hat{\gamma}(j) := n^{-1} \sum_{l=1}^{n-j} X_{l+j}X_l, \quad j = 0, 1, \ldots, n-1. \quad (5.2.4)$$

Note that the divisor $n$ is not equal to the number $n - j$ of terms in the sum. Thus, the sample autocovariance is a biased estimator. On the other
hand, this divisor ensures that an estimate corresponds to some second-order stationary series. (For all our purposes the divisor \(n - j\) may be used as well.)

Then, according to (5.2.2), if one wants to estimate a spectral density, a natural step is to plug in the sample autocovariance function in place of an unknown autocovariance function. The resulting estimator (up to the factor \(1/2\pi\)) is known as a periodogram,

\[
I(\lambda) := \hat{\gamma}(0) + 2 \sum_{j=1}^{n-1} \hat{\gamma}(j) \cos(j\lambda) = n^{-1} \left| \sum_{l=1}^{n} X_l e^{-i\lambda l} \right|^2.
\]

Here \(i\) is the imaginary unit, i.e., \(i^2 = -1\), \(e^{ix} = \cos(x) + i \sin(x)\), and the periodogram is defined at the so-called Fourier frequencies \(\lambda_k := 2\pi k/n\), where \(k\) are integers satisfying \(-\pi < \lambda_k \leq \pi\). Examples of periodograms are given below.

This simple tool for spectral-domain analysis, invented in the late nineteenth century, has been both the glory and the curse of this analysis. The glory, because many interesting practical problems were solved at a time when no computers were available. The curse, because the periodogram, which had demonstrated its value for locating periodicities, proved to be an erratic and inconsistent estimator.

The reason for the failure of the periodogram is clear from the point of view of nonparametric curve estimation theory discussed in Chapter 3. Indeed, based on \(n\) observations, the periodogram estimates \(n\) Fourier coefficients (values of an underlying autocovariance function) and then just plugs them in. This explains the erratic performance and inconsistency.

Thus, it is no surprise that in the 1940s interest in frequency-domain inference was reawakened by ideas of averaging (smoothing) the periodogram in the neighborhood of each Fourier frequency (today known as kernel smoothing, discussed in Chapter 8) and by procedures of orthogonal series estimation, in which the sample autocovariance function is smoothed. In particular, the latter approach led to lag-window Tukey estimators

\[
\tilde{f}(\lambda) := (2\pi)^{-1} \hat{\gamma}(0) + \pi^{-1} \sum_{j=1}^{J} w(j/J) \hat{\gamma}(j) \cos(j\lambda),
\]

which are the cosine series estimators familiar from the previous chapters. Here the lag window function \(w(x)\) is such that \(|w(x)| \leq 1\) and \(w(x) = 0\) for \(x > 1\), and \(J\) is called the window width or cutoff. For instance, the simplest lag window function is rectangular, where \(w(x) = 1\) for \(x \leq 1\), and this implies a truncated estimator.

This series estimator is the most apparent application of the orthogonal series approach, since the spectral density is defined via the cosine series.

Thus, for the problem of estimation of the spectral density, the universal data-driven estimator (3.1.15) of Section 3.1 may be employed
straightforwardly with \( \hat{\gamma}(j) \) used in place of \( \hat{\theta}_j \) and where the coefficient of difficulty,

\[
d := 2\pi \int_{-\pi}^{\pi} f^2(\lambda) d\lambda = \gamma^2(0) + 2 \sum_{j=1}^{\infty} \hat{\gamma}^2(j),
\]

(5.2.7)
is estimated by

\[
\hat{d}_n := \hat{\gamma}^2(0) + 2 \sum_{j=1}^{J_n} \hat{\gamma}^2(j).
\]

(5.2.8)

Here the sequence \( J_n \) is the same as in Section 3.1.

Exercise 5.2.8 shows that if an underlying time series is a causal ARMA process with bounded fourth moments, then

\[
E\{ (\hat{\gamma}(j) - \gamma(j))^2 \} = d n^{-1} (1 + r_{nj}),
\]

(5.2.9)

where \( r_{nj} \to 0 \) as both \( n \) and \( j \) increase. Relation (5.2.9) explains formula (5.2.7) for the coefficient of difficulty of estimation of the spectral density.

Figure 5.3 illustrates the performance of the estimator for an underlying Gaussian ARMA(1, 1) time series \( Y_t - 0.4Y_{t-1} = 0.5(Z_t + 0.5Z_{t-1}) \). The top diagram shows a particular realization that “slowly” oscillates over time. This is because here the covariance between \( Y_t \) and \( Y_{t-1} \) is positive. This follows from the following formula for calculating the autocovariance function of the causal ARMA(1, 1) process \( Y_t - aY_{t-1} = \sigma(Z_t + bZ_{t-1}), \) \(|a| < 1:\)

\[
\gamma(0) = \frac{\sigma^2[(a + b)^2 + 1 - a^2]}{(1 - a^2)}, \quad \gamma(1) = \frac{\sigma^2(a + b)(1 + ab)}{(1 - a^2)},
\]

\[
\gamma(j) = a^{j-1} \gamma(1), \quad j \geq 2.
\]

(5.2.10)

See sketch of the proof in Exercise 5.2.9. Note that if \( a > 0 \) and \( b > 0 \), then \( \gamma(1) > 0 \), and a realization will “slowly” change over time. On the other hand, if \( a + b < 0 \) and \( 1 + ab > 0 \) (for instance, consider a moving average MA(1) process \( Y_t = \sigma(Z_t + bZ_{t-1}) \) with negative \( b \)), then a realization may change its sign almost every time. Thus, depending on \( a \) and \( b \), we may see either slow or fast oscillations in a realization of an ARMA(1, 1) process.

Figure 5.3.2 shows the underlying theoretical spectral density of the ARMA(1, 1) process. As we see, because here both \( a \) and \( b \) are positive, in the spectral domain low frequencies dominate high frequencies. (To look at the inverse situation, the MA(1) process mentioned earlier may be considered.) The formula for calculating the spectral density is \( f(\lambda) = \sigma^2 |1 + be^{i\lambda}|^2 / [2\pi |1 - ae^{i\lambda}|^2] \), and it is a particular case of the
The top diagram shows a particular realization of a Gaussian ARMA(1, 1) time series $Y_t - a Y_{t-1} = \sigma (Z_t + b Z_{t-1})$, $t = 1, 2, \ldots, n$, where $a = 0.4$, $b = 0.5$, $\sigma = 0.5$, and $n = 120$. The diagram below shows the spectral density of this ARMA process. The two bottom diagrams show the periodogram estimate and the universal spectral density estimate. \{The length $n$ of a realization is controlled by the argument $n$. The parameters of an ARMA(1, 1) process are controlled by the arguments $\sigma$, $a$, and $b$. Use $|a| < 1$. All the other arguments control the coefficients of the universal estimator (3.1.15), and they are explained in the caption of Figure 3.2. Note that the string $sp$ is added to these arguments to indicate that they control the coefficients of the universal spectral density estimator.\} 

\[ n=120, \sigma=.5, a=.4, b= .5, cJ0sp=4, cJ1sp=.5, \\
\text{cJM}sp=6, cJTsp=4, cBsp=2 \]

following formula for a causal ARMA($p,q$) process defined at (5.1.2),

$$f(\lambda) = \frac{\sigma^2 \left| 1 + \sum_{j=1}^{q} b_j e^{-ij\lambda} \right|^2}{2\pi \left| 1 - \sum_{j=1}^{p} a_j e^{-ij\lambda} \right|^2}. \quad (5.2.11)$$

Now let us see how the periodogram (5.2.5) and the universal nonparametric estimator show us the underlying spectral density. (The interesting feature of positively correlated time series is that it may create an illusion of a seasonal component; see Figure 5.3.1. Thus, it will be of a special interest to watch how the nonparametric estimates handle such a realization.) The periodogram is shown in Figure 5.3.3. As we see, it does not resemble the underlying spectral density, and moreover, its mode at frequency $\lambda^* \approx 0.55$ indicates the possibility of a seasonal component with period.
11 (the formula (5.2.3) was used to find this period). It is easy to believe in this conclusion after visualizing the series in Figure 5.3.1, but we know that this is just an illusion and there is no seasonal component. This is why a periodogram cannot be used as a reliable tool for searching for cyclical components.

The bottom diagram exhibits the universal estimate, which correctly shows the absence of any seasonal component (there are no modes in the frequency region of possible seasonal components). Also, the estimate nicely resembles the underlying spectral density.

Now let us discuss the following important question, which always arises when one uses a nonparametric estimator for the case of a parametric underlying model. Assume that we have some information about an underlying time series, for instance, that it is an ARMA($p, q$) process. Then, is it worthwhile to use a nonparametric estimator that ignores this information?

To answer this question, let us make some preliminary comments about parametric spectral density estimators. For the case of a Gaussian AR time series, a well-known parametric adaptive spectral density estimator, supported by S–PLUS, is an estimator based on Akaike’s information criterion (AIC). In short, this is a parametric penalized maximum likelihood estimator of the order $p$; see more in Chapter 8 about the method. We do not discuss this parametric estimate in more detail because it is supported by S–PLUS and we can use it as a given tool. (If an underlying process is ARMA($p', q$), then S–PLUS recommends approximating it by an AR($p$) process, that is, again use that parametric estimate.) The only information that is required by this estimator is the largest possible value of $p$.

Thus, let us explain how to compare our universal nonparametric estimator with this parametric one. We perform a Monte Carlo study that should be absolutely favorable to the parametric estimate; here this means that an underlying time series is a Gaussian AR($p$) process with $p \leq 7$, and this maximal order 7 is given to the parametric estimator. Then, the parametric estimator is used as an oracle (because it knows the underlying model) for the nonparametric one, and their performances are compared. (Note that the idea of this experiment resembles the experiments with oracles discussed in Sections 3.2–3.3.)

Our particular experiment is as follows. For each pair $(p, n)$ of $p \in \{1, 2, 3, 4, 5, 6, 7\}$ and $n \in \{30, 50, 100, 300, 500, 1000\}$, 1,000 independent Monte Carlo simulations of a causal Gaussian AR($p$) time series are performed where roots of the autoregressive polynomials are iid uniform with absolute values between 2 and 10. Then, for each realization from the set of 1,000 simulations, the parametric oracle’s estimate and the universal estimate are calculated, and the ratios of their integrated squared errors (ISE) are computed. Table 5.1 displays the sample medians of these ratios. If the ratio is larger than 1, then the oracle (Akaike’s parametric estimator, which knows that the underlying model is AR($p$) and $p \leq 7$) is better than the universal estimator, and vice versa.
Table 5.1. Median Ratios of Sample ISE: Universal/Oracle

<table>
<thead>
<tr>
<th>p</th>
<th>n = 30</th>
<th>n = 50</th>
<th>n = 100</th>
<th>n = 300</th>
<th>n = 500</th>
<th>n = 1000</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1.2</td>
<td>1.5</td>
<td>1.5</td>
<td>1.7</td>
<td>2.8</td>
<td>2.1</td>
</tr>
<tr>
<td>2</td>
<td>1.3</td>
<td>1.3</td>
<td>1.5</td>
<td>1.6</td>
<td>1.7</td>
<td>1.8</td>
</tr>
<tr>
<td>3</td>
<td>0.9</td>
<td>1.3</td>
<td>1.3</td>
<td>1.3</td>
<td>1.5</td>
<td>1.5</td>
</tr>
<tr>
<td>4</td>
<td>0.9</td>
<td>1.2</td>
<td>1.3</td>
<td>1.3</td>
<td>1.4</td>
<td>1.4</td>
</tr>
<tr>
<td>5</td>
<td>1.1</td>
<td>1.0</td>
<td>1.0</td>
<td>1.1</td>
<td>1.1</td>
<td>1.5</td>
</tr>
<tr>
<td>6</td>
<td>1.0</td>
<td>0.9</td>
<td>1.0</td>
<td>1.2</td>
<td>1.1</td>
<td>1.5</td>
</tr>
<tr>
<td>7</td>
<td>1.0</td>
<td>1.0</td>
<td>1.0</td>
<td>1.2</td>
<td>1.1</td>
<td>1.5</td>
</tr>
</tbody>
</table>

There is no surprise that the outcome is favorable to the parametric oracle, especially for large \( n \) and small \( p \); after all, the oracle “knows” the underlying model up to a fixed number of parameters, whereas the nonparametric estimates are based only on data. Nonetheless, the outcome of the experiment is promising, because the ratios are very reasonable, especially for the smallest sample sizes (which we are primarily interested in) and larger orders (more complicated models). In short, even if one knows the underlying AR process up to several parameters, the best parametric estimator does not significantly outperform the universal data-driven estimator for the case of small sample sizes.

To shed further light on the issue, consider a similar experiment only for some specific Gaussian AR(1) processes \( X_t - aX_{t-1} = Z_t \) with \( a \) equal to 0.5, 0.1, 0.05, and 0. For \( n = 100 \), the ratios are 1.6, 0.95, 0.92, and 0.87. Thus for the case of small \( a \) (including a white noise time series), the nonparametric estimator outperforms the parametric ones. Also note that if an underlying process is not AR(1) but, for instance, a Gaussian MA(1) process \( X_t = Z_t + 0.5Z_{t-1} \), then the ratio is 0.41, that is, the nonparametric estimator dramatically outperforms the parametric one.

Thus, it is fair to conclude that only for the cases where a practitioner is absolutely sure in an underlying parametric dynamic model is there an incentive to use only a parametric estimator. Otherwise, it is wiser to begin with a nonparametric estimate as a “first look at the data at hand” and then, if it confirms a prior opinion about an underlying parametric model, use a parametric estimator. Such a conservative approach allows one to avoid inconsistent estimation due to a wrong prior assumption.

### 5.3 Example of the Nonparametric Analysis of a Time Series

Let us combine all the earlier steps in the nonparametric analysis of a time series and explore them together via an example. The example and all steps are illustrated by Figure 5.4. Because everything in this section is about this figure, this is the only section where it is also discussed how to repeat
5.3 Example of the Nonparametric Analysis of a Time Series 195

![Diagrams](image)

**FIGURE 5.4.** The comprehensive nonparametric analysis of a time series. In diagrams 3, 8, and 10 the estimates are shown by dashed lines and the underlying functions by solid lines. In diagram 6 the conventional estimate is shown by squares, and its smoothing by the universal estimate is shown by the solid line. The subtitle to diagram 7 shows which estimate, c - conventional or u - universal, was used. The subtitle to diagram 10 shows the coefficients of the ARMA(1, 1) process that give the best fit to the time series of rescaled residuals. {Use $|a| < 1.$} $\text{n}=120, \text{trendf}=3, \text{scalef}=2, \text{sigmasc}=0.5, \text{ss}=1, \text{sc}=1, a = -.3, b = -.5, \text{TMAX}=35, \text{Tseas}=10, \text{ManualPer}=F, \text{seasest} = "c", \text{set.period}=c(8,12), \text{set.lambda}=c(0,2), \text{lbscale}=1, s0=0.5, s1=0.5, cJ0=4, cJ1=5, cJM=6, cT=4, r=2, cB=2, cJ0sp=4, cJ1sp=5, cJMsp=6, cJTsp=4, cBsp=2}

this figure using the software (it is simply impossible to discuss all details in the caption).

The underlying deterministic part $f(t) + S(t)$, $1 \leq t \leq n$, is shown in Figure 5.4.1, and it resembles many practical examples. Here the trend $f(t)$ is the Bimodal corner function (with domain $[1, n]$), and the seasonal component is a trigonometric function $S(t) := s_s \sin(2\pi t/T_{seas}) + s_c \cos(2\pi t/T_{seas})$ with the period $T_{seas}$. The length of observations $n$ is controlled by the argument $n$ with the default value 120. The trend component is chosen by the argument `trendf`, and the seasonal component is set by the arguments...
ss, sc, and Tseas; the default values of these arguments are trendf = 3, ss = 1, sc = 1, and Tseas = 10.

The stationary stochastic term is generated by a normed ARMA(1, 1) process \( \varepsilon_t = \varepsilon'_t/(E\{\varepsilon'_t^2\})^{1/2} \), where \( \varepsilon'_t - a\varepsilon'_t = Z_t + bZ_{t-1} \) and \( \{Z_t\} \) is a standard Gaussian white noise. The default values are \( a = -0.3 \) and \( b = -0.5 \).

We discussed an ARMA(1, 1) process earlier; thus we may predict that this particular stochastic component will be highly oscillatory and its spectral density should monotonically increase in frequency. Then this stationary stochastic term is multiplied by a scale function. The scale function is a coefficient \( \sigma_{sc} \) times 1 plus the Normal corner function with the domain \([1, n]\), i.e., the scale function is \( \sigma_{sc}(1 + f_2(l/n)) \), where \( f_2(x) \) is the Normal corner function. The choice of a corner function, used in the scale function, is controlled by the argument scalef with the default value 2, and the factor \( \sigma_{sc} \) is controlled by the argument sigmasc with the default value 0.5.

Data are generated by adding the scaled stochastic term to the deterministic one. A particular realization is shown by dots in Figure 5.4.2, and this is the data set (time series) at hand. Can you realize the underlying trend, seasonal component, scale function, and the structure of the noise from the data? The answer is probably “no,” so let us see how the nonparametric data-driven procedures discussed earlier handle this data set.

The first step is the nonparametric estimation of the trend. Recall that according to Section 5.1, the trend and seasonal components are separated in the frequency domain, see (5.1.5), and the boundary \( J_{max} \) is defined via a manually chosen \( T_{max} \). For this data set we choose the default \( T_{max} = 35 \), which according to (5.1.8) implies \( J_{max} = 7 \); the choice of \( T_{max} \) is controlled by the argument TMAX. By choosing this default value we assume that a cosine approximation (5.1.5) with \( J_{max} = 7 \) may approximate well an underlying trend and, at the same time, does not touch a possible seasonal component, which, by the assumption, has a period less than 35.

The nonparametric estimate of the trend (the dashed line) is shown in Figure 5.4.3. It clearly oversmooths the underlying trend, but it is necessary to be fair toward this estimate. Yes, this estimate is much worse than Binomial’s best estimates, which we saw in the previous chapters. On the other hand, now the problem is much more complicated: The setting is heteroscedastic with the pronounced scale function, the errors are dependent, and there is a significant seasonal component whose period and magnitude are comparable with the distance and the difference between the modes of the underlying Bimodal trend shown by the solid line. This is what causes the trend estimate to be essentially smoother than the underlying Bimodal trend. Actually, even by visualizing the first diagram where the deterministic part is shown, it is not an easy task to realize the modes of the underlying Bimodal model, and the situation becomes much more complicated with the noisy time series exhibited in the second diagram.

The next step is to detrend the data (subtract the estimated trend from the original data), and the result is shown in the fourth diagram (Figure
5.3 Example of the Nonparametric Analysis of a Time Series

Now, based on this time series, one must recover the underlying seasonal component. Can you recognize the seasonal component in this detrended data? Because we know that the seasonal component is a trigonometric function, we can see it in the right tail, less so in the left tail, but in the main middle part of the time series the seasonal component is absolutely blurred. This is what makes the heteroscedastic setting so complicated. (One of the options is to use a Box–Cox transformation discussed in the subsection “Estimation of Scale Function” of Section 5.1; we do not use it here because we would like to see how the “pure” nonparametric methods will perform.)

Now let us see how our nonparametric analysis performs. The nonparametric spectral density estimate of the detrended data is shown in Figure 5.4.5. (Recall that as in Section 5.2, arguments of the spectral density estimate have the attached string \( sp \), for instance, \( cJ0sp \) is the argument that controls the coefficient \( c_{J0} \) of this estimate. This allows one to use separate arguments for the regression estimate, which recovers the trend and scale functions, and the spectral density estimate.)

Diagram 5 indicates that the detrended data have a spectral density with a pronounced mode at the frequency about 0.6. The period 9.62 (the estimated period) calculated according to (5.2.3) is given in the subtitle. The corresponding rounded (to the nearest integer) period is 10, and this is exactly the underlying period.

While for these particular data the rounded estimated period has been determined correctly, this is not always the case. The small sample sizes and large errors may take their toll and lead to an incorrect estimate of the period. We shall discuss such a case a bit later.

Then the rounded estimated period is used to estimate the underlying seasonal component. Squares in Figure 5.4.6 show the conventional estimate (5.1.6); the solid line shows how the universal nonparametric estimate smooths the conventional estimate. As we see, the conventional estimate is not perfect, but it is fairly good for the setting considered. Note that its magnitude is correct, and the phase is shown absolutely correctly. Keep in mind that each point is the average of just 12 observations, so even for a parametric setting this would be considered a small sample size. The nonparametric estimate apparently oversmooths the data because the period 10 is too small; recall the discussion in Section 5.1.

The next step is to deseasonalize the detrended data, that is, to subtract an estimate of the seasonal component. The argument \( seast \) (which is shorthand for seasonal estimate) allows one to use either the conventional or the universal nonparametric estimate of the seasonal component by setting \( seast = "c" \) or \( seast = "u" \), respectively. The data obtained are shown in Figure 5.4.7, and the argument used is given in the subtitle. Note that at this step the data may be referred to as the time series of residuals because the original data set is detrended and deseasonalized (the estimated deterministic part is removed).
The detrended and deseasonalized time series is clearly not stationary, since its variability in the middle part is essentially larger than in the tails. This conclusion is supported by the estimate of the underlying scale function (which is $\sigma_{sc}(1 + f_2(t/n)$ with $f_2(x)$, $0 \leq x \leq 1$, being the Normal corner function and $\sigma_{sc} = 0.5$). The scale function is shown by the solid line in Figure 5.4.8. The estimate (dashed line) is almost perfect in the middle part, but the tails “spoil” the outcome. This is explained by the cumulative effect of a not perfect estimate of the deterministic part, the small sample size, and dependent errors. Nevertheless, under the circumstances and because the range of the underlying scale function is shown just perfectly, it is fair to rate this particular scale estimate as a good one.

The next step is to rescale the residuals shown in Figure 5.4.7 to obtain a stationary noise. The rescaled residuals are simply the residuals divided by the estimated scale function. To avoid a zero divisor, the estimate is truncated from below by the argument $\text{lbscale}$; the default value is 0.1.

Thus, here we divide the detrended and deseasonalized data shown in Figure 5.4.7 by the estimated scale function shown in Figure 5.4.8. The result is shown in Figure 5.4.9. The hope is that these data are stationary and that they correspond to a simple stochastic process like an ARMA($p$, $q$) with small orders $p$ and $q$. Visual analysis shows that there is no apparent trend, or a seasonal component, or a scale function. Thus, our final step is to look at the spectral density estimate of the rescaled residuals. The estimate (the dashed line) is shown in Figure 5.4.10. As we see, the estimate exhibits no pronounced modes (which can indicate the presence of deterministic periodic components), and we see that in this time series high frequencies dominate low frequencies. Thus, this time series looks like a stationary one, and with the help of the experience gained from Figure 5.3, we may conjecture that an ARMA(1, 1) process $\varepsilon_t - a\varepsilon_{t-1} = \sigma(Z_t + bZ_{t-1})$ with negative $a$ and $b$ may be a good bet on an underlying stochastic term. Indeed, the underlying spectral density (the solid line) has a similar shape, and the fact that it is below the estimate tells us that the rescaled residuals have a larger variance than a typical realization from the underlying ARMA(1, 1) process where $a = -0.3$ and $b = -0.5$. Also, the subtitle shows us the estimated parameters of the ARMA(1, 1) process that gives the best fit to the data. They are obtained using the S–PLUS function (parametric maximum likelihood estimate) arima.mle.

This finishes our analysis of this particular time series.

Now let us return to Figure 5.4.5. Here the frequency of the mode correctly defines the period by the formula (5.2.3), but this is not always the case. First, there may be several local modes created by both a seasonal component and a stochastic component, and large errors may produce a wrong global mode. As a result, the period will be estimated incorrectly. One of the possibilities to avoid such a complication is to use prior information about the domain of possible periods. To play around with this possibility, two arguments are added to Figure 5.4, namely, $\text{set.period}$ and
set.lambda. The first one, set.period = c(T1,T2), allows one to skip estimation of a seasonal component whenever an estimated period is beyond the interval [T1,T2]. The second argument, set.lambda = c(\lambda_1, \lambda_2), allows one to restrict the search for the mode to this particular frequency interval. While these two arguments do a similar job, they are good tools for gaining the necessary experience in dealing with the time and frequency domains. {Graphics 6 and 7 are skipped if the estimated period is beyond the interval [T1,T2], in which case a warning statement is issued.}

The second reason for the failure of the estimation of the period is that due to large noise and small sample size, the mode of an estimated spectral density may be relatively flat. As a result, even if a spectral density estimate is close to an underlying density in the sense of integrated squared error, locations of the estimated mode and the underlying mode may differ significantly. To understand why, consider, as an example, frequencies \( \lambda_1^* = 0.6, \lambda_2^* = 0.59, \) and \( \lambda_3^* = 0.54. \) Then, according to (5.2.3), the corresponding periods are \( T_1^* = 2\pi/0.6 = 10.47, \) \( T_2^* = 2\pi/0.59 = 10.64, \) and \( T_3^* = 2\pi/0.54 = 11.63, \) which imply the rounded periods 10, 11, and 12, respectively. Thus, due to the rounding a relatively small error in the location of a mode may imply a significant error in the estimated period.

Two questions immediately arise: how to detect such a case and how to correct the mistake. To answer the first question, let us look at another realization of Figure 5.4 (i.e., another realization of the noise term), shown in Figure 5.5. As we see, here the estimated period (see the subtitle for Figure 5.5.5) is 10.87, and this leads to the wrong period, 11. Let us assess the consequences of using this wrongly estimated period. First, the estimated seasonal component in no way resembles the underlying one. While this rather chaotic estimate cannot be the indicator of a wrongly estimated period, it should raise a flag of suspicion. Then, we see that the rescaled residuals in Figure 5.5.9 apparently exhibit a cyclical component. This is a one more reason to suspect the mistake. The estimated scale function (the dashed line in diagram 8) is dramatically oversmoothed because now the subtracted estimated seasonal component plays the role of an extra additive noise.

Finally, the estimated spectral density (the dashed line) in Figure 5.5.10 indicates that the seasonal component was not removed. Indeed, we see that the shape of this estimate resembles the shape of the spectral density of detrended data shown in Figure 5.5.5. This spectral density of rescaled residuals is the most reliable indicator of a wrongly estimated period.

The obvious method to cure such a mistake is to try a different period for estimation of a seasonal component, and here the apparent choice of the period is \( T = 10, \) which is the rounded-down estimated period. {To do this, set the argument ManualPer = T (in S–PLUS “T” stands for “True” and “F” for “False”). This stops the calculations at diagram 5, and the first 5 diagrams are displayed. Then the program prompts for entering a period from the keyboard. At the prompt 1: enter a period (here it should
be 10, but any integer period may be tried) from the keyboard and then press Return; then at the prompt 2: just press Return. This completes the procedure, and the seasonal component will be calculated with the period entered. The period will be shown in the subtitle of diagram 6.} We do not illustrate this procedure by a separate figure because the result is similar to diagrams shown in Figures 5.4.6–5.4.10.

Another useful practical comment is as follows. In many cases a spectral density estimate of rescaled residuals has a relatively large left tail while an underlying theoretical spectral density does not. A particular example will be given in the next section. One of the typical reasons for such a mistake is a poorly estimated trend. Unfortunately, for the cases of small sample sizes and relatively large errors there is no cure for this “disease,” but knowledge of this phenomenon may help in explaining a particular outcome.
5.4 Case Study: Missing Observations

In the previous section we have considered the case of a classical realization $X_1, X_2, \ldots, X_n$ of a time series $\{X_t\}$. In many practical situations some of the observations may be skipped (missing). This happens due to stochastic circumstances or because there is no way to obtain realizations at some particular moments. Classical examples of the second case are as follows. Suppose that a researcher should collect some daily data about a group of students at a particular school. Since schools are closed on weekends, every sixth and seventh observation will be missed. Another example is observations of an object from a satellite that periodically “loses” the object. We shall see that a case of deterministically skipped data may be essentially worse than a case of data skipped at random. Thus, the more difficult case of deterministically skipped data will be of our primary interest.

Another interesting practical interpretation of the setting is the case of spatial data (recall the discussion in Section 5.1) that are not collected at a regular grid but may be approximated by a model of data at a regular grid with skipped (missing) observations. For such a setting, for instance, geostatistics considers the problem of interpolation of an underlying trend at skipped points as one of the most important.

What are the necessary changes in our estimates to consider a case of missing observations? Let us again consider the general problem illustrated by Figure 5.4 and discuss a similar analysis for the particular example where every sixth and seventh observation is missed. This mimics a time series of weekly observations with missing weekends.

Figure 5.6 illustrate this setting. The first diagram shows by dots the unobserved deterministic part, which is the same as in Figure 5.4.1, only here every sixth and seventh observation (“weekends”) is skipped. Figure 5.6.2 shows observed noisy observations (data at hand). Note that in this time series of length 120 only 86 observations are available and 34 are missing. Thus, the quality of estimation should be worse than for the case considered in Figure 5.4 simply because the sample size is smaller.

The smaller number of observations is not the only issue to worry about. Let us consider a case where every other realization is skipped. Then there is no way to estimate an underlying autocovariance function (just think about estimation of $\gamma(1) = E\{X_{t+1}X_t\}$). Similarly, if a seasonal component has a period equal (or for small samples even close) to the period of missing realizations, then a problem of estimating such a seasonal component becomes impossible. As an example, if for the case of missing weekends a seasonal component has period equal to 7 (a weekly seasonal component), then there is no way to estimate values of this seasonal component at weekends because there are no such observations. Note that this is not the case if the periods are different. For instance, if a seasonal component has period 10, then during the first 10 days the sixth and seventh observations of the seasonal component are missing, during the second 10 days the third and
In other words, the universal estimator is robust to missing observations. Because a heteroscedastic regression allows any spacing of predictors, no changes are needed at this step because a heteroscedastic regression allows any spacing of predictors.

Since we use the nonparametric estimator of Section 4.2, that is, the estimator for a heteroscedastic regression model, no changes are needed at this step because a heteroscedastic regression allows any spacing of predictors. In other words, the universal estimator is robust to missing observations.

The sequence of available and missing observations is controlled by the argument \texttt{set.obs} with the default value \texttt{set.obs=c(1,1,1,1,0,0)}, which implies weekly observations with missing weekends. \{\texttt{n=120, set.obs=c(1,1,1,1,1,0,0), trendf=3, scalef=2, sigmasc=.5, ss=1, sc=1, a = -.3,b= -.5, TMAX=35, Tseas=10, ManualPer=F, seasest= "c", set.period=c(8,12), set.lambda=c(0.2), lbscale=.1, s0=.5, s1=.5, cJ0=4, cJ1=.5, cJM=6, cT=4, r=2, cB=2, cJ0sp=4, cJ1sp=.5, cJMsp=6, cJTsp=4, cBsp=2}\}

FIGURE 5.6. Nonparametric time series analysis for the case of missing observations. The structure of this figure is the same as that of Figure 5.4. \{The sequence of available and missing observations is controlled by the argument \texttt{set.obs} with the default value \texttt{set.obs=c(1,1,1,1,0,0)}, which implies weekly observations with missing weekends.\}
The estimate is shown in Figure 5.6.3, and it is relatively good. Note that here the estimated trend is based on a smaller sample size than the estimates in Figures 5.4.4 and 5.5.4, and furthermore, the observations are regularly missing.

The detrended time series is shown in Figure 5.6.4. Note that here again every sixth and seventh realization is skipped.

The next step is to estimate the spectral density of the detrended data. The estimator of Section 5.2 cannot be used here because the sample autocovariance estimator (5.2.4) requires all \( n \) observations. However, it is not difficult to find a reasonable substitution for the sample autocovariance estimator. Recall that 
\[
\hat{\gamma}(j) := E\{X_{t+j}X_t\},
\]
so an unbiased estimator is
\[
\tilde{\gamma}(j) := \frac{1}{\hat{m}_j} \sum_{l \in \hat{M}_j} X_{l+j}X_l,
\]
where \( \hat{M}_j \) is a random set of \( l \in \{1, 2, \ldots, n\} \) such that pairs \((X_{l+j}, X_l)\) are observed (i.e., both \( X_{l+j} \) and \( X_l \) are not missing) and \( \hat{m}_j \) is the number of such pairs. Then the estimator of Section 5.2 may be used straightforwardly with \( \tilde{\gamma}(j) \) in place of \( \hat{\gamma}(j) \) and the number \( \hat{m}_0 \) of available observations in place of \( n \).

A particular spectrum estimate is shown in Figure 5.6.5. Here it depicts the location of the mode correctly (it implies the period 10.05), but look at the huge right tail. Here the estimator ignores it because, as we discussed in the previous section, the argument \( set\_lambda \) restricts the search after the period of seasonal component to the frequencies \([0, 2]\).

When the period is found, all the following steps until Figure 5.6.10 are the same as in Figure 5.4. In particular, in Figure 5.6.8 we see that the estimate of the scale function (the dashed line) oversmooths the underlying scale function (the solid line). On the other hand, this estimate correctly exhibits the symmetric shape as well as the correct minimal value of the scale function. In Figure 5.6.10 we use the same modified spectrum estimator as in Figure 5.6.6. The particular spectral density estimate is not perfect, but it indicates that there is no pronounced seasonal component. Note that the left tail, which shows the presence of low-frequency harmonics in the rescaled residuals, is due to imperfect estimation of the trend.

5.5 Case Study: Hidden Components

In the previous sections we discussed the problem of estimating a trend component and a seasonal component in the deterministic part of a time series. There was no problem in separating these two components, since by definition, they have different spectrum domains.

Here we would like to consider a more complicated case where a trend is a linear combination of several low-frequency components. We begin with
a rather simple model of a time series with no seasonal component or a
nuisance hidden component (the general case will be considered later)
\[ Y_l := f(l) + \sigma(l)\varepsilon_l, \quad l = 1, 2, \ldots, n, \quad (5.5.1) \]
where the trend \( f(l) \) is a weighted sum of \( K \) hidden additive components,
\[ f(l) := \sum_{k=1}^{K} w_k \psi'_k(l). \quad (5.5.2) \]

The problem is to estimate either the hidden additive components \( \psi'_k(t) \),
\( k = 1, \ldots, K \), when the weights \( \{w_k\} \) are given or the weights \( w_k, k = 1, \ldots, K \),
when \( \{\psi'_k\} \) are given. Below we consider both these problems.

- **Estimation of Hidden Components.** First of all, it is apparent
that in general to estimate the hidden components one needs at least \( K \)
realizations like (5.5.1) with different weights. Thus, let us assume that
\( K \) realizations like (5.5.1) are given with \( K \) different vectors of weights
\( W_s := (w_{s1}, \ldots, w_{sK}), s = 1, \ldots, K \). Thus, we observe \( K \) different noisy
combinations of additive components,
\[ Y_{sl} := \sum_{k=1}^{K} w_{sk} \psi'_k(l) + \sigma(l)\varepsilon_{sl}, \quad l = 1, 2, \ldots, n, \quad s = 1, 2, \ldots, K, \quad (5.5.3) \]
where \( \{\varepsilon_{sl}, l = 1, 2, \ldots, n\}, s = 1, \ldots, K, \) are \( K \) independent realizations
of a second-order stationary time series.

Let us apply our orthogonal series approach to solve this problem. Define
\( \psi_k(x) := \psi'_k(xn) \) where \( \psi_k(x) \) is a function supported on \([0, 1] \), and recall
our traditional notation \( \{\varphi_j(x)\} \) for elements of the cosine basis on \([0, 1] \).
Then the problem of estimating \( \psi'_k(x) \) is equivalent to estimating \( \psi_k(x) \),
which may be solved via estimation of the Fourier coefficients
\[ u_{kj} := \int_0^1 \psi_k(x)\varphi_j(x)dx. \quad (5.5.4) \]

Using observations (5.5.3) we can estimate the Fourier coefficients
\[ \theta_{sj} := \int_0^1 f_s(x)\varphi_j(x)dx \quad (5.5.5) \]
of the trends
\[ f_s(x) := \sum_{k=1}^{K} w_{sk} \psi_k(x) \quad (5.5.6) \]
(rescaled onto the unit interval) by a sample mean estimate
\[ \hat{\theta}_{sj} := n^{-1} \sum_{l=1}^{n} Y_{sl}\varphi_j(l/n). \quad (5.5.7) \]
Let us use uppercase letters for denoting $K$-component column vectors with corresponding lowercase entries, for instance, $\Theta_j := (\theta_{1j}, \theta_{2j}, \ldots, \theta_{Kj})^\prime$, and by $W$ the $K \times K$ matrix with entries $w_{sk}$. Then, (5.5.4)–(5.5.6) imply the following system of linear equations:

$$\theta_{sj} = \sum_{k=1}^{K} w_{sk} u_{kj}, \quad 1 \leq s \leq K. \quad (5.5.8)$$

The system of linear equations (5.5.8) may be compactly written as the following matrix equation:

$$\Theta_j = W U_j. \quad (5.5.9)$$

Assume that the matrix $W$ is invertible and denote its inverse by $W^{-1}$. Then

$$U_j = W^{-1} \Theta_j. \quad (5.5.10)$$

Thus, since the entries of the matrix $\Theta_j$ may be estimated by the sample mean estimate (5.5.7), we simply plug the estimates into (5.5.10) and then get estimates $\hat{u}_{kj}$. Recall that

$$\psi_k(x) = \sum_{j=0}^{\infty} u_{kj} \varphi_j(x),$$

and because the Fourier coefficients $u_{kj}$ are estimated, our universal nonparametric estimator may be used straightforwardly.

Figure 5.7 illustrates both the setting and how the estimator performs for the case $K = 3$ and the hidden components being the Normal, the Strata, and the Monotone. First, let us look at the left column of diagrams. The top time series “First Noisy Composition” shows a particular realization of (5.5.1)–(5.5.2) with the weights shown in the subtitle. The error term is a Gaussian ARMA(1, 1) process $\varepsilon_t - 0.4 \varepsilon_{t-1} = 0.5(Z_t + 0.3Z_{t-1})$, see examples in Figure 5.1, multiplied by a scale function. The scale function is equal to 1 plus the Normal corner function with the domain $[1, n]$.

Similarly, the second and third time series are shown in the diagrams below. The analysis of these three time series reveals that even the knowledge of the underlying components and the weights does not help to realize them. This is a rather typical situation with linear combinations of functions. So let us see how the estimator, which has at hand only these 3 realizations and the corresponding weights (in other words, the data shown in the left column), solves this puzzle. Estimates of the components (dashed lines) are shown in the right column of Figure 5.7. As we see, the estimates are pretty good and allow us easily to realize the shape of the underlying hidden components (solid lines).

In practical applications the matrix of weights $W$ may not be known exactly. Figure 5.8 allows us to analyze how this can affect the estimates. It is assumed that a given matrix $\tilde{W}$ is equal to an underlying matrix
FIGURE 5.7. Recovery of hidden components. The left column of diagrams shows the time series of observations (noisy compositions); the known underlying weights are shown in the corresponding subtitles. The right column of diagrams shows the estimated components (dashed lines) and the underlying components (solid lines). The default hidden components are the Normal, Strata, and Monotone corner functions with the domain [1, n]; their choice is controlled by the argument set.adc. It is possible to consider K equal to 2, 3, or 4, and then K components should be chosen using set.adc. The underlying weights for the jth composition are controlled by the argument wj. The error term is a Gaussian ARMA(1, 1) process $\varepsilon_t - b\varepsilon_{t-1} = \sigma(Z_t + bZ_{t-1})$ multiplied by a scale function. The parameters of the ARMA(1, 1) process are controlled by the arguments a, b, and $sigma$. The scale function is equal to 1 plus a corner function whose choice is controlled by the argument scalef. All other arguments control the coefficients of the universal estimator. \[ n=120, set.adc=c(2,4,7), w1=c(2,1.5,1), w2=c(1,2,1.7), w3=c(1.4, 1.5, 2), w4=c(1,1,2), scalef=2, a=.4, b=.3, sigma=.5, s0=.5, s1=.5, cJ0=4, cJ1=.5, cJM=6, cT=4, r=2, cB=2 \]

$W$ plus a random matrix with entries being independent standard normal variables multiplied by $\sigma_1/n^{1/2}$. In other words, this mimics the case where the entries are measured with normal $N(0, \sigma_1^2/n)$ additive errors. Figure 5.8 shows three columns of estimates (dashed lines) obtained for different values of $\sigma_1$. Each column is obtained similarly to the right column in Figure 5.7, and all the estimates are based on the same data set. Thus, the only difference between the columns is that different noisy matrices $\tilde{W}$ are used. The corresponding $\sigma_1$ may be seen in the subtitles. Note that the first column, which corresponds to the case $\sigma_1 = 0$, shows estimates with the correctly known matrix of weights, the two others with noisy matrices.
FIGURE 5.8. Recovery of hidden components (solid lines) with a noisy matrix of weights. The estimator is the same as the one used in Figure 5.7, and the estimates are shown by dashed lines. The difference with Figure 5.7 is that here normal $N(0, \sigma_1^2/n)$ errors are added to the underlying weights, and these noisy weights are then used by the estimator. Each column corresponds to a specific noise level shown in the subtitle. Because the first column corresponds to the case $\sigma_1 = 0$ (the weights are known correctly), it allows one to see the effect of noisy weights on the recovery of hidden components. \{The set of $\sigma_1$ used is controlled by the argument set.sigma1.\} [set.sigma1=c(0,2,5), n=120, set.adc=c(2,4,7), w1=c(2,1.5,1), w2=c(1,2,1.7), w3=c(1.4, 1.5, 2), w4=c(1,1.1,2), scalef=2, a=.5, b=.3, sigma=.5, s0=.5, s1=.5, cJ0=4, cJ1=.5, cJM=6, cT=4, r=2, cB=2]

As we see, incorrect information about weights may lead to a wrong estimation. Figure 5.8 is a useful tool to get first-hand experience in understanding how random errors in $W$ may affect the recovery of hidden components.

- **Learning Machine for Estimating Weights.** The problem of estimating weights $\{w_k\}$ of a noisy composition (5.5.1)–(5.5.2) arises in many applications where the main issue is not to recover components but to estimate weights. For instance, in applied spectroscopy weights may be considered as concentrations of mixed substances with different spectral profiles.

  A typical complication of such a problem is that the components $\psi_k(x)$ are not known as well, so a learning machine should be used (recall the discussion of learning machines in Section 4.10). Here we consider the case where a training data set consists of noisy compositions with known weights (similar to those shown in Figure 5.7), and then a composition with un-
known weights is given (we shall refer to a composition whose weights should be estimated as the composition).

The underlying idea of a learning machine is as follows. As in (5.5.8) we may write for the Fourier coefficients \( \{ \theta_j \} \) of a trend \( f(x) \) of the composition,

\[
\theta_j = \sum_{k=1}^{K} w_k u_{kj}, \quad j = 0, 1, \ldots, \tag{5.5.11}
\]

where recall that \( \{ u_{kj}, j = 0, 1, \ldots \} \) are the Fourier coefficients (5.5.4) of the \( k \)th component \( \psi_k(x) \).

Note that were the Fourier coefficients \( \theta_j \) and \( u_{kj} \) known, then (5.5.11) implies a classical regression problem with respect to the weights \( \{ w_k \} \). In our case the Fourier coefficients are unknown, but we may estimate them from given noisy compositions and then plug them into (5.5.11). There are some complications that arise by using such a plugging-in because we get a problem of linear regression with errors in predictors (recall Section 4.11). Here we do not discuss an optimal solution but simply restrict the set of Fourier coefficients in (5.5.11) to \( j = 0, 1, 2, \ldots, J_W \) in the hope that the first Fourier coefficients are typically large and thus the errors will be relatively small. (Recall that Figure 5.8 gave us some feeling and experience in dealing with such a situation.) Set \( J_W = 5 \), and note that it must be at least \( K - 1 \). Then the corresponding linear regression problem may be solved by standard methods, here the S–PLUS function \texttt{lm} is used.

The performance of this learning machine is illustrated in Figure 5.9. The training set of 3 noisy compositions with known weights (but unknown components) is shown in the left column, the data are simulated as in Figure 5.7, and the same notation is used. The top diagram in the right column shows the simulated noisy composition; the corresponding weights are shown in the subtitle and they should be estimated. Thus the learning machine knows the 4 sets of data shown (3 training noisy compositions plus the main one), and it knows the weights for the training compositions shown in the left column.

The diagram “Estimated Components” exhibits the estimated hidden additive components; they are obtained similarly to those shown in the right column of Figure 5.7 and based only on the training sets shown in the left column. As we see, these particular estimates are not perfect but give us a fair impression about the shapes of the underlying Normal, Strata, and Monotone corner functions.

The bottom diagram shows the estimate of the composition. Roughly speaking, the learning machine then tries to fit this estimate by weighted compositions of the estimated components shown above. The important technical detail is that the learning machine does it solely via the first \( 1 + J_W \) Fourier coefficients of these 4 curves. The estimated weights are shown in the subtitle for the right bottom diagram, and this is the “answer”
given by the learning machine. For this particular set of data the estimates are 0.9, 1.9, and 3.1, and this is a good outcome for the case of 3 hidden components, the sample size 120, and dependent errors.

- Extra Nuisance Component. Consider a more general setting where in model (5.5.1) an extra nuisance additive component $G'(l)$ is presented, namely, when the time series is

$$Y_l = f(l) + G'(l) + \sigma(l)\varepsilon_l, \quad l = 1, 2, \ldots, n, \quad (5.5.12)$$

and (5.5.2) holds.

Our estimators can easily handle this case under the following assumption: $G'(l)$ is the same in all the experiments. In other words, in all the experiments the deterministic components are $f_s + G$. Under this assump-
tion, the nuisance additive component $G'$ becomes an extra $(K + 1)$th additive component with the constant weight equal to 1 for all the experiments. In other words, if we set $\psi'_{K+1}(l) := G'(l)$, $w_{K+1} := 1$, and consider the case of $K + 1$ hidden components, then the problem is reduced to the previously discussed ones with just an additional $(K + 1)$th experiment. Then, for instance, the problem of estimating the additive components is solved based on $K + 1$ experiments with different weights for the first $K$ components. {To simulate the situation, use Figure 5.7 and set the last elements in the vectors $w_s$ to 1; then the $K$th component may be considered as a nuisance one. Similar changes are needed in Figures 5.8–5.9.}

Another useful comment is as follows. Suppose that this nuisance component is a seasonal component and its frequency is beyond $J_{max}$, that is, $J_{max} + 1 \geq J_W$. Then this nuisance seasonal component has no effect on our nonparametric estimators because they perform in the low-frequency domain.

5.6 Case Study: Bivariate Time Series

In many practical situations it is necessary to analyze a pair of time series. For instance, the relationship between the price and supply of a commodity is of a central interest for econometrics, and the relationship between the number of police officers on the streets and the level of crime is of a central interest for a government.

We begin our discussion with one simple but very informative example of a bivariate time series $\{(X_t, Y_t)\}$ defined by

$$
X_t = \sigma_1 Z^X_t, \quad t = 1, 2, \ldots, n, \quad (5.6.1)
$$

$$
Y_t = bX_{t-k} + \sigma_2 Z^Y_t, \quad t = 1, 2, \ldots, n. \quad (5.6.2)
$$

Here $Z^X_t$ and $Z^Y_t$ are independent standard Gaussian white noises (that is, these time series are iid standard normal), the coefficients $\sigma_1$ and $\sigma_2$ are nonnegative, $b$ is real, and the parameter $k$, which is called a delay, is an integer and may be either positive or negative.

The important feature of this bivariate time series is that for a positive $k$ the time series $\{X_t\}$ leads the time series $\{Y_t\}$, while for negative values of $k$ the situation is reversed. Also, it is apparent that these two univariate time series have a linear relationship, since the second time series is simply a lagged multiple of the first time series with added noise.

Is it possible to realize such a structure of the bivariate time series via just visualizing its realization? Let us check this. The two top diagrams in Figure 5.10 show particular realizations of the first univariate time series $\{X_t\}$ and the second univariate time series $\{Y_t\}$. As we see, it is not an easy task to realize from visualizing these time series that they are related
in any sense. In short, due to the relatively large additive noise in (5.6.2), visualization is not fruitful.

Thus, let us try to solve this problem using a statistical approach. Recall (see Section 5.2) that for a zero-mean and second-order univariate stationary time series \( \{X_t\} \) its main characteristics are the autocovariance function

\[
\gamma_{XX}(h) := E\{X_{t+h}X_t\} \tag{5.6.3}
\]

and, under a mild assumption like \( \sum_{h=\infty}^{\infty} |\gamma_{XX}(h)| < \infty \), the corresponding spectral density

\[
f_{XX}(\lambda) := (2\pi)^{-1} \sum_{h=\infty}^{\infty} \gamma_{XX}(h)e^{-ih\lambda}, \quad \lambda \in (-\pi, \pi]. \tag{5.6.4}
\]

Because the autocovariance function is symmetric, i.e., \( \gamma(h) = \gamma(-h) \), the spectral density is a real and even function.

To analyze the relationship between two zero-mean and second-order stationary sequences \( \{X_t\} \) and \( \{Y_t\} \), we shall use very similar notions of the cross-covariance

\[
\gamma_{XY}(h) := E\{X_{t+h}Y_t\} \tag{5.6.5}
\]
and, under a mild assumption like \( \sum_{h=\infty}^{\infty} |\gamma_{XY}(h)| < \infty \), the cross spectral density or simply cross spectrum

\[
 f_{XY}(\lambda) := (2\pi)^{-1} \sum_{h=-\infty}^{\infty} \gamma_{XY}(h)e^{-ih\lambda}, \quad \lambda \in (-\pi, \pi]. \tag{5.6.6}
\]

The similarity between the auto-characteristics and cross-characteristics is striking, but there is one very important difference that is necessary to know. While any autocovariance function is always symmetric and thus any spectral density is real, a cross-covariance may be asymmetric, that is, \( \gamma_{XY}(h) \) may differ from \( \gamma_{XY}(-h) \), and this implies a complex cross spectrum. To see this, let us calculate the cross-covariance function for the example (5.6.1)–(5.6.2). Using the assumption that \( Z^X_t \) and \( Z^Y_t \) are independent standard white noises, a simple calculation shows that

\[
 \gamma_{XY}(h) = b\sigma^2 I_{\{h=-k\}}. \tag{5.6.7}
\]

Recall that \( I_{\{A\}} \) is the indicator function of an event \( A \). Thus, the cross-covariance between the series (5.6.1) and (5.6.2) is not zero only for \( h = -k \), thus it is not symmetric in \( h \). As a result, the corresponding cross spectral density (5.6.6) becomes complex and is defined by the formula

\[
 f_{XY}(\lambda) = (2\pi)^{-1} b\sigma^2 (\cos(k\lambda) + i\sin(k\lambda)). \tag{5.6.8}
\]

Thus only in the case of the zero delay \( k = 0 \) is the cross spectrum real.

Since it is not very convenient to analyze a complex function directly, we shall use the following approach. First, let us recall that any complex number may be expressed in polar coordinates. Correspondingly, a complex spectrum \( f_{XY}(\lambda) := f_r(\lambda) + if_{im}(\lambda) \) may be written as

\[
 f_{XY}(\lambda) = \alpha_{XY}(\lambda)e^{i\phi(\lambda)},
\]

where \( \alpha_{XY}(\lambda) := [f_r^2(\lambda) + f_{im}^2(\lambda)]^{1/2} \) is called the amplitude spectrum and \( \phi(\alpha) := \text{arg}(f_r(\lambda) + if_{im}(\lambda)) \in (-\pi, \pi] \) is called the phase spectrum. Note that by definition the phase spectrum lies between \( -\pi \) and \( \pi \).

Second, recall the notion of a correlation coefficient between two zero-mean random variables \( U \) and \( V \), \( \rho_{UV} := E\{UV\}/[E\{U^2\}E\{V^2\}]^{1/2} \). The correlation coefficient varies between \(-1\) and \(1\) and measures the extent to which these random variables are linearly related, namely, the larger the absolute value of \( \rho_{UV} \) the stronger the linear relationship between \( U \) and \( V \). For the case of the spectrum, we can introduce the similar notion of the absolute coherency

\[
 \mathcal{K}_{XY}(\lambda) := \frac{\alpha_{XY}(\lambda)}{[f_{XX}(\lambda)f_{YY}(\lambda)]^{1/2}}. \tag{5.6.9}
\]

The absolute coherency lies between 0 and 1, and like the coefficient of correlation it measures the extent to which these two series are linearly related at frequency \( \lambda \).
As an example, let us calculate the above-defined characteristics for the bivariate time series (5.6.1)–(5.6.2). Simple calculations show that the amplitude spectrum, the phase spectrum, and the absolute coherency are defined by the following formulae (recall that the notion of module was introduced in Section 3.5):

\[
\alpha_{XY}(\lambda) = (2\pi)^{-1}|b|\sigma_1^2, \quad (5.6.10)
\]
\[
\phi_{XY}(\lambda) = (k\lambda + \pi)[\text{mod } 2\pi] - \pi, \quad (5.6.11)
\]
\[
K(\lambda) = |b|\sigma_1/(b^2\sigma_1^2 + \sigma_2^2)^{1/2}. \quad (5.6.12)
\]

These results are the key to understanding statistical methods for the analysis of a bivariate time series. First, we see that the delay \(k\) is exactly the slope of the phase spectrum \(\phi_{XY}(\lambda)\) because this phase spectrum is piecewise linear with constant slope \(k\) (like the dotted line in the bottom diagram in Figure 5.10). Of course, this will not be the case for an arbitrary bivariate time series. However, the derivative (slope) \(d\phi_{XY}(\lambda)/d\lambda\) of the phase spectrum can still be regarded as a measure of the phase lag of \(Y_t\) behind \(X_t\) at frequency \(\lambda\). This explains why the derivative (slope) of a phase spectrum is called the group delay. (The derivative may be negative, and this indicates that \(Y_t\) leads \(X_t\).) Thus, visualizing the phase spectrum allows one to reveal which time series is the leader and which one is the follower. Second, the absolute coherency (5.6.12) becomes closer to 1 if either \(b\sigma_1\) increases or \(\sigma_2\) decreases. These conclusions are well understood because in both these cases the effect of the additive noise \(\sigma_2Z_t^Y\) on \(Y_t\) in (5.6.2) becomes smaller. Thus, the absolute coherency is indeed a notion that is similar to the correlation coefficient, and it shows how strong a linear relationship between \(\{X_t\}\) and \(\{Y_t\}\) is at frequency \(\lambda\).

These are the reasons why both the absolute coherency and the phase spectrum are the two primary characteristics used in the spectrum analysis of bivariate time series.

Now let us explain how to estimate these characteristics. Since the only new function here is the cross spectral density (5.6.6), we note that a partial sum for (5.6.6) should be written as

\[
f_{XY}(\lambda, J_1, J_2) := (2\pi)^{-1}\sum_{j=-J_1}^{J_2} \gamma_{XY}(j)e^{-ij\lambda}. \quad (5.6.13)
\]

In contrast to estimating a spectral density where \(J_1 = J_2\) (see (5.2.6)), here \(J_1\) and \(J_2\) may be different because the cross-covariance in general is not symmetric. Apart from this, the estimator of Section 5.2 may be used straightforwardly with the only modification that the estimated cutoffs are
defined by the formula

\[
(\hat{J}_1, \hat{J}_2) := \arg\min_{j_1, j_2} \left( \sum_{j=-j_1}^{j_2} \left( 2dn^{-1} - \hat{\gamma}_{XY}^2(j) \right) \right), \quad 0 \leq j_1, j_2 \leq J_n
\]

(5.6.14)

where

\[
\hat{\gamma}_{XY}(j) := \frac{n-j}{n} \sum_{l=1}^{n-j} X_{l+j} Y_l
\]

(5.6.15)

is the sample cross-covariance, which is used in place of the sample covariance, and

\[
\hat{d} := \sum_{j=-j_n}^{J_n} \hat{\gamma}_{XX}(j) \hat{\gamma}_{YY}(j)
\]

(5.6.16)

is the estimated coefficient of difficulty.

Estimates of the absolute coherency and the phase spectrum, calculated for the bivariate time series shown in Figure 5.10, are exhibited in the bottom diagram of that figure. These estimates are almost perfect. The slope of the estimated phase spectrum is approximately 3 at all frequencies. The estimated absolute coherency also correctly shows that the linear relationship between these two time series is practically the same at all frequencies. Also note that the coherency is far from 1, and this is absolutely right because the variance of the independent additive noise in \(Y_t\) is equal to the variance of \(X_t\). To get the absolute coherency close to 1, the coefficients of the model should be changed, as has been discussed above. Also, it is very useful to look at the estimates when the delay is negative. Thus, it is highly recommended to do Exercise 5.6.4.

Now let us apply our methodology and nonparametric universal estimators to an econometrics model that defines a bivariate time series with the first component \(\{P_t\}\) being the mean corrected price of a commodity and the second component \(\{S_t\}\) being the supply of this commodity at time \(t\). The model is defined as

\[
P_t = -b_P S_t + \sigma_P Z^P_t, \quad S_t = b_S P_{t-1} + \sigma_S Z^S_t, \quad t = 1, \ldots, n,
\]

(5.6.17)

where \(0 < b_P, b_S < 1\), time series \(\{Z^P_t\}\) and \(\{Z^S_t\}\) are independent standard Gaussian white noises, and the initial value of the price is \(P_0 = 1\).

Figure 5.11 shows a particular realization of this econometrics model; see the top two diagrams. Here \(b_P = 0.4, b_S = 0.8, \sigma_P = 1\), and \(\sigma_S = 0.5\).

It is not an easy task to analyze these realizations manually, so let us see what our nonparametric estimates, shown in the bottom two diagrams, tell us about this bivariate time series. The estimated absolute coherency reveals that a linear relationship between price and supply is strongest at high frequencies. Thus, our next step is to understand who leads whom at high frequencies. This we do with the help of the estimated phase spectrum,
which clearly indicates that price leads supply because the slope of the estimate is positive. Moreover, the slope at high frequencies is about 1, so we see that even the delay may be correctly defined via analyzing the estimated phase spectrum. These conclusions give us some insight into the relationship between price and supply based on this bivariate time series, and they do correspond to the underlying model.

5.7 Case Study: Dynamic Model and Forecasting

Consider the nonlinear dynamic model

\[ Y_t := f(Y_{t-1}) + s(Y_{t-1})\varepsilon_t, \quad Y_0 := \xi, \quad t = 1, 2, \ldots, n. \]  

(5.7.1)

Here \( Y_t \) is called a state of the model, \( f \) is called an iterative map, and \( s \) is called a scale map. The noise \( \varepsilon_t \) is a stationary time series, for instance an ARMA process, and \( \xi \) is an initial state of the model. Note that if \( s(y) = 0 \), then \( Y_t = f(Y_{t-1}) \), i.e., a current state of this dynamic model is defined solely by its previous state (the states are iterated). This explains the name of \( f \).
At first glance, the dynamic model (5.7.1) may resemble a classical time series model \( X_t := f(t) + \sigma(t)\varepsilon_t \), where \( f(t) \) is the deterministic part (trend/seasonal component) and \( \sigma(t) \) is a scale function. However, these models are absolutely different: The deterministic component in a classical time series decomposition is a function of time, while in a dynamic model the deterministic component is a function of the previous realization.

Dynamic systems naturally arise in applications where one believes that a current state is defined primarily by its previous state and a current “noise.” They are also used to approximate stochastic differential equations, for instance the equation for a continuous-in-time diffusion process \( y_t \),

\[
dy_t = \psi(y_t)dt + \sigma(y_t)dB(t), \quad t \geq 0, \quad y_0 = \xi.
\]

Here \( B(t) \) is a Brownian process (the definition will be given in Section 7.2), \( \psi \) is called a drift function, and \( \sigma \) is called a volatility function. A famous example is the Black–Scholes model for the stock price \( S_t \),

\[
dS_t = (\mu + \nu^2/2)S_t dt + \nu S_t dB(t), \quad t \geq 0, \quad S_0 = \xi.
\]

The parameters \( \mu \) and \( \nu \) are the so-called stock drift and volatility.

To explain a relationship between (5.7.1) and (5.7.2), consider equidistant observations of \( y_t \) with the sampling interval \( \delta := 1/n \). For large \( n \) these observations may be approximately written as the Euler scheme

\[
y_t = n^{-1}\psi(y_{t-\delta}) + y_{t-\delta} + n^{-1/2}\sigma(y_{t-\delta})Z_t, \quad t = \delta, 2\delta, \ldots, \quad y_0 = \xi, \quad (5.7.4)
\]

where \( Z_t \) are iid normal random variables. Thus, if we set \( f(y) = n^{-1}\psi(y) + y, \quad s(t) = n^{-1/2}\sigma(y) \), and consider standard normal \( \varepsilon_t \), then the relationship becomes transparent.

Let us explain how the universal estimate may be used for finding iterative and scale maps \( f \) and \( s \). Define \( X_t := Y_{t-1} \) and rewrite (5.7.1) as

\[
Y_t := f(X_t) + s(X_t)\varepsilon_t, \quad X_1 = \xi, \quad t = 1, 2, \ldots, n.
\]

This equation, at least formally, resembles the classical heteroscedastic regression problem discussed in Sections 4.2–4.3, where \( f \) was called the regression function and \( s \) the scale function. Thus, we may try to use the universal estimates of those sections for estimation of both \( f \) and \( s \).

Figure 5.12 illustrates how a dynamic model iterates and how the universal estimator performs. The top diagram shows a particular realization of states simulated by a dynamic model (5.7.1). Here the iterative map is \( f(y) = 2y/(1 + 2y^2) \), and the scale map is 2 times a standard normal density. The noise term is a Gaussian ARMA(1, 1) process \( \varepsilon_t + 0.3\varepsilon_{t-1} = Z_t + 0.4Z_{t-1} \). The initial state \( Y_0 \) is a realization of a uniform random variable \( \xi \) on \((0, 1)\).

The analysis of this particular realization, based on methods discussed in Section 5.1, shows that there is no visible trend, and a seasonal component