Chapter 1

Stationarity and Related Notions

The first chapter explains the basic notions and highlights the objectives of time series analysis. In Section 1.1 we give several important examples, discuss their characteristic features and deduce a general approach to the data analysis. In Section 1.2, stationary processes are identified as a reasonably broad class of random variables which are able to capture the main features extracted from the examples. Finally, we discuss how to treat deterministic trends and seasonal components in Sections 1.3 and 1.4, and assess the residuals in Section 1.5. Section 1.6 concludes.

1.1 Introduction and Examples

The first definition clarifies the notion *time series analysis*.

Definition 1.1.1 (Time Series) *Let $T \neq \emptyset$ be an index set. A family $(X_t\colon t \in T)$ of random variables on a probability space $(\Omega, \mathcal{A}, P)$ is called a stochastic process. If $\omega \in \Omega$ is fixed, then the mapping $t \mapsto X_t(\omega) =: x_t$ is referred to as time series (also sample path or realization) of the stochastic process $(X_t\colon t \in T)$. We will use the notation $(x_t\colon t \in T)$ in the discourse.*

The most common choices for the index set T include the integers $\mathbb{Z}$, the positive integers $\mathbb{N}$, the nonnegative integers $\mathbb{N}_0$, the real numbers $\mathbb{R}$ and the positive halfline $\mathbb{R}_+ = [0, \infty)$. In this class, we are mainly concerned with the first three cases which are subsumed under the notion *discrete time series analysis*.

Throughout, we assume without proof that the stochastic processes introduced do indeed exist. To actually verify this statement, certain consistency criteria have to be imposed on the family of finite-dimensional distributions of $(X_t\colon t \in T)$ which then, in turn, determine the distribution of the process. The result is known as Kolmogorov's theorem and is presented, for example, in Section 1.7 of Brockwell and Davis, 1991, and Section 36 of Billingsley, 1995.

Oftentimes the stochastic process $(X_t\colon t \in T)$ is itself referred to as a time series, in the sense that a realization is identified with the underlying probabilistic mechanism. The objective of time series analysis is to gain knowledge of the underlying random phenomenon through examining one (and typically only one) realization.

We start with a number of well known examples.

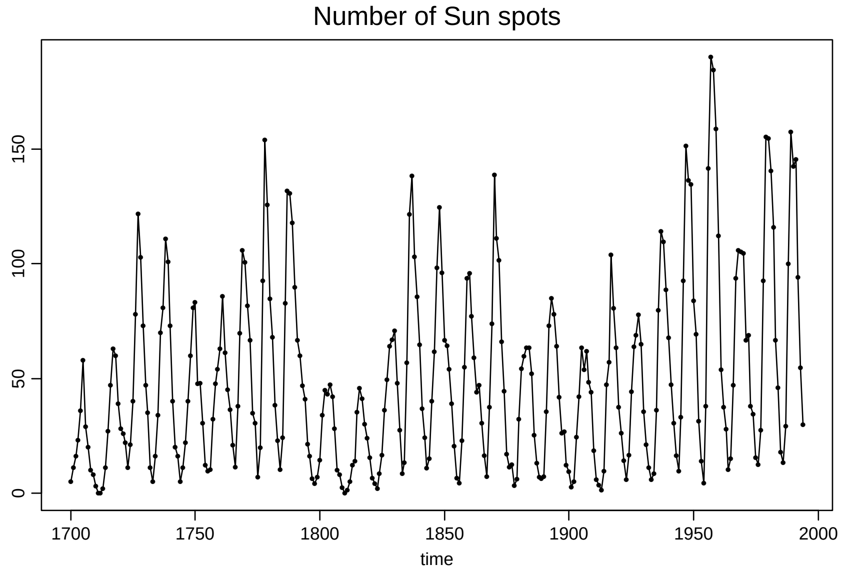

Figure 1.1: Wölfer's sunspot numbers from 1700 to 1994.

Example 1.1.1 (Wölfer's sunspot numbers) In Figure 1.1, the number of sunspots (that is, dark spots observed on the surface of the sun) observed annually are plotted against time. The horizontal axis labels time in years, while the vertical axis represents the observed values x_t of the random variable

$$X_t = \# \text{ of sunspots at time } t, \qquad t = 1700, \dots, 1994.$$

The figure is called a *time series plot*. It is a useful device for a preliminary analysis. Sunspot numbers are used to explain magnetic oscillations on the sun surface.

Example 1.1.2 (Canadian lynx data) The time series plot in Figure 1.2 comes from a biological data set. It contains the annual returns of lynx at auction in London by the Hudson Bay Company from 1821–1934 (on a $\log_{10}$ scale). Here, we have observations of the stochastic process

$$X_t = \log_{10}(\text{number of lynx trapped at time } 1820 + t), \qquad t = 1, \dots, 114.$$

The data is used as an estimate for the number of lynx trapped along the MacKenzie River in Canada. This estimate is often used as a proxy for the true population size of the lynx. A similar time series plot could be obtained for the snowshoe rabbit, the primary food source of the Canadian lynx, hinting at an intricate predator-prey relationship.

Example 1.1.3 (Treasury bills) Another important field of application for time series analysis lies in the area of finance. To hedge the risks of portfolios, investors commonly use short-term risk-free interest rates such as the yields of three-month, six-month, and twelve-month Treasury bills plotted in Figure 1.3. The (multivariate) data displayed consists of 2,386 weekly observations from July 17, 1959, to December 31, 1999. Here,

$$X_t = (X_{t,1}, X_{t,2}, X_{t,3}), \qquad t = 1, \dots, 2386,$$

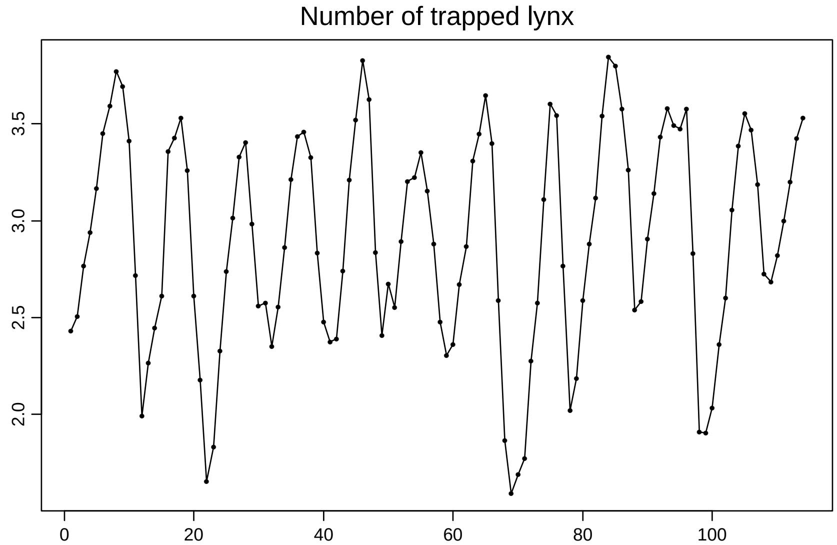

Figure 1.2: Number of lynx trapped in the MacKenzie River district between 1821 and 1934.

where $X_{t,1}$, $X_{t,2}$ and $X_{t,3}$ denote the three-month, six-month, and twelve-month yields at time t, respectively. It can be seen from the graph that all three Treasury bills are moving very similarly over time, implying a high correlation between the components of X_t.

Example 1.1.4 (S&P 500) The Standard and Poor's 500 index (S&P 500) is a value-weighted index based on the prices of 500 stocks that account for approximately 70% of the U.S. equity market capitalization. It is a leading economic indicator and is also used to hedge market portfolios. Figure 1.4 contains the 7,076 daily S&P 500 closing prices from January 3, 1972, to December 31, 1999, on a natural logarithm scale. We are consequently looking at the time series plot of the process

$$X_t = \ln(\text{closing price of S\&P 500 at time } t), \qquad t = 1, \ldots, 7076.$$

Note that the logarithm transform has been applied to make the returns directly comparable to the percentage of investment return.

There are countless other examples from all areas of science. To develop a theory capable of handling broad applications, the statistician needs to rely on a mathematical framework that can explain phenomena such as

- trends (apparent in Example 1.1.4);
- seasonal or cyclical effects (apparent in Examples 1.1.1 and 1.1.2);
- random fluctuations (all Examples);
- dependence (all Examples?).

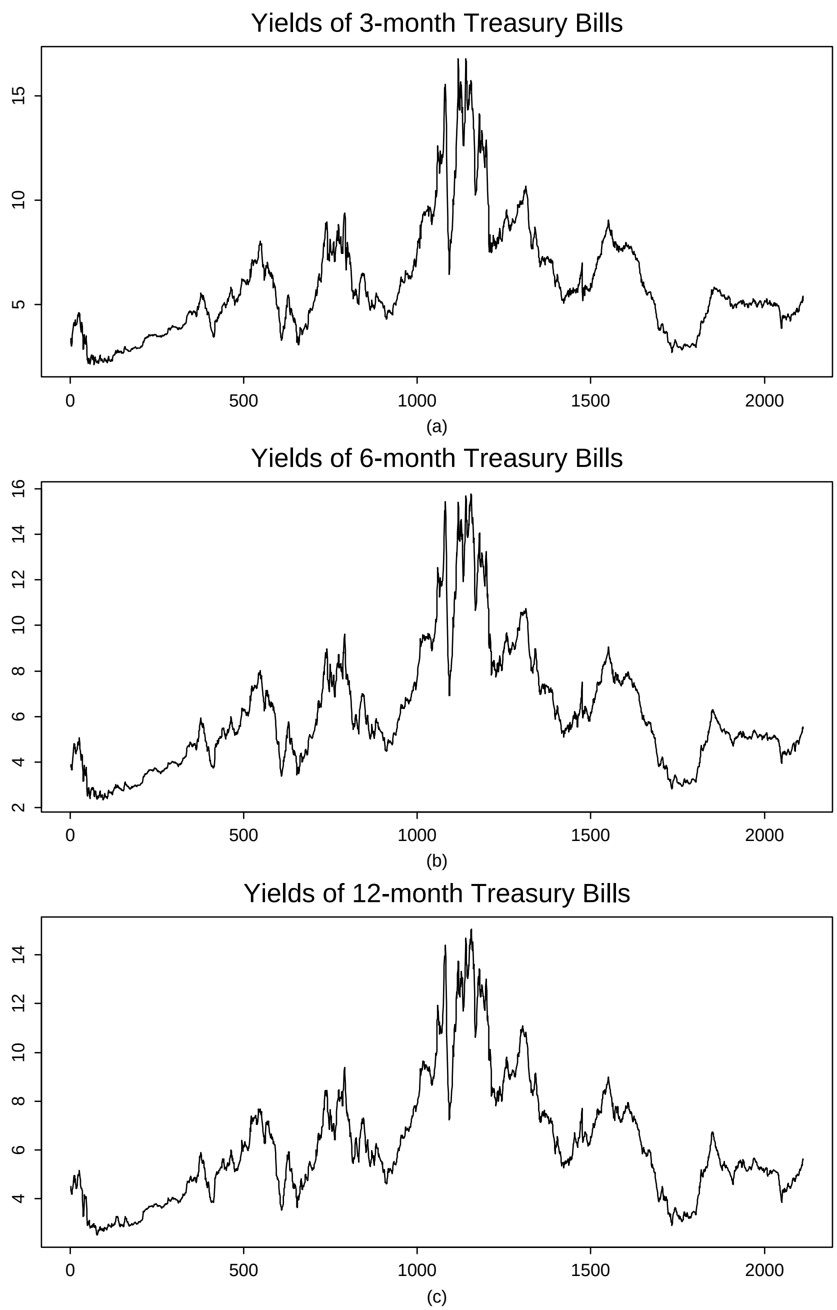

Figure 1.3: Yields of Treasury bills from July 17, 1959, to December 31, 1999.

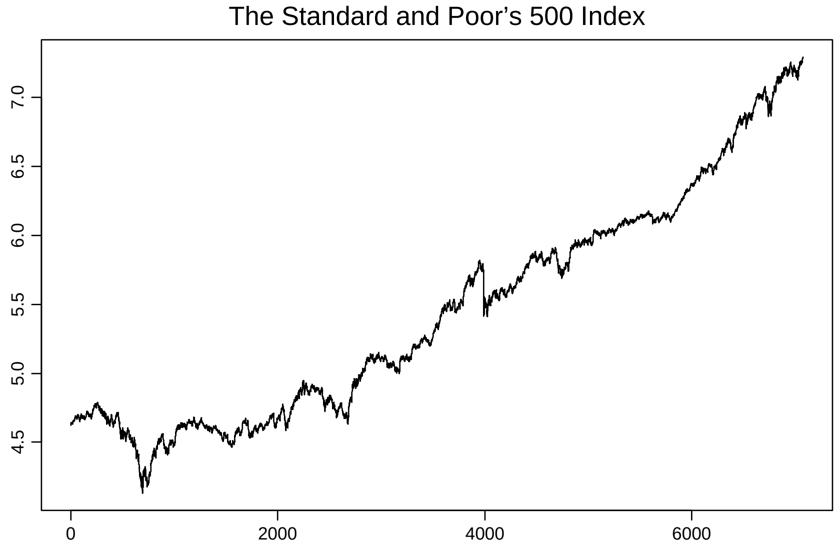

Figure 1.4: S&P 500 from January 3, 1972, to December 31, 1999.

The classical approach taken in time series analysis is to postulate that the stochastic process $(X_t\colon t \in T)$ under investigation can be divided into deterministic trend and seasonal components plus a centered random component, giving rise to the model

$$X_t = m_t + s_t + Y_t, \qquad t \in T, \tag{1.1.1}$$

where $(m_t\colon t \in T)$ denotes the trend function ("mean component"), $(s_t\colon t \in T)$ the seasonal effects and $(Y_t\colon t \in T)$ a (zero mean) stochastic process. The rest of this chapter will be devoted to introducing the classes of strictly and weakly stationary stochastic processes (in Section 1.2) and to providing tools to eliminate trends and seasonal components from a given time series (in Section 1.3).

1.2 Stationary Time Series

Fitting solely independent and identically distributed random variables to data is too narrow a concept. While on one hand, they allow for a somewhat nice and easy mathematical treatment, their use is, on the other hand, often hard to justify in applications. Our goal is therefore to introduce a concept that keeps some of the desirable properties of independent and identically distributed random variables ("regularity"), but that also considerably enlarges the class of stochastic processes to choose from by allowing dependence as well as varying distributions. Dependence between two random variables X and Y is usually measured in terms of the *covariance function*

$$\mathrm{Cov}(X, Y) = E\big[(X - E[X])(Y - E[Y])\big]$$

and the *correlation function*

$$\mathrm{Corr}(X, Y) = \frac{\mathrm{Cov}(X, Y)}{\sqrt{\mathrm{Var}(X)\mathrm{Var}(Y)}}.$$

With these notations, we can now introduce the classes of strictly and weakly dependent stochastic processes.

Definition 1.2.1 (Strict Stationarity) *A stochastic process* $(X_t\colon t \in T)$ *is called strictly stationary if, for all* $t_1, \ldots, t_n \in T$ *and* h *such that* $t_1 + h, \ldots, t_n + h \in T$*, it holds that*

$$(X_{t_1}, \ldots, X_{t_n}) \stackrel{\mathcal{D}}{=} (X_{t_1+h}, \ldots, X_{t_n+h}).$$

That is, the finite-dimensional distributions of the process are invariant under time shifts. Here $=^{\mathcal{D}}$ *indicates equality in distribution.*

Definition 1.2.2 (Weak Stationarity) *A stochastic process* $(X_t\colon t \in T)$ *is called weakly stationary if*

- *the second moments are finite:* $E[X_t^2] < \infty$ *for all* $t \in T$*;*
- *the means are constant:* $E[X_t] = m$ *for all* $t \in T$*;*
- *the covariance of* X_t *and* X_{t+h} *depends on* h *only:*

  $$\gamma(h) = \gamma_X(h) = \mathrm{Cov}(X_t, X_{t+h}), \qquad h \in T,$$

  is independent of $t \in T$ *and is called the autocovariance function (ACVF). Moreover,*

  $$\rho(h) = \rho_X(h) = \frac{\gamma(h)}{\gamma(0)}, \qquad h \in T,$$

  is called the autocorrelation function (ACF).

Remark 1.2.1 If $(X_t\colon t \in T)$ is a strictly stationary stochastic process with finite second moments, then it is also weakly stationary. The converse is not necessarily true. If $(X_t\colon t \in T)$, however, is weakly stationary and Gaussian, then it is also strictly stationary. Recall that a stochastic process is called Gaussian if, for any $t_1, \ldots, t_n \in T$, the random vector $(X_{t_1}, \ldots, X_{t_n})$ is multivariate normally distributed.

This section is concluded with examples of stationary and nonstationary stochastic processes.

Example 1.2.1 (White Noise) Let $(Z_t\colon t \in \mathbb{Z})$ be a sequence of real-valued, pairwise uncorrelated random variables with $E[Z_t] = 0$ and $0 < \mathrm{Var}(Z_t) = \sigma^2 < \infty$ for all $t \in \mathbb{Z}$. Then $(Z_t\colon t \in \mathbb{Z})$ is called *white noise*. It defines a centered, weakly stationary process with ACVF and ACF given by

$$\gamma(h) = \begin{cases} \sigma^2, & h = 0, \\ 0, & h \neq 0, \end{cases} \qquad \text{and} \qquad \rho(h) = \begin{cases} 1, & h = 0, \\ 0, & h \neq 0, \end{cases}$$

respectively.

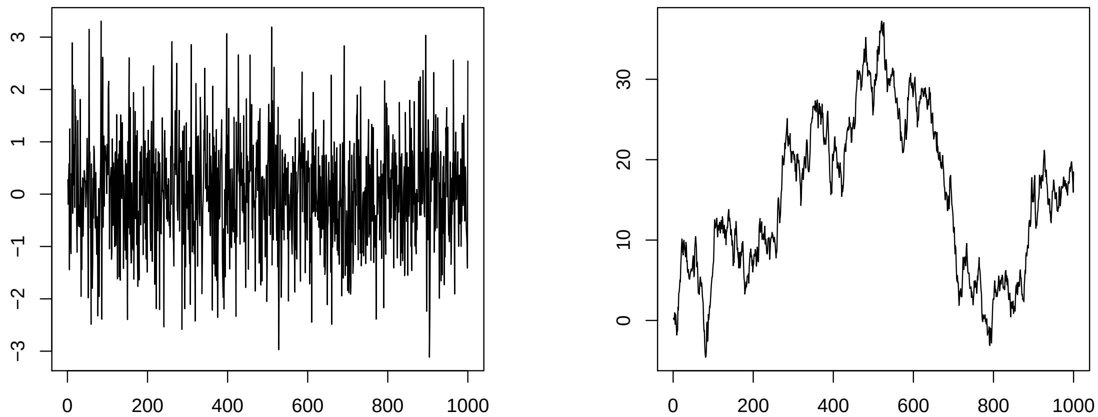

Figure 1.5: 1000 simulated values of iid $\mathcal{N}(0,1)$ noise (left panel) and a random walk with iid $\mathcal{N}(0,1)$ innovations (right panel).

Example 1.2.2 (Cyclical Time Series) Let A and B be uncorrelated random variables with zero mean and variances $\mathrm{Var}(A) = \mathrm{Var}(B) = \sigma^2$, and let $\lambda \in \mathbb{R}$ be a frequency parameter. Define

$$X_t = A\cos(\lambda t) + B\sin(\lambda t), \qquad t \in \mathbb{R}.$$

The resulting stochastic process $(X_t : t \in \mathbb{R})$ is then weakly stationary. Since $\sin(\lambda t+\varphi) = \sin(\varphi)\cos(\lambda t) + \cos(\varphi)\sin(\lambda t)$, the process can be represented as

$$X_t = R\sin(\lambda t + \varphi), \qquad t \in \mathbb{R},$$

so that R is the stochastic amplitude and $\varphi \in [-\pi, \pi]$ the stochastic phase of a *sinusoid*.

Example 1.2.3 (Random Walk) Let $(Z_t : t \in \mathbb{N}) \sim \mathrm{WN}(0, \sigma^2)$. Let $S_0 = 0$ and

$$S_t = Z_1 + \ldots + Z_t, \qquad t \in \mathbb{N}.$$

The resulting stochastic process $(S_t : t \in \mathbb{N}_0)$ is called a *random walk* and is the most important nonstationary time series. Indeed, it holds here that, for $h > 0$,

$$\mathrm{Cov}(S_t, S_{t+h}) = \mathrm{Cov}(S_t, S_t + R_{t,h}) = t\sigma^2,$$

where $R_{t,h} = Z_{t+1} + \ldots + Z_{t+h}$, and the ACVF obviously depends on t.

In the next chapter, we shall discuss in detail so-called autoregressive moving average processes which have become a centerpiece in *linear* time series analysis. Later on, we shall also deal with the highly popular ARCH and GARCH processes which are widely applied in the areas of finance and econometrics. They are the most prominent processes used in *nonlinear* time series analysis.

In general, however, the true parameters of a stationary stochastic process $(X_t\colon t \in T)$ are unknown to the statistician. Therefore, they have to be estimated from a realization $x_1, \ldots, x_n$. We shall mainly work with the following set of estimators. The *sample mean* of $x_1, \ldots, x_n$ is defined as

$$\bar{x} = \frac{1}{n}\sum_{t=1}^{n} x_t.$$

The *sample autocovariance function (sample ACVF)* is given by

$$\hat{\gamma}(h) = \frac{1}{n}\sum_{t=1}^{n-h}(x_{t+h} - \bar{x})(x_t - \bar{x}), \qquad h = 0, 1, \ldots, n-1. \tag{1.2.1}$$

Finally, the *sample autocorrelation function (sample ACF)* is

$$\hat{\rho}(h) = \frac{\hat{\gamma}(h)}{\hat{\gamma}(0)}, \qquad h = 0, 1, \ldots, n-1.$$

Example 1.2.4 Let $(Z_t\colon t \in \mathbb{Z})$ be a sequence of independent standard normally distributed random variables (see the left panel of Figure 1.5 for a typical realization of size $n = 1{,}000$). Then, clearly, $\gamma(0) = \rho(0) = 1$ and $\gamma(h) = \rho(h) = 0$ whenever $h \neq 0$. Table 1.1 gives the corresponding estimated values $\hat{\gamma}(h)$ and $\hat{\rho}(h)$ for $h = 0, 1, \ldots, 5$. The

h	0	1	2	3	4	5
$\hat{\gamma}(h)$	1.069632	0.072996	−0.000046	−0.000119	0.024282	0.0013409
$\hat{\rho}(h)$	1.000000	0.068244	−0.000043	−0.000111	0.022700	0.0012529

Table 1.1: Estimated ACVF and ACF for selected values of h.

estimated values are all very close to the true ones, indicating that the estimators work reasonably well for $n = 1{,}000$. Indeed it can be shown that they are asymptotically unbiased and consistent. Moreover, the sample autocorrelations $\hat{\rho}(h)$ are approximately normal with zero mean and variance $1/1000$.

1.3 Eliminating Trend Components

In this section we develop three different methods to estimate the trend of a time series model. We assume that it makes sense to postulate the model (1.1.1) with $s_t = 0$ for all $t \in T$, that is,

$$X_t = m_t + Y_t, \qquad t \in T, \tag{1.3.1}$$

where (without loss of generality) $E[Y_t] = 0$. In particular, we will discuss three different methods, (1) the least squares estimation of m_t, (2) smoothing by means of moving averages and (3) differencing.

Method 1 (Least squares estimation) It is often useful to assume that a trend component can be modeled appropriately by a polynomial,

$$m_t = b_0 + b_1 t + \ldots + b_p t^p, \qquad p \in \mathbb{N}_0.$$

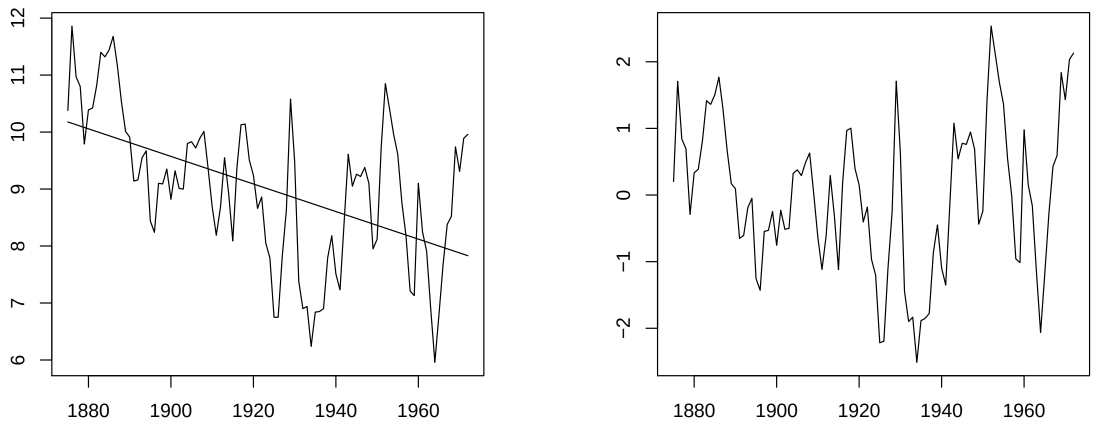

Figure 1.6: Annual water levels of Lake Huron (left panel) and the residual plot obtained from fitting a linear trend to the data (right panel).

In this case, the unknown parameters $b_0, \ldots, b_p$ can be estimated by the least squares method. Combined, they yield the estimated polynomial trend

$$\hat{m}_t = \hat{b}_0 + \hat{b}_1 t + \ldots + \hat{b}_p t^p, \qquad t \in T,$$

where $\hat{b}_0, \ldots, \hat{b}_p$ denote the corresponding least squares estimates. Note that we do not estimate the order p. It has to be selected by the statistician—for example, by inspecting the time series plot or, more generally, by model selection techniques. The residuals $\hat{Y}_t$ can be obtained as

$$\hat{Y}_t = X_t - \hat{m}_t = X_t - \hat{b}_0 - \hat{b}_1 t - \ldots - \hat{b}_p t^p, \qquad t \in T.$$

How to assess the goodness of fit of the fitted trend will be subject of Section 1.5 below.

Example 1.3.1 (Level of Lake Huron) The left panel of Figure 1.6 contains the time series of the annual average water levels in feet (reduced by 570) of Lake Huron from 1875 to 1972. We are dealing with a realization of the process

$$X_t = (\text{Average water level of Lake Huron in the year } 1874 + t) - 570, \qquad t = 1, \ldots, 98.$$

There seems to be a linear decline in the water level and it is therefore reasonable to fit a polynomial of order one to the data. Evaluating the least squares estimators provides us with the values

$$\hat{b}_0 = 10.202 \qquad \text{and} \qquad \hat{b}_1 = -0.0242$$

for the intercept and the slope, respectively. The resulting observed residuals $\hat{y}_t = \hat{Y}_t(\omega)$ are plotted against time in the right panel of Figure 1.6. There is no apparent trend left in the data. On the other hand, the plot does not strongly support the stationarity of the residuals. Additionally, there is evidence of dependence in the data.

Method 2 (Smoothing with Moving Averages) Let $(X_t \colon t \in \mathbb{Z})$ be a stochastic process following model (1.3.1). Choose $q \in \mathbb{N}_0$ and define the *two-sided moving average*

$$W_t = \frac{1}{2q+1} \sum_{j=-q}^{q} X_{t+j}, \qquad t \in \mathbb{Z}. \tag{1.3.2}$$

The random variables W_t can be utilized to estimate the trend component m_t in the following way. First note that

$$W_t = \frac{1}{2q+1} \sum_{j=-q}^{q} m_{t+j} + \frac{1}{2q+1} \sum_{j=-q}^{q} Y_{t+j} \approx m_t,$$

assuming that the trend is locally approximately linear and that the average of the Y_t over the interval $[t-q, t+q]$ is close to zero. Therefore, m_t can be estimated by

$$\hat{m}_t = W_t, \qquad t = q+1, \ldots, n-q.$$

Notice that there is no possibility of estimating the first q and last $n-q$ drift terms due to the two-sided nature of the moving averages. In contrast, one can also define *one-sided moving averages* by letting

$$\hat{m}_1 = X_1, \qquad \hat{m}_t = aX_t + (1-a)\hat{m}_{t-1}, \quad t = 2, \ldots, n.$$

Figure 1.7 contains estimators $\hat{m}_t$ based on the two-sided moving averages for the Lake Huron data of Example 1.3.1 for selected choices of q (upper panel) and the corresponding estimated residuals (lower panel).

More general versions of the moving average smoothers can be obtained in the following way. Observe that in the case of the two-sided version W_t each variable $X_{t-q}, \ldots, X_{t+q}$ obtains a "weight" $a_j = (2q+1)^{-1}$. The sum of all weights thus equals one. The same is true for the one-sided moving averages with weights a and $1-a$. Generally, one can hence define a smoother by letting

$$\hat{m}_t = \sum_{j=-q}^{q} a_j X_{t+j}, \qquad t = q+1, \ldots, n-q, \tag{1.3.3}$$

where $a_{-q} + \ldots + a_q = 1$. These general moving averages (two-sided and one-sided) are commonly referred to as *linear filters*. There are countless choices for the weights. The one here, $a_j = (2q+1)^{-1}$, has the advantage that linear trends pass undistorted. In the next example, we introduce a filter which passes cubic trends without distortion.

Example 1.3.2 (Spencer's 15-point moving average) Suppose that the filter in display (1.3.3) is defined by weights satisfying $a_j = 0$ if $|j| > 7$, $a_j = a_{-j}$ and

$$(a_0, a_1, \ldots, a_7) = \frac{1}{320}(74, 67, 46, 21, 3, -5, -6, -3).$$

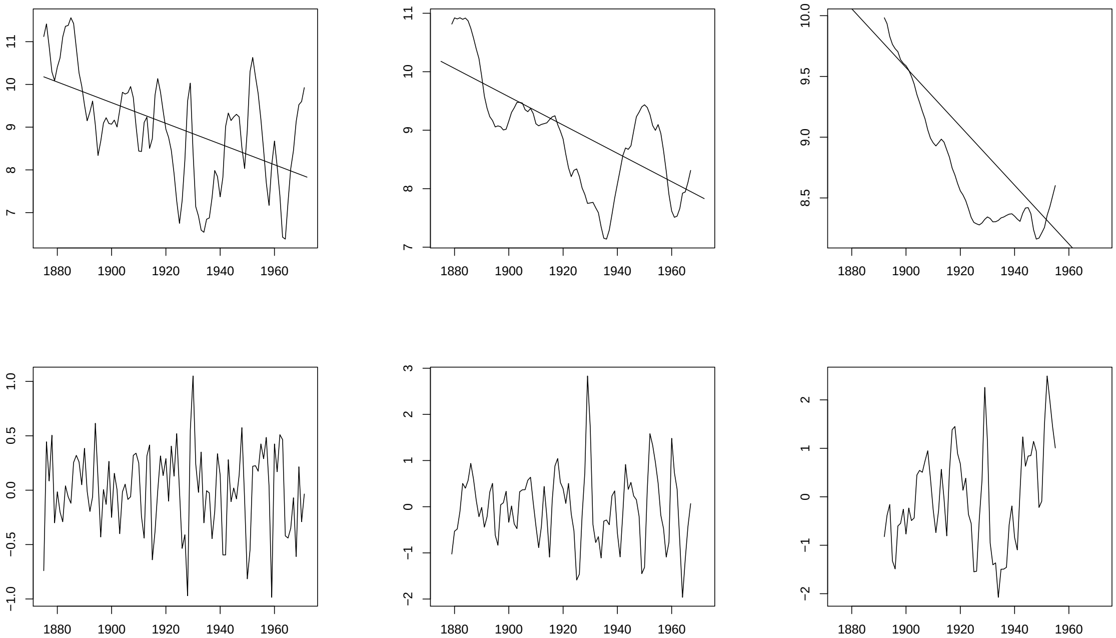

Figure 1.7: The two-sided moving average filters W_t for the Lake Huron data (upper panel) and their residuals (lower panel) with bandwith $q = 2$ (left), $q = 10$ (middle) and $q = 35$ (right).

Then, the corresponding filters passes cubic trends $m_t = b_0 + b_1 t + b_2 t^2 + b_3 t^3$ undistorted. To see this, observe that

$$\sum_{j=-7}^{7} a_j = 1 \qquad \text{and} \qquad \sum_{j=-7}^{7} j^r a_j = 0, \qquad r = 1, 2, 3.$$

Now apply Proposition 1.3.1 below to arrive at the conclusion.

Proposition 1.3.1 *A linear filter (1.3.3) passes a polynomial of degree p if and only if*

$$\sum_j a_j = 1 \qquad \textit{and} \qquad \sum_j j^r a_j = 0, \qquad r = 1, \ldots, p.$$

Proof. It suffices to show that $\sum_j a_j (t+j)^r = t^r$ for $r = 0, \ldots, p$. Using the binomial theorem, we can write

$$\begin{aligned} \sum_j a_j (t+j)^r &= \sum_j a_j \sum_{k=0}^{r} \binom{r}{k} t^k j^{r-k} \\ &= \sum_{k=0}^{r} \binom{r}{k} t^k \left(\sum_j a_j j^{r-k} \right) \\ &= t^r \end{aligned}$$

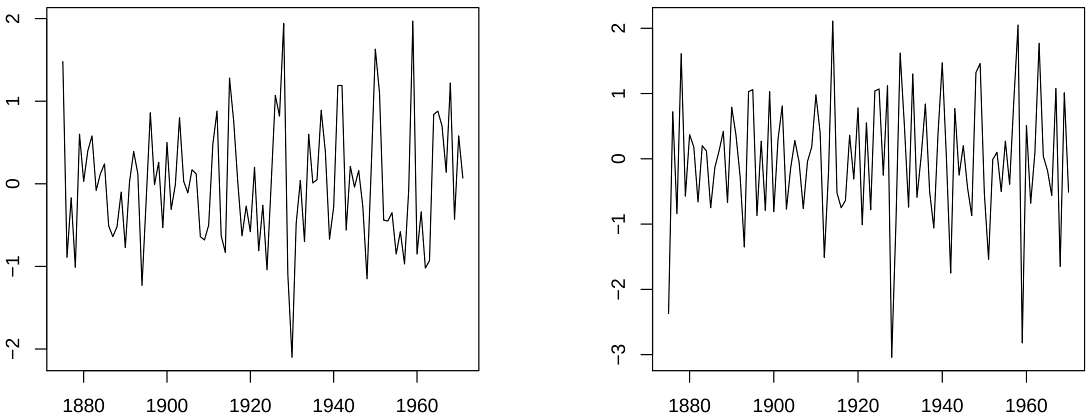

Figure 1.8: Time series plots of the observed sequences (∇x_t) in the left panel and $(\nabla^2 x_t)$ in the right panel of the differenced Lake Huron data described in Example 1.3.1.

for any $r = 0, \ldots, p$ if and only if the above conditions hold. This completes the proof. $\Box$

Method 3 (Differencing) A third possibility to remove drift terms from a given time series is differencing. To this end, we introduce the *difference operator* ∇ as

$$\nabla X_t = X_t - X_{t-1} = (1 - B)X_t, \qquad t \in T,$$

where B denotes the backshift operator $BX_t = X_{t-1}$. Repeated application of ∇ is defined in the intuitive way:

$$\nabla^2 X_t = \nabla(\nabla X_t) = \nabla(X_t - X_{t-1}) = X_t - 2X_{t-1} + X_{t-2}$$

and, recursively, the representations follow also for higher powers of ∇. Suppose that you are applying the difference operator to a linear trend $m_t = b_0 + b_1 t$, then you obtain

$$\nabla m_t = m_t - m_{t-1} = b_0 + b_1 t - b_0 - b_1(t - 1) = b_1$$

which is a constant. Inductively, this leads to the conclusion that for a polynomial drift of degree p, namely $m_t = \sum_{j=0}^{p} b_j t^j$, we have that $\nabla^p m_t = p!b_p$ and thus constant. Applying this technique to a stochastic process of the form (1.3.1) with a polynomial drift m_t, yields then

$$\nabla^p X_t = p!b_p + \nabla^p Y_t, \qquad t \in T.$$

This is a stationary process with mean $p!b_p$. The plots in Figure 1.8 contain the first and second differences for the Lake Huron data.

The next example shows that the difference operator can also be applied to a random walk to create stationary data.

Example 1.3.3 Let $(S_t \colon t \in \mathbb{N}_0)$ be the random walk of Example 1.2.3. If we apply the difference operator ∇ to this stochastic process, we obtain

$$\nabla S_t = S_t - S_{t-1} = Z_t, \qquad t \in \mathbb{N}.$$

In other words, ∇ does nothing else but recover the original white noise sequence that was used to build the random walk.

1.4 Eliminating Trend and Seasonal Components

Let us go back to the classical decomposition (1.1.1),

$$X_t = m_t + s_t + Y_t, \qquad t \in T,$$

with $E[Y_t] = 0$. In this section, we shall discuss three methods that aim at estimating both the trend and seasonal components in the data. As additional requirement on $(s_t\colon t \in T)$, we assume that

$$s_{t+d} = s_t, \qquad \sum_{j=1}^{d} s_j = 0,$$

where d denotes the period of the seasonal component. (If we are dealing with yearly data sampled monthly, then obviously $d = 12$.) It is convenient to relabel the observations $x_1, \ldots, x_n$ in terms of the seasonal period d as

$$x_{j,k} = x_{k+d(j-1)}.$$

In the case of yearly data, observation $x_{j,k}$ thus represents the data point observed for the kth month of the jth year. For convenience we shall always refer to the data in this fashion even if the acutal period is something other than 12.

Method 1 (Small trend method) If the changes in the drift term appear to be small, then it is reasonable to assume that the drift in year j, say, m_j is constant. As a natural estimator we can therefore apply

$$\hat{m}_j = \frac{1}{d}\sum_{k=1}^{d} x_{j,k}.$$

To estimate the seasonality in the data, one can in a second step utilize the quantities

$$\hat{s}_k = \frac{1}{N}\sum_{j=1}^{N}(x_{j,k} - \hat{m}_j),$$

where N is determined by the equation $n = Nd$, provided that data has been collected over N *full* cycles. Direct calculations show that these estimators possess the property $\hat{s}_1 + \ldots + \hat{s}_d = 0$ (as in the case of the true seasonal components s_t). To further assess the quality of the fit, one needs to analyze the observed residuals

$$\hat{y}_{j,k} = x_{j,k} - \hat{m}_j - \hat{s}_k.$$

Note that due to the relabeling of the observations and the assumption of a slowly changing trend, the drift component is solely described by the "annual" subscript j, while the seasonal component only contains the "monthly" subscript k.

Example 1.4.1 (Australian wine sales) The left panel of Figure 1.9 shows the monthly sales of red wine (in kiloliters) in Australia from January 1980 to October 1991. Since there is an apparent increase in the fluctuations over time, the right panel of the same

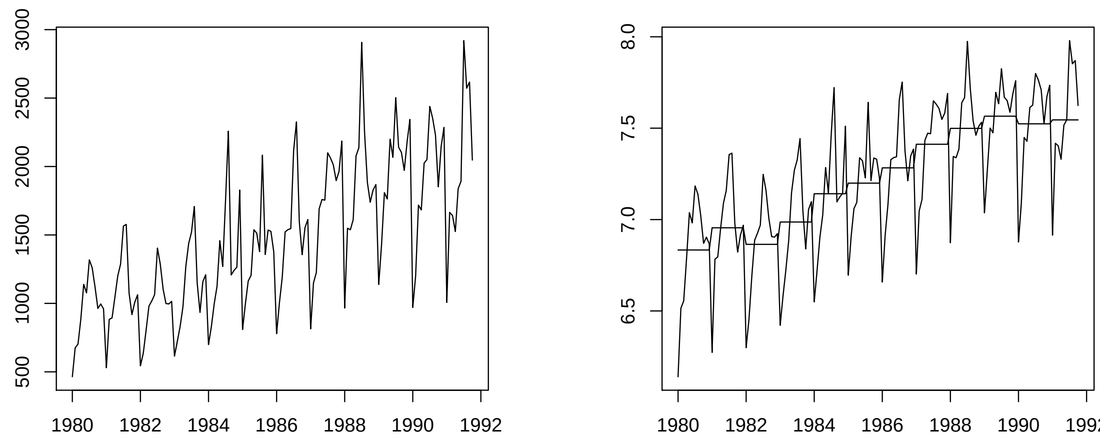

Figure 1.9: Time series plots of the red wine sales in Australia from January 1980 to October 1991 (left) and its log transformation with yearly mean estimates (right).

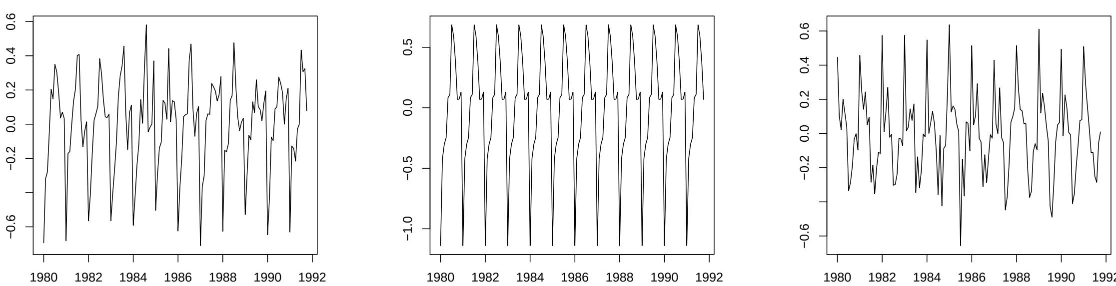

Figure 1.10: The detrended log series (left), the estimated seasonal component (center) and the corresponding residuals series (right) of the Australian red wine sales data.

figure shows the natural logarithm transform of the data. There is clear evidence of both trend and seasonality. In the following, we will continue to work with the log transformed data. Using the small trend method as described above, we first estimate the annual means, which are already incorporated in the right time series plot of Figure 1.9. Note that there are only ten months of data available for the year 1991, so that the estimation has to be adjusted accordingly. The detrended data is shown in the left panel of Figure 1.10. The middle plot in the same figure shows the estimated seasonal component, while the right panel displays the residuals. Even though the assumption of small changes in the drift is somewhat questionable, the residuals appear to look quite nice. They indicate that there is dependence in the data (see Section 1.5 below for more on this subject).

Method 2 (Moving average estimation) This method is to be preferred over the first one whenever the underlying trend component is not constant. Three steps are to be applied to the data.

1st Step: Trend estimation. At first, we focus on the removal of the trend component with the linear filters discussed in the previous section. If the period d is odd, then we

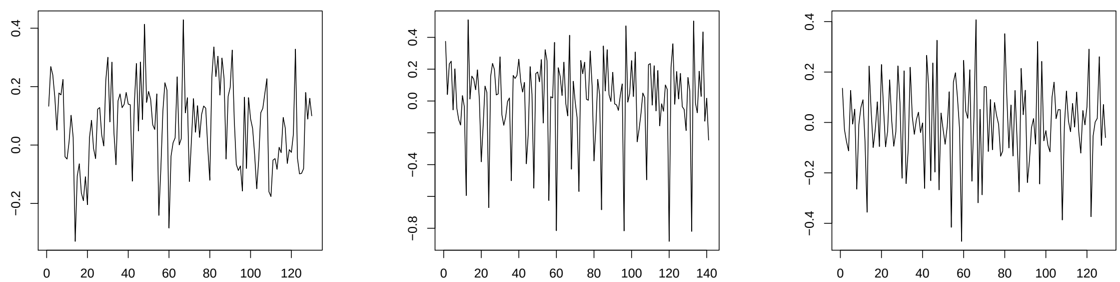

Figure 1.11: The differenced observed series $\nabla_{12}x_t$ (left), ∇x_t (middle) and $\nabla\nabla_{12}x_t = \nabla_{12}\nabla x_t$ (right) for the Australian red wine sales data.

can directly use $\hat{m}_t = W_t$ as in (1.3.2) with q specified by the equation $d = 2q + 1$. If the period $d = 2q$ is even, then we slightly modify W_t and use

$$\hat{m}_t = \frac{1}{d}(.5x_{t-q} + x_{t-q+1} + \ldots + x_{t+q-1} + .5x_{t+q}), \qquad t = q+1, \ldots, n-q.$$

2nd Step: Seasonality estimation. To estimate the seasonal component, let

$$\mu_k = \frac{1}{N-1}\sum_{j=2}^{N}(x_{k+d(j-1)} - \hat{m}_{k+d(j-1)}), \qquad k = 1, \ldots, q,$$

$$\mu_k = \frac{1}{N-1}\sum_{j=1}^{N-1}(x_{k+d(j-1)} - \hat{m}_{k+d(j-1)}), \qquad k = q+1, \ldots, d.$$

Define now

$$\hat{s}_k = \mu_k - \frac{1}{d}\sum_{\ell=1}^{d}\mu_\ell, \qquad k = 1, \ldots, d,$$

and set $\hat{s}_k = \hat{s}_{k-d}$ whenever $k > d$. This will provide us with *deseasonalized data* which can be examined further. In the final step, any remaining trend can be removed from the data.

3rd Step: Trend Reestimation. Apply any of the methods from Section 1.3.

Method 3 (Differencing at lag d) Introducing the *lag-d difference operator* ∇_d, defined by letting

$$\nabla_d X_t = X_t - X_{t-d} = (1 - B^d)X_t, \qquad t = d+1, \ldots, n,$$

and assuming model (1.1.1), we arrive at the transformed random variables

$$\nabla_d X_t = m_t - m_{t-d} + Y_t - Y_{t-d}, \qquad t = d+1, \ldots, n.$$

Note that the seasonality is removed, since $s_t = s_{t-d}$. The remaining noise variables $Y_t - Y_{t-d}$ are stationary and have zero mean. The new trend component $m_t - m_{t-d}$ can be eliminated using any of the methods developed in Section 1.3.

Example 1.4.2 (Australian wine sales) We revisit the Australian red wine sales data of Example 1.4.1 and apply the differencing techniques just established. The left plot of Figure 1.11 shows the the data after an application of the operator ∇_{12}. If we decide to estimate the remaining trend in the data with the differencing method from Section 1.3, we arrive at the residual plot given in the right panel of Figure 1.11. Note that the order of application does not change the residuals, that is, $\nabla\nabla_{12}x_t = \nabla_{12}\nabla x_t$. The middle panel of Figure 1.11 displays the differenced data which still contains the seasonal component.

1.5 Assessing the Residuals

In this subsection, we introduce several goodness-of-fit tests to further analyze the residuals obtained after the elimination of trend and seasonal components. The main objective is to determine whether or not these residuals can be regarded as obtained from a sequence of independent, identically distributed random variables or if there is dependence in the data. Throughout we denote by $Y_1, \ldots, Y_n$ the residuals and by $y_1, \ldots, y_n$ a typical realization.

Method 1 (The sample ACF) We have seen in Example 1.2.4 that the estimators $\hat{\rho}(j)$ of the ACF $\rho(j)$ are asymptotically independent and normally distributed with mean zero and variance n^{-1}, provided the underlying residuals are independent and identically distributed with a finite variance. Therefore, plotting the sample ACF for a certain number of lags, say h, we expect that approximately 95% of these values are within the bounds $\pm 1.96/\sqrt{n}$.

Method 2 (The Portmanteau test) The Portmanteau test is based on the test statistic

$$Q = n \sum_{j=1}^{h} \hat{\rho}^2(j).$$

Using the fact that the variables $\sqrt{n}\hat{\rho}(j)$ are asymptotically standard normal, it becomes apparent that Q itself can be approximated with a chi-squared distribution possessing h degrees of freedom. We now reject the hypothesis of independent and identically distributed residuals at the level α if $Q > \chi^2_{1-\alpha}(h)$, where $\chi^2_{1-\alpha}(h)$ is the $1-\alpha$ quantile of the chi-squared distribution with h degrees of freedom. Several refinements of the original Portmanteau test have been established in the literature. We refer here only to the papers Ljung and Box (1978), and McLeod and Li (1983) for further information on this topic.

Method 3 (The rank test) This test is very useful for finding linear trends. Denote by

$$\Pi = \#\{(i,j) : Y_i > Y_j,\ i > j,\ i = 2, \ldots, n\}$$

the random number of pairs (i,j) satisfying the conditions $Y_i > Y_j$ and $i > j$. Clearly, there are $\binom{n}{2} = \frac{1}{2}n(n-1)$ pairs (i,j) such that $i > j$. If $Y_1, \ldots, Y_n$ are independent and identically distributed, then $P(Y_i > Y_j) = 1/2$ (assuming a continuous distribution). Now it follows that $\mu_\Pi = E[\Pi] = \frac{1}{4}n(n-1)$ and, similarly, $\sigma^2_\Pi = \mathrm{Var}(\Pi) = \frac{1}{72}n(n-1)(2n+5)$. Moreover, for large enough sample sizes n, Π has an approximate normal distribution with mean μ_Π and variance σ_Π^2. Consequently, one would reject the hypothesis of independent, identically distributed data at the level α if

$$P = \frac{|\Pi - \mu_\Pi|}{\sigma_\Pi} > z_{1-\alpha/2},$$

where $z_{1-\alpha/2}$ denotes the $1-\alpha/2$ quantile of the standard normal distribution.

Method 4 (Tests for normality) If there is evidence that the data are generated by Gaussian random variables, one can create the *qq plot* to check for normality. It is based on a visual inspection of the data. To this end, denote by $Y_{(1)} < \ldots < Y_{(n)}$ the order statistics of the residuals $Y_1, \ldots, Y_n$ which are normally distributed with expected value μ and variance σ^2. It holds that

$$E[Y_{(j)}] = \mu + \sigma E[X_{(j)}], \tag{1.5.1}$$

where $X_{(1)} < \ldots < X_{(n)}$ are the order statistics of a standard normal distribution. The qq plot is defined as the graph of the pairs $(E[X_{(1)}], Y_{(1)}), \ldots, (E[X_{(n)}], Y_{(n)})$. According to display (1.5.1), the resulting graph will be approximately linear with the squared correlation R^2 of the points being close to 1. The assumption of normality will thus be rejected if R^2 is "too" small. It is common to approximate $E[X_{(j)}] \approx \Phi_j = \Phi^{-1}((j-.5)/n)$ (Φ being the distribution function of the standard normal distribution) and the previous statement is made precise by letting

$$R^2 = \frac{\left[\sum_{j=1}^{n}(Y_{(j)} - \bar{Y})\Phi_j\right]^2}{\sum_{j=1}^{n}(Y_{(j)} - \bar{Y})^2 \sum_{j=1}^{n} \Phi_j^2},$$

where $\bar{Y} = \frac{1}{n}(Y_1 + \ldots + Y_n)$. The critical values for R^2 are tabulated and can be found, for example in Shapiro and Francia (1972).

1.6 Summary

In this chapter, we have introduced the classical decomposition (1.1.1) of a time series into a drift component, a seasonal component and a sequence of residuals. We have provided methods to estimate the drift and the seasonality. Moreover, we have identified the class of stationary processes as a reasonably broad class of random variables. We have introduced several ways to check whether or not the resulting residuals can be considered to be independent, identically distributed. In the Chapter 3, we will discuss in depth the class of autoregressive moving average (ARMA) processes, a parametric class of random variables that are at the center of linear time series analysis because they are able to capture a wide range of dependence structures and allow for a thorough mathematical treatment.