Part II

Time Series

12 Introduction

This Part is mainly a summary of the book of Brockwell and Davis (2002). Additionally the textbook Shumway and Stoffer (2010) can be recommended.[1]

Our purpose is to study techniques drawing inferences from time series. Before we can do this, it is necessary to set up a hypothetical probability model to represent the data. After an appropriate family of models has been chosen, it is then possible to estimate parameters, check for goodness of fit to the data, and possibly to use the fitted model to enhance our understanding of the mechanism generating the series. Once a satisfactory model has been developed, it may be used in a variety of ways depending on the particular field of application.

12.1 Definitions and Examples

Definition 12.1.1. A time series is a set of observations x_t each one being recorded at a specific time t. A discrete time series is one in which the set T_0 of times at which observations are made is a discrete set. Continuous time series are obtained when observations are recorded continuously over some time interval.

Example. Some examples of discrete univariate time series from climate sciences and energy markets are shown on pages 12-2 to 12-4.

Example (Basel, p. 1-6)**.** We come back to the Basel temperature time series which starts in 1755 and is the longest temperature time series in Switzerland. We will analyze this data set and try to find an adequate time series model. To get an overview we start with Figure 12.4 showing the annual mean temperatures from 1755 to 2010. With a boxplot (see Figure 12.5, p. 12-6) the monthly variability, called seasonality, but also the intermonthly variability can be shown. Finally Figure 12.6, p. 12-7, shows the monthly temperature time series from 1991 to 2010.

An important part of the analysis of a time series is the selection of a suitable probability model for the data. To allow for the possibly unpredictable nature of future observations it is natural to suppose that each observation x_t is a realized value of a certain random variable X_t.

Definition 12.1.2. A time series model for the observed data $\{x_t\}$ is a specification of the joint distribution (or possibly only the means and covariances) of a sequence of random variables $\{X_t\}$ of which $\{x_t\}$ is postulated to be a realization.

Remark. We shall frequently use the term time series to mean both the data and the process of which it is a realization. A complete probabilistic time series model for the

[1] http://www.stat.pitt.edu/stoffer/tsa3/ (12.04.2015).

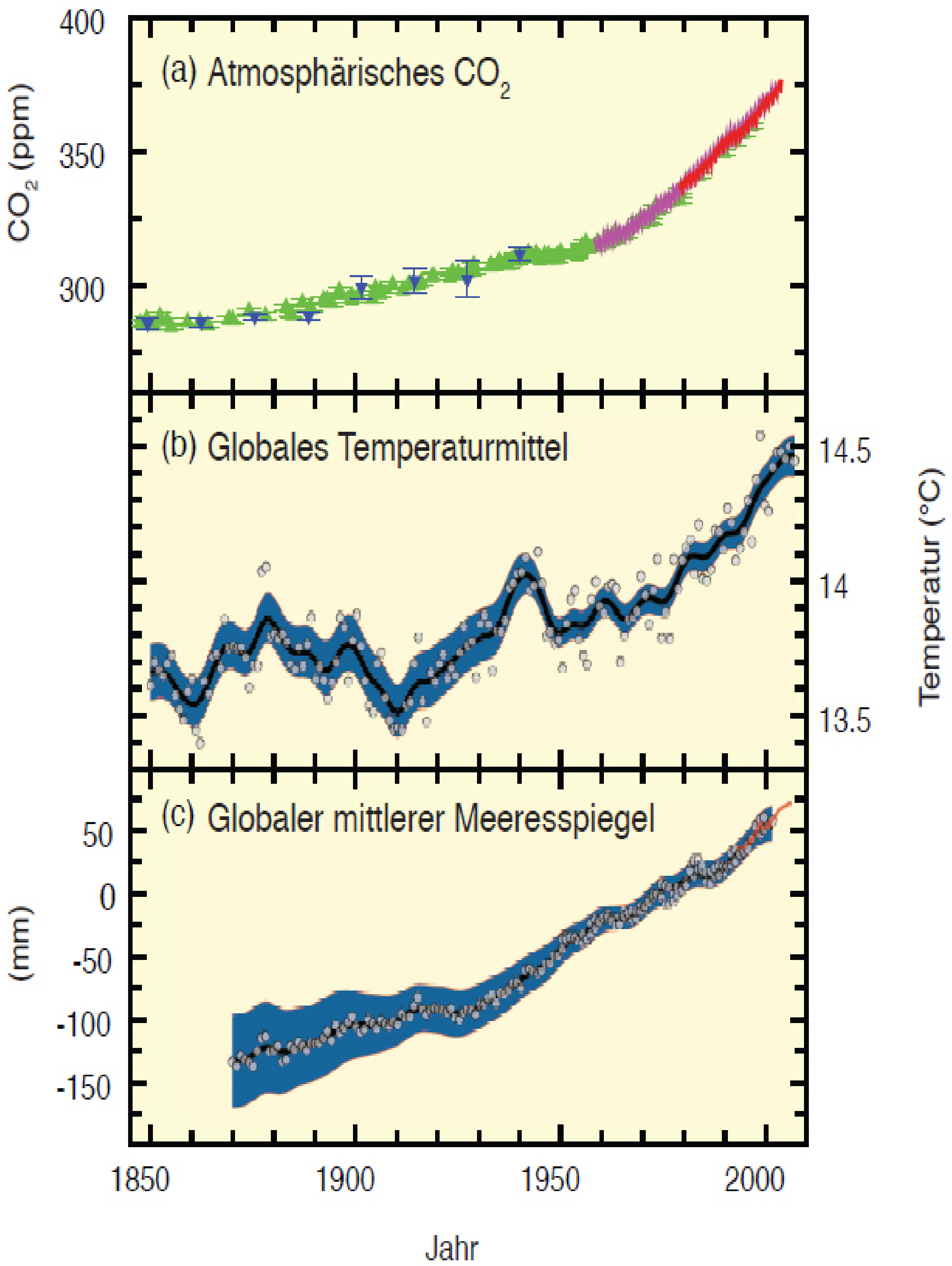
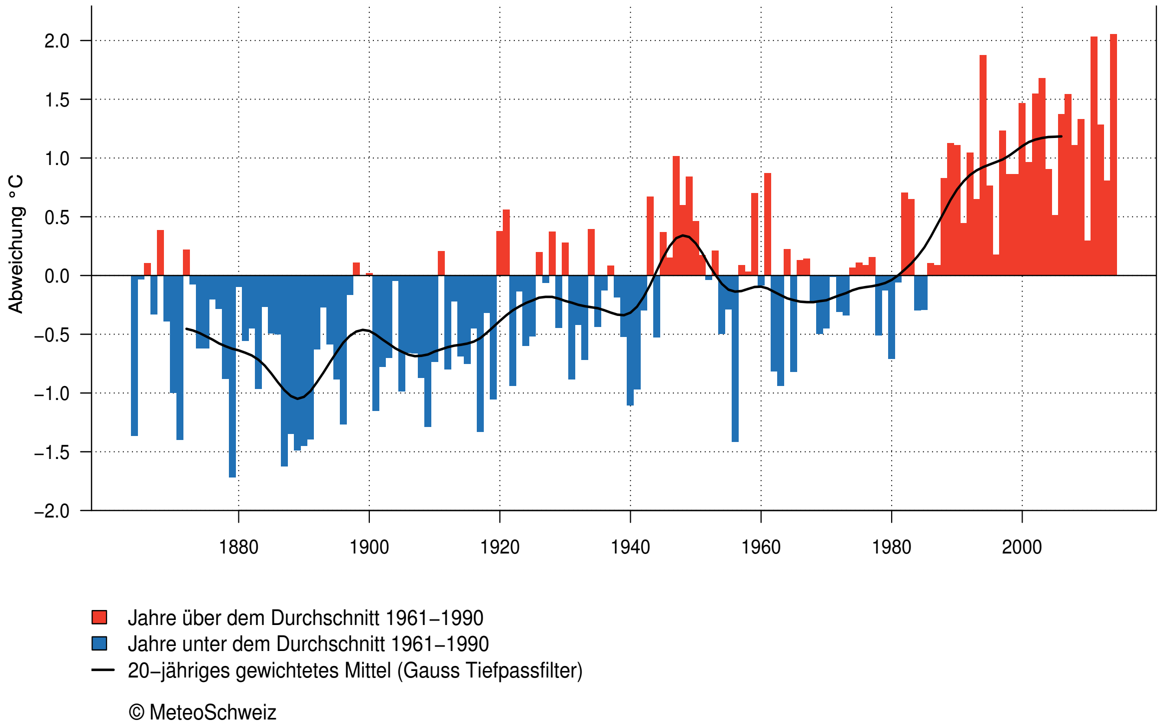


Figure 12.1: Global development of the atmospheric CO_2 concentration, annual mean temperature and sea level from 1850 to 2010. Source: OcCC (2008).



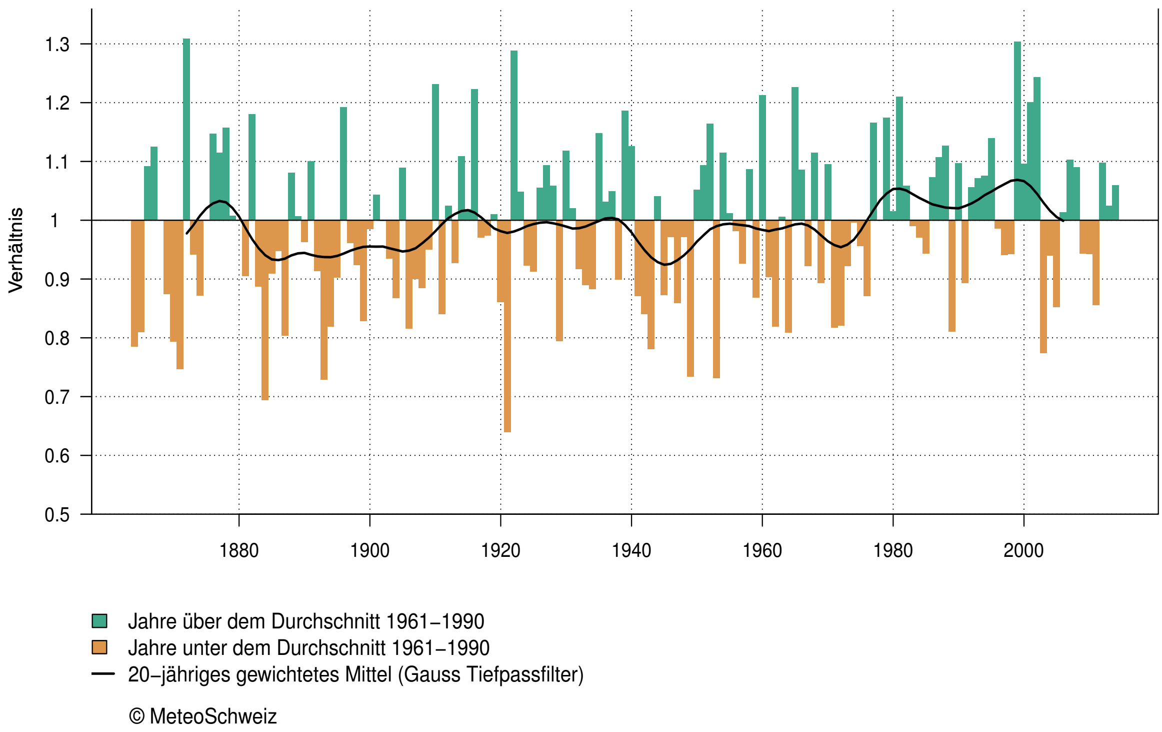


Figure 12.2: Time series from 1864 to 2014 of the anomalies from the reference period 1961-1990 of the annual mean temperature in Switzerland (top) and the precipitation in Southern Switzerland (bottom). Source: MeteoSchweiz.

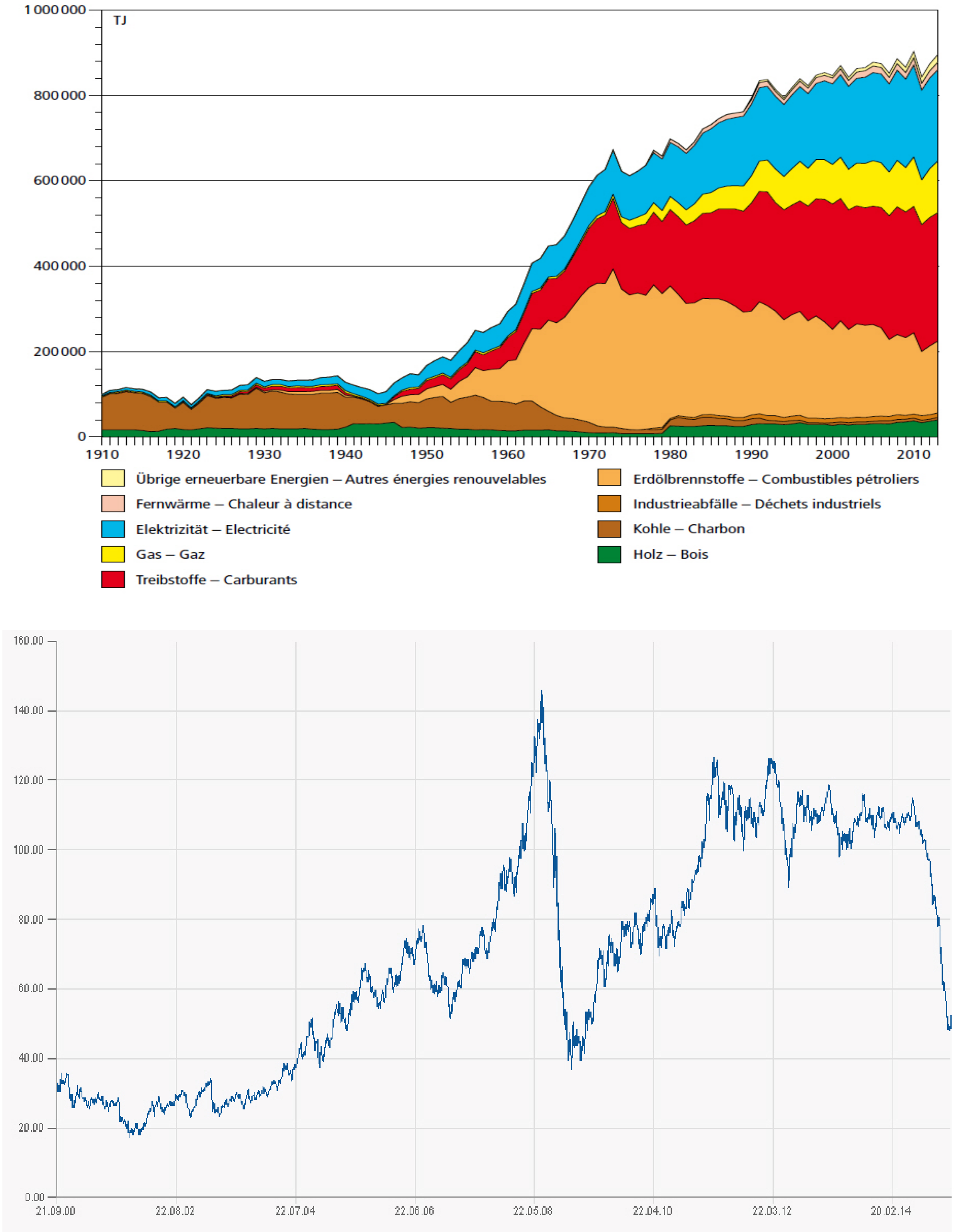


Figure 12.3: Top: Energy demand in Switzerland from 1910 to 2013. Source: Swiss Federal Office of Energy (2013). Bottom: Brent Crude oil prices from 2000 to 2014 in USD per barrel. Source: www.finanzen.ch.

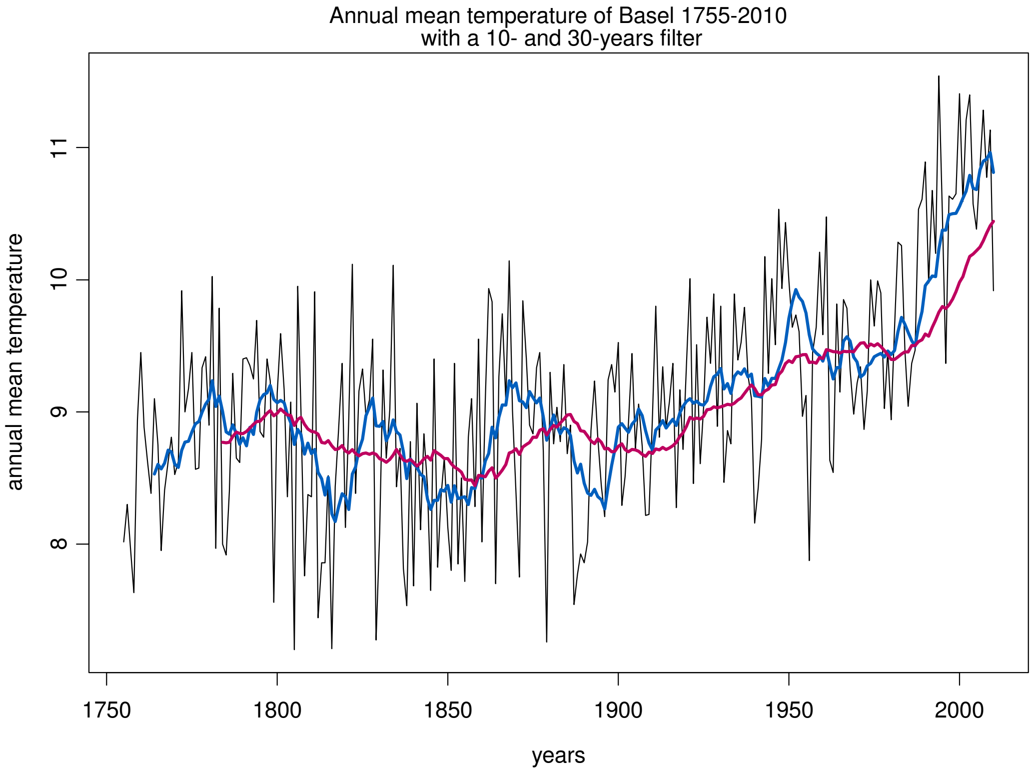


Figure 12.4: Basel annual mean temperature time series from 1755 to 2010. Data set: Basel, p. 1-6.

sequence or random variables $\{X_1, X_2, \ldots\}$ would specify all of the joint distributions of the random vectors $(X_1, \ldots, X_n)'$, $n = 1, 2, \ldots$, or equivalently all of the probabilities

$$P(X_1 \leq x_1, \ldots, X_n \leq x_n), \quad -\infty < x_1, \ldots, x_n < \infty, \quad n = 1, 2, \ldots.$$

Such a specification is rarely used in time series analysis, since in general it will contain far too many parameters to be estimated from the available data. Instead we specify only the first- and second order moments of the joint distributions, i.e., the expected values $\mathrm{E}X_t$ and the expected products $\mathrm{E}(X_{t+h}X_t)$, $t = 1, 2, \ldots$, $h = 0, 1, \ldots$, focusing on the properties of the sequence $\{X_t\}$ that depend only on these. Such properties are referred as second-order properties. In the particular case where all the joint distributions are multivariate normal, the second-order properties of $\{X_t\}$ completely determine the joint distributions and hence give a complete probabilistic characterization of the sequence.

Definition 12.1.3. $\{X_t\}$ is a Gaussian time series if all of its joint distributions are multivariate normal, i.e., if for any collection of integers $i_1, \ldots, i_n$, the random vector $(X_{i_1}, \ldots, X_{i_n})'$ has a multivariate normal distribution.

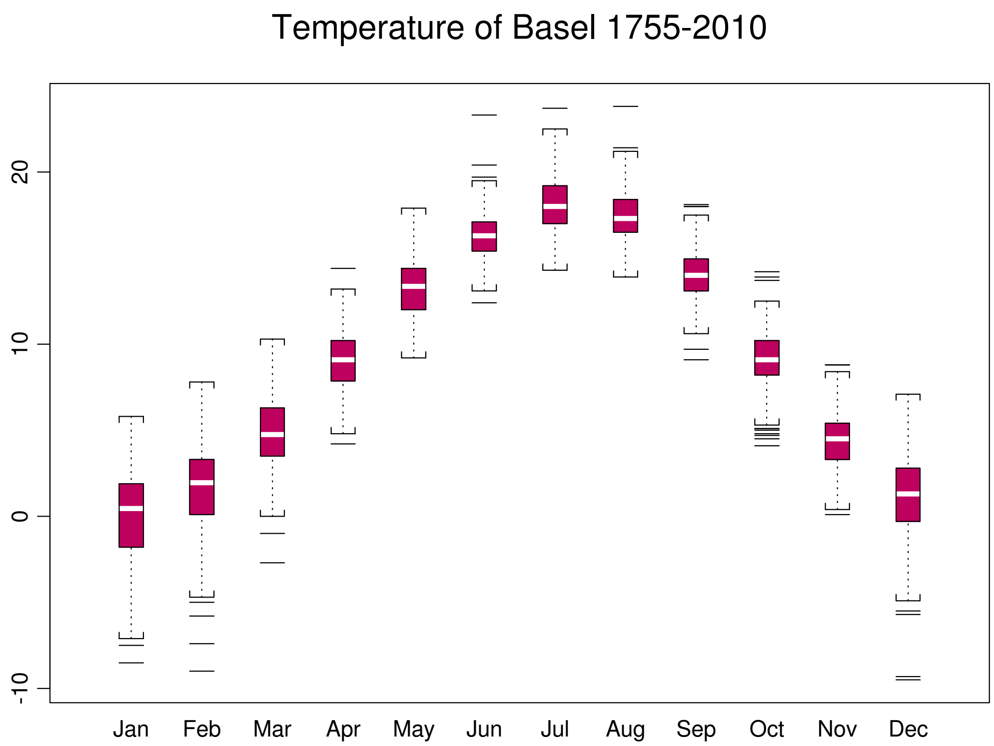


Figure 12.5: Boxplot of the Basel monthly mean temperature time series from 1755 to 2010. Data set: Basel, p. 1-6.

12.2 Simple Time Series Models

12.2.1 Zero-mean Models

We introduce some important time series models.

- Independent and identically distributed (iid) noise: Perhaps the simplest model for a time series is one in which there is no trend or seasonal component and in which the observations are simply independent and identically distributed (iid) random variables with zero mean. We refer to such a sequence of random variables $X_1, X_2, \ldots$ as iid noise. By definition we can write, for any positive integer n and real numbers $x_1, \ldots, x_n$,

$$P(X_1 \leq x_1, \ldots, X_n \leq x_n) = P(X_1 \leq x_1) \cdot \ldots \cdot P(X_n \leq x_n) = F(x_1) \cdots F(x_n),$$

where $F(\cdot)$ is the cumulative distribution function of each of the identically distributed random variables $X_1, X_2, \ldots$. In this model there is no dependence between observations. In particular, for all $h \geq 1$ and all $x, x_1, \ldots, x_n$,

$$P(X_{n+h} \leq x | X_1 = x_1, \ldots, X_n = x_n) = P(X_{n+h} \leq x),$$

showing that knowledge of $X_1, \ldots, X_n$ is of no value for predicting the behavior of X_{n+h}.

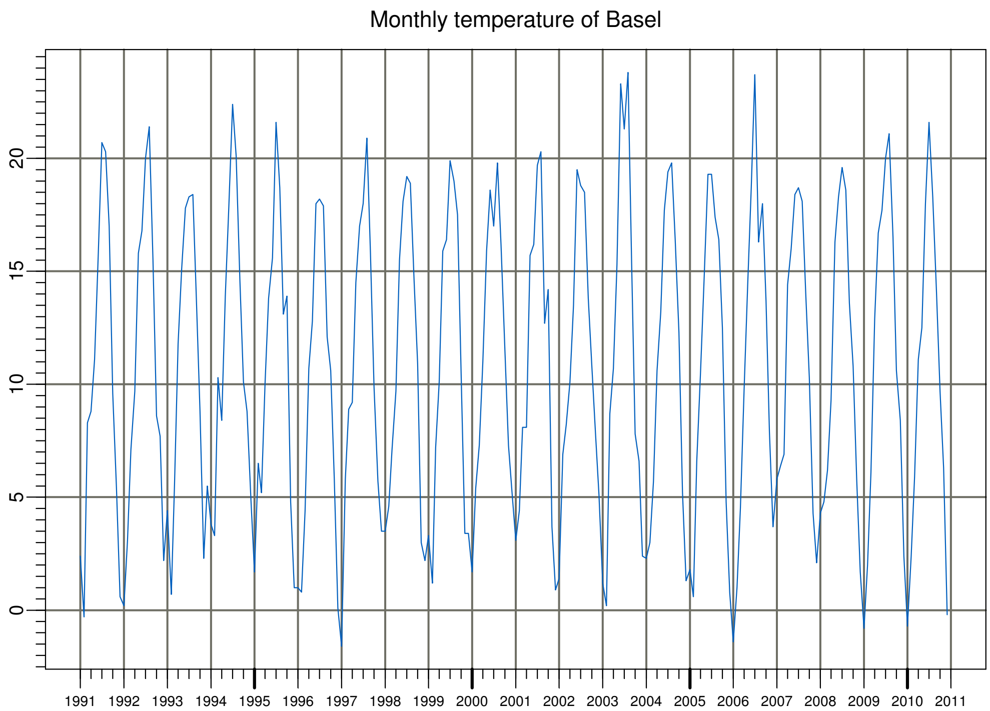


Figure 12.6: Basel monthly mean temperature time series from 1991 to 2010. Data set: Basel, p. 1-6.

Remark. We shall use the notation

$$\{X_t\} \sim \text{IID}(0, \sigma^2)$$

to indicate that the random variables X_t are independent and identically distributed random variables, each with mean 0 and variance σ^2.

Although iid noise is a rather uninteresting process for forecasting, it plays an important role as a building block for more complicated time series models.

- Binary process: Consider the sequence of iid random variables $\{X_t, t = 1, 2, \ldots\}$ with

$$P(X_t = 1) = p, \qquad P(X_t = -1) = 1 - p,$$

where $p = 1/2$.

- Random walk: The random walk $\{S_t, t = 0, 1, 2, \ldots\}$ is obtained by cumulatively summing iid random variables. Thus a random walk with zero mean is obtained by defining

$$S_0 = 0 \quad \text{and} \quad S_t = X_1 + \ldots + X_t \quad \text{for } t = 1, 2, \ldots,$$

where $\{X_t\}$ is iid noise.

Example. Figure 12.7 shows different examples of zero-mean time series models.

12.2.2 Models with Trend and Seasonality

In several of the time series there is a clear trend in the data. In these cases a zero-mean model for the data is clearly inappropriate.

- Trend: $X_t = m_t + Y_t$, where m_t is a slowly changing function known as the trend component and Y_t has zero mean. A useful technique for estimating m_t is the method of least squares.

  Example. In the least squares procedure we attempt to fit a parametric family of functions, e.g., $m_t = a_0 + a_1 t + a_2 t^2$ to the data $\{x_1, \ldots, x_n\}$ by choosing the parameters a_0, a_1 and a_2 to minimize

  $$\sum_{t=1}^{n} (x_t - m_t)^2$$

- Seasonality: In order to represent a seasonal effect, allowing for noise but assuming no trend, we can use the simple model

  $$X_t = s_t + Y_t,$$

  where s_t is a periodic function of t with period d, i.e., $s_{t-d} = s_t$.

  Example. A convenient choice for s_t is a sum of harmonics given by

  $$s_t = a_0 + \sum_{j=1}^{k} (a_j \cos(\lambda_j t) + b_j \sin(\lambda_j t))$$

  where $a_0, \ldots, a_k$ and $b_1, \ldots, b_k$ are unknown parameters and $\lambda_1, \ldots, \lambda_k$ are fixed frequencies, each being some integer multiple of $2\pi/d$.

12.3 General Approach to Time Series Modeling

The examples of the previous section illustrate a general approach to time series analysis. Before introducing the ideas of dependence and stationarity, the following outline provides the reader with an overview of the way in which the various ideas fit together:

1. Plot the series and examine the main features of the graph, checking in particular whether there is
   (a) a trend,
   (b) a seasonal component,

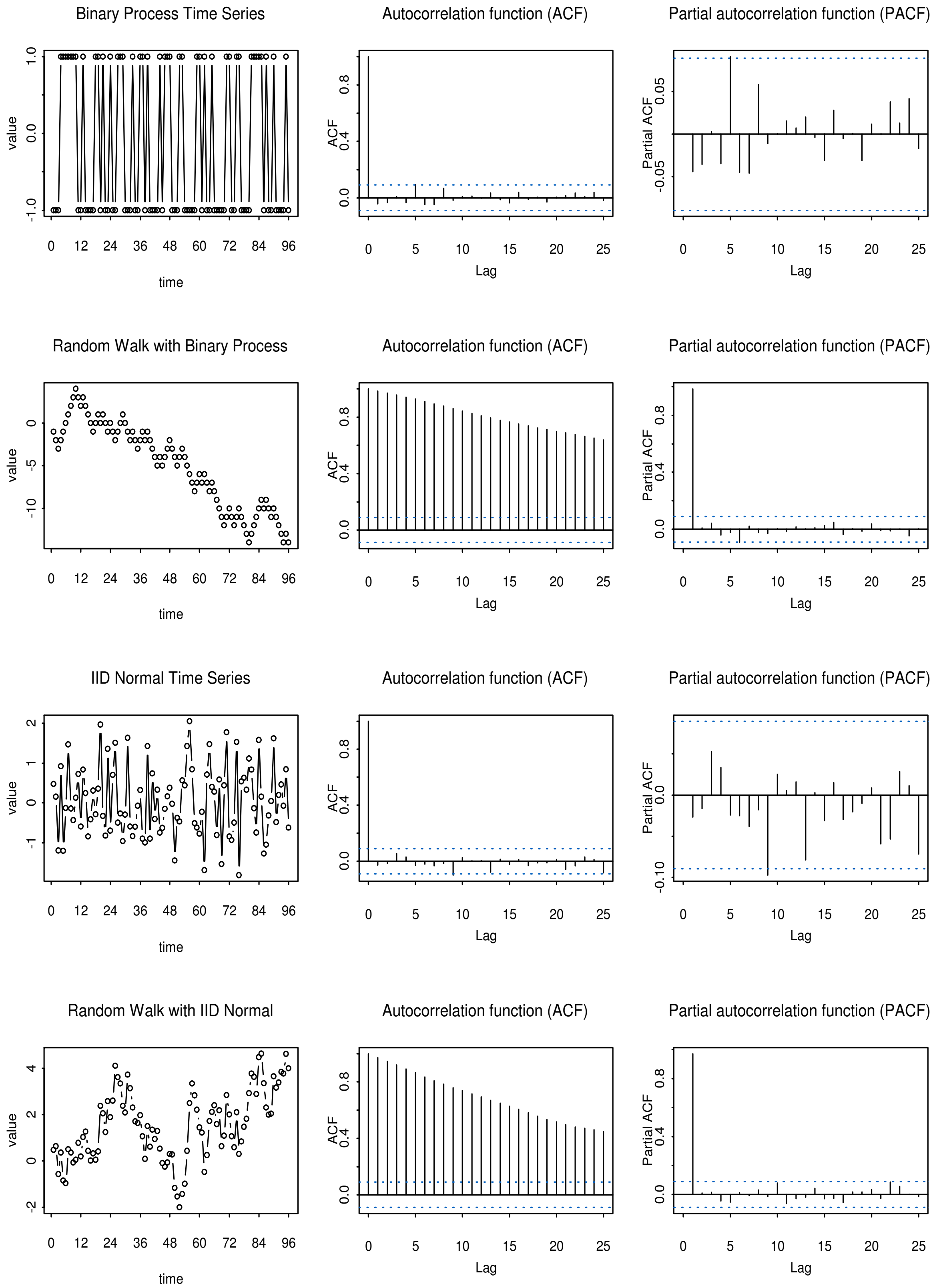


Figure 12.7: Examples of elementary time series models. The left column shows an excerpt ($m = 96$) of the whole time series ($n = 480$).

(c) any apparent sharp changes in behavior,

(d) any outlying observations.

Example (Basel, p. 1-6)**.** The Basel annual mean temperature time series (Figure 12.4, p. 12-5) does not have an obvious trend for the first period till 1900 while for the second period, starting with the 20th century a linear upward trend can be observed. Furthermore the Basel monthly mean temperature time series (Figure 12.5, p. 12-6) shows a strong seasonal component.

2. Remove the trend and seasonal components to get stationary residuals (see Section 12.4). To achieve this goal it may sometimes be necessary to apply a preliminary transformation to the data. There are several ways in which trend and seasonality can be removed (see Section 12.5).

3. Choose a model to fit the residuals.

4. Forecasting will be achieved by forecasting the residuals and then inverting the transformations described above to arrive at forecasts of the original series $\{X_t\}$.

12.4 Stationary Models and the Autocorrelation Function

Loosely speaking, a time series $\{X_t, t = 0, \pm 1, \pm 2, \ldots\}$ is said to be stationary if it has statistical properties similar to those of the "time-shifted" series $\{X_{t+h}, t = 0, \pm 1, \pm 2, \ldots\}$, for each integer h. So stationarity implies that the model parameters do not vary with time. Restricting attention to those properties that depend only on the first- and second-order moments of $\{X_t\}$, we can make this idea precise with the following definitions.

Definition 12.4.1. Let $\{X_t\}$ be a time series with $\mathrm{E}(X_t^2) < \infty$. The mean function of $\{X_t\}$ is

$$\mu_X(t) = \mathrm{E}(X_t).$$

The covariance function of $\{X_t\}$ is

$$\gamma_X(r, s) = \mathrm{Cov}(X_r, X_s) = \mathrm{E}((X_r - \mu_X(r))(X_s - \mu_X(s)))$$

for all integers r and s.

Definition 12.4.2. $\{X_t\}$ is (weakly) stationary if

- the mean value function $\mu_X(t)$ is constant and does not depend on time t and
- the autocovariance function $\gamma_X(r, s)$ depends on r and s only through their difference $|r - s|$, or in other words $\gamma_X(t + h, t)$ is independent of t for each h.

Remark. A strictly stationary time series is one for which the probabilistic behavior of every collection of values $\{x_1, \dots x_n\}$ is identical to that of the time shifted set $\{x_{1+h}, \dots x_{n+h}\}$. That is, strict stationarity of a time series $\{X_t, t = 0, \pm 1, \pm 2, \dots\}$ is defined by the condition that $(X_1, \dots, X_n)$ and $(X_{1+h}, \dots, X_{n+h})$ have the same joint distributions for all integers h and $n > 0$. It can be checked that if $\{X_t\}$ is strictly stationary and $\mathrm{E}X_t^2 < \infty$ for all t, then $\{X_t\}$ is also weakly stationary. Whenever we use the term stationary we shall mean weakly stationary as in Definition 12.4.2.

Remark. If a Gaussian time series is weakly stationary it is also strictly stationary.

Remark. Whenever we use the term covariance function with reference to a stationary time series $\{X_t\}$ we define

$$\gamma_X(h) := \gamma_X(h, 0) = \gamma_X(t + h, t).$$

The function $\gamma_X(\cdot)$ will be referred to as the autocovariance function and $\gamma_X(h)$ as its value at lag h.

Definition 12.4.3. Let $\{X_t\}$ be a stationary time series. The autocovariance function (ACFV) of $\{X_t\}$ at lag h is

$$\gamma_X(h) = \mathrm{Cov}(X_{t+h}, X_t) = \mathrm{E}(X_{t+h}X_t) - \mathrm{E}X_{t+h}\mathrm{E}X_t.$$

The autocorrelation function (ACF) of $\{X_t\}$ at lag h is

$$\rho_X(h) := \frac{\gamma_X(h)}{\gamma_X(0)} = \mathrm{Cor}(X_{t+h}, X_t).$$

Example. Let's have a look to some zero-mean time series models (compare Figure 12.7):

- iid noise: If $\{X_t\}$ is iid noise and $\mathrm{E}(X_t^2) = \sigma^2 < \infty$, then the first requirement of Definition 12.4.2 is satisfied since $\mathrm{E}(X_t) = 0$, for all t. By the assumed independence

  $$\gamma_X(t + h, t) = \begin{cases} \sigma^2, & \text{if } h = 0, \\ 0, & \text{if } h \neq 0. \end{cases}$$

  which does not depend on t. Hence iid noise with finite second moment is stationary. We shall use the notation

  $$\{X_t\} \sim \mathrm{IID}(0, \sigma^2)$$

  to indicate that the random variables X_t are independent and identically distributed random variables, each with mean 0 and variance σ^2.

- White noise: If $\{X_t\}$ is a sequence of uncorrelated random variables, each with zero mean and variance σ^2, then $\{X_t\}$ is stationary with the same covariance function as the iid noise. Such a sequence is referred to as white noise (with mean 0 and variance σ^2). This is indicated by the notation

  $$\{X_t\} \sim \mathrm{WN}(0, \sigma^2).$$

Remark. Every IID$(0,\sigma^2)$ sequence is WN$(0,\sigma^2)$ but not conversely: Let $\{Z_t\} \sim$ N$(0,1)$ and define

$$X_t = \begin{cases} Z_t, & t \text{ even}, \\ (Z_{t-1}^2 - 1)/\sqrt{2}, & t \text{ odd}, \end{cases}$$

then $\{X_t\} \sim$ WN$(0,1)$ but not IID$(0,1)$.

- Random walk: We find $\mathrm{E}S_t = 0$, $\mathrm{E}S_t^2 = t\sigma^2 < \infty$ for all t, and, for $h \geq 0$,

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

  Since $\gamma_S(t+h,t)$ depends on t, the series $\{S_t\}$ is not stationary.

Example. Figure 12.8 shows different examples of time series models with trend and seasonal components.

Example. First-order moving average or MA(1) process: Consider the series defined by the equation

$$X_t = Z_t + \theta Z_{t-1}, \quad t = 0, \pm 1, \pm 2, \ldots, \tag{12.1}$$

where $\{Z_t\} \sim$ WN$(0,\sigma^2)$ and θ is a real-valued constant. We find $\mathrm{E}X_t = 0$ and $\mathrm{E}X_t^2 = \sigma^2(1+\theta^2) < \infty$. The autocovariance function can be calculated as

$$\gamma_X(t+h,t) = \begin{cases} \sigma^2(1+\theta^2), & \text{if } h = 0, \\ \sigma^2\theta, & \text{if } h = \pm 1, \\ 0, & \text{if } |h| > 1. \end{cases} \tag{12.2}$$

Therefore $\{X_t\}$ is stationary.

Example. Note that the processes $X_t = Z_t + 5Z_{t-1}$, where $\{Z_t\} \sim$ WN$(0,1)$ and $Y_t = W_t + \frac{1}{5}W_{t-1}$, where $\{W_t\} \sim$ WN$(0,25)$ have the same autocovariance functions. Since we can only observe the time series X_t or Y_t and not the noise Z_t or W_t we cannot distinguish between the models. But as we will see later on, the second model is invertible, while the first is not.

Example. First-order autoregression or AR(1) process: Let us *assume* that the series $\{X_t\}$ defined by the equation

$$X_t = \phi X_{t-1} + Z_t, \quad t = 0, \pm 1, \pm 2, \ldots, \tag{12.3}$$

is stationary, where $\{Z_t\} \sim$ WN$(0,\sigma^2)$, Z_t is uncorrelated with X_s for each $s < t$ and $|\phi| < 1$.

By taking expectations on each side of (12.3) and using the fact that $\mathrm{E}Z_t = 0$, we see at once that $\mathrm{E}X_t = 0$. To find the autocovariance function of $\{X_t\}$ multiply each side of (12.3) by X_{t-h} and then take expectations to get

$$\begin{aligned} \gamma_X(h) &= \mathrm{Cov}(X_t, X_{t-h}) = \mathrm{Cov}(\phi X_{t-1}, X_{t-h}) + \mathrm{Cov}(Z_t, X_{t-h}) \\ &= \phi\gamma_X(h-1) = \ldots = \phi^h\gamma_X(0). \end{aligned}$$

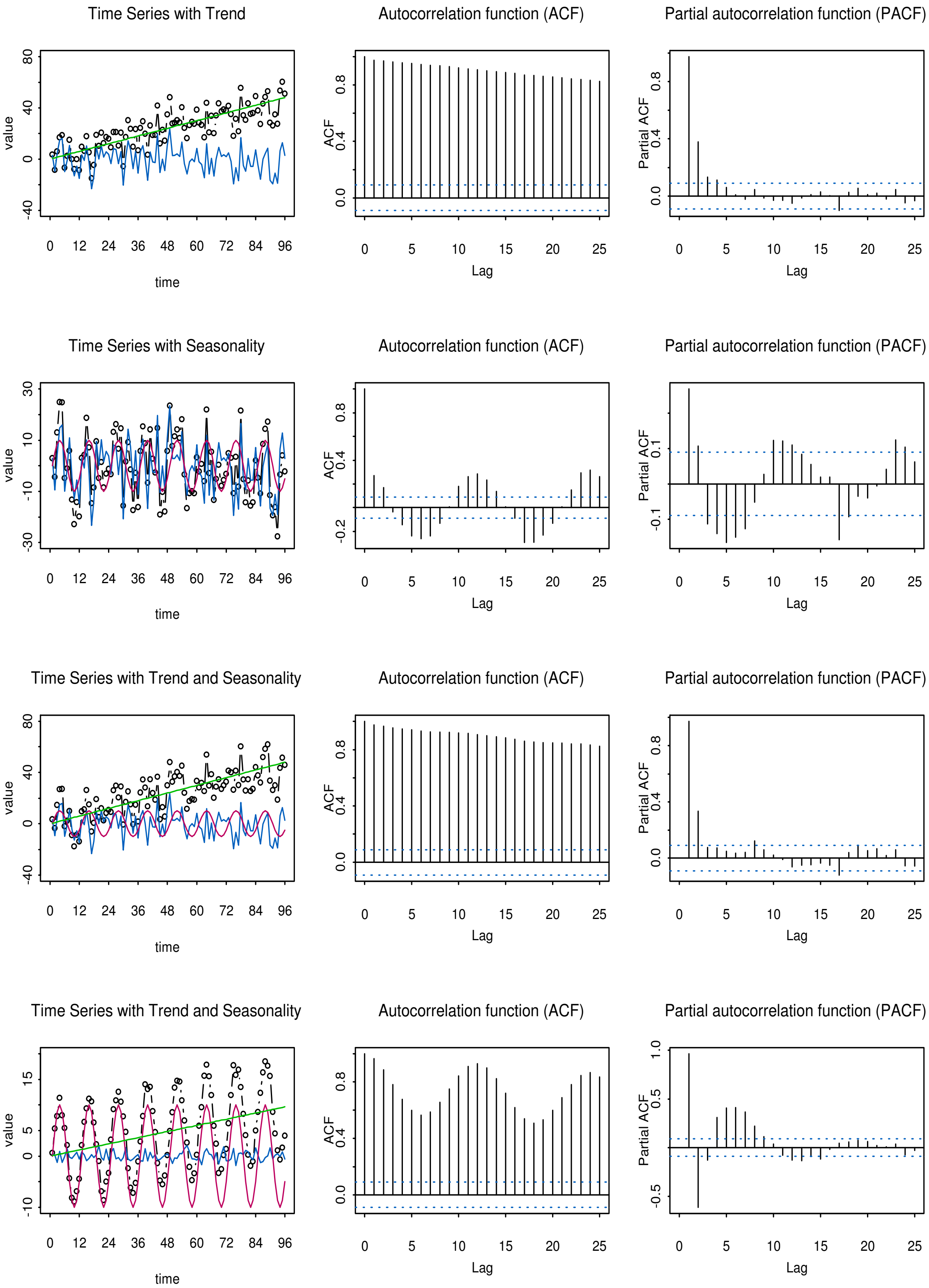


Figure 12.8: Examples of time series models with trend and seasonal components. The left column shows an excerpt ($m = 96$) of the whole time series ($n = 480$).

Observing that $\gamma(h) = \gamma(-h)$, we find that

$$\rho_X(h) = \frac{\gamma_X(h)}{\gamma_X(0)} = \phi^{|h|}, \quad h = 0, \pm 1, \ldots.$$

It follows from the linearity of the covariance function and the fact that Z_t is uncorrelated with X_{t-1} that

$$\gamma_X(0) = \mathrm{Cov}(X_t, X_t) = \mathrm{Cov}(\phi X_{t-1} + Z_t, \phi X_{t-1} + Z_t) = \phi^2 \gamma_X(0) + \sigma^2$$

and hence that

$$\gamma_X(0) = \frac{\sigma^2}{1 - \phi^2}.$$

Combined we have

$$\gamma_X(h) = \phi^{|h|} \frac{\sigma^2}{1 - \phi^2}.$$

For more details on AR(1) processes see page 13-6.

Definition 12.4.4. Let $x_1, \ldots, x_n$ be observations of a time series. The sample mean of $x_1, \ldots, x_n$ is

$$\overline{x} = \frac{1}{n} \sum_{j=1}^{n} x_j.$$

For $|h| < n$ we define the sample autocovariance function

$$\hat{\gamma}(h) := \frac{1}{n} \sum_{j=1}^{n-|h|} (x_{j+|h|} - \overline{x})(x_j - \overline{x}),$$

and the sample autocorrelation function

$$\hat{\rho}(h) = \frac{\hat{\gamma}(h)}{\hat{\gamma}(0)}.$$

Remark. For data containing a trend, $|\hat{\rho}(h)|$ will exhibit slow decay as h increases, and for data with a substantial deterministic periodic component, $|\hat{\rho}(h)|$ will exhibit similar behavior with the same periodicity (see Figure 12.8). Thus $\hat{\rho}(\cdot)$ can be useful as an indicator of nonstationarity.

Example (Basel, p. 1-6)**.** Figure 12.9 shows the Basel monthly mean temperatures from 1900 to 2010 with the corresponding autocorrelation function and partial autocorrelation function (definition see Section 14.3, p. 14-7). The waves of the autocorrelation plot are a typical indication for a seasonal dependency within the time series. Since the seasonal component is very strong the trend component can not be detected.

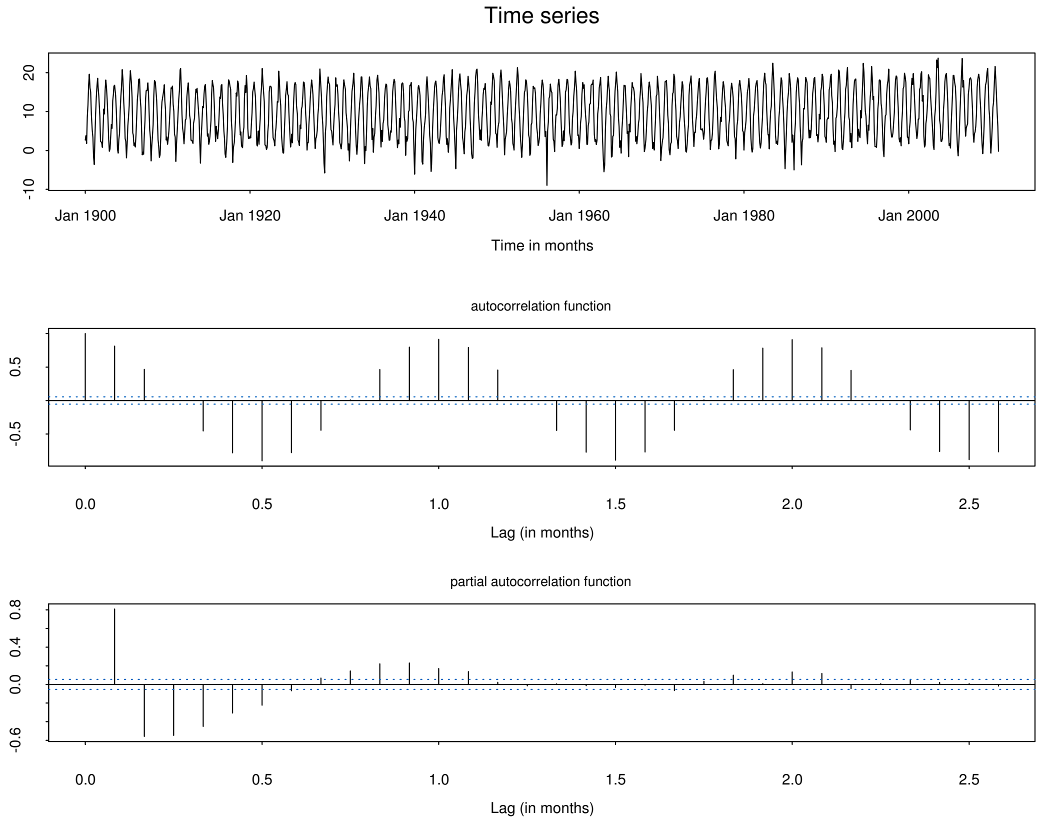


Figure 12.9: Basel monthly mean temperature time series from 1900 to 2010 with corresponding autocorrelation function and partial autocorrelation function. Data set: Basel, p. 1-6.

12.5 Estimation and Elimination of Trend and Seasonal Components

The first step in the analysis of any time series is to plot the data. If there are any apparent discontinuities in the series, such as a sudden change of level, it may be advisable to analyze the series by first breaking it into homogeneous segments. If there are outlying observations, they should be studied carefully to check whether there is any justification for discarding them. Inspection of a graph may also suggest the possibility of representing the data as a realization of the classical decomposition model

$$X_t = m_t + s_t + Y_t, \tag{12.4}$$

where m_t is a slowly changing function known as a trend component, s_t is a function with known period d referred as a seasonal decomposition, and Y_t is a random noise component that is stationary. If the seasonal and noise fluctuations appear to increase with the level of the process, then a preliminary transformation of the data is often used to make the transformed data more compatible with the model (12.4).

Our aim is to estimate and extract the deterministic components m_t and s_t in the hope that the residual or noise component Y_t will turn out to be a stationary time series (method 1). We can then use the theory of such processes to find a satisfactory probabilistic model for the process Y_t, to analyze its properties, and to use it in conjunction with m_t and s_t for purposes of prediction and simulation of $\{X_t\}$.

Another approach is to apply differencing operators repeatedly to the series $\{X_t\}$ until the differenced observations resemble a realization of some stationary time series $\{W_t\}$ (method 2). We can then use the theory of stationary processes for the modeling, analysis, and prediction of $\{W_t\}$ and hence of the original process.

12.5.1 Nonseasonal Model with Trend

In the absence of a seasonal component the model (12.4) becomes

$$X_t = m_t + Y_t, \quad t = 1, \dots, n, \tag{12.5}$$

where $\mathrm{E}Y_t = 0$.

Method 1: Trend estimation

A lot of methods can be found in literature, here some examples.

a) Smoothing with a finite moving average filter. Let q be a non-negative integer and consider the two-sided moving average

$$W_t = (2q+1)^{-1} \sum_{j=-q}^{q} X_{t-j}$$

of the process $\{X_t\}$ defined in (12.5). Then for $q+1 \le t \le n-q$, we find

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

assuming that m_t is approximately linear over the interval $[t-q, t+q]$ and that the average of the error terms over this interval is close to zero. The moving average thus provides us with the estimates

$$\hat{m}_t = (2q+1)^{-1} \sum_{j=-q}^{q} X_{t-j}, \quad q+1 \le t \le n-q. \tag{12.6}$$

b) Exponential smoothing. For any fixed $\alpha \in [0,1]$, the one-sided moving averages $\hat{m}_t$, $t = 1, \dots, n$, are defined by the recursions

$$\hat{m}_t = \alpha X_t + (1-\alpha)\hat{m}_{t-1}, \quad t = 2, \dots, n,$$

and

$$\hat{m}_1 = X_1.$$

c) Smoothing by eliminating high-frequency components.

d) Polynomial fitting. This method can be used to estimate higher-order polynomial trends.

Method 2: Trend Elimination by Differencing

The trend term is eliminated by differencing. We define the lag-1 difference operator ∇ by

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

where B is the backward shift operator,

$$BX_t = X_{t-1}.$$

Powers of the operators B and ∇ are defined in the obvious way, i.e.,

$$\begin{aligned} &B^j(X_t) = X_{t-j}, \\ &\nabla^j(X_t) = \nabla(\nabla^{j-1}(X_t)), \quad j \geq 1, \text{ with } \nabla^0(X_t) = X_t. \end{aligned}$$

Example. If the time series $\{X_t\}$ in (12.5) has a polynomial trend of degree k, it can be eliminated by application of the operator ∇^k:

$$\nabla^k X_t = k!\, c_k + \nabla^k Y_t,$$

which gives a stationary process with mean $k!\, c_k$.

12.5.2 Seasonal Model with Trend

The methods described for the estimation and elimination of trend can be adapted in a natural way to eliminate both trend and seasonality in the general model, specified as follows.

$$X_t = m_t + s_t + Y_t, \quad t = 1, \ldots, n$$

where

$$\mathrm{E}Y_t = 0, \quad s_{t+d} = s_t, \quad \sum_{j=1}^{d} s_j = 0.$$

Method 1: Estimation of trend and seasonal component

Suppose we have observations $\{x_1, \ldots, x_n\}$. The trend is first estimated by applying a moving average filter specially chosen to eliminate the seasonal component and to dampen the noise. If the period d is even, say $d = 2q$, then we use

$$\hat{m}_t = (0.5x_{t-q} + x_{t-q+1} + \ldots + x_{t+q-1} + 0.5x_{t+q})/d, \quad q < t \leq n - q.$$

If the period is odd, say $d = 2q + 1$, then we use the simple moving average (12.6). The second step is to estimate the seasonal component.

Method 2: Elimination of trend and seasonal components by differencing

The technique of differencing that was applied to nonseasonal data can be adapted to deal with seasonality of period d by introducing the lag-d differencing operator ∇_d defined by

$$\nabla_d X_t = X_t - X_{t-d} = (1 - B^d)X_t.$$

Applying the operator ∇_d to the model

$$X_t = m_t + s_t + Y_t,$$

where $\{s_t\}$ has period d, we obtain

$$\nabla_d X_t = m_t - m_{t-d} + Y_t - Y_{t-d},$$

which gives a decomposition of the difference $\nabla_d X_t$ into a trend component $(m_t - m_{t-d})$ and a noise term $(Y_t - Y_{t-d})$. The trend, $m_t - m_{t-d}$, can then be eliminated using the methods already described, in particular by applying a power of the operator ∇.

Example (Basel, p. 1-6)**.** Figure 12.10 shows the Basel monthly mean temperature time series from 1900 to 2010 after eliminating the seasonal component by differencing at lag 12 and the linear trend by differencing at lag 1, with the corresponding autocorrelation function and partial autocorrelation function. Comparing this figure to Figure 12.9, p. 12-15, the autocorrelation function plot does not show any waves any longer. But we see from the autocorrelation function (and partial autocorrelation function) that there is a negative correlation of the remaining time series at lag 12. This observation is confirmed by considering the lag-plot (see Figure 12.11).

12.6 Testing the Estimated Noise Sequence

The objective of the data transformations described in Section 12.5 is to produce a series with no apparent deviations from stationarity, and in particular with no apparent trend or seasonality. Assuming that this has been done, the next step is to model the estimated noise sequence (i.e., the residuals obtained either by differencing the data or by estimating and subtracting the trend and seasonal components). If there is no dependence among between these residuals, then we can regard them as observations of independent random variables, and there is no further modeling to be done except to estimate their mean and variance. However, if there is significant dependence among the residuals, then we need to look for a more complex stationary time series model for the noise that accounts for the dependence. This will be to our advantage, since dependence means in particular that past observations of the noise sequence can assist in predicting future values.

In this section we examine some tests for checking the hypothesis that the residuals from Section 12.5 are observed values of independent and identically distributed random variables. If they are, then our work is done. If not, the theory of stationary processes (see Chapter 13) must be used to find a more appropriate model. An overview of some tests:

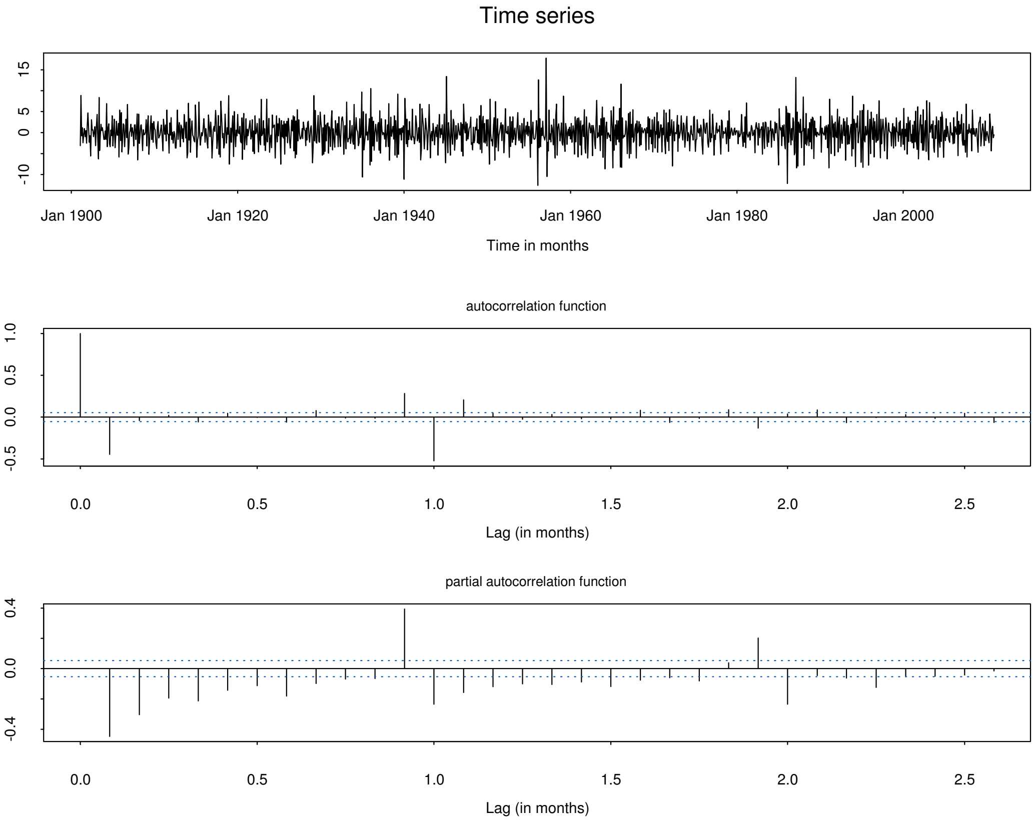


Figure 12.10: Basel monthly mean temperature time series from 1900 to 2010 after differencing at lag 12 for eliminating the seasonal component and at lag 1 for eliminating the linear trend with the corresponding autocorrelation function and partial autocorrelation function.

a) Sample autocorrelation function (SACF): For large n, the SACF of an iid sequence $Y_1, \ldots, Y_n$ with finite variance are approximately iid with $N(0, 1/n)$. Hence, if $y_1, \ldots, y_n$ is a realization of such an iid sequence, about 95% of the sample autocorrelations should fall between the bounds $\pm 1.96/\sqrt{n}$. If we compute the sample autocorrelations up to lag 40 and find that more than two or three values fall outside the bounds, or that one value falls far outside the bounds, we therefore reject the iid hypothesis.

b) Portmanteau test: Consider the statistic

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

where $\hat{\rho}$ is the sample autocorrelation function. If $Y_1, \ldots, Y_n$ is a finite-variance iid sequence, then Q is approximately distributed as the sum of squares of the independent $N(0, 1)$ random variables, $\sqrt{n}\hat{\rho}(j)$, $j = 1, \ldots, h$, i.e., as chi-squared with

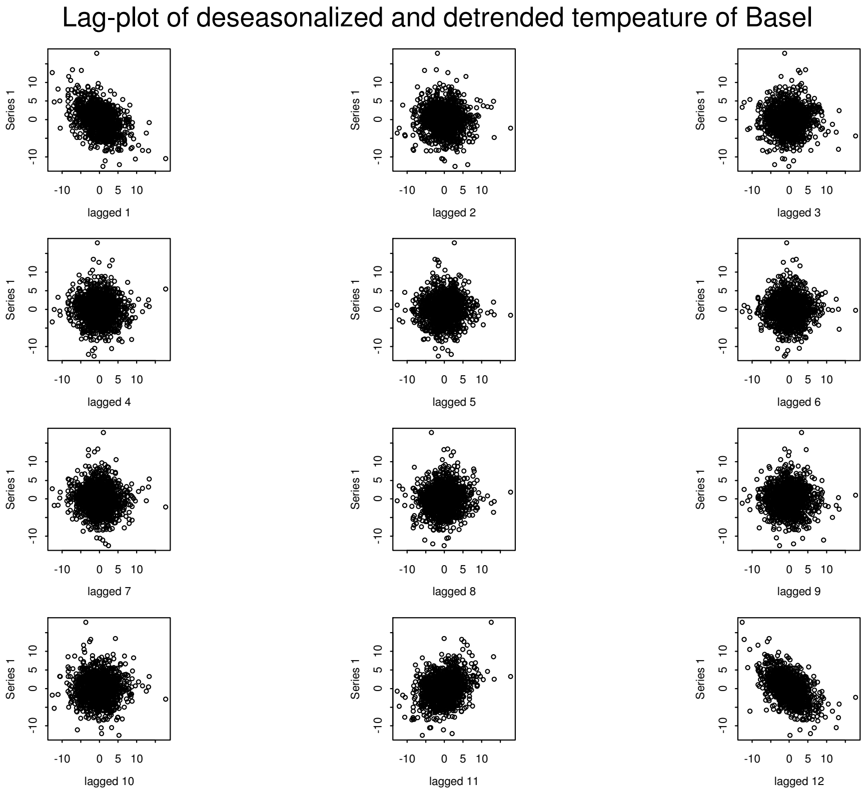


Figure 12.11: Lag-plot of the Basel monthly mean temperature time series from 1900 to 2010 after differencing at lag 12 for eliminating the seasonal component and at lag 1 for eliminating the linear trend.

h degrees of freedom. A large value of Q suggests that the sample autocorrelations of the data are too large for the data to be a sample from an iid sequence. We therefore reject the iid hypothesis at 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.

There exists a refinement of this test, formulated by Ljung and Box, in which Q is replaced by

$$Q_{LB} = n(n+2)\sum_{j=1}^{h}\frac{\hat{\rho}^2(j)}{n-j},$$

whose distribution is better approximated by the chi-squared distribution with h degrees of freedom.

Example (Basel, p. 1-6)**.** Figure 17.1, p. 17-3, shows the Ljung-Box test for the residuals of the Basel monthly mean temperature time series.

c) The turning point test: If $y_1, \ldots, y_n$ is a sequence of observations, we say that there is a turning point at time i, if $y_{i-1} < y_i$ and $y_i > y_{i+1}$ or if $y_{i-1} > y_i$ and

$y_i < y_{i+1}$. If T is the number of turning points of an iid sequence of length n, then, since the probability of a turning point at time i is $2/3$, the expected value of T is

$$\mu_T = \mathrm{E}(T) = \frac{2}{3}(n-2).$$

It can be shown for an iid sequence that the variance of T is

$$\sigma_T^2 = \mathrm{Var}(T) = \frac{16n-29}{90}.$$

A large value of $T - \mu_T$ indicates that the series is fluctuating more rapidly than expected for an iid sequence. On the other hand, a value $T - \mu_T$ much smaller than zero indicates a positive correlation between neighboring observations. For an iid sequence with n large, it can be shown that

$$T \overset{approx}{\sim} \mathrm{N}(\mu_T, \sigma_T^2).$$

This means we carry out a test of the iid hypothesis, rejecting it at level α if $|T - \mu_T|/\sigma_T > \Phi_{1-\alpha/2}$, where $\Phi_{1-\alpha/2}$ is the $1-\alpha/2$ quantile of the standard normal distribution.

d) The difference-sign test: For this test we count the number S of values of i such that $y_i > y_{i-1}$, $i = 2, \ldots, n$, or equivalently the number of times the differenced series $y_i - y_{i-1}$ is positive. For an iid sequence we see that

$$\mu_S = \frac{n-1}{2} \quad \text{and} \quad \sigma_S^2 = \frac{n+1}{12},$$

and for large n,

$$S \overset{approx}{\sim} \mathrm{N}(\mu_S, \sigma_S^2).$$

A large positive (or negative) value of $S - \mu_S$ indicates the presence of an increasing (or decreasing) trend in the data. We therefore reject the assumption of no trend in the data if $|S - \mu_S|/\sigma_S > \Phi_{1-\alpha/2}$. The difference-sign test must be used with caution. A set of observations exhibiting a strong cyclic component will pass the difference-sign test for randomness, since roughly half of the observations will be points of increase.

e) The rank test: This test is particularly useful for detecting a linear trend in the data. Define P to be the number of pairs (i, j) such that $y_j > y_i$ and $j > i$, $i = 1, \ldots, n-1$. The mean of P is $\mu_p = \frac{1}{4}n(n-1)$. A large positive (negative) value of $P - \mu_P$ indicates the presence of an increasing (decreasing) trend in the data.

Remark. The general strategy in applying the tests is to check them all and to proceed with caution if any of them suggests a serious deviation from the iid hypothesis.