Springer Texts in Statistics

Robert H. Shumway
David S. Stoffer

Time Series Analysis and Its Applications

With R Examples

Fourth Edition

Springer

Robert H. Shumway
David S. Stoffer

Time Series Analysis and Its Applications

With R Examples

Fourth Edition

live free or bark

Preface to the Fourth Edition

The fourth edition follows the general layout of the third edition but includes some modernization of topics as well as the coverage of additional topics. The preface to the third edition—which follows—still applies, so we concentrate on the differences between the two editions here. As in the third edition, R code for each example is given in the text, even if the code is excruciatingly long. Most of the examples with seemingly endless coding are in the latter chapters. The R package for the text, `astsa`, is still supported and details may be found in Appendix R. A number of data sets have been updated. For example, the global temperature deviation series have been updated to 2015 and are included in the newest version of the package; the corresponding examples and problems have been updated accordingly.

Chapter 1 of this edition is similar to the previous edition, but we have included the definition of trend stationarity and the the concept of prewhitening when using cross-correlation. The New York Stock Exchange data set, which focused on an old financial crisis, was replaced with a more current series of the Dow Jones Industrial Average, which focuses on a newer financial crisis. In Chapter 2, we rewrote some of the regression review, changed the smoothing examples from the mortality data example to the Southern Oscillation Index and finding El Niño. We also expanded the discussion of lagged regression to Chapter 3 to include the possibility of autocorrelated errors.

In Chapter 3, we removed normality from definition of ARMA models; while the assumption is not necessary for the definition, it is essential for inference and prediction. We added a section on regression with ARMA errors and the corresponding problems; this section was previously in Chapter 5. Some of the examples have been modified and we added some examples in the seasonal ARMA section.

In Chapter 4, we improved and added some examples. The idea of modulated series is discussed using the classic star magnitude data set. We moved some of the filtering section forward for easier access to information when needed. We removed the reliance on `spec.pgram` (from the `stats` package) to `mvspec` (from the `astsa` package) so we can avoid having to spend pages explaining the quirks of `spec.pgram`, which tended to take over the narrative. The section on wavelets was removed because

there are so many accessible texts available. The spectral representation theorems are discussed in a little more detail using examples based on simple harmonic processes.

The general layout of Chapter 5 and of Chapter 7 is the same, although we have revised some of the examples. As previously mentioned, we moved regression with ARMA errors to Chapter 3.

Chapter 6 sees the biggest change in this edition. We have added a section on smoothing splines, and a section on hidden Markov models and switching autoregressions. The Bayesian section is completely rewritten and is on linear Gaussian state space models only. The nonlinear material in the previous edition is removed because it was old, and the newer material is in Douc, Moulines, and Stoffer (2014). Many of the examples have been rewritten to make the chapter more accessible. Our goal was to be able to have a course on state space models based primarily on the material in Chapter 6.

The Appendices are similar, with some minor changes to Appendix A and Appendix B. We added material to Appendix C, including a discussion of Riemann–Stieltjes and stochastic integration, a proof of the fact that the spectra of autoregressive processes are dense in the space of spectral densities, and a proof of the fact that spectra are approximately the eigenvalues of the covariance matrix of a stationary process.

We tweaked, rewrote, improved, or revised some of the exercises, but the overall ordering and coverage is roughly the same. And, of course, we moved regression with ARMA errors problems to Chapter 3 and removed the Chapter 4 wavelet problems. The exercises for Chapter 6 have been updated accordingly to reflect the new and improved version of the chapter.

Davis, CA
Pittsburgh, PA
September 2016

Robert H. Shumway
David S. Stoffer

Preface to the Third Edition

The goals of this book are to develop an appreciation for the richness and versatility of modern time series analysis as a tool for analyzing data, and still maintain a commitment to theoretical integrity, as exemplified by the seminal works of Brillinger (1975) and Hannan (1970) and the texts by Brockwell and Davis (1991) and Fuller (1995). The advent of inexpensive powerful computing has provided both real data and new software that can take one considerably beyond the fitting of simple time domain models, such as have been elegantly described in the landmark work of Box and Jenkins (1970). This book is designed to be useful as a text for courses in time series on several different levels and as a reference work for practitioners facing the analysis of time-correlated data in the physical, biological, and social sciences.

We have used earlier versions of the text at both the undergraduate and graduate levels over the past decade. Our experience is that an undergraduate course can be accessible to students with a background in regression analysis and may include Section 1.1–Section 1.5, Section 2.1–Section 2.3, the results and numerical parts of Section 3.1–Section 3.9, and briefly the results and numerical parts of Section 4.1–Section 4.4. At the advanced undergraduate or master's level, where the students have some mathematical statistics background, more detailed coverage of the same sections, with the inclusion of extra topics from Chapter 5 or Chapter 6 can be used as a one-semester course. Often, the extra topics are chosen by the students according to their interests. Finally, a two-semester upper-level graduate course for mathematics, statistics, and engineering graduate students can be crafted by adding selected theoretical appendices. For the upper-level graduate course, we should mention that we are striving for a broader but less rigorous level of coverage than that which is attained by Brockwell and Davis (1991), the classic entry at this level.

The major difference between this third edition of the text and the second edition is that we provide R code for almost all of the numerical examples. An R package called `astsa` is provided for use with the text; see Section R.2 for details. R code is provided simply to enhance the exposition by making the numerical examples reproducible.

We have tried, where possible, to keep the problem sets in order so that an instructor may have an easy time moving from the second edition to the third edition.

However, some of the old problems have been revised and there are some new problems. Also, some of the data sets have been updated. We added one section in Chapter 5 on unit roots and enhanced some of the presentations throughout the text. The exposition on state-space modeling, ARMAX models, and (multivariate) regression with autocorrelated errors in Chapter 6 have been expanded. In this edition, we use standard R functions as much as possible, but we use our own scripts (included in `astsa`) when we feel it is necessary to avoid problems with a particular R function; these problems are discussed in detail on the website for the text under R Issues.

We thank John Kimmel, Executive Editor, Springer Statistics, for his guidance in the preparation and production of this edition of the text. We are grateful to Don Percival, University of Washington, for numerous suggestions that led to substantial improvement to the presentation in the second edition, and consequently in this edition. We thank Doug Wiens, University of Alberta, for help with some of the R code in Chapter 4 and Chapter 7, and for his many suggestions for improvement of the exposition. We are grateful for the continued help and advice of Pierre Duchesne, University of Montreal, and Alexander Aue, University of California, Davis. We also thank the many students and other readers who took the time to mention typographical errors and other corrections to the first and second editions. Finally, work on the this edition was supported by the National Science Foundation while one of us (D.S.S.) was working at the Foundation under the Intergovernmental Personnel Act.

Davis, CA — *Robert H. Shumway*
Pittsburgh, PA — *David S. Stoffer*
September 2010

Contents

Chapter 1

Characteristics of Time Series

The analysis of experimental data that have been observed at different points in time leads to new and unique problems in statistical modeling and inference. The obvious correlation introduced by the sampling of adjacent points in time can severely restrict the applicability of the many conventional statistical methods traditionally dependent on the assumption that these adjacent observations are independent and identically distributed. The systematic approach by which one goes about answering the mathematical and statistical questions posed by these time correlations is commonly referred to as time series analysis.

The impact of time series analysis on scientific applications can be partially documented by producing an abbreviated listing of the diverse fields in which important time series problems may arise. For example, many familiar time series occur in the field of economics, where we are continually exposed to daily stock market quotations or monthly unemployment figures. Social scientists follow population series, such as birthrates or school enrollments. An epidemiologist might be interested in the number of influenza cases observed over some time period. In medicine, blood pressure measurements traced over time could be useful for evaluating drugs used in treating hypertension. Functional magnetic resonance imaging of brain-wave time series patterns might be used to study how the brain reacts to certain stimuli under various experimental conditions.

In our view, the first step in any time series investigation always involves careful examination of the recorded data plotted over time. This scrutiny often suggests the method of analysis as well as statistics that will be of use in summarizing the information in the data. Before looking more closely at the particular statistical methods, it is appropriate to mention that two separate, but not necessarily mutually exclusive, approaches to time series analysis exist, commonly identified as the *time domain approach* and the *frequency domain approach*. The time domain approach views the investigation of lagged relationships as most important (e.g., how does what happened today affect what will happen tomorrow), whereas the frequency domain approach views the investigation of cycles as most important (e.g., what is the economic cycle through periods of expansion and recession). We will explore both types of approaches in the following sections.

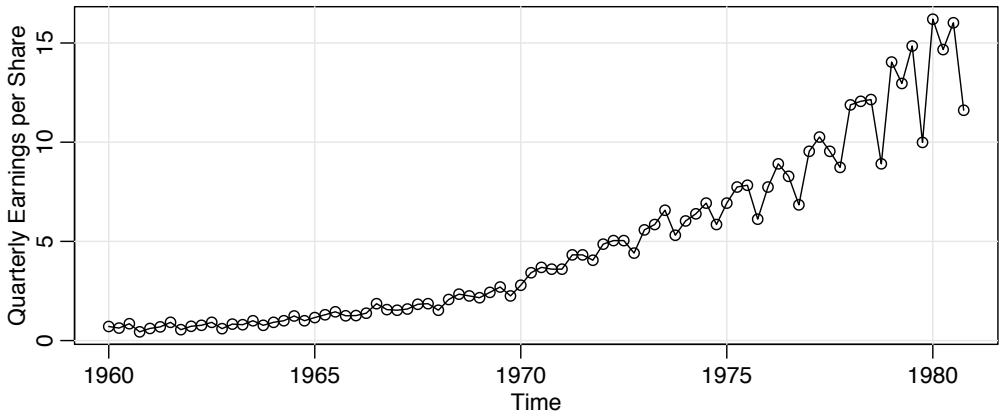

Fig. 1.1. *Johnson & Johnson quarterly earnings per share, 84 quarters, 1960-I to 1980-IV.*

1.1 The Nature of Time Series Data

Some of the problems and questions of interest to the prospective time series analyst can best be exposed by considering real experimental data taken from different subject areas. The following cases illustrate some of the common kinds of experimental time series data as well as some of the statistical questions that might be asked about such data.

Example 1.1 Johnson & Johnson Quarterly Earnings

Figure 1.1 shows quarterly earnings per share for the U.S. company Johnson & Johnson, furnished by Professor Paul Griffin (personal communication) of the Graduate School of Management, University of California, Davis. There are 84 quarters (21 years) measured from the first quarter of 1960 to the last quarter of 1980. Modeling such series begins by observing the primary patterns in the time history. In this case, note the gradually increasing underlying trend and the rather regular variation superimposed on the trend that seems to repeat over quarters. Methods for analyzing data such as these are explored in Chapter 2 and Chapter 6. To plot the data using the R statistical package, type the following:[1.1]

```
library(astsa)     # SEE THE FOOTNOTE
plot(jj, type="o", ylab="Quarterly Earnings per Share")
```

Example 1.2 Global Warming

Consider the global temperature series record shown in Figure 1.2. The data are the global mean land–ocean temperature index from 1880 to 2015, with the base period 1951-1980. In particular, the data are deviations, measured in degrees centigrade, from the 1951-1980 average, and are an update of Hansen et al. (2006). We note an apparent upward trend in the series during the latter part of the twentieth century that has been used as an argument for the global warming hypothesis. Note also the leveling off at about 1935 and then another rather sharp upward trend at about

[1.1] Throughout the text, we assume that the R package for the book, `astsa`, has been installed and loaded. See Section R.2 for further details.

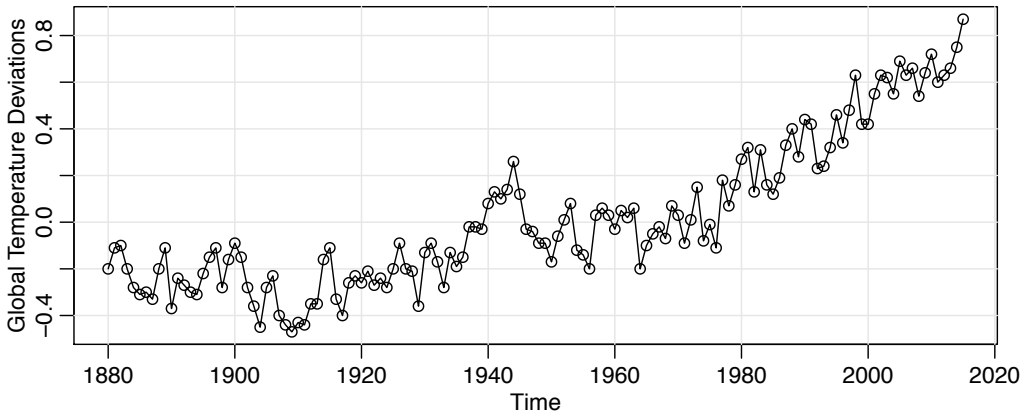

Fig. 1.2. *Yearly average global temperature deviations (1880–2015) in degrees centigrade.*

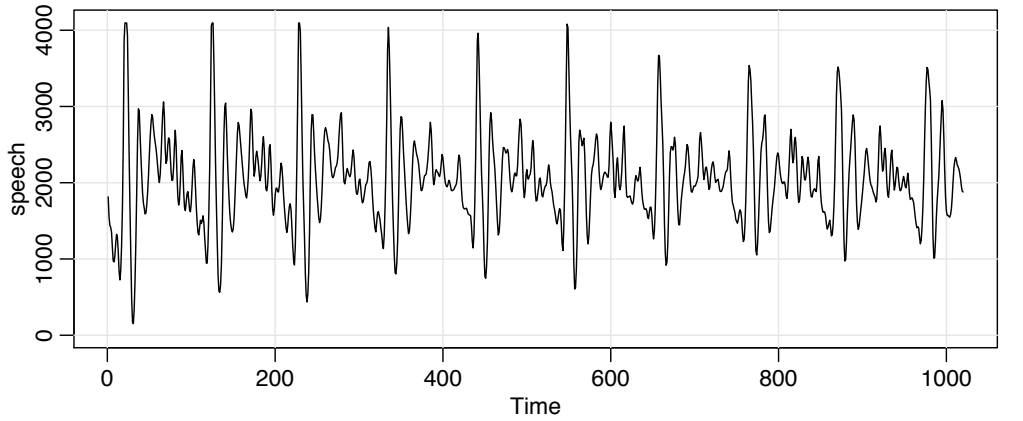

Fig. 1.3. *Speech recording of the syllable* $aaa\cdots hhh$ *sampled at 10,000 points per second with* $n = 1020$ *points.*

1970. The question of interest for global warming proponents and opponents is whether the overall trend is natural or whether it is caused by some human-induced interface. Problem 2.8 examines 634 years of glacial sediment data that might be taken as a long-term temperature proxy. Such percentage changes in temperature do not seem to be unusual over a time period of 100 years. Again, the question of trend is of more interest than particular periodicities. The R code for this example is similar to the code in Example 1.1:

```
plot(globtemp, type="o", ylab="Global Temperature Deviations")
```

Example 1.3 Speech Data

Figure 1.3 shows a small .1 second (1000 point) sample of recorded speech for the phrase $aaa\cdots hhh$, and we note the repetitive nature of the signal and the rather regular periodicities. One current problem of great interest is computer recognition of speech, which would require converting this particular signal into the recorded phrase $aaa\cdots hhh$. Spectral analysis can be used in this context to produce a signature of this phrase that can be compared with signatures of various

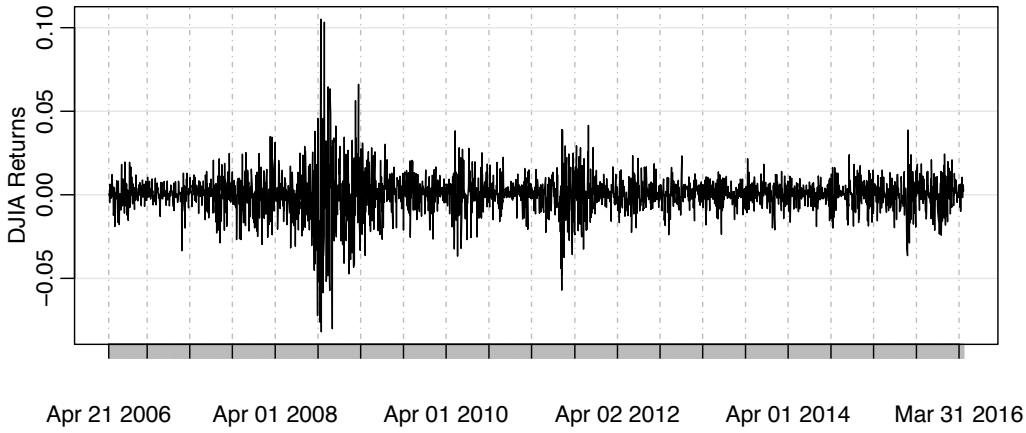

Fig. 1.4. *The daily returns of the Dow Jones Industrial Average (DJIA) from April 20, 2006 to April 20, 2016.*

library syllables to look for a match. One can immediately notice the rather regular repetition of small wavelets. The separation between the packets is known as the *pitch period* and represents the response of the vocal tract filter to a periodic sequence of pulses stimulated by the opening and closing of the glottis. In R, you can reproduce Figure 1.3 using `plot(speech)`.

Example 1.4 Dow Jones Industrial Average

As an example of financial time series data, Figure 1.4 shows the daily *returns* (or percent change) of the Dow Jones Industrial Average (DJIA) from April 20, 2006 to April 20, 2016. It is easy to spot the financial crisis of 2008 in the figure. The data shown in Figure 1.4 are typical of return data. The mean of the series appears to be stable with an average return of approximately zero, however, highly volatile (variable) periods tend to be clustered together. A problem in the analysis of these type of financial data is to forecast the volatility of future returns. Models such as *ARCH* and *GARCH* models (Engle, 1982; Bollerslev, 1986) and *stochastic volatility* models (Harvey, Ruiz and Shephard, 1994) have been developed to handle these problems. We will discuss these models and the analysis of financial data in Chapter 5 and Chapter 6. The data were obtained using the Technical Trading Rules (TTR) package to download the data from Yahoo^{TM} and then plot it. We then used the fact that if x_t is the actual value of the DJIA and $r_t = (x_t - x_{t-1})/x_{t-1}$ is the return, then $1 + r_t = x_t/x_{t-1}$ and $\log(1 + r_t) = \log(x_t/x_{t-1}) = \log(x_t) - \log(x_{t-1}) \approx r_t$.[1.2] The data set is also available in `astsa`, but `xts` must be loaded.

```
# library(TTR)
# djia  = getYahooData("^DJI", start=20060420, end=20160420, freq="daily")
library(xts)
djiar = diff(log(djia$Close))[-1] # approximate returns
plot(djiar, main="DJIA Returns", type="n")
lines(djiar)
```

[1.2] $\log(1 + p) = p - \frac{p^2}{2} + \frac{p^3}{3} - \cdots$ for $-1 < p \leq 1$. If p is near zero, the higher-order terms in the expansion are negligible.

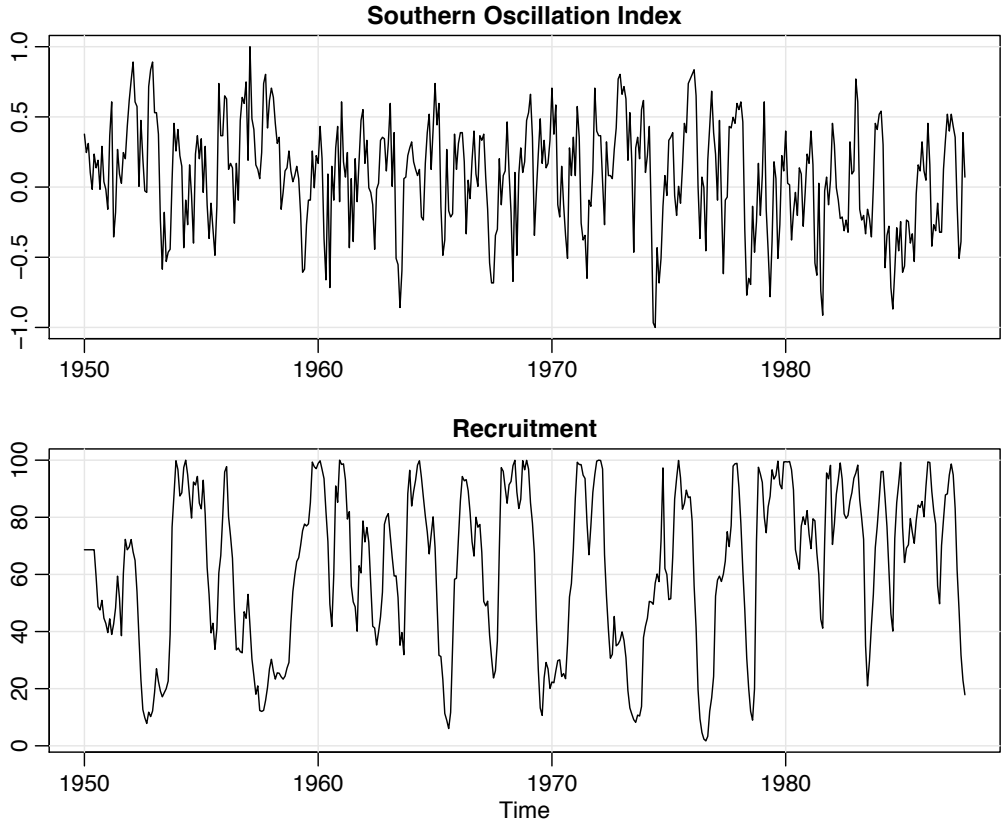

Fig. 1.5. *Monthly SOI and Recruitment (estimated new fish), 1950-1987.*

Example 1.5 El Niño and Fish Population

We may also be interested in analyzing several time series at once. Figure 1.5 shows monthly values of an environmental series called the *Southern Oscillation Index* (SOI) and associated Recruitment (number of new fish) furnished byDr. Roy Mendelssohn of the Pacific Environmental Fisheries Group (personal communication). Both series are for a period of 453 months ranging over the years 1950–1987. The SOI measures changes in air pressure, related to sea surface temperatures in the central Pacific Ocean. The central Pacific warms every three to seven years due to the El Niño effect, which has been blamed for various global extreme weather events. Both series in Figure 1.5 exhibit repetitive behavior, with regularly repeating *cycles* that are easily visible. This periodic behavior is of interest because underlying processes of interest may be regular and the rate or *frequency* of oscillation characterizing the behavior of the underlying series would help to identify them. The series show two basic oscillations types, an obvious annual cycle (hot in the summer, cold in the winter), and a slower frequency that seems to repeat about every 4 years. The study of the kinds of cycles and their strengths is the subject of Chapter 4. The two series are also related; it is easy to imagine the fish population is dependent on the ocean temperature. This possibility suggests trying some version of regression analysis as a procedure for relating the two series. *Transfer function*

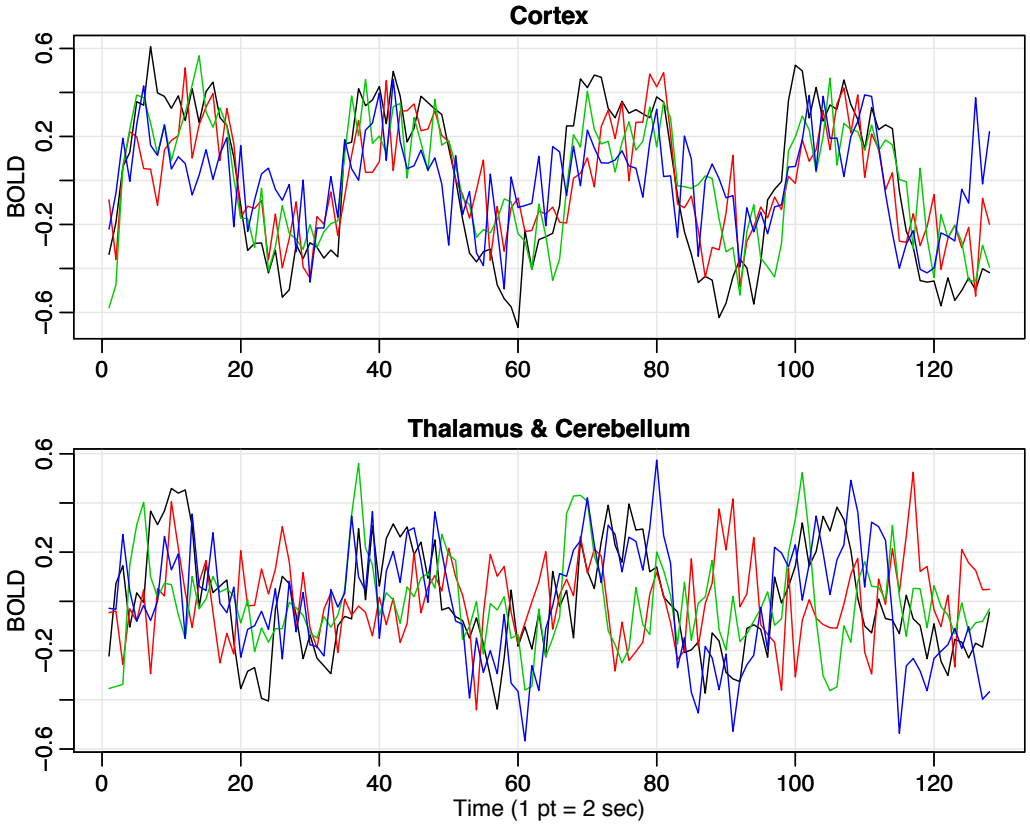

Fig. 1.6. *fMRI data from various locations in the cortex, thalamus, and cerebellum; $n = 128$ points, one observation taken every 2 seconds.*

modeling, as considered in Chapter 5, can also be applied in this case. The following R code will reproduce Figure 1.5:

```
par(mfrow = c(2,1))  # set up the graphics
plot(soi, ylab="", xlab="", main="Southern Oscillation Index")
plot(rec, ylab="", xlab="", main="Recruitment")
```

Example 1.6 fMRI Imaging

A fundamental problem in classical statistics occurs when we are given a collection of independent series or vectors of series, generated under varying experimental conditions or treatment configurations. Such a set of series is shown in Figure 1.6, where we observe data collected from various locations in the brain via functional magnetic resonance imaging (fMRI). In this example, five subjects were given periodic brushing on the hand. The stimulus was applied for 32 seconds and then stopped for 32 seconds; thus, the signal period is 64 seconds. The sampling rate was one observation every 2 seconds for 256 seconds ($n = 128$). For this example, we averaged the results over subjects (these were evoked responses, and all subjects were in phase). The series shown in Figure 1.6 are consecutive measures of blood oxygenation-level dependent (BOLD) signal intensity, which measures areas of activation in the brain. Notice that the periodicities appear strongly in the motor cortex

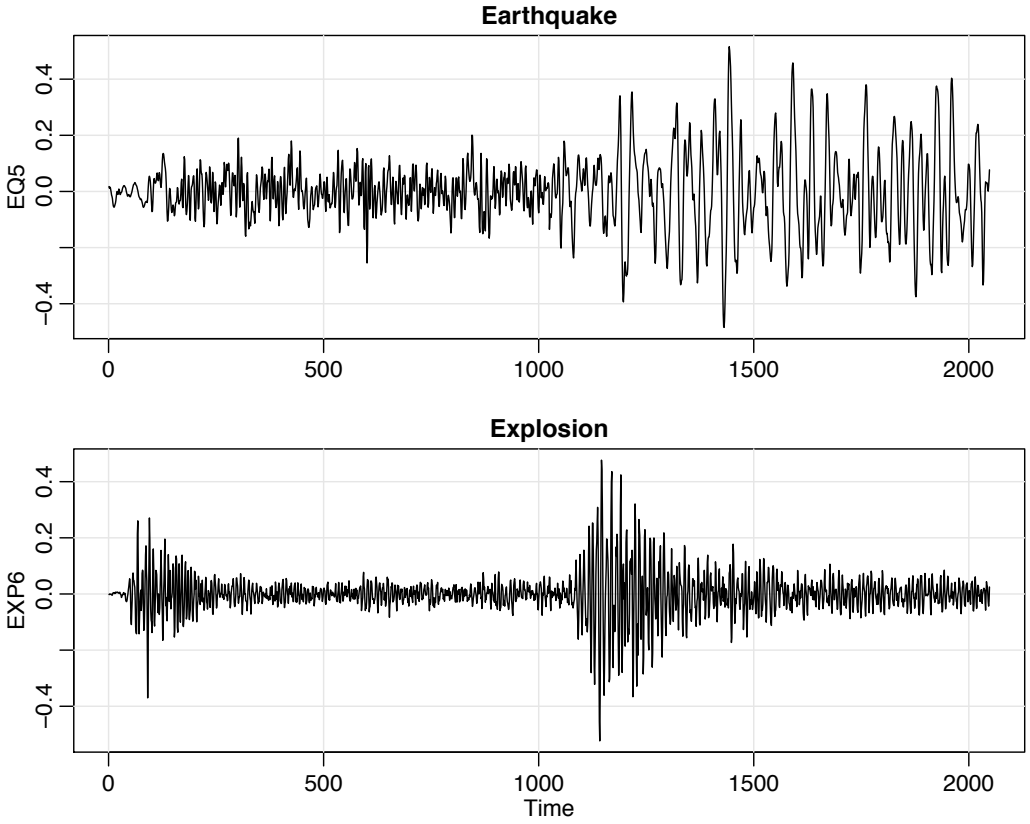

Fig. 1.7. *Arrival phases from an earthquake (top) and explosion (bottom) at 40 points per second.*

series and less strongly in the thalamus and cerebellum. The fact that one has series from different areas of the brain suggests testing whether the areas are responding differently to the brush stimulus. Analysis of variance techniques accomplish this in classical statistics, and we show in Chapter 7 how these classical techniques extend to the time series case, leading to a spectral analysis of variance. The following R commands can be used to plot the data:

```
par(mfrow=c(2,1))
ts.plot(fmri1[,2:5], col=1:4, ylab="BOLD", main="Cortex")
ts.plot(fmri1[,6:9], col=1:4, ylab="BOLD", main="Thalamus & Cerebellum")
```

Example 1.7 Earthquakes and Explosions

As a final example, the series in Figure 1.7 represent two phases or arrivals along the surface, denoted by P ($t = 1, \ldots, 1024$) and S ($t = 1025, \ldots, 2048$), at a seismic recording station. The recording instruments in Scandinavia are observing earthquakes and mining explosions with one of each shown in Figure 1.7. The general problem of interest is in distinguishing or discriminating between waveforms generated by earthquakes and those generated by explosions. Features that may be important are the rough amplitude ratios of the first phase P to the second phase S, which tend to be smaller for earthquakes than for explosions. In the case of the

two events in Figure 1.7, the ratio of maximum amplitudes appears to be somewhat less than .5 for the earthquake and about 1 for the explosion. Otherwise, note a subtle difference exists in the periodic nature of the S phase for the earthquake. We can again think about spectral analysis of variance for testing the equality of the periodic components of earthquakes and explosions. We would also like to be able to classify future P and S components from events of unknown origin, leading to the *time series discriminant analysis* developed in Chapter 7.

To plot the data as in this example, use the following commands in R:

```
par(mfrow=c(2,1))
plot(EQ5, main="Earthquake")
plot(EXP6, main="Explosion")
```

1.2 Time Series Statistical Models

The primary objective of time series analysis is to develop mathematical models that provide plausible descriptions for sample data, like that encountered in the previous section. In order to provide a statistical setting for describing the character of data that seemingly fluctuate in a random fashion over time, we assume a *time series* can be defined as a collection of random variables indexed according to the order they are obtained in time. For example, we may consider a time series as a sequence of random variables, $x_1, x_2, x_3, \ldots$, where the random variable x_1 denotes the value taken by the series at the first time point, the variable x_2 denotes the value for the second time period, x_3 denotes the value for the third time period, and so on. In general, a collection of random variables, $\{x_t\}$, indexed by t is referred to as a *stochastic process*. In this text, t will typically be discrete and vary over the integers $t = 0, \pm 1, \pm 2, \ldots$, or some subset of the integers. The observed values of a stochastic process are referred to as a *realization* of the stochastic process. Because it will be clear from the context of our discussions, we use the term *time series* whether we are referring generically to the process or to a particular realization and make no notational distinction between the two concepts.

It is conventional to display a sample time series graphically by plotting the values of the random variables on the vertical axis, or ordinate, with the time scale as the abscissa. It is usually convenient to connect the values at adjacent time periods to reconstruct visually some original hypothetical continuous time series that might have produced these values as a discrete sample. Many of the series discussed in the previous section, for example, could have been observed at any continuous point in time and are conceptually more properly treated as *continuous time series*. The approximation of these series by *discrete time parameter series* sampled at equally spaced points in time is simply an acknowledgment that sampled data will, for the most part, be discrete because of restrictions inherent in the method of collection. Furthermore, the analysis techniques are then feasible using computers, which are limited to digital computations. Theoretical developments also rest on the idea that a continuous parameter time series should be specified in terms of finite-dimensional *distribution functions* defined over a finite number of points in time. This is not to

say that the selection of the sampling interval or rate is not an extremely important consideration. The appearance of data can be changed completely by adopting an insufficient sampling rate. We have all seen wheels in movies appear to be turning backwards because of the insufficient number of frames sampled by the camera. This phenomenon leads to a distortion called *aliasing* (see Section 4.1).

The fundamental visual characteristic distinguishing the different series shown in Example 1.1–Example 1.7 is their differing degrees of smoothness. One possible explanation for this smoothness is that it is being induced by the supposition that adjacent points in time are *correlated*, so the value of the series at time t, say, x_t, depends in some way on the past values $x_{t-1}, x_{t-2}, \ldots$. This model expresses a fundamental way in which we might think about generating realistic-looking time series. To begin to develop an approach to using collections of random variables to model time series, consider Example 1.8.

Example 1.8 White Noise (3 flavors)

A simple kind of generated series might be a collection of uncorrelated random variables, w_t, with mean 0 and finite variance σ_w^2. The time series generated from uncorrelated variables is used as a model for noise in engineering applications, where it is called *white noise*; we shall denote this process as $w_t \sim \text{wn}(0, \sigma_w^2)$. The designation white originates from the analogy with white light and indicates that all possible periodic oscillations are present with equal strength.

We will sometimes require the noise to be independent and identically distributed (iid) random variables with mean 0 and variance σ_w^2. We distinguish this by writing $w_t \sim \text{iid}(0, \sigma_w^2)$ or by saying *white independent noise* or *iid noise*. A particularly useful white noise series is *Gaussian white noise*, wherein the w_t are independent normal random variables, with mean 0 and variance σ_w^2; or more succinctly, $w_t \sim \text{iid}\,\text{N}(0, \sigma_w^2)$. Figure 1.8 shows in the upper panel a collection of 500 such random variables, with $\sigma_w^2 = 1$, plotted in the order in which they were drawn. The resulting series bears a slight resemblance to the explosion in Figure 1.7 but is not smooth enough to serve as a plausible model for any of the other experimental series. The plot tends to show visually a mixture of many different kinds of oscillations in the white noise series.

If the stochastic behavior of all time series could be explained in terms of the white noise model, classical statistical methods would suffice. Two ways of introducing serial correlation and more smoothness into time series models are given in Example 1.9 and Example 1.10.

Example 1.9 Moving Averages and Filtering

We might replace the white noise series w_t by a *moving average* that smooths the series. For example, consider replacing w_t in Example 1.8 by an average of its current value and its immediate neighbors in the past and future. That is, let

$$v_t = \tfrac{1}{3}\left(w_{t-1} + w_t + w_{t+1}\right), \tag{1.1}$$

which leads to the series shown in the lower panel of Figure 1.8. Inspecting the series shows a smoother version of the first series, reflecting the fact that the slower

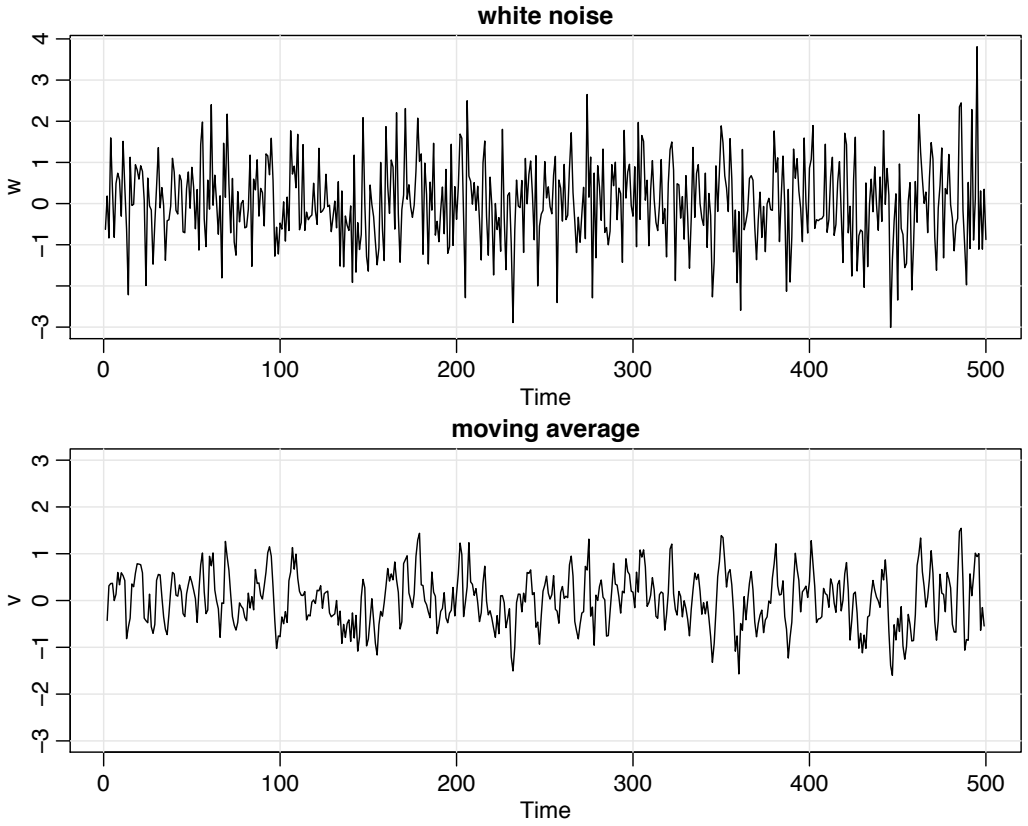

Fig. 1.8. *Gaussian white noise series (top) and three-point moving average of the Gaussian white noise series (bottom).*

oscillations are more apparent and some of the faster oscillations are taken out. We begin to notice a similarity to the SOI in Figure 1.5, or perhaps, to some of the fMRI series in Figure 1.6.

A linear combination of values in a time series such as in (1.1) is referred to, generically, as a filtered series; hence the command `filter` in the following code for Figure 1.8.

```
w = rnorm(500,0,1)                    # 500 N(0,1) variates
v = filter(w, sides=2, filter=rep(1/3,3))  # moving average
par(mfrow=c(2,1))
plot.ts(w, main="white noise")
plot.ts(v, ylim=c(-3,3), main="moving average")
```

The speech series in Figure 1.3 and the Recruitment series in Figure 1.5, as well as some of the MRI series in Figure 1.6, differ from the moving average series because one particular kind of oscillatory behavior seems to predominate, producing a sinusoidal type of behavior. A number of methods exist for generating series with this quasi-periodic behavior; we illustrate a popular one based on the autoregressive model considered in Chapter 3.

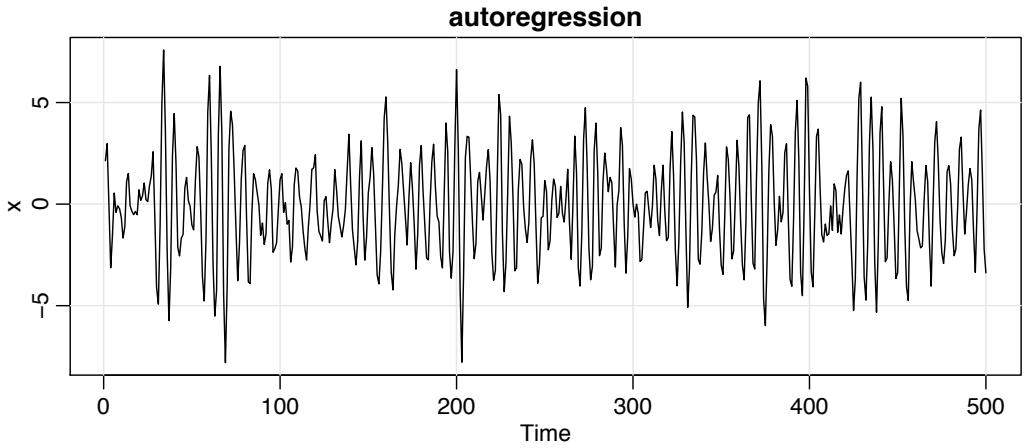

Fig. 1.9. *Autoregressive series generated from model* (1.2).

Example 1.10 Autoregressions

Suppose we consider the white noise series w_t of Example 1.8 as input and calculate the output using the second-order equation

$$x_t = x_{t-1} - .9x_{t-2} + w_t \tag{1.2}$$

successively for $t = 1, 2, \ldots, 500$. Equation (1.2) represents a regression or prediction of the current value x_t of a time series as a function of the past two values of the series, and, hence, the term *autoregression* is suggested for this model. A problem with startup values exists here because (1.2) also depends on the initial conditions x_0 and x_{-1}, but assuming we have the values, we generate the succeeding values by substituting into (1.2). The resulting output series is shown in Figure 1.9, and we note the periodic behavior of the series, which is similar to that displayed by the speech series in Figure 1.3. The autoregressive model above and its generalizations can be used as an underlying model for many observed series and will be studied in detail in Chapter 3.

As in the previous example, the data are obtained by a filter of white noise. The function `filter` uses zeros for the initial values. In this case, $x_1 = w_1$, and $x_2 = x_1 + w_2 = w_1 + w_2$, and so on, so that the values do not satisfy (1.2). An easy fix is to run the filter for longer than needed and remove the initial values.

```
w = rnorm(550,0,1)                    # 50 extra to avoid startup problems
x = filter(w, filter=c(1,-.9), method="recursive")[-(1:50)] # remove first 50
plot.ts(x, main="autoregression")
```

Example 1.11 Random Walk with Drift

A model for analyzing trend such as seen in the global temperature data in Figure 1.2, is the *random walk with drift* model given by

$$x_t = \delta + x_{t-1} + w_t \tag{1.3}$$

for $t = 1, 2, \ldots$, with initial condition $x_0 = 0$, and where w_t is white noise. The constant δ is called the *drift*, and when $\delta = 0$, (1.3) is called simply a *random walk*.

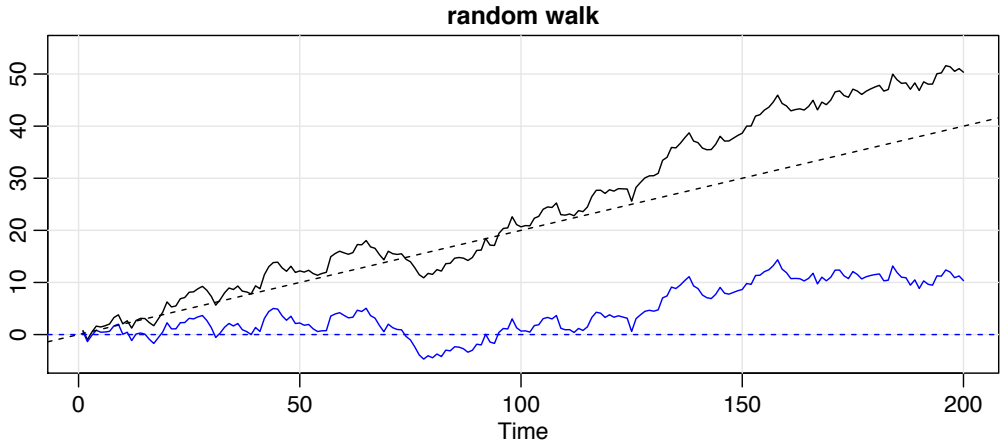

Fig. 1.10. *Random walk, $\sigma_w = 1$, with drift $\delta = .2$ (upper jagged line), without drift, $\delta = 0$ (lower jagged line), and straight (dashed) lines with slope δ.*

The term random walk comes from the fact that, when $\delta = 0$, the value of the time series at time t is the value of the series at time $t-1$ plus a completely random movement determined by w_t. Note that we may rewrite (1.3) as a cumulative sum of white noise variates. That is,

$$x_t = \delta\, t + \sum_{j=1}^{t} w_j \tag{1.4}$$

for $t = 1, 2, \ldots$; either use induction, or plug (1.4) into (1.3) to verify this statement. Figure 1.10 shows 200 observations generated from the model with $\delta = 0$ and .2, and with $\sigma_w = 1$. For comparison, we also superimposed the straight line $.2t$ on the graph. To reproduce Figure 1.10 in R use the following code (notice the use of multiple commands per line using a semicolon).

```
set.seed(154)            # so you can reproduce the results
w = rnorm(200);  x = cumsum(w)    # two commands in one line
wd = w +.2;      xd = cumsum(wd)
plot.ts(xd, ylim=c(-5,55), main="random walk", ylab='')
lines(x, col=4); abline(h=0, col=4, lty=2); abline(a=0, b=.2, lty=2)
```

Example 1.12 Signal in Noise

Many realistic models for generating time series assume an underlying signal with some consistent periodic variation, contaminated by adding a random noise. For example, it is easy to detect the regular cycle fMRI series displayed on the top of Figure 1.6. Consider the model

$$x_t = 2\cos(2\pi \tfrac{t+15}{50}) + w_t \tag{1.5}$$

for $t = 1, 2, \ldots, 500$, where the first term is regarded as the signal, shown in the upper panel of Figure 1.11. We note that a sinusoidal waveform can be written as

$$A\cos(2\pi\omega t + \phi), \tag{1.6}$$

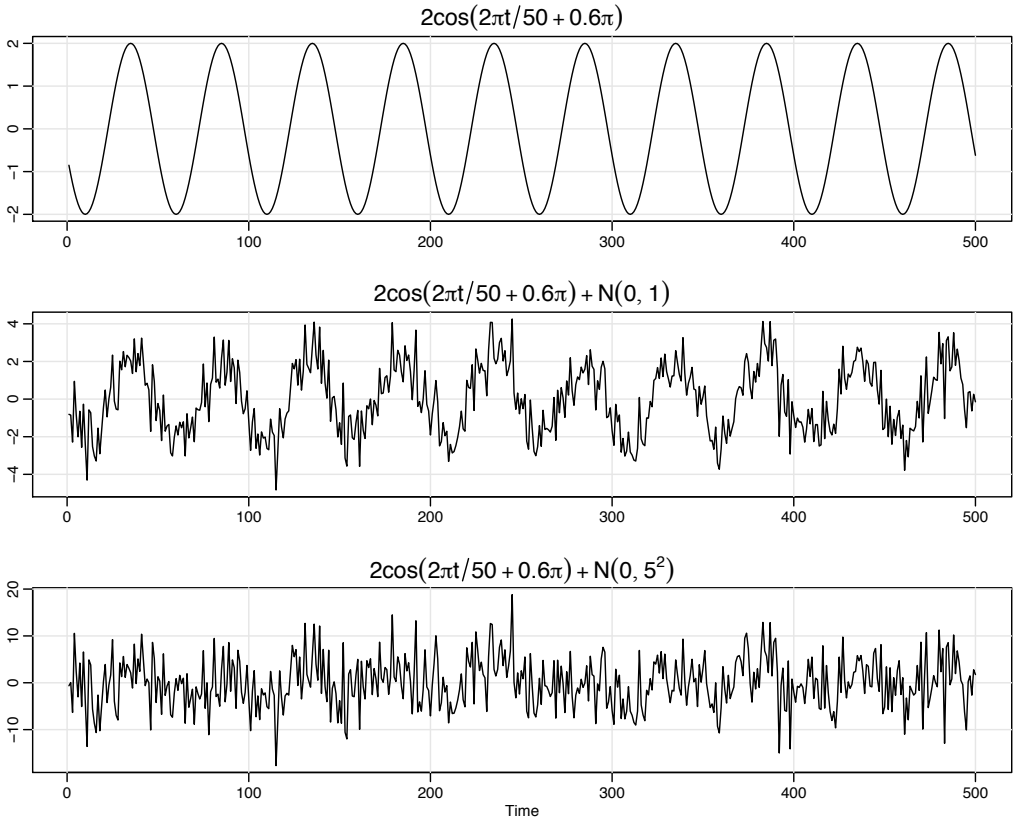

Fig. 1.11. *Cosine wave with period 50 points (top panel) compared with the cosine wave contaminated with additive white Gaussian noise, $\sigma_w = 1$ (middle panel) and $\sigma_w = 5$ (bottom panel); see (1.5).*

where A is the amplitude, ω is the frequency of oscillation, and ϕ is a phase shift. In (1.5), $A = 2$, $\omega = 1/50$ (one cycle every 50 time points), and $\phi = 2\pi 15/50 = .6\pi$.

An additive noise term was taken to be white noise with $\sigma_w = 1$ (middle panel) and $\sigma_w = 5$ (bottom panel), drawn from a normal distribution. Adding the two together obscures the signal, as shown in the lower panels of Figure 1.11. Of course, the degree to which the signal is obscured depends on the amplitude of the signal and the size of σ_w. The ratio of the amplitude of the signal to σ_w (or some function of the ratio) is sometimes called the *signal-to-noise ratio (SNR)*; the larger the SNR, the easier it is to detect the signal. Note that the signal is easily discernible in the middle panel of Figure 1.11, whereas the signal is obscured in the bottom panel. Typically, we will not observe the signal but the signal obscured by noise.

To reproduce Figure 1.11 in R, use the following commands:

```
cs = 2*cos(2*pi*1:500/50 + .6*pi);  w = rnorm(500,0,1)
par(mfrow=c(3,1), mar=c(3,2,2,1), cex.main=1.5)
plot.ts(cs, main=expression(2*cos(2*pi*t/50+.6*pi)))
plot.ts(cs+w, main=expression(2*cos(2*pi*t/50+.6*pi) + N(0,1)))
plot.ts(cs+5*w, main=expression(2*cos(2*pi*t/50+.6*pi) + N(0,25)))
```

In Chapter 4, we will study the use of *spectral analysis* as a possible technique for detecting regular or periodic signals, such as the one described in Example 1.12. In general, we would emphasize the importance of simple additive models such as given above in the form

$$x_t = s_t + v_t, \tag{1.7}$$

where s_t denotes some unknown signal and v_t denotes a time series that may be white or correlated over time. The problems of detecting a signal and then in estimating or extracting the waveform of s_t are of great interest in many areas of engineering and the physical and biological sciences. In economics, the underlying signal may be a trend or it may be a seasonal component of a series. Models such as (1.7), where the signal has an autoregressive structure, form the motivation for the state-space model of Chapter 6.

In the above examples, we have tried to motivate the use of various combinations of random variables emulating real time series data. Smoothness characteristics of observed time series were introduced by combining the random variables in various ways. Averaging independent random variables over adjacent time points, as in Example 1.9, or looking at the output of difference equations that respond to white noise inputs, as in Example 1.10, are common ways of generating correlated data. In the next section, we introduce various theoretical measures used for describing how time series behave. As is usual in statistics, the complete description involves the multivariate distribution function of the jointly sampled values $x_1, x_2, \ldots, x_n$, whereas more economical descriptions can be had in terms of the mean and autocorrelation functions. Because correlation is an essential feature of time series analysis, the most useful descriptive measures are those expressed in terms of covariance and correlation functions.

1.3 Measures of Dependence

A complete description of a time series, observed as a collection of n random variables at arbitrary time points $t_1, t_2, \ldots, t_n$, for any positive integer n, is provided by the joint distribution function, evaluated as the probability that the values of the series are jointly less than the n constants, $c_1, c_2, \ldots, c_n$; i.e.,

$$F_{t_1,t_2,\ldots,t_n}(c_1, c_2, \ldots, c_n) = \Pr\big(x_{t_1} \leq c_1, x_{t_2} \leq c_2, \ldots, x_{t_n} \leq c_n\big). \tag{1.8}$$

Unfortunately, these multidimensional distribution functions cannot usually be written easily unless the random variables are jointly normal, in which case the joint density has the well-known form displayed in (1.33).

Although the joint distribution function describes the data completely, it is an unwieldy tool for displaying and analyzing time series data. The distribution function (1.8) must be evaluated as a function of n arguments, so any plotting of the corresponding multivariate density functions is virtually impossible. The marginal distribution functions

$$F_t(x) = P\{x_t \leq x\}$$

or the corresponding marginal density functions

$$f_t(x) = \frac{\partial F_t(x)}{\partial x},$$

when they exist, are often informative for examining the marginal behavior of a series.[1.3] Another informative marginal descriptive measure is the mean function.

Definition 1.1 *The* **mean function** *is defined as*

$$\mu_{xt} = \mathrm{E}(x_t) = \int_{-\infty}^{\infty} x f_t(x)\, dx, \tag{1.9}$$

provided it exists, where E *denotes the usual expected value operator. When no confusion exists about which time series we are referring to, we will drop a subscript and write* μ_{xt} *as* μ_t.

Example 1.13 Mean Function of a Moving Average Series

If w_t denotes a white noise series, then $\mu_{wt} = \mathrm{E}(w_t) = 0$ for all t. The top series in Figure 1.8 reflects this, as the series clearly fluctuates around a mean value of zero. Smoothing the series as in Example 1.9 does not change the mean because we can write

$$\mu_{vt} = \mathrm{E}(v_t) = \tfrac{1}{3}[\mathrm{E}(w_{t-1}) + \mathrm{E}(w_t) + \mathrm{E}(w_{t+1})] = 0.$$

Example 1.14 Mean Function of a Random Walk with Drift

Consider the random walk with drift model given in (1.4),

$$x_t = \delta\, t + \sum_{j=1}^{t} w_j, \qquad t = 1, 2, \ldots .$$

Because $\mathrm{E}(w_t) = 0$ for all t, and δ is a constant, we have

$$\mu_{xt} = \mathrm{E}(x_t) = \delta\, t + \sum_{j=1}^{t} \mathrm{E}(w_j) = \delta\, t$$

which is a straight line with slope δ. A realization of a random walk with drift can be compared to its mean function in Figure 1.10.

[1.3] If x_t is Gaussian with mean μ_t and variance σ_t^2, abbreviated as $x_t \sim \mathrm{N}(\mu_t, \sigma_t^2)$, the marginal density is given by $f_t(x) = \dfrac{1}{\sigma_t\sqrt{2\pi}} \exp\left\{-\frac{1}{2\sigma_t^2}(x-\mu_t)^2\right\}$, $x \in \mathbb{R}$.

Example 1.15 Mean Function of Signal Plus Noise

A great many practical applications depend on assuming the observed data have been generated by a fixed signal waveform superimposed on a zero-mean noise process, leading to an additive signal model of the form (1.5). It is clear, because the signal in (1.5) is a fixed function of time, we will have

$$\begin{aligned}\mu_{xt} &= \mathrm{E}(x_t) = \mathrm{E}\left[2\cos(2\pi\tfrac{t+15}{50}) + w_t\right] \\ &= 2\cos(2\pi\tfrac{t+15}{50}) + \mathrm{E}(w_t) \\ &= 2\cos(2\pi\tfrac{t+15}{50}),\end{aligned}$$

and the mean function is just the cosine wave.

The lack of independence between two adjacent values x_s and x_t can be assessed numerically, as in classical statistics, using the notions of covariance and correlation. Assuming the variance of x_t is finite, we have the following definition.

Definition 1.2 *The* **autocovariance function** *is defined as the second moment product*

$$\gamma_x(s,t) = \mathrm{cov}(x_s, x_t) = \mathrm{E}[(x_s - \mu_s)(x_t - \mu_t)], \tag{1.10}$$

for all s and t. When no possible confusion exists about which time series we are referring to, we will drop the subscript and write $\gamma_x(s,t)$ as $\gamma(s,t)$. Note that $\gamma_x(s,t) = \gamma_x(t,s)$ for all time points s and t.

The autocovariance measures the *linear* dependence between two points on the same series observed at different times. Very smooth series exhibit autocovariance functions that stay large even when the t and s are far apart, whereas choppy series tend to have autocovariance functions that are nearly zero for large separations. Recall from classical statistics that if $\gamma_x(s,t) = 0$, x_s and x_t are not linearly related, but there still may be some dependence structure between them. If, however, x_s and x_t are bivariate normal, $\gamma_x(s,t) = 0$ ensures their independence. It is clear that, for $s = t$, the autocovariance reduces to the (assumed finite) *variance*, because

$$\gamma_x(t,t) = \mathrm{E}[(x_t - \mu_t)^2] = \mathrm{var}(x_t). \tag{1.11}$$

Example 1.16 Autocovariance of White Noise

The white noise series w_t has $\mathrm{E}(w_t) = 0$ and

$$\gamma_w(s,t) = \mathrm{cov}(w_s, w_t) = \begin{cases}\sigma_w^2 & s = t, \\ 0 & s \neq t.\end{cases} \tag{1.12}$$

A realization of white noise with $\sigma_w^2 = 1$ is shown in the top panel of Figure 1.8.

We often have to calculate the autocovariance between filtered series. A useful result is given in the following proposition.

Property 1.1 Covariance of Linear Combinations

If the random variables

$$U = \sum_{j=1}^{m} a_j X_j \quad \text{and} \quad V = \sum_{k=1}^{r} b_k Y_k$$

are linear combinations of (finite variance) random variables $\{X_j\}$ *and* $\{Y_k\}$*, respectively, then*

$$\operatorname{cov}(U, V) = \sum_{j=1}^{m} \sum_{k=1}^{r} a_j b_k \operatorname{cov}(X_j, Y_k). \tag{1.13}$$

Furthermore, $\operatorname{var}(U) = \operatorname{cov}(U, U)$.

Example 1.17 Autocovariance of a Moving Average

Consider applying a three-point moving average to the white noise series w_t of the previous example as in Example 1.9. In this case,

$$\gamma_v(s,t) = \operatorname{cov}(v_s, v_t) = \operatorname{cov}\left\{\tfrac{1}{3}(w_{s-1} + w_s + w_{s+1}), \tfrac{1}{3}(w_{t-1} + w_t + w_{t+1})\right\}.$$

When $s = t$ we have

$$\begin{aligned}\gamma_v(t,t) &= \tfrac{1}{9}\operatorname{cov}\{(w_{t-1} + w_t + w_{t+1}), (w_{t-1} + w_t + w_{t+1})\} \\ &= \tfrac{1}{9}[\operatorname{cov}(w_{t-1}, w_{t-1}) + \operatorname{cov}(w_t, w_t) + \operatorname{cov}(w_{t+1}, w_{t+1})] \\ &= \tfrac{3}{9}\sigma_w^2.\end{aligned}$$

When $s = t + 1$,

$$\begin{aligned}\gamma_v(t+1,t) &= \tfrac{1}{9}\operatorname{cov}\{(w_t + w_{t+1} + w_{t+2}), (w_{t-1} + w_t + w_{t+1})\} \\ &= \tfrac{1}{9}[\operatorname{cov}(w_t, w_t) + \operatorname{cov}(w_{t+1}, w_{t+1})] \\ &= \tfrac{2}{9}\sigma_w^2,\end{aligned}$$

using (1.12). Similar computations give $\gamma_v(t-1,t) = 2\sigma_w^2/9$, $\gamma_v(t+2,t) = \gamma_v(t-2,t) = \sigma_w^2/9$, and 0 when $|t-s| > 2$. We summarize the values for all s and t as

$$\gamma_v(s,t) = \begin{cases} \frac{3}{9}\sigma_w^2 & s = t, \\ \frac{2}{9}\sigma_w^2 & |s-t| = 1, \\ \frac{1}{9}\sigma_w^2 & |s-t| = 2, \\ 0 & |s-t| > 2. \end{cases} \tag{1.14}$$

Example 1.17 shows clearly that the smoothing operation introduces a covariance function that decreases as the separation between the two time points increases and disappears completely when the time points are separated by three or more time points. This particular autocovariance is interesting because it only depends on the time separation or *lag* and not on the absolute location of the points along the series. We shall see later that this dependence suggests a mathematical model for the concept of *weak stationarity*.

Example 1.18 Autocovariance of a Random Walk
For the random walk model, $x_t = \sum_{j=1}^{t} w_j$, we have

$$\gamma_x(s,t) = \text{cov}(x_s, x_t) = \text{cov}\left(\sum_{j=1}^{s} w_j, \sum_{k=1}^{t} w_k\right) = \min\{s,t\}\,\sigma_w^2,$$

because the w_t are uncorrelated random variables. Note that, as opposed to the previous examples, the autocovariance function of a random walk depends on the particular time values s and t, and not on the time separation or lag. Also, notice that the variance of the random walk, $\text{var}(x_t) = \gamma_x(t,t) = t\,\sigma_w^2$, increases without bound as time t increases. The effect of this variance increase can be seen in Figure 1.10 where the processes start to move away from their mean functions $\delta\, t$ (note that $\delta = 0$ and .2 in that example).

As in classical statistics, it is more convenient to deal with a measure of association between -1 and 1, and this leads to the following definition.

Definition 1.3 *The* **autocorrelation function (ACF)** *is defined as*

$$\rho(s,t) = \frac{\gamma(s,t)}{\sqrt{\gamma(s,s)\gamma(t,t)}}. \tag{1.15}$$

The ACF measures the linear predictability of the series at time t, say x_t, using only the value x_s. We can show easily that $-1 \le \rho(s,t) \le 1$ using the Cauchy–Schwarz inequality.[1.4] If we can predict x_t *perfectly* from x_s through a linear relationship, $x_t = \beta_0 + \beta_1 x_s$, then the correlation will be $+1$ when $\beta_1 > 0$, and -1 when $\beta_1 < 0$. Hence, we have a rough measure of the ability to forecast the series at time t from the value at time s.

Often, we would like to measure the predictability of another series y_t from the series x_s. Assuming both series have finite variances, we have the following definition.

Definition 1.4 *The* **cross-covariance function** *between two series, x_t and y_t, is*

$$\gamma_{xy}(s,t) = \text{cov}(x_s, y_t) = \text{E}[(x_s - \mu_{xs})(y_t - \mu_{yt})]. \tag{1.16}$$

There is also a scaled version of the cross-covariance function.

Definition 1.5 *The* **cross-correlation function (CCF)** *is given by*

$$\rho_{xy}(s,t) = \frac{\gamma_{xy}(s,t)}{\sqrt{\gamma_x(s,s)\gamma_y(t,t)}}. \tag{1.17}$$

[1.4] The Cauchy–Schwarz inequality implies $|\gamma(s,t)|^2 \le \gamma(s,s)\gamma(t,t)$.

We may easily extend the above ideas to the case of more than two series, say, $x_{t1}, x_{t2}, \ldots, x_{tr}$; that is, *multivariate time series* with r components. For example, the extension of (1.10) in this case is

$$\gamma_{jk}(s,t) = \mathrm{E}[(x_{sj} - \mu_{sj})(x_{tk} - \mu_{tk})] \qquad j,k = 1,2,\ldots,r. \tag{1.18}$$

In the definitions above, the autocovariance and cross-covariance functions may change as one moves along the series because the values depend on both s and t, the locations of the points in time. In Example 1.17, the autocovariance function depends on the separation of x_s and x_t, say, $h = |s-t|$, and not on where the points are located in time. As long as the points are separated by h units, the location of the two points does not matter. This notion, called *weak stationarity*, when the mean is constant, is fundamental in allowing us to analyze sample time series data when only a single series is available.

1.4 Stationary Time Series

The preceding definitions of the mean and autocovariance functions are completely general. Although we have not made any special assumptions about the behavior of the time series, many of the preceding examples have hinted that a sort of regularity may exist over time in the behavior of a time series. We introduce the notion of regularity using a concept called *stationarity*.

Definition 1.6 *A* **strictly stationary** *time series is one for which the probabilistic behavior of every collection of values*

$$\{x_{t_1}, x_{t_2}, \ldots, x_{t_k}\}$$

is identical to that of the time shifted set

$$\{x_{t_1+h}, x_{t_2+h}, \ldots, x_{t_k+h}\}.$$

That is,

$$\Pr\{x_{t_1} \le c_1, \ldots, x_{t_k} \le c_k\} = \Pr\{x_{t_1+h} \le c_1, \ldots, x_{t_k+h} \le c_k\} \tag{1.19}$$

for all $k = 1, 2, \ldots$, all time points $t_1, t_2, \ldots, t_k$, all numbers $c_1, c_2, \ldots, c_k$, and all time shifts $h = 0, \pm 1, \pm 2, \ldots$.

If a time series is strictly stationary, then all of the multivariate distribution functions for subsets of variables must agree with their counterparts in the shifted set for all values of the shift parameter h. For example, when $k = 1$, (1.19) implies that

$$\Pr\{x_s \le c\} = \Pr\{x_t \le c\} \tag{1.20}$$

for any time points s and t. This statement implies, for example, that the probability the value of a time series sampled hourly is negative at 1 AM is the same as at 10 AM.

In addition, if the mean function, μ_t, of the series exists, (1.20) implies that $\mu_s = \mu_t$ for all s and t, and hence μ_t must be constant. Note, for example, that a random walk process with drift is *not* strictly stationary because its mean function changes with time; see Example 1.14.

When $k = 2$, we can write (1.19) as

$$\Pr\{x_s \le c_1, x_t \le c_2\} = \Pr\{x_{s+h} \le c_1, x_{t+h} \le c_2\} \tag{1.21}$$

for any time points s and t and shift h. Thus, if the variance function of the process exists, (1.20)–(1.21) imply that the autocovariance function of the series x_t satisfies

$$\gamma(s,t) = \gamma(s+h, t+h)$$

for all s and t and h. We may interpret this result by saying the autocovariance function of the process depends only on the time difference between s and t, and not on the actual times.

The version of stationarity in Definition 1.6 is too strong for most applications. Moreover, it is difficult to assess strict stationarity from a single data set. Rather than imposing conditions on all possible distributions of a time series, we will use a milder version that imposes conditions only on the first two moments of the series. We now have the following definition.

Definition 1.7 *A* **weakly stationary** *time series, x_t, is a finite variance process such that*

(i) the mean value function, μ_t, defined in (1.9) is constant and does not depend on time t, and

(ii) the autocovariance function, $\gamma(s,t)$, defined in (1.10) depends on s and t only through their difference $|s-t|$.

Henceforth, we will use the term **stationary** *to mean weakly stationary; if a process is stationary in the strict sense, we will use the term strictly stationary.*

Stationarity requires regularity in the mean and autocorrelation functions so that these quantities (at least) may be estimated by averaging. It should be clear from the discussion of strict stationarity following Definition 1.6 that a strictly stationary, finite variance, time series is also stationary. The converse is not true unless there are further conditions. One important case where stationarity implies strict stationarity is if the time series is Gaussian [meaning all finite distributions, (1.19), of the series are Gaussian]. We will make this concept more precise at the end of this section.

Because the mean function, $\mathrm{E}(x_t) = \mu_t$, of a stationary time series is independent of time t, we will write

$$\mu_t = \mu. \tag{1.22}$$

Also, because the autocovariance function, $\gamma(s,t)$, of a stationary time series, x_t, depends on s and t only through their difference $|s-t|$, we may simplify the notation. Let $s = t + h$, where h represents the time shift or *lag*. Then

$$\gamma(t+h, t) = \mathrm{cov}(x_{t+h}, x_t) = \mathrm{cov}(x_h, x_0) = \gamma(h, 0)$$

because the time difference between times $t+h$ and t is the same as the time difference between times h and 0. Thus, the autocovariance function of a stationary time series does not depend on the time argument t. Henceforth, for convenience, we will drop the second argument of $\gamma(h, 0)$.

Definition 1.8 *The* **autocovariance function of a stationary time series** *will be written as*

$$\gamma(h) = \text{cov}(x_{t+h}, x_t) = \text{E}[(x_{t+h} - \mu)(x_t - \mu)]. \tag{1.23}$$

Definition 1.9 *The* **autocorrelation function (ACF) of a stationary time series** *will be written using* (1.15) *as*

$$\rho(h) = \frac{\gamma(t+h, t)}{\sqrt{\gamma(t+h, t+h)\gamma(t, t)}} = \frac{\gamma(h)}{\gamma(0)}. \tag{1.24}$$

The Cauchy–Schwarz inequality shows again that $-1 \leq \rho(h) \leq 1$ for all h, enabling one to assess the relative importance of a given autocorrelation value by comparing with the extreme values -1 and 1.

Example 1.19 Stationarity of White Noise

The mean and autocovariance functions of the white noise series discussed in Example 1.8 and Example 1.16 are easily evaluated as $\mu_{wt} = 0$ and

$$\gamma_w(h) = \text{cov}(w_{t+h}, w_t) = \begin{cases} \sigma_w^2 & h = 0, \\ 0 & h \neq 0. \end{cases}$$

Thus, white noise satisfies the conditions of Definition 1.7 and is weakly stationary or stationary. If the white noise variates are also normally distributed or Gaussian, the series is also strictly stationary, as can be seen by evaluating (1.19) using the fact that the noise would also be iid. The autocorrelation function is given by $\rho_w(0) = 1$ and $\rho(h) = 0$ for $h \neq 0$.

Example 1.20 Stationarity of a Moving Average

The three-point moving average process of Example 1.9 is stationary because, from Example 1.13 and Example 1.17, the mean and autocovariance functions $\mu_{vt} = 0$, and

$$\gamma_v(h) = \begin{cases} \frac{3}{9}\sigma_w^2 & h = 0, \\ \frac{2}{9}\sigma_w^2 & h = \pm 1, \\ \frac{1}{9}\sigma_w^2 & h = \pm 2, \\ 0 & |h| > 2 \end{cases}$$

are independent of time t, satisfying the conditions of Definition 1.7.

The autocorrelation function is given by

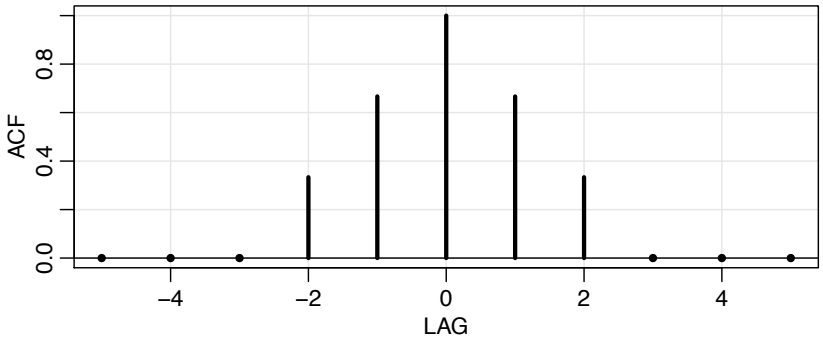

Fig. 1.12. *Autocorrelation function of a three-point moving average.*

$$\rho_v(h) = \begin{cases} 1 & h = 0, \\ \frac{2}{3} & h = \pm 1, \\ \frac{1}{3} & h = \pm 2, \\ 0 & |h| > 2. \end{cases}$$

Figure 1.12 shows a plot of the autocorrelations as a function of lag h. Note that the ACF is symmetric about lag zero.

Example 1.21 A Random Walk is Not Stationary

A random walk is not stationary because its autocovariance function, $\gamma(s,t) = \min\{s,t\}\sigma_w^2$, depends on time; see Example 1.18 and Problem 1.8. Also, the random walk with drift violates both conditions of Definition 1.7 because, as shown in Example 1.14, the mean function, $\mu_{xt} = \delta t$, is also a function of time t.

Example 1.22 Trend Stationarity

For example, if $x_t = \alpha + \beta t + y_t$, where y_t is stationary, then the mean function is $\mu_{x,t} = \mathrm{E}(x_t) = \alpha + \beta t + \mu_y$, which is not independent of time. Therefore, the process is not stationary. The autocovariance function, however, is independent of time, because $\gamma_x(h) = \mathrm{cov}(x_{t+h}, x_t) = \mathrm{E}[(x_{t+h} - \mu_{x,t+h})(x_t - \mu_{x,t})] = \mathrm{E}[(y_{t+h} - \mu_y)(y_t - \mu_y)] = \gamma_y(h)$. Thus, the model may be considered as having stationary behavior around a linear trend; this behavior is sometimes called *trend stationarity*. An example of such a process is the price of chicken series displayed in Figure 2.1.

The autocovariance function of a stationary process has several special properties. First, $\gamma(h)$ is *non-negative definite* (see Problem 1.25) ensuring that variances of linear combinations of the variates x_t will never be negative. That is, for any $n \geq 1$, and constants $a_1, \ldots, a_n$,

$$0 \leq \mathrm{var}(a_1 x_1 + \cdots + a_n x_n) = \sum_{j=1}^{n} \sum_{k=1}^{n} a_j a_k \gamma(j-k), \tag{1.25}$$

using Property 1.1. Also, the value at $h = 0$, namely

$$\gamma(0) = \mathrm{E}[(x_t - \mu)^2] \tag{1.26}$$

is the variance of the time series and the Cauchy–Schwarz inequality implies

$$|\gamma(h)| \le \gamma(0).$$

A final useful property, noted in a previous example, is that the autocovariance function of a stationary series is symmetric around the origin; that is,

$$\gamma(h) = \gamma(-h) \tag{1.27}$$

for all h. This property follows because

$$\gamma((t+h) - t) = \mathrm{cov}(x_{t+h}, x_t) = \mathrm{cov}(x_t, x_{t+h}) = \gamma(t - (t+h)),$$

which shows how to use the notation as well as proving the result.

When several series are available, a notion of stationarity still applies with additional conditions.

Definition 1.10 *Two time series, say, x_t and y_t, are said to be* **jointly stationary** *if they are each stationary, and the* cross-covariance function

$$\gamma_{xy}(h) = \mathrm{cov}(x_{t+h}, y_t) = \mathrm{E}[(x_{t+h} - \mu_x)(y_t - \mu_y)] \tag{1.28}$$

is a function only of lag h.

Definition 1.11 *The* **cross-correlation function (CCF)** *of jointly stationary time series x_t and y_t is defined as*

$$\rho_{xy}(h) = \frac{\gamma_{xy}(h)}{\sqrt{\gamma_x(0)\gamma_y(0)}}. \tag{1.29}$$

Again, we have the result $-1 \le \rho_{xy}(h) \le 1$ which enables comparison with the extreme values -1 and 1 when looking at the relation between x_{t+h} and y_t. The cross-correlation function is not generally symmetric about zero, i.e., typically $\rho_{xy}(h) \ne \rho_{xy}(-h)$. This is an important concept; it should be clear that $\mathrm{cov}(x_2, y_1)$ and $\mathrm{cov}(x_1, y_2)$ need not be the same. It is the case, however, that

$$\rho_{xy}(h) = \rho_{yx}(-h), \tag{1.30}$$

which can be shown by manipulations similar to those used to show (1.27).

Example 1.23 Joint Stationarity

Consider the two series, x_t and y_t, formed from the sum and difference of two successive values of a white noise process, say,

$$x_t = w_t + w_{t-1} \quad \text{and} \quad y_t = w_t - w_{t-1},$$

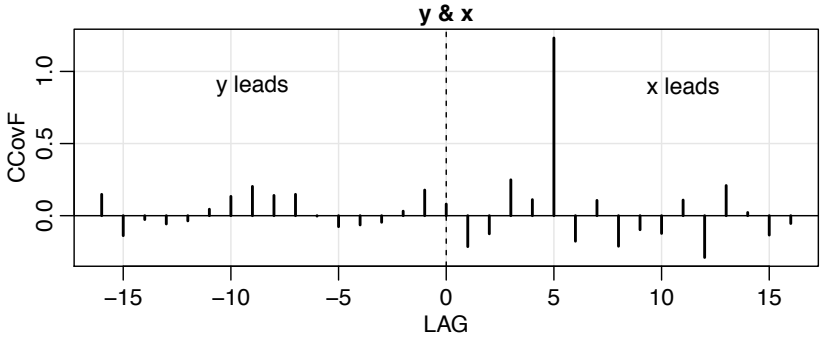

Fig. 1.13. *Demonstration of the results of Example 1.24 when $\ell = 5$. The title shows which side leads.*

where w_t are independent random variables with zero means and variance σ_w^2. It is easy to show that $\gamma_x(0) = \gamma_y(0) = 2\sigma_w^2$ and $\gamma_x(1) = \gamma_x(-1) = \sigma_w^2, \gamma_y(1) = \gamma_y(-1) = -\sigma_w^2$. Also,

$$\gamma_{xy}(1) = \operatorname{cov}(x_{t+1}, y_t) = \operatorname{cov}(w_{t+1} + w_t, w_t - w_{t-1}) = \sigma_w^2$$

because only one term is nonzero. Similarly, $\gamma_{xy}(0) = 0, \gamma_{xy}(-1) = -\sigma_w^2$. We obtain, using (1.29),

$$\rho_{xy}(h) = \begin{cases} 0 & h = 0, \\ 1/2 & h = 1, \\ -1/2 & h = -1, \\ 0 & |h| \geq 2. \end{cases}$$

Clearly, the autocovariance and cross-covariance functions depend only on the lag separation, h, so the series are jointly stationary.

Example 1.24 Prediction Using Cross-Correlation

As a simple example of cross-correlation, consider the problem of determining possible leading or lagging relations between two series x_t and y_t. If the model

$$y_t = Ax_{t-\ell} + w_t$$

holds, the series x_t is said to *lead* y_t for $\ell > 0$ and is said to *lag* y_t for $\ell < 0$. Hence, the analysis of leading and lagging relations might be important in predicting the value of y_t from x_t. Assuming that the noise w_t is uncorrelated with the x_t series, the cross-covariance function can be computed as

$$\begin{aligned} \gamma_{yx}(h) &= \operatorname{cov}(y_{t+h}, x_t) = \operatorname{cov}(Ax_{t+h-\ell} + w_{t+h}, x_t) \\ &= \operatorname{cov}(Ax_{t+h-\ell}, x_t) = A\gamma_x(h - \ell). \end{aligned}$$

Since (Cauchy–Schwarz) the largest absolute value of $\gamma_x(h - \ell)$ is $\gamma_x(0)$, i.e., when $h = \ell$, the cross-covariance function will look like the autocovariance of the input series x_t, and it will have a peak on the positive side if x_t leads y_t and a peak on the negative side if x_t lags y_t. Below is the R code of an example where x_t is white noise, $\ell = 5$, and with $\hat{\gamma}_{yx}(h)$ shown in Figure 1.13.

```
x = rnorm(100)
y = lag(x, -5) + rnorm(100)
ccf(y, x, ylab='CCovF', type='covariance')
```

The concept of weak stationarity forms the basis for much of the analysis performed with time series. The fundamental properties of the mean and autocovariance functions (1.22) and (1.23) are satisfied by many theoretical models that appear to generate plausible sample realizations. In Example 1.9 and Example 1.10, two series were generated that produced stationary looking realizations, and in Example 1.20, we showed that the series in Example 1.9 was, in fact, weakly stationary. Both examples are special cases of the so-called linear process.

Definition 1.12 *A* **linear process**, x_t, *is defined to be a linear combination of white noise variates* w_t, *and is given by*

$$x_t = \mu + \sum_{j=-\infty}^{\infty} \psi_j w_{t-j}, \qquad \sum_{j=-\infty}^{\infty} |\psi_j| < \infty. \tag{1.31}$$

For the linear process (see Problem 1.11), we may show that the autocovariance function is given by

$$\gamma_x(h) = \sigma_w^2 \sum_{j=-\infty}^{\infty} \psi_{j+h}\psi_j \tag{1.32}$$

for $h \geq 0$; recall that $\gamma_x(-h) = \gamma_x(h)$. This method exhibits the autocovariance function of the process in terms of the lagged products of the coefficients. We only need $\sum_{j=-\infty}^{\infty} \psi_j^2 < \infty$ for the process to have finite variance, but we will discuss this further in Chapter 5. Note that, for Example 1.9, we have $\psi_0 = \psi_{-1} = \psi_1 = 1/3$ and the result in Example 1.20 comes out immediately. The autoregressive series in Example 1.10 can also be put in this form, as can the general autoregressive moving average processes considered in Chapter 3.

Notice that the linear process (1.31) is dependent on the future ($j < 0$), the present ($j = 0$), and the past ($j > 0$). For the purpose of forecasting, a future dependent model will be useless. Consequently, we will focus on processes that do not depend on the future. Such models are called *causal*, and a causal linear process has $\psi_j = 0$ for $j < 0$; we will discuss this further in Chapter 3.

Finally, as previously mentioned, an important case in which a weakly stationary series is also strictly stationary is the normal or Gaussian series.

Definition 1.13 *A process,* $\{x_t\}$, *is said to be a* **Gaussian process** *if the n-dimensional vectors* $x = (x_{t_1}, x_{t_2}, \ldots, x_{t_n})'$, *for every collection of distinct time points* $t_1, t_2, \ldots, t_n$, *and every positive integer n, have a multivariate normal distribution.*

Defining the $n \times 1$ mean vector $\mathrm{E}(x) \equiv \mu = (\mu_{t_1}, \mu_{t_2}, \ldots, \mu_{t_n})'$ and the $n \times n$ covariance matrix as $\mathrm{var}(x) \equiv \Gamma = \{\gamma(t_i, t_j);\ i, j = 1, \ldots, n\}$, which is assumed to be positive definite, the multivariate normal density function can be written as

$$f(x) = (2\pi)^{-n/2}|\Gamma|^{-1/2} \exp\left\{-\frac{1}{2}(x-\mu)'\Gamma^{-1}(x-\mu)\right\}, \tag{1.33}$$

for $x \in \mathbb{R}^n$, where $|\cdot|$ denotes the determinant.

We list some important items regarding linear and Gaussian processes.

- If a Gaussian time series, $\{x_t\}$, is weakly stationary, then μ_t is constant and $\gamma(t_i, t_j) = \gamma(|t_i - t_j|)$, so that the vector μ and the matrix Γ are independent of time. These facts imply that all the finite distributions, (1.33), of the series $\{x_t\}$ depend only on time lag and not on the actual times, and hence the series must be strictly stationary. In a sense, weak stationarity and normality go hand-in-hand in that we will base our analyses on the idea that it is enough for the first two moments to behave nicely. We use the multivariate normal density in the form given above as well as in a modified version, applicable to complex random variables throughout the text.
- A result called the *Wold Decomposition* (Theorem B.5) states that a stationary non-deterministic time series is a causal linear process (but with $\sum \psi_j^2 < \infty$). A linear process need not be Gaussian, but if a time series is Gaussian, then it is a causal linear process with $w_t \sim$ iid $\mathrm{N}(0, \sigma_w^2)$. Hence, stationary Gaussian processes form the basis of modeling many time series.
- It is not enough for the marginal distributions to be Gaussian for the process to be Gaussian. It is easy to construct a situation where X and Y are normal, but (X, Y) is not bivariate normal; e.g., let X and Z be independent normals and let $Y = Z$ if $XZ > 0$ and $Y = -Z$ if $XZ \leq 0$.

1.5 Estimation of Correlation

Although the theoretical autocorrelation and cross-correlation functions are useful for describing the properties of certain hypothesized models, most of the analyses must be performed using sampled data. This limitation means the sampled points $x_1, x_2, \ldots, x_n$ only are available for estimating the mean, autocovariance, and autocorrelation functions. From the point of view of classical statistics, this poses a problem because we will typically not have iid copies of x_t that are available for estimating the covariance and correlation functions. In the usual situation with only one realization, however, the assumption of stationarity becomes critical. Somehow, we must use averages over this single realization to estimate the population means and covariance functions.

Accordingly, if a time series is stationary, the mean function (1.22) $\mu_t = \mu$ is constant so that we can estimate it by the *sample mean*,

$$\bar{x} = \frac{1}{n} \sum_{t=1}^{n} x_t. \tag{1.34}$$

In our case, $\mathrm{E}(\bar{x}) = \mu$, and the standard error of the estimate is the square root of $\mathrm{var}(\bar{x})$, which can be computed using first principles (recall Property 1.1), and is given by