

**Introduction to
Time Series and
Forecasting,
Second Edition**

*Peter J. Brockwell
Richard A. Davis*

Springer

Springer Texts in Statistics

Advisors:

George Casella Stephen Fienberg Ingram Olkin

Springer

New York

Berlin

Heidelberg

Barcelona

Hong Kong

London

Milan

Paris

Singapore

Tokyo

This page intentionally left blank

Peter J. Brockwell Richard A. Davis

Introduction to Time Series and Forecasting

Second Edition

With 126 Illustrations



Includes CD-ROM



Springer

Peter J. Brockwell
Department of Statistics
Colorado State University
Fort Collins, CO 80523
USA
pjbrock@stat.colostate.edu

Richard A. Davis
Department of Statistics
Colorado State University
Fort Collins, CO 80523
USA
rdavis@stat.colostate.edu

Editorial Board

George Casella
Department of Statistics
Griffin-Floyd Hall
University of Florida
P.O. Box 118545
Gainesville, FL 32611-8545
USA

Stephen Fienberg
Department of Statistics
Carnegie Mellon University
Pittsburgh, PA 15213-3890
USA

Ingram Olkin
Department of Statistics
Stanford University
Stanford, CA 94305
USA

Library of Congress Cataloging-in-Publication Data
Brockwell, Peter J.

Introduction to time series and forecasting / Peter J. Brockwell and Richard A. Davis.—2nd ed.
p. cm. — (Springer texts in statistics)

Includes bibliographical references and index.

ISBN 0-387-95351-5 (alk. paper)

I. Time-series analysis. I. Davis, Richard A. II. Title. III. Series.

QA280.B757 2002

519.5'5—dc21

2001049262

Printed on acid-free paper.

© 2002, 1996 Springer-Verlag New York, Inc.

All rights reserved. This work may not be translated or copied in whole or in part without the written permission of the publishers (Springer-Verlag New York, Inc., 175 Fifth Avenue, New York, NY 10010, USA), except for brief excerpts in connection with reviews or scholarly analysis. Use in connection with any form of information storage and retrieval, electronic adaptation, computer software, or by similar or dissimilar methodology now known or hereafter developed is forbidden.

The use of general descriptive names, trade names, trademarks, etc., in this publication, even if the former are not especially identified, is not to be taken as a sign that such names, as understood by the Trade Marks and Merchandise Marks Act, may accordingly be used freely by anyone.

Production managed by MaryAnn Brickner; manufacturing supervised by Joe Quatela.

Typeset by The Bartlett Press, Inc., Marietta, GA.

Printed and bound by R.R. Donnelley and Sons, Harrisonburg, VA.

Printed in the United States of America.

9 8 7 6 5 4 3 2 1

ISBN 0-387-95351-5

SPIN 10850334

Springer-Verlag New York Berlin Heidelberg

A member of BertelsmannSpringer Science+Business Media GmbH

Disclaimer:

This eBook does not include the ancillary media that was packaged with the original printed version of the book.

To Pam and Patti

This page intentionally left blank

Preface

This book is aimed at the reader who wishes to gain a working knowledge of time series and forecasting methods as applied in economics, engineering and the natural and social sciences. Unlike our earlier book, *Time Series: Theory and Methods*, referred to in the text as TSTM, this one requires only a knowledge of basic calculus, matrix algebra and elementary statistics at the level (for example) of Mendenhall, Wackerly and Scheaffer (1990). It is intended for upper-level undergraduate students and beginning graduate students.

The emphasis is on methods and the analysis of data sets. The student version of the time series package ITSM2000, enabling the reader to reproduce most of the calculations in the text (and to analyze further data sets of the reader's own choosing), is included on the CD-ROM which accompanies the book. The data sets used in the book are also included. The package requires an IBM-compatible PC operating under Windows 95, NT version 4.0, or a later version of either of these operating systems. The program ITSM can be run directly from the CD-ROM or installed on a hard disk as described at the beginning of Appendix D, where a detailed introduction to the package is provided.

Very little prior familiarity with computing is required in order to use the computer package. Detailed instructions for its use are found in the on-line help files which are accessed, when the program ITSM is running, by selecting the menu option `Help>Contents` and selecting the topic of interest. Under the heading `Data` you will find information concerning the data sets stored on the CD-ROM. The book can also be used in conjunction with other computer packages for handling time series. Chapter 14 of the book by Venables and Ripley (1994) describes how to perform many of the calculations using S-plus.

There are numerous problems at the end of each chapter, many of which involve use of the programs to study the data sets provided.

To make the underlying theory accessible to a wider audience, we have stated some of the key mathematical results without proof, but have attempted to ensure that the logical structure of the development is otherwise complete. (References to proofs are provided for the interested reader.)

Since the upgrade to ITSM2000 occurred after the first edition of this book appeared, we have taken the opportunity, in this edition, to coordinate the text with the new software, to make a number of corrections pointed out by readers of the first edition and to expand on several of the topics treated only briefly in the first edition.

Appendix D, the software tutorial, has been rewritten in order to be compatible with the new version of the software.

Some of the other extensive changes occur in (i) Section 6.6, which highlights the role of the innovations algorithm in generalized least squares and maximum likelihood estimation of regression models with time series errors, (ii) Section 6.4, where the treatment of forecast functions for ARIMA processes has been expanded and (iii) Section 10.3, which now includes GARCH modeling and simulation, topics of considerable importance in the analysis of financial time series. The new material has been incorporated into the accompanying software, to which we have also added the option *Autofit*. This streamlines the modeling of time series data by fitting maximum likelihood ARMA(p, q) models for a specified range of (p, q) values and automatically selecting the model with smallest AICC value.

There is sufficient material here for a full-year introduction to univariate and multivariate time series and forecasting. Chapters 1 through 6 have been used for several years in introductory one-semester courses in univariate time series at Colorado State University and Royal Melbourne Institute of Technology. The chapter on spectral analysis can be excluded without loss of continuity by readers who are so inclined.

We are greatly indebted to the readers of the first edition and especially to Matthew Calder, coauthor of the new computer package, and Anthony Brockwell for their many valuable comments and suggestions. We also wish to thank Colorado State University, the National Science Foundation, Springer-Verlag and our families for their continuing support during the preparation of this second edition.

Fort Collins, Colorado
August 2001

Peter J. Brockwell
Richard A. Davis

Contents

Preface	vii
1. Introduction	1
1.1. Examples of Time Series	1
1.2. Objectives of Time Series Analysis	6
1.3. Some Simple Time Series Models	7
1.3.1. Some Zero-Mean Models	8
1.3.2. Models with Trend and Seasonality	9
1.3.3. A General Approach to Time Series Modeling	14
1.4. Stationary Models and the Autocorrelation Function	15
1.4.1. The Sample Autocorrelation Function	18
1.4.2. A Model for the Lake Huron Data	21
1.5. Estimation and Elimination of Trend and Seasonal Components	23
1.5.1. Estimation and Elimination of Trend in the Absence of Seasonality	24
1.5.2. Estimation and Elimination of Both Trend and Seasonality	31
1.6. Testing the Estimated Noise Sequence Problems	35 40
2. Stationary Processes	45
2.1. Basic Properties	45
2.2. Linear Processes	51
2.3. Introduction to ARMA Processes	55
2.4. Properties of the Sample Mean and Autocorrelation Function	57
2.4.1. Estimation of μ	58
2.4.2. Estimation of $\gamma(\cdot)$ and $\rho(\cdot)$	59
2.5. Forecasting Stationary Time Series	63
2.5.1. The Durbin–Levinson Algorithm	69
2.5.2. The Innovations Algorithm	71
2.5.3. Prediction of a Stationary Process in Terms of Infinitely Many Past Values	75

2.6. The Wold Decomposition	77
Problems	78
3. ARMA Models	83
3.1. ARMA(p, q) Processes	83
3.2. The ACF and PACF of an ARMA(p, q) Process	88
3.2.1. Calculation of the ACVF	88
3.2.2. The Autocorrelation Function	94
3.2.3. The Partial Autocorrelation Function	94
3.2.4. Examples	96
3.3. Forecasting ARMA Processes	100
Problems	108
4. Spectral Analysis	111
4.1. Spectral Densities	112
4.2. The Periodogram	121
4.3. Time-Invariant Linear Filters	127
4.4. The Spectral Density of an ARMA Process	132
Problems	134
5. Modeling and Forecasting with ARMA Processes	137
5.1. Preliminary Estimation	138
5.1.1. Yule–Walker Estimation	139
5.1.2. Burg’s Algorithm	147
5.1.3. The Innovations Algorithm	150
5.1.4. The Hannan–Rissanen Algorithm	156
5.2. Maximum Likelihood Estimation	158
5.3. Diagnostic Checking	164
5.3.1. The Graph of $\{\hat{R}_t, t = 1, \dots, n\}$	165
5.3.2. The Sample ACF of the Residuals	166
5.3.3. Tests for Randomness of the Residuals	166
5.4. Forecasting	167
5.5. Order Selection	169
5.5.1. The FPE Criterion	170
5.5.2. The AICC Criterion	171
Problems	174
6. Nonstationary and Seasonal Time Series Models	179
6.1. ARIMA Models for Nonstationary Time Series	180
6.2. Identification Techniques	187

6.3. Unit Roots in Time Series Models	193
6.3.1. Unit Roots in Autoregressions	194
6.3.2. Unit Roots in Moving Averages	196
6.4. Forecasting ARIMA Models	198
6.4.1. The Forecast Function	200
6.5. Seasonal ARIMA Models	203
6.5.1. Forecasting SARIMA Processes	208
6.6. Regression with ARMA Errors	210
6.6.1. OLS and GLS Estimation	210
6.6.2. ML Estimation	213
Problems	219
7. Multivariate Time Series	223
7.1. Examples	224
7.2. Second-Order Properties of Multivariate Time Series	229
7.3. Estimation of the Mean and Covariance Function	234
7.3.1. Estimation of μ	234
7.3.2. Estimation of $\Gamma(h)$	235
7.3.3. Testing for Independence of Two Stationary Time Series	237
7.3.4. Bartlett's Formula	238
7.4. Multivariate ARMA Processes	241
7.4.1. The Covariance Matrix Function of a Causal ARMA Process	244
7.5. Best Linear Predictors of Second-Order Random Vectors	244
7.6. Modeling and Forecasting with Multivariate AR Processes	246
7.6.1. Estimation for Autoregressive Processes Using Whittle's Algorithm	247
7.6.2. Forecasting Multivariate Autoregressive Processes	250
7.7. Cointegration	254
Problems	256
8. State-Space Models	259
8.1. State-Space Representations	260
8.2. The Basic Structural Model	263
8.3. State-Space Representation of ARIMA Models	267
8.4. The Kalman Recursions	271
8.5. Estimation For State-Space Models	277
8.6. State-Space Models with Missing Observations	283
8.7. The EM Algorithm	289
8.8. Generalized State-Space Models	292
8.8.1. Parameter-Driven Models	292

8.8.2. Observation-Driven Models	299
Problems	311
9. Forecasting Techniques	317
9.1. The ARAR Algorithm	318
9.1.1. Memory Shortening	318
9.1.2. Fitting a Subset Autoregression	319
9.1.3. Forecasting	320
9.1.4. Application of the ARAR Algorithm	321
9.2. The Holt–Winters Algorithm	322
9.2.1. The Algorithm	322
9.2.2. Holt–Winters and ARIMA Forecasting	324
9.3. The Holt–Winters Seasonal Algorithm	326
9.3.1. The Algorithm	326
9.3.2. Holt–Winters Seasonal and ARIMA Forecasting	328
9.4. Choosing a Forecasting Algorithm	328
Problems	330
10. Further Topics	331
10.1. Transfer Function Models	331
10.1.1. Prediction Based on a Transfer Function Model	337
10.2. Intervention Analysis	340
10.3. Nonlinear Models	343
10.3.1. Deviations from Linearity	344
10.3.2. Chaotic Deterministic Sequences	345
10.3.3. Distinguishing Between White Noise and iid Sequences	347
10.3.4. Three Useful Classes of Nonlinear Models	348
10.3.5. Modeling Volatility	349
10.4. Continuous-Time Models	357
10.5. Long-Memory Models	361
Problems	365
A. Random Variables and Probability Distributions	369
A.1. Distribution Functions and Expectation	369
A.2. Random Vectors	374
A.3. The Multivariate Normal Distribution	377
Problems	381

B. Statistical Complements	383
B.1. Least Squares Estimation	383
B.1.1. The Gauss-Markov Theorem	385
B.1.2. Generalized Least Squares	386
B.2. Maximum Likelihood Estimation	386
B.2.1. Properties of Maximum Likelihood Estimators	387
B.3. Confidence Intervals	388
B.3.1. Large-Sample Confidence Regions	388
B.4. Hypothesis Testing	389
B.4.1. Error Probabilities	390
B.4.2. Large-Sample Tests Based on Confidence Regions	390
C. Mean Square Convergence	393
C.1. The Cauchy Criterion	393
D. An ITSM Tutorial	395
D.1. Getting Started	396
D.1.1. Running ITSM	396
D.2. Preparing Your Data for Modeling	396
D.2.1. Entering Data	397
D.2.2. Information	397
D.2.3. Filing Data	397
D.2.4. Plotting Data	398
D.2.5. Transforming Data	398
D.3. Finding a Model for Your Data	403
D.3.1. Autofit	403
D.3.2. The Sample ACF and PACF	403
D.3.3. Entering a Model	404
D.3.4. Preliminary Estimation	406
D.3.5. The AICC Statistic	408
D.3.6. Changing Your Model	408
D.3.7. Maximum Likelihood Estimation	409
D.3.8. Optimization Results	410
D.4. Testing Your Model	411
D.4.1. Plotting the Residuals	412
D.4.2. ACF/PACF of the Residuals	412
D.4.3. Testing for Randomness of the Residuals	414
D.5. Prediction	415
D.5.1. Forecast Criteria	415
D.5.2. Forecast Results	415

D.6. Model Properties	416
D.6.1. ARMA Models	417
D.6.2. Model ACF, PACF	418
D.6.3. Model Representations	419
D.6.4. Generating Realizations of a Random Series	420
D.6.5. Spectral Properties	421
D.7. Multivariate Time Series	421
References	423
Index	429

1

Introduction

- 1.1 Examples of Time Series
- 1.2 Objectives of Time Series Analysis
- 1.3 Some Simple Time Series Models
- 1.4 Stationary Models and the Autocorrelation Function
- 1.5 Estimation and Elimination of Trend and Seasonal Components
- 1.6 Testing the Estimated Noise Sequence

In this chapter we introduce some basic ideas of time series analysis and stochastic processes. Of particular importance are the concepts of stationarity and the autocovariance and sample autocovariance functions. Some standard techniques are described for the estimation and removal of trend and seasonality (of known period) from an observed time series. These are illustrated with reference to the data sets in Section 1.1. The calculations in all the examples can be carried out using the time series package ITSM, the student version of which is supplied on the enclosed CD. The data sets are contained in files with names ending in .TSM. For example, the Australian red wine sales are filed as WINE.TSM. Most of the topics covered in this chapter will be developed more fully in later sections of the book. The reader who is not already familiar with random variables and random vectors should first read Appendix A, where a concise account of the required background is given.

1.1 Examples of Time Series

A **time series** is a set of observations x_t , each one being recorded at a specific time t . A *discrete-time time series* (the type to which this book is primarily devoted) is one in which the set T_0 of times at which observations are made is a discrete set, as is the

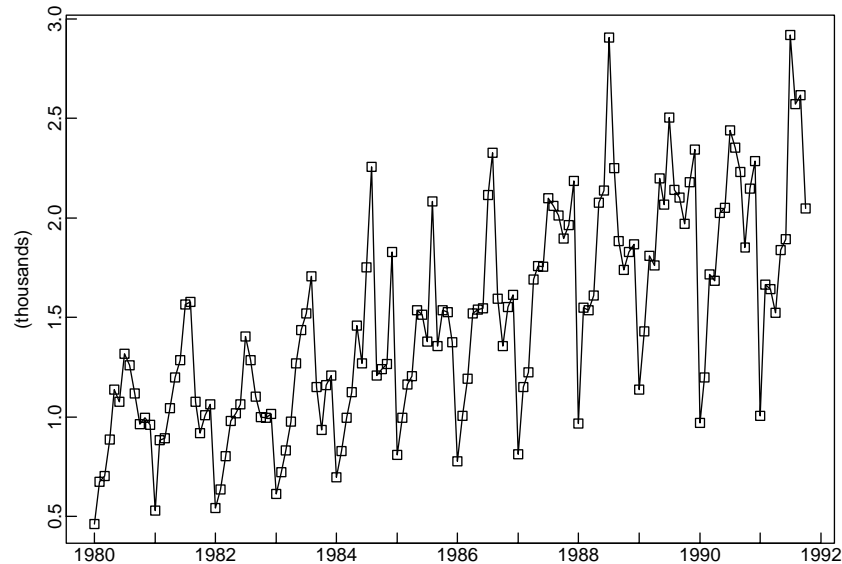


Figure 1-1

The Australian red wine sales, Jan. '80 – Oct. '91.

case, for example, when observations are made at fixed time intervals. *Continuous-time time series* are obtained when observations are recorded continuously over some time interval, e.g., when $T_0 = [0, 1]$.

Example 1.1.1 Australian red wine sales; WINE.TSM

Figure 1.1 shows the monthly sales (in kiloliters) of red wine by Australian winemakers from January 1980 through October 1991. In this case the set T_0 consists of the 142 times $\{(\text{Jan. 1980}), (\text{Feb. 1980}), \dots, (\text{Oct. 1991})\}$. Given a set of n observations made at uniformly spaced time intervals, it is often convenient to rescale the time axis in such a way that T_0 becomes the set of integers $\{1, 2, \dots, n\}$. In the present example this amounts to measuring time in months with (Jan. 1980) as month 1. Then T_0 is the set $\{1, 2, \dots, 142\}$. It appears from the graph that the sales have an upward trend and a seasonal pattern with a peak in July and a trough in January. To plot the data using ITSM, run the program by double-clicking on the ITSM icon and then select the option File>Project>Open>Univariate, click OK, and select the file WINE.TSM. The graph of the data will then appear on your screen. \square

Example 1.1.2 All-star baseball games, 1933–1995

Figure 1.2 shows the results of the all-star games by plotting x_t , where

$$x_t = \begin{cases} 1 & \text{if the National League won in year } t, \\ -1 & \text{if the American League won in year } t. \end{cases}$$

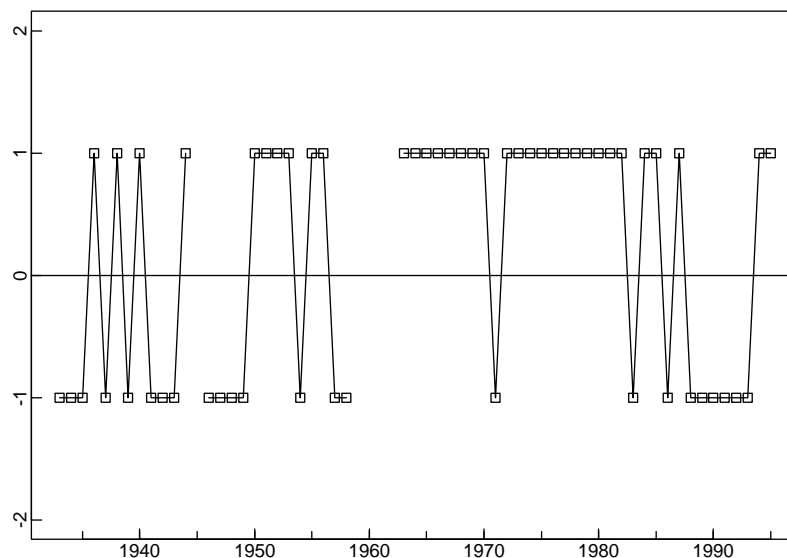


Figure 1-2
Results of the
all-star baseball
games, 1933–1995.

This is a series with only two possible values, ± 1 . It also has some missing values, since no game was played in 1945, and two games were scheduled for each of the years 1959–1962. □

Example 1.1.3 Accidental deaths, U.S.A., 1973–1978; DEATHS.TSM

Like the red wine sales, the monthly accidental death figures show a strong seasonal pattern, with the maximum for each year occurring in July and the minimum for each year occurring in February. The presence of a trend in Figure 1.3 is much less apparent than in the wine sales. In Section 1.5 we shall consider the problem of representing the data as the sum of a trend, a seasonal component, and a residual term. □

Example 1.1.4 A signal detection problem; SIGNAL.TSM

Figure 1.4 shows simulated values of the series

$$X_t = \cos\left(\frac{t}{10}\right) + N_t, \quad t = 1, 2, \dots, 200,$$

where $\{N_t\}$ is a sequence of independent normal random variables, with mean 0 and variance 0.25. Such a series is often referred to as *signal plus noise*, the signal being the smooth function, $S_t = \cos(\frac{t}{10})$ in this case. Given only the data X_t , how can we determine the unknown signal component? There are many approaches to this general problem under varying assumptions about the signal and the noise. One simple approach is to *smooth* the data by expressing X_t as a sum of sine waves of various frequencies (see Section 4.2) and eliminating the high-frequency components. If we do this to the values of $\{X_t\}$ shown in Figure 1.4 and retain only the lowest

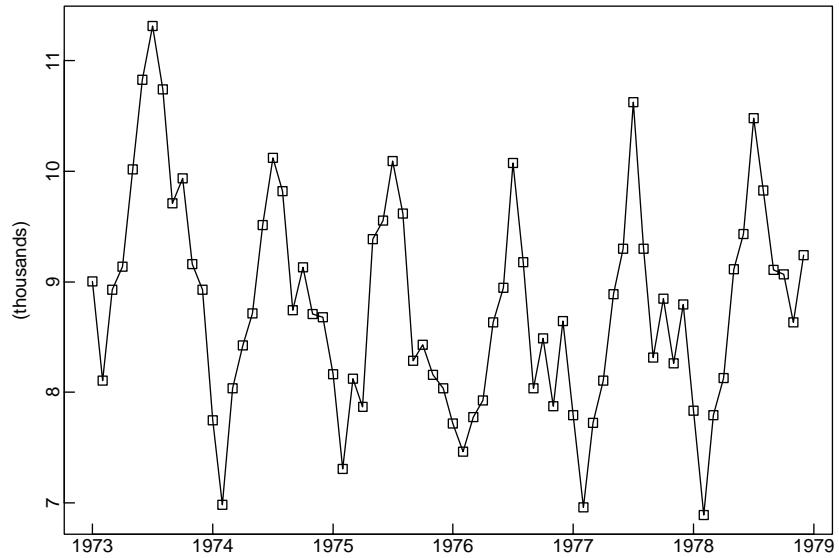


Figure 1-3
The monthly accidental deaths data, 1973–1978.

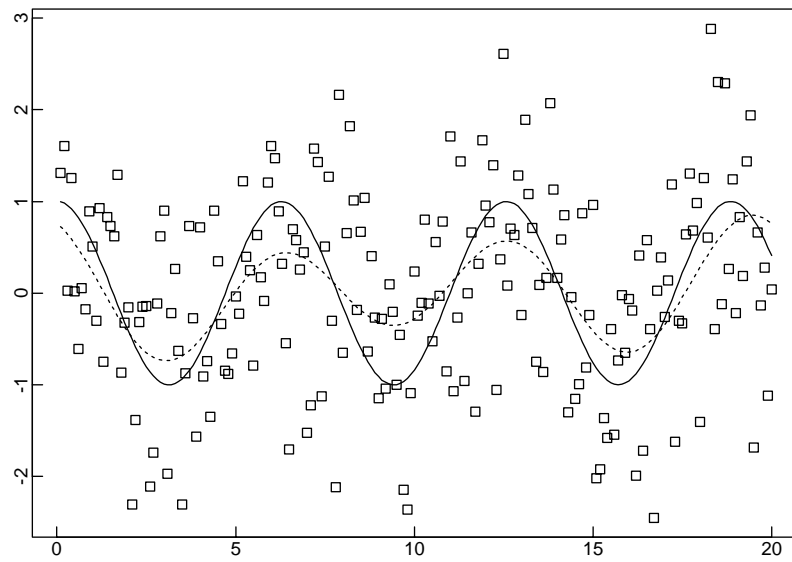


Figure 1-4
The series $\{X_t\}$ of
Example 1.1.4.

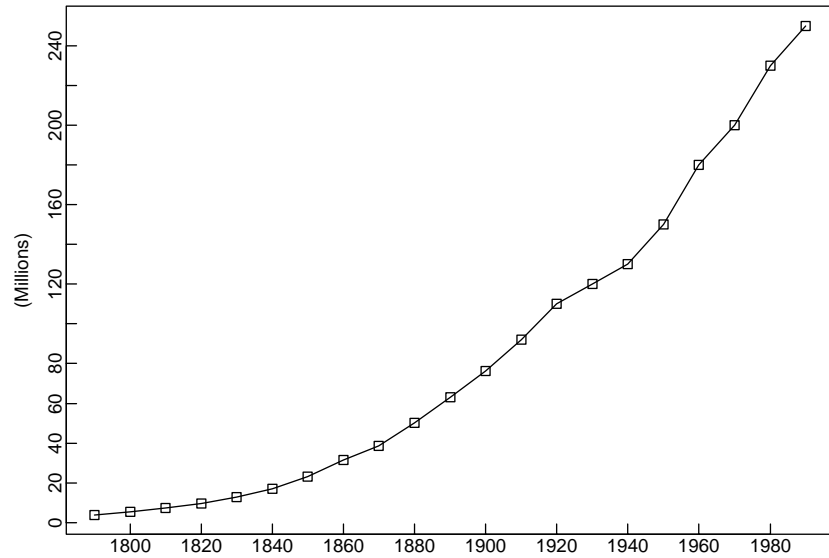


Figure 1-5
Population of the
U.S.A. at ten-year
intervals, 1790–1990.

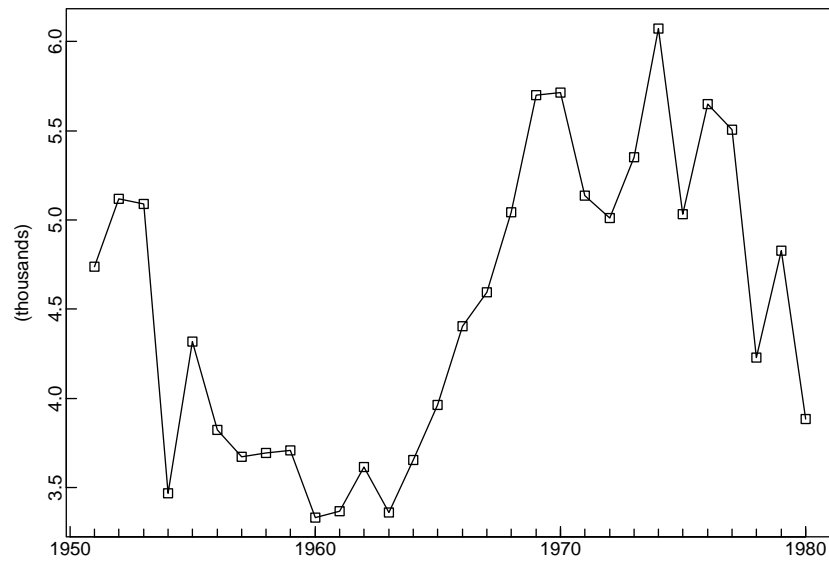


Figure 1-6
Strikes in the
U.S.A., 1951–1980.

3.5% of the frequency components, we obtain the estimate of the signal also shown in Figure 1.4. The waveform of the signal is quite close to that of the true signal in this case, although its amplitude is somewhat smaller. □

Example 1.1.5 Population of the U.S.A., 1790–1990; USPOP.TSM

The population of the U.S.A., measured at ten-year intervals, is shown in Figure 1.5. The graph suggests the possibility of fitting a quadratic or exponential trend to the data. We shall explore this further in Section 1.3. □

Example 1.1.6 Number of strikes per year in the U.S.A., 1951–1980; STRIKES.TSM

The annual numbers of strikes in the U.S.A. for the years 1951–1980 are shown in Figure 1.6. They appear to fluctuate erratically about a slowly changing level. □

1.2 Objectives of Time Series Analysis

The examples considered in Section 1.1 are an extremely small sample from the multitude of time series encountered in the fields of engineering, science, sociology, and economics. Our purpose in this book is to study techniques for drawing inferences from such series. Before we can do this, however, 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.

The model may be used simply to provide a compact description of the data. We may, for example, be able to represent the accidental deaths data of Example 1.1.3 as the sum of a specified trend, and seasonal and random terms. For the interpretation of economic statistics such as unemployment figures, it is important to recognize the presence of seasonal components and to remove them so as not to confuse them with long-term trends. This process is known as **seasonal adjustment**. Other applications of time series models include separation (or filtering) of noise from signals as in Example 1.1.4, prediction of future values of a series such as the red wine sales in Example 1.1.1 or the population data in Example 1.1.5, testing hypotheses such as global warming using recorded temperature data, predicting one series from observations of another, e.g., predicting future sales using advertising expenditure data, and controlling future values of a series by adjusting parameters. Time series models are also useful in simulation studies. For example, the performance of a reservoir depends heavily on the random daily inputs of water to the system. If these are modeled as a time series, then we can use the fitted model to simulate a large number of independent sequences of daily inputs. Knowing the size and mode of operation

of the reservoir, we can determine the fraction of the simulated input sequences that cause the reservoir to run out of water in a given time period. This fraction will then be an estimate of the probability of emptiness of the reservoir at some time in the given period.

1.3 Some Simple Time Series Models

An important part of the analysis of a time series is the selection of a suitable probability model (or class of models) 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 1.3.1

A **time series model** for the observed data $\{x_t\}$ is a specification of the joint distributions (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 sequence of random variables $\{X_1, X_2, \dots\}$ would specify all of the **joint distributions** of the random vectors $(X_1, \dots, X_n)'$, $n = 1, 2, \dots$, or equivalently all of the probabilities

$$P[X_1 \leq x_1, \dots, X_n \leq x_n], \quad -\infty < x_1, \dots, x_n < \infty, \quad n = 1, 2, \dots$$

Such a specification is rarely used in time series analysis (unless the data are generated by some well-understood simple mechanism), 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 $E X_t$ and the expected products $E(X_{t+h} X_t)$, $t = 1, 2, \dots$, $h = 0, 1, 2, \dots$, focusing on properties of the sequence $\{X_t\}$ that depend only on these. Such properties of $\{X_t\}$ are referred to 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. In general we shall lose a certain amount of information by looking at time series “through second-order spectacles”; however, as we shall see in Chapter 2, the theory of minimum mean squared error linear prediction depends only on the second-order properties, thus providing further justification for the use of the second-order characterization of time series models.

Figure 1.7 shows one of many possible realizations of $\{S_t, t = 1, \dots, 200\}$, where $\{S_t\}$ is a sequence of random variables specified in Example 1.3.3 below. In most practical problems involving time series we see only *one* realization. For example,

there is only one available realization of Fort Collins's annual rainfall for the years 1900–1996, but we imagine it to be one of the many sequences that *might* have occurred. In the following examples we introduce some simple time series models. One of our goals will be to expand this repertoire so as to have at our disposal a broad range of models with which to try to match the observed behavior of given data sets.

1.3.1 Some Zero-Mean Models

Example 1.3.1 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, \dots as iid noise. By definition we can write, for any positive integer n and real numbers x_1, \dots, x_n ,

$$P[X_1 \leq x_1, \dots, X_n \leq x_n] = P[X_1 \leq x_1] \cdots P[X_n \leq x_n] = F(x_1) \cdots F(x_n),$$

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

$$P[X_{n+h} \leq x | X_1 = x_1, \dots, X_n = x_n] = P[X_{n+h} \leq x],$$

showing that knowledge of X_1, \dots, X_n is of no value for predicting the behavior of X_{n+h} . Given the values of X_1, \dots, X_n , the function f that minimizes the mean squared error $E[(X_{n+h} - f(X_1, \dots, X_n))^2]$ is in fact identically zero (see Problem 1.2). Although this means that iid noise is a rather uninteresting process for forecasters, it plays an important role as a building block for more complicated time series models. \square

Example 1.3.2 A binary process

As an example of iid noise, consider the sequence of iid random variables $\{X_t, t = 1, 2, \dots\}$ with

$$P[X_t = 1] = p, \quad P[X_t = -1] = 1 - p,$$

where $p = \frac{1}{2}$. The time series obtained by tossing a penny repeatedly and scoring +1 for each head and -1 for each tail is usually modeled as a realization of this process. A priori we might well consider the same process as a model for the all-star baseball games in Example 1.1.2. However, even a cursory inspection of the results from 1963–1982, which show the National League winning 19 of 20 games, casts serious doubt on the hypothesis $P[X_t = 1] = \frac{1}{2}$. \square

Example 1.3.3 Random walk

The random walk $\{S_t, t = 0, 1, 2, \dots\}$ (starting at zero) is obtained by cumulatively summing (or “integrating”) iid random variables. Thus a random walk with zero mean is obtained by defining $S_0 = 0$ and

$$S_t = X_1 + X_2 + \dots + X_t, \quad \text{for } t = 1, 2, \dots,$$

where $\{X_t\}$ is iid noise. If $\{X_t\}$ is the binary process of Example 1.3.2, then $\{S_t, t = 0, 1, 2, \dots\}$ is called a **simple symmetric random walk**. This walk can be viewed as the location of a pedestrian who starts at position zero at time zero and at each integer time tosses a fair coin, stepping one unit to the right each time a head appears and one unit to the left for each tail. A realization of length 200 of a simple symmetric random walk is shown in Figure 1.7. Notice that the outcomes of the coin tosses can be recovered from $\{S_t, t = 0, 1, \dots\}$ by differencing. Thus the result of the t th toss can be found from $S_t - S_{t-1} = X_t$. \square

1.3.2 Models with Trend and Seasonality

In several of the time series examples of Section 1.1 there is a clear trend in the data. An increasing trend is apparent in both the Australian red wine sales (Figure 1.1) and the population of the U.S.A. (Figure 1.5). In both cases a zero-mean model for the data is clearly inappropriate. The graph of the population data, which contains no apparent periodic component, suggests trying a model of the form

$$X_t = m_t + Y_t,$$

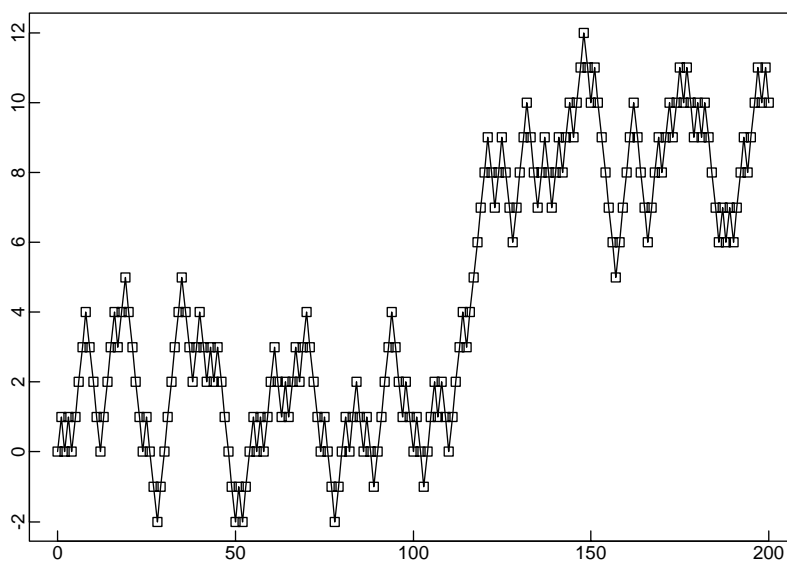


Figure 1-7
One realization of a
simple random walk
 $\{S_t, t = 0, 1, 2, \dots, 200\}$

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 (some other methods are considered in Section 1.5).

In the least squares procedure we attempt to fit a parametric family of functions, e.g.,

$$m_t = a_0 + a_1t + a_2t^2, \quad (1.3.1)$$

to the data $\{x_1, \dots, x_n\}$ by choosing the parameters, in this illustration a_0, a_1 , and a_2 , to minimize $\sum_{t=1}^n (x_t - m_t)^2$. This method of curve fitting is called **least squares regression** and can be carried out using the program ITSM and selecting the Regression option.

Example 1.3.4 Population of the U.S.A., 1790–1990

To fit a function of the form (1.3.1) to the population data shown in Figure 1.5 we relabel the time axis so that $t = 1$ corresponds to 1790 and $t = 21$ corresponds to 1990. Run ITSM, select File>Project>Open>Univariate, and open the file US-POP.TSM. Then select Regression>Specify, choose Polynomial Regression with order equal to 2, and click OK. Then select Regression>Estimation>Least Squares, and you will obtain the following estimated parameter values in the model (1.3.1):

$$\begin{aligned} \hat{a}_0 &= 6.9579 \times 10^6, \\ \hat{a}_1 &= -2.1599 \times 10^6, \end{aligned}$$

and

$$\hat{a}_2 = 6.5063 \times 10^5.$$

A graph of the fitted function is shown with the original data in Figure 1.8. The estimated values of the noise process Y_t , $1 \leq t \leq 21$, are the residuals obtained by subtraction of $\hat{m}_t = \hat{a}_0 + \hat{a}_1t + \hat{a}_2t^2$ from x_t .

The estimated trend component \hat{m}_t furnishes us with a natural predictor of future values of X_t . For example, if we estimate the noise Y_{22} by its mean value, i.e., zero, then (1.3.1) gives the estimated U.S. population for the year 2000 as

$$\hat{m}_{22} = 6.9579 \times 10^6 - 2.1599 \times 10^6 \times 22 + 6.5063 \times 10^5 \times 22^2 = 274.35 \times 10^6.$$

However, if the residuals $\{Y_t\}$ are highly correlated, we may be able to use their values to give a better estimate of Y_{22} and hence of the population X_{22} in the year 2000. \square

Example 1.3.5 Level of Lake Huron 1875–1972; LAKE.DAT

A graph of the level in feet of Lake Huron (reduced by 570) in the years 1875–1972 is displayed in Figure 1.9. Since the lake level appears to decline at a roughly linear rate, ITSM was used to fit a model of the form

$$X_t = a_0 + a_1t + Y_t, \quad t = 1, \dots, 98 \quad (1.3.2)$$

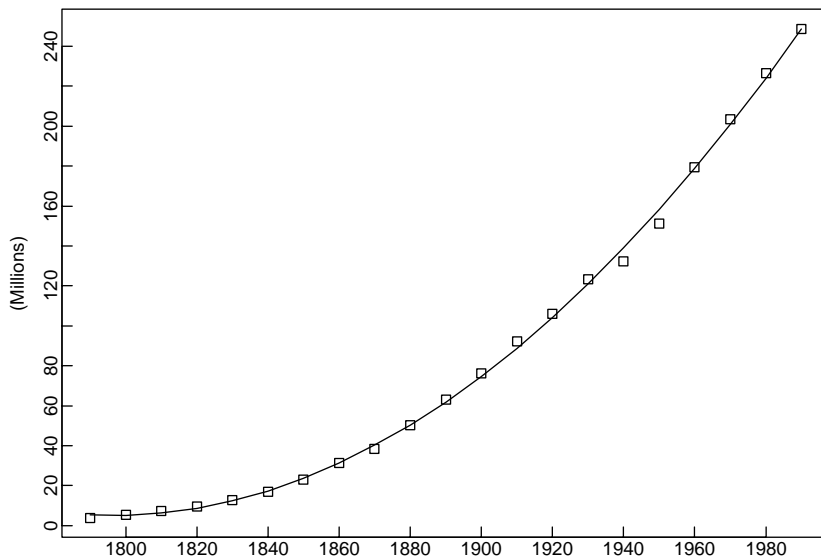


Figure 1-8
Population of the U.S.A.
showing the quadratic trend
fitted by least squares.

(with the time axis relabeled as in Example 1.3.4). The least squares estimates of the parameter values are

$$\hat{a}_0 = 10.202 \quad \text{and} \quad \hat{a}_1 = -.0242.$$

(The resulting least squares line, $\hat{a}_0 + \hat{a}_1 t$, is also displayed in Figure 1.9.) The estimates of the noise, Y_t , in the model (1.3.2) are the residuals obtained by subtracting the least squares line from x_t and are plotted in Figure 1.10. There are two interesting

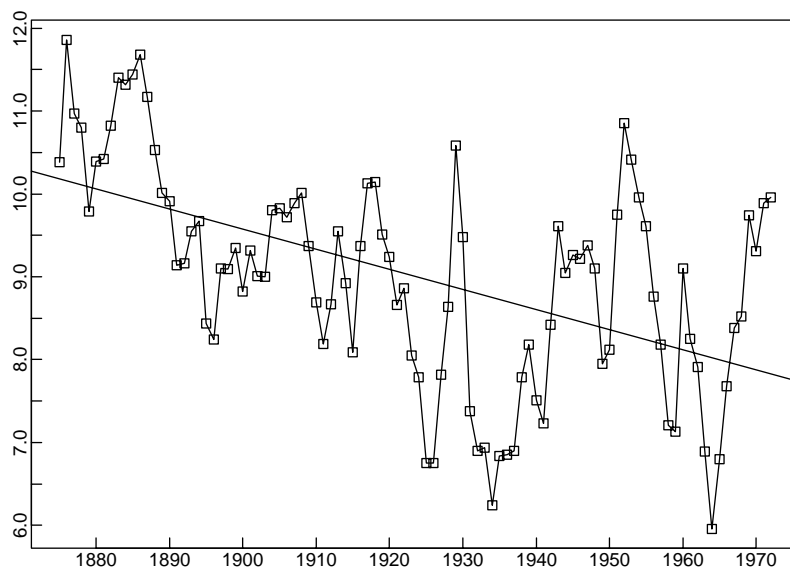


Figure 1-9
Level of Lake Huron
1875–1972 showing the
line fitted by least squares.

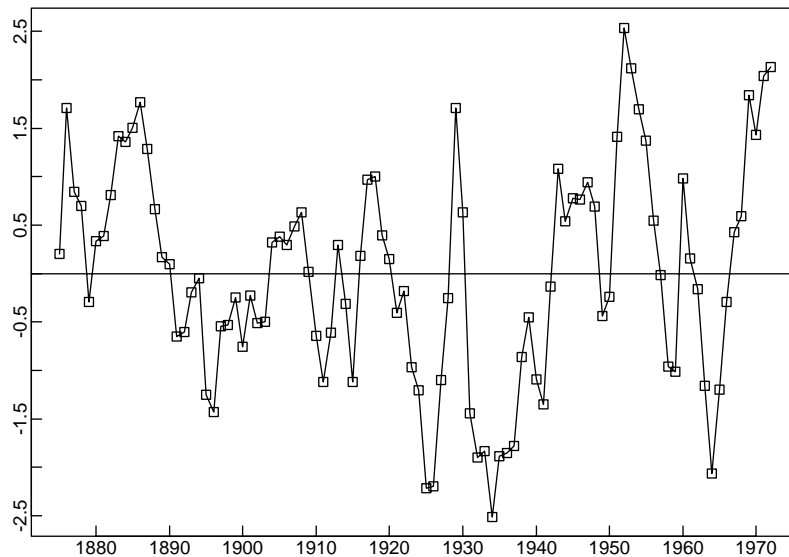


Figure 1-10
Residuals from fitting a
line to the Lake Huron
data in Figure 1.9.

features of the graph of the residuals. The first is the absence of any discernible trend. The second is the smoothness of the graph. (In particular, there are long stretches of residuals that have the same sign. This would be very unlikely to occur if the residuals were observations of iid noise with zero mean.) Smoothness of the graph of a time series is generally indicative of the existence of some form of dependence among the observations.

Such dependence can be used to advantage in forecasting future values of the series. If we were to assume the validity of the fitted model with iid residuals $\{Y_t\}$, then the minimum mean squared error predictor of the next residual (Y_{99}) would be zero (by Problem 1.2). However, Figure 1.10 strongly suggests that Y_{99} will be positive.

How then do we quantify dependence, and how do we construct models for forecasting that incorporate dependence of a particular type? To deal with these questions, Section 1.4 introduces the autocorrelation function as a measure of dependence, and stationary processes as a family of useful models exhibiting a wide variety of dependence structures. \square

Harmonic Regression

Many time series are influenced by seasonally varying factors such as the weather, the effect of which can be modeled by a periodic component with fixed known period. For example, the accidental deaths series (Figure 1.3) shows a repeating annual pattern with peaks in July and troughs in February, strongly suggesting a seasonal factor with period 12. In order to represent such 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 ($s_{t-d} = s_t$). A convenient choice for s_t is a sum of harmonics (or sine waves) given by

$$s_t = a_0 + \sum_{j=1}^k (a_j \cos(\lambda_j t) + b_j \sin(\lambda_j t)), \quad (1.3.3)$$

where a_0, a_1, \dots, a_k and b_1, \dots, b_k are unknown parameters and $\lambda_1, \dots, \lambda_k$ are fixed frequencies, each being some integer multiple of $2\pi/d$. To carry out harmonic regression using ITSM, select `Regression>Specify` and check `Include intercept term` and `Harmonic Regression`. Then specify the number of harmonics (k in (1.3.3)) and enter k integer-valued Fourier indices f_1, \dots, f_k . For a sine wave with period d , set $f_1 = n/d$, where n is the number of observations in the time series. (If n/d is not an integer, you will need to delete a few observations from the beginning of the series to make it so.) The other $k - 1$ Fourier indices should be positive integer multiples of the first, corresponding to harmonics of the fundamental sine wave with period d . Thus to fit a single sine wave with period 365 to 365 daily observations we would choose $k = 1$ and $f_1 = 1$. To fit a linear combination of sine waves with periods $365/j$, $j = 1, \dots, 4$, we would choose $k = 4$ and $f_j = j$, $j = 1, \dots, 4$. Once k and f_1, \dots, f_k have been specified, click `OK` and then select `Regression>Estimation>Least Squares` to obtain the required regression coefficients. To see how well the fitted function matches the data, select `Regression>Show fit`.

Example 1.3.6 Accidental deaths

To fit a sum of two harmonics with periods twelve months and six months to the monthly accidental deaths data x_1, \dots, x_n with $n = 72$, we choose $k = 2$, $f_1 =$

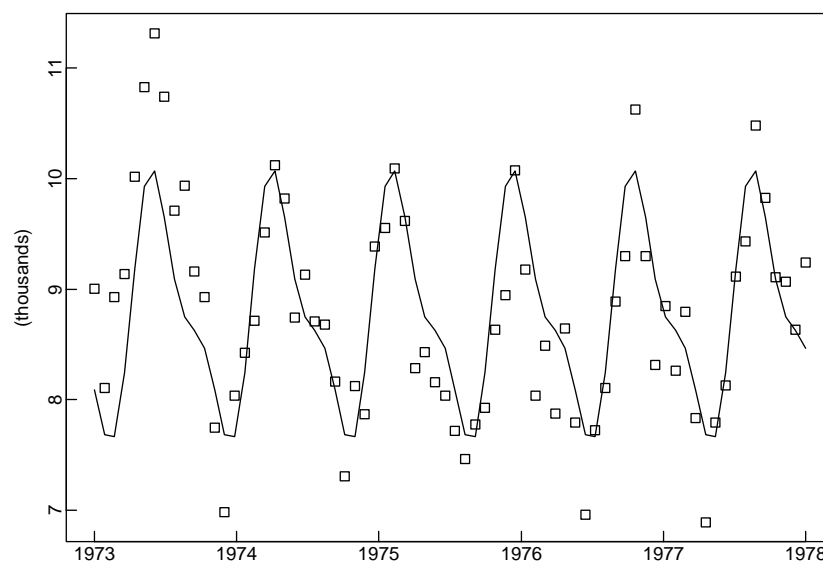


Figure 1-11

The estimated harmonic component of the accidental deaths data from ITSM.

$n/12 = 6$, and $f_2 = n/6 = 12$. Using ITSM as described above, we obtain the fitted function shown in Figure 1.11. As can be seen from the figure, the periodic character of the series is captured reasonably well by this fitted function. In practice, it is worth experimenting with several different combinations of harmonics in order to find a satisfactory estimate of the seasonal component. The program ITSM also allows fitting a linear combination of harmonics and polynomial trend by checking both `Harmonic Regression` and `Polynomial Regression` in the `Regression>Specification` dialog box. Other methods for dealing with seasonal variation in the presence of trend are described in Section 1.5. \square

1.3.3 A General Approach to Time Series Modeling

The examples of the previous section illustrate a general approach to time series analysis that will form the basis for much of what is done in this book. Before introducing the ideas of dependence and stationarity, we outline this approach to provide the reader with an overview of the way in which the various ideas of this chapter fit together.

- Plot the series and examine the main features of the graph, checking in particular whether there is
 - (a) a trend,
 - (b) a seasonal component,
 - (c) any apparent sharp changes in behavior,
 - (d) any outlying observations.
- Remove the trend and seasonal components to get *stationary* residuals (as defined in Section 1.4). To achieve this goal it may sometimes be necessary to apply a preliminary transformation to the data. For example, if the magnitude of the fluctuations appears to grow roughly linearly with the level of the series, then the transformed series $\{\ln X_1, \dots, \ln X_n\}$ will have fluctuations of more constant magnitude. See, for example, Figures 1.1 and 1.17. (If some of the data are negative, add a positive constant to each of the data values to ensure that all values are positive before taking logarithms.) There are several ways in which trend and seasonality can be removed (see Section 1.5), some involving estimating the components and subtracting them from the data, and others depending on *differencing* the data, i.e., replacing the original series $\{X_t\}$ by $\{Y_t := X_t - X_{t-d}\}$ for some positive integer d . Whichever method is used, the aim is to produce a stationary series, whose values we shall refer to as residuals.
- Choose a model to fit the residuals, making use of various sample statistics including the sample autocorrelation function to be defined in Section 1.4.
- Forecasting will be achieved by forecasting the residuals and then inverting the transformations described above to arrive at forecasts of the original series $\{X_t\}$.

- An extremely useful alternative approach touched on only briefly in this book is to express the series in terms of its Fourier components, which are sinusoidal waves of different frequencies (cf. Example 1.1.4). This approach is especially important in engineering applications such as signal processing and structural design. It is important, for example, to ensure that the resonant frequency of a structure does not coincide with a frequency at which the loading forces on the structure have a particularly large component.

1.4 Stationary Models and the Autocorrelation Function

Loosely speaking, a time series $\{X_t, t = 0, \pm 1, \dots\}$ is said to be stationary if it has statistical properties similar to those of the “time-shifted” series $\{X_{t+h}, t = 0, \pm 1, \dots\}$, for each integer h . 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 1.4.1

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

$$\mu_X(t) = E(X_t).$$

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

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

for all integers r and s .

Definition 1.4.2

$\{X_t\}$ is **(weakly) stationary** if

- (i) $\mu_X(t)$ is independent of t ,

and

- (ii) $\gamma_X(t+h, t)$ is independent of t for each h .

Remark 1. Strict stationarity of a time series $\{X_t, t = 0, \pm 1, \dots\}$ 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 is easy to check that if $\{X_t\}$ is strictly stationary and $E X_t^2 < \infty$ for all t , then $\{X_t\}$ is also weakly stationary (Problem 1.3). Whenever we use the term *stationary* we shall mean weakly stationary as in Definition 1.4.2, unless we specifically indicate otherwise. \square

Remark 2. In view of condition (ii), whenever we use the term covariance function with reference to a *stationary* time series $\{X_t\}$ we shall mean the function γ_X of one

variable, defined by

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

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

Definition 1.4.3

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

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

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

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

In the following examples we shall frequently use the easily verified **linearity property of covariances**, that if $EX^2 < \infty$, $EY^2 < \infty$, $EZ^2 < \infty$ and a , b , and c are any real constants, then

$$\text{Cov}(aX + bY + c, Z) = a \text{Cov}(X, Z) + b \text{Cov}(Y, Z).$$

Example 1.4.1 iid noise

If $\{X_t\}$ is iid noise and $E(X_t^2) = \sigma^2 < \infty$, then the first requirement of Definition 1.4.2 is obviously satisfied, since $E(X_t) = 0$ 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

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

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

Example 1.4.2 White noise

If $\{X_t\}$ is a sequence of uncorrelated random variables, each with zero mean and variance σ^2 , then clearly $\{X_t\}$ is stationary with the same covariance function as the iid noise in Example 1.4.1. Such a sequence is referred to as **white noise** (with mean 0 and variance σ^2). This is indicated by the notation

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

Clearly, every $\text{IID}(0, \sigma^2)$ sequence is $\text{WN}(0, \sigma^2)$ but not conversely (see Problem 1.8 and the ARCH(1) process of Section 10.3). \square

Example 1.4.3 The random walk

If $\{S_t\}$ is the random walk defined in Example 1.3.3 with $\{X_t\}$ as in Example 1.4.1, then $ES_t = 0$, $E(S_t^2) = t\sigma^2 < \infty$ for all t , and, for $h \geq 0$,

$$\begin{aligned}\gamma_S(t+h, t) &= \text{Cov}(S_{t+h}, S_t) \\ &= \text{Cov}(S_t + X_{t+1} + \cdots + X_{t+h}, S_t) \\ &= \text{Cov}(S_t, S_t) \\ &= t\sigma^2.\end{aligned}$$

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

Example 1.4.4 First-order moving average or MA(1) process

Consider the series defined by the equation

$$X_t = Z_t + \theta Z_{t-1}, \quad t = 0, \pm 1, \dots, \quad (1.4.1)$$

where $\{Z_t\} \sim \text{WN}(0, \sigma^2)$ and θ is a real-valued constant. From (1.4.1) we see that $EX_t = 0$, $EX_t^2 = \sigma^2(1 + \theta^2) < \infty$, and

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

Thus the requirements of Definition 1.4.2 are satisfied, and $\{X_t\}$ is stationary. The autocorrelation function of $\{X_t\}$ is

$$\rho_X(h) = \begin{cases} 1, & \text{if } h = 0, \\ \theta / (1 + \theta^2), & \text{if } h = \pm 1, \\ 0, & \text{if } |h| > 1. \end{cases} \quad \square$$

Example 1.4.5 First-order autoregression or AR(1) process

Let us *assume* now that $\{X_t\}$ is a stationary series satisfying the equations

$$X_t = \phi X_{t-1} + Z_t, \quad t = 0, \pm 1, \dots, \quad (1.4.2)$$

where $\{Z_t\} \sim \text{WN}(0, \sigma^2)$, $|\phi| < 1$, and Z_t is uncorrelated with X_s for each $s < t$. (We shall show in Section 2.2 that there is in fact exactly one such solution of (1.4.2).) By taking expectations on each side of (1.4.2) and using the fact that $EZ_t = 0$, we see

at once that

$$EX_t = 0.$$

To find the autocorrelation function of $\{X_t\}$ we multiply each side of (1.4.2) by X_{t-h} ($h > 0$) and then take expectations to get

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

Observing that $\gamma(h) = \gamma(-h)$ and using Definition 1.4.3, we find that

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

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

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

and hence that $\gamma_X(0) = \sigma^2 / (1 - \phi^2)$. □

1.4.1 The Sample Autocorrelation Function

Although we have just seen how to compute the autocorrelation function for a few simple time series models, in practical problems we do not start with a model, but with *observed data* $\{x_1, x_2, \dots, x_n\}$. To assess the degree of dependence in the data and to select a model for the data that reflects this, one of the important tools we use is the **sample autocorrelation function** (sample ACF) of the data. If we believe that the data are realized values of a stationary time series $\{X_t\}$, then the sample ACF will provide us with an estimate of the ACF of $\{X_t\}$. This estimate may suggest which of the many possible stationary time series models is a suitable candidate for representing the dependence in the data. For example, a sample ACF that is close to zero for all nonzero lags suggests that an appropriate model for the data might be iid noise. The following definitions are natural *sample* analogues of those for the autocovariance and autocorrelation functions given earlier for stationary time series *models*.

Definition 1.4.4

Let x_1, \dots, x_n be observations of a time series. The **sample mean** of x_1, \dots, x_n is

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

The **sample autocovariance function** is

$$\hat{\gamma}(h) := n^{-1} \sum_{t=1}^{n-|h|} (x_{t+|h|} - \bar{x})(x_t - \bar{x}), \quad -n < h < n.$$

The **sample autocorrelation function** is

$$\hat{\rho}(h) = \frac{\hat{\gamma}(h)}{\hat{\gamma}(0)}, \quad -n < h < n.$$

Remark 3. For $h \geq 0$, $\hat{\gamma}(h)$ is approximately equal to the sample covariance of the $n - h$ pairs of observations $(x_1, x_{1+h}), (x_2, x_{2+h}), \dots, (x_{n-h}, x_n)$. The difference arises from use of the divisor n instead of $n - h$ and the subtraction of the *overall* mean, \bar{x} , from each factor of the summands. Use of the divisor n ensures that the sample covariance matrix $\hat{\Gamma}_n := [\hat{\gamma}(i - j)]_{i,j=1}^n$ is nonnegative definite (see Section 2.4.2). \square

Remark 4. Like the sample covariance matrix defined in Remark 3, the sample correlation matrix $\hat{R}_n := [\hat{\rho}(i - j)]_{i,j=1}^n$ is nonnegative definite. Each of its diagonal elements is equal to 1, since $\hat{\rho}(0) = 1$. \square

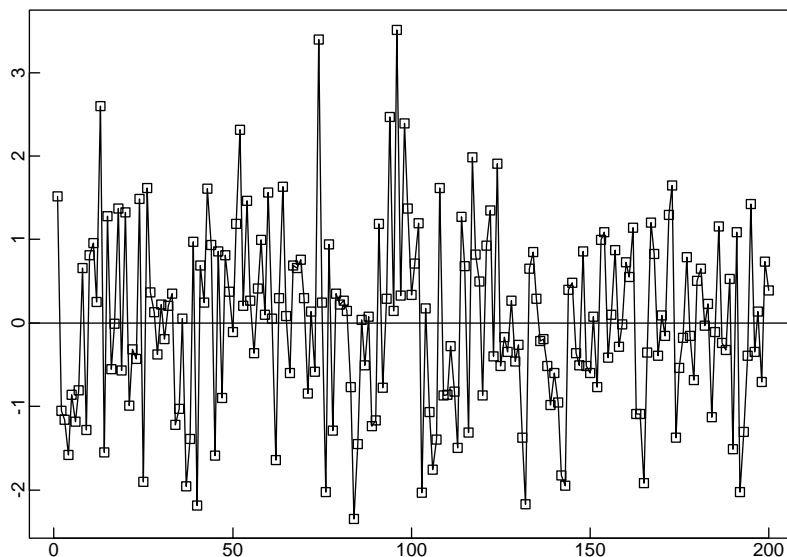


Figure 1-12
200 simulated values
of iid $N(0,1)$ noise.

Example 1.4.6 Figure 1.12 shows 200 simulated values of normally distributed iid $(0, 1)$, denoted by IID $N(0, 1)$, noise. Figure 1.13 shows the corresponding sample autocorrelation function at lags $0, 1, \dots, 40$. Since $\rho(h) = 0$ for $h > 0$, one would also expect the corresponding sample autocorrelations to be near 0. It can be shown, in fact, that for iid noise with finite variance, the sample autocorrelations $\hat{\rho}(h)$, $h > 0$, are approximately IID $N(0, 1/n)$ for n large (see TSTM p. 222). Hence, approximately 95% of the sample autocorrelations should fall between the bounds $\pm 1.96/\sqrt{n}$ (since 1.96 is the .975 quantile of the standard normal distribution). Therefore, in Figure 1.13 we would expect roughly $40(.05) = 2$ values to fall outside the bounds. To simulate 200 values of IID $N(0, 1)$ noise using ITSM, select `File>Project>New>Univariate` then `Model>Simulate`. In the resulting dialog box, enter 200 for the required Number of Observations. (The remaining entries in the dialog box can be left as they are, since the model assumed by ITSM, until you enter another, is IID $N(0, 1)$ noise. If you wish to reproduce exactly the same sequence at a later date, record the Random Number Seed for later use. By specifying different values for the random number seed you can generate independent realizations of your time series.) Click on `OK` and you will see the graph of your simulated series. To see its sample autocorrelation function together with the autocorrelation function of the model that generated it, click on the third yellow button at the top of the screen and you will see the two graphs superimposed (with the latter in red.) The horizontal lines on the graph are the bounds $\pm 1.96/\sqrt{n}$. \square

Remark 5. The sample autocovariance and autocorrelation functions can be computed for *any* data set $\{x_1, \dots, x_n\}$ and are not restricted to observations from a

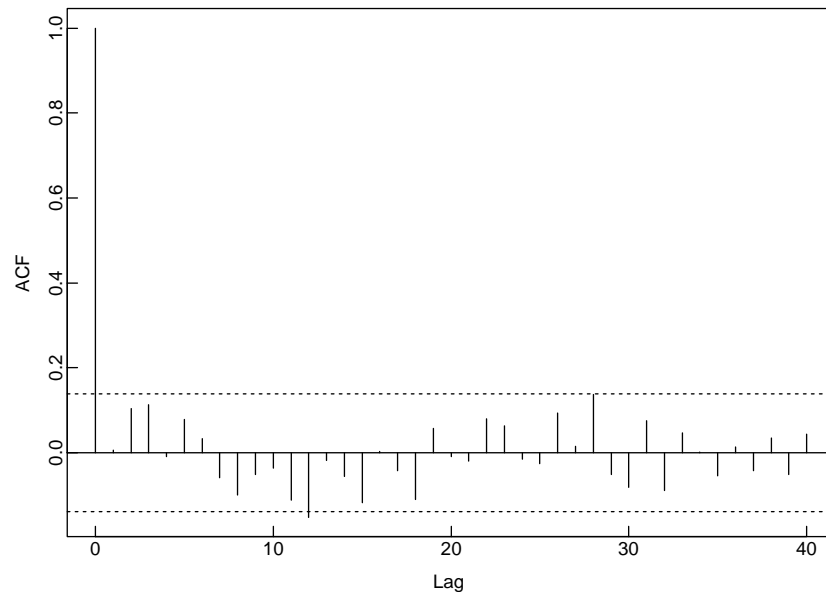


Figure 1-13
The sample autocorrelation function for the data of Figure 1.12 showing the bounds $\pm 1.96/\sqrt{n}$.

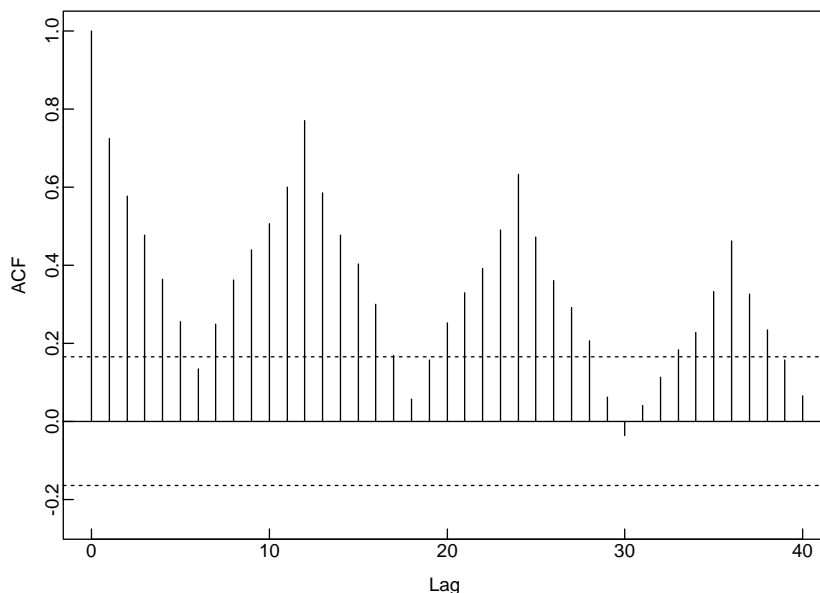


Figure 1-14

The sample autocorrelation function for the Australian red wine sales showing the bounds $\pm 1.96/\sqrt{n}$.

stationary time series. 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 the sample ACF of the Australian red wine sales in Figure 1.14 and Problem 1.9.) Thus $\hat{\rho}(\cdot)$ can be useful as an indicator of nonstationarity (see also Section 6.1). \square

1.4.2 A Model for the Lake Huron Data

As noted earlier, an iid noise model for the residuals $\{y_1, \dots, y_{98}\}$ obtained by fitting a straight line to the Lake Huron data in Example 1.3.5 appears to be inappropriate. This conclusion is confirmed by the sample ACF of the residuals (Figure 1.15), which has three of the first forty values well outside the bounds $\pm 1.96/\sqrt{98}$.

The roughly geometric decay of the first few sample autocorrelations (with $\hat{\rho}(h+1)/\hat{\rho}(h) \approx 0.7$) suggests that an AR(1) series (with $\phi \approx 0.7$) might provide a reasonable model for these residuals. (The form of the ACF for an AR(1) process was computed in Example 1.4.5.)

To explore the appropriateness of such a model, consider the points (y_1, y_2) , (y_2, y_3) , \dots , (y_{97}, y_{98}) plotted in Figure 1.16. The graph does indeed suggest a linear relationship between y_t and y_{t-1} . Using simple least squares estimation to fit a straight line of the form $y_t = ay_{t-1}$, we obtain the model

$$Y_t = .791Y_{t-1} + Z_t, \quad (1.4.3)$$

where $\{Z_t\}$ is iid noise with variance $\sum_{t=2}^{98} (y_t - .791y_{t-1})^2/97 = .5024$. The sample ACF of the estimated noise sequence $z_t = y_t - .791y_{t-1}$, $t = 2, \dots, 98$, is slightly

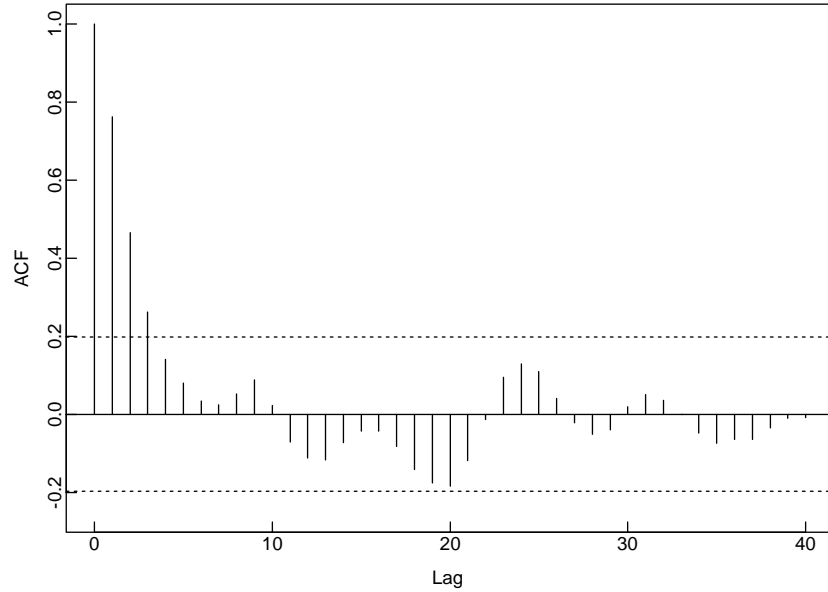


Figure 1-15

The sample autocorrelation function for the Lake Huron residuals of Figure 1.10 showing the bounds $\pm 1.96/\sqrt{n}$.

outside the bounds $\pm 1.96/\sqrt{97}$ at lag 1 ($\hat{\rho}(1) = .216$), but it is inside the bounds for all other lags up to 40. This check that the estimated noise sequence is consistent with the iid assumption of (1.4.3) reinforces our belief in the fitted model. More *goodness of fit* tests for iid noise sequences are described in Section 1.6. The estimated noise sequence $\{z_t\}$ in this example passes them all, providing further support for the model (1.4.3).

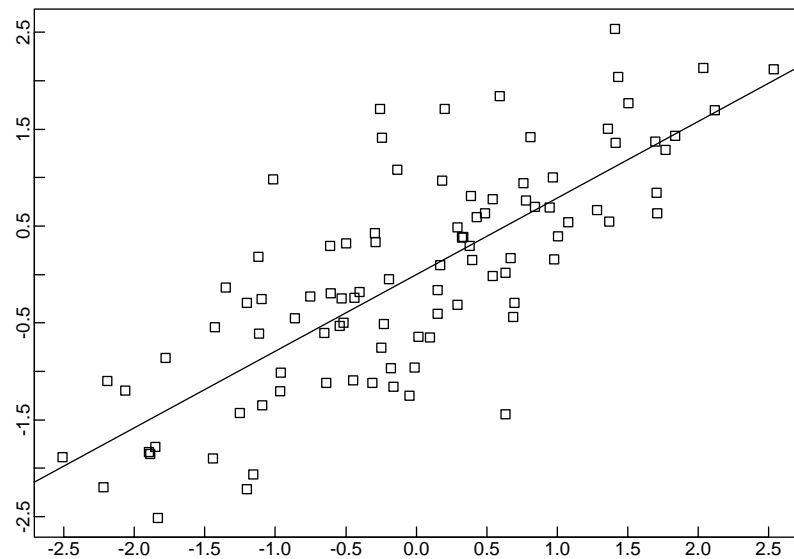


Figure 1-16

Scatter plot of (y_{t-1}, y_t) , $t = 2, \dots, 98$, for the data in Figure 1.10 showing the least squares regression line $y = .791x$.

A better fit to the residuals in equation (1.3.2) is provided by the second-order autoregression

$$Y_t = \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + Z_t, \quad (1.4.4)$$

where $\{Z_t\}$ is iid noise with variance σ^2 . This is analogous to a linear model in which Y_t is regressed on the previous *two* values Y_{t-1} and Y_{t-2} of the time series. The least squares estimates of the parameters ϕ_1 and ϕ_2 , found by minimizing $\sum_{t=3}^{98} (y_t - \phi_1 y_{t-1} - \phi_2 y_{t-2})^2$, are $\hat{\phi}_1 = 1.002$ and $\hat{\phi}_2 = -.2834$. The estimate of σ^2 is $\hat{\sigma}^2 = \sum_{t=3}^{98} (y_t - \hat{\phi}_1 y_{t-1} - \hat{\phi}_2 y_{t-2})^2 / 96 = .4460$, which is approximately 11% smaller than the estimate of the noise variance for the AR(1) model (1.4.3). The improved fit is indicated by the sample ACF of the estimated residuals, $y_t - \hat{\phi}_1 y_{t-1} - \hat{\phi}_2 y_{t-2}$, which falls well within the bounds $\pm 1.96/\sqrt{96}$ for *all* lags up to 40.

1.5 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 (as for example if an observation has been incorrectly recorded). Inspection of a graph may also suggest the possibility of representing the data as a realization of the process (the **classical decomposition** model)

$$X_t = m_t + s_t + Y_t, \quad (1.5.1)$$

where m_t is a slowly changing function known as a **trend component**, s_t is a function with known period d referred to as a **seasonal component**, and Y_t is a **random noise component** that is stationary in the sense of Definition 1.4.2. 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 (1.5.1). Compare, for example, the red wine sales in Figure 1.1 with the transformed data, Figure 1.17, obtained by applying a logarithmic transformation. The transformed data do not exhibit the increasing fluctuation with increasing level that was apparent in the original data. This suggests that the model (1.5.1) is more appropriate for the transformed than for the original series. In this section we shall assume that the model (1.5.1) is appropriate (possibly after a preliminary transformation of the data) and examine some techniques for estimating the components m_t , s_t , and Y_t in the model.

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. We can then use the theory of such processes to find a satisfactory probabilistic

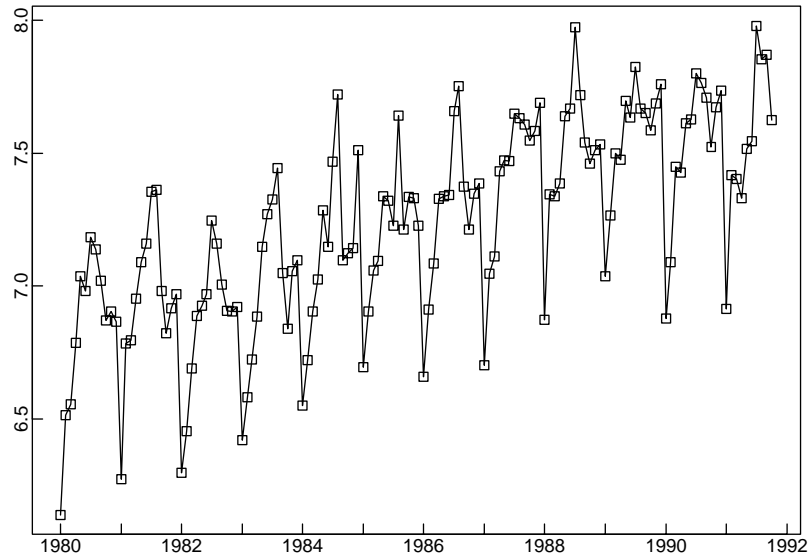


Figure 1-17
The natural logarithms
of the red wine data.

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, developed extensively by Box and Jenkins (1976), is to apply differencing operators repeatedly to the series $\{X_t\}$ until the differenced observations resemble a realization of some stationary time series $\{W_t\}$. We can then use the theory of stationary processes for the modeling, analysis, and prediction of $\{W_t\}$ and hence of the original process. The various stages of this procedure will be discussed in detail in Chapters 5 and 6.

The two approaches to trend and seasonality removal, (1) by estimation of m_t and s_t in (1.5.1) and (2) by differencing the series $\{X_t\}$, will now be illustrated with reference to the data introduced in Section 1.1.

1.5.1 Estimation and Elimination of Trend in the Absence of Seasonality

In the absence of a seasonal component the model (1.5.1) becomes the following.

Nonseasonal Model with Trend:

$$X_t = m_t + Y_t, \quad t = 1, \dots, n, \quad (1.5.2)$$

where $EY_t = 0$.

(If $EY_t \neq 0$, then we can replace m_t and Y_t in (1.5.2) with $m_t + EY_t$ and $Y_t - EY_t$, respectively.)

Method 1: Trend Estimation

Moving average and spectral smoothing are essentially nonparametric methods for trend (or signal) estimation and not for model building. Special smoothing filters can also be designed to remove periodic components as described under Method S1 below. The choice of smoothing filter requires a certain amount of subjective judgment, and it is recommended that a variety of filters be tried in order to get a good idea of the underlying trend. Exponential smoothing, since it is based on a moving average of *past* values only, is often used for forecasting, the smoothed value at the present time being used as the forecast of the next value.

To construct a *model* for the data (with no seasonality) there are two general approaches, both available in ITSM. One is to fit a polynomial trend (by least squares) as described in Method 1(d) below, then to subtract the fitted trend from the data and to find an appropriate stationary time series model for the residuals. The other is to eliminate the trend by differencing as described in Method 2 and then to find an appropriate stationary model for the differenced series. The latter method has the advantage that it usually requires the estimation of fewer parameters and does not rest on the assumption of a trend that remains fixed throughout the observation period. The study of the residuals (or of the differenced series) is taken up in Section 1.6.

(a) *Smoothing with a finite moving average filter.* Let q be a nonnegative integer and consider the two-sided moving average

$$W_t = (2q + 1)^{-1} \sum_{j=-q}^q X_{t-j} \quad (1.5.3)$$

of the process $\{X_t\}$ defined by (1.5.2). Then for $q + 1 \leq t \leq n - q$,

$$W_t = (2q + 1)^{-1} \sum_{j=-q}^q m_{t-j} + (2q + 1)^{-1} \sum_{j=-q}^q Y_{t-j} \approx m_t, \quad (1.5.4)$$

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 (see Problem 1.11).

The moving average thus provides us with the estimates

$$\hat{m}_t = (2q + 1)^{-1} \sum_{j=-q}^q X_{t-j}, \quad q + 1 \leq t \leq n - q. \quad (1.5.5)$$

Since X_t is not observed for $t \leq 0$ or $t > n$, we cannot use (1.5.5) for $t \leq q$ or $t > n - q$. The program ITSM deals with this problem by defining $X_t := X_1$ for $t < 1$ and $X_t := X_n$ for $t > n$.

Example 1.5.1 The result of applying the moving-average filter (1.5.5) with $q = 2$ to the strike data of Figure 1.6 is shown in Figure 1.18. The estimated noise terms $\hat{Y}_t = X_t - \hat{m}_t$ are shown in Figure 1.19. As expected, they show no apparent trend. To apply this filter using ITSM, open the project STRIKES.TSM, select Smooth>Moving Average, specify

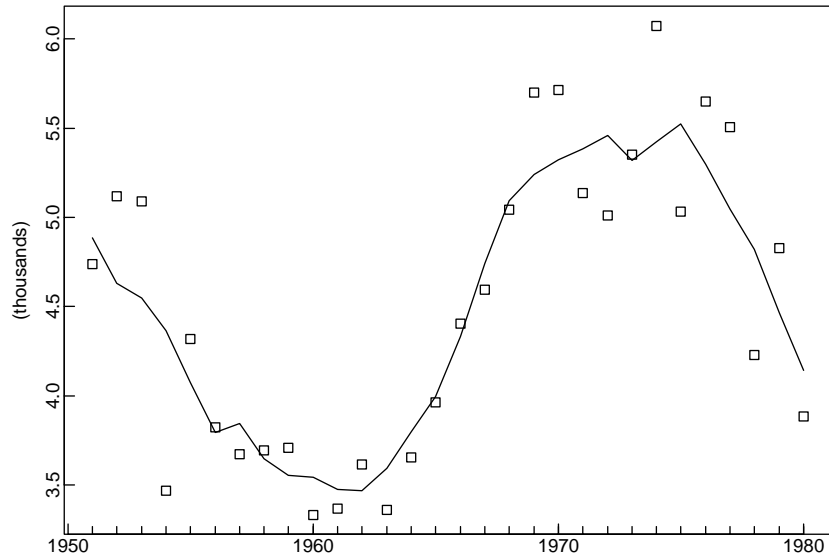


Figure 1-18

Simple 5-term moving average \hat{m}_t of the strike data from Figure 1.6.

2 for the filter order, and enter the weights 1,1,1 for Theta(0), Theta(1), and Theta(2) (these are automatically normalized so that the sum of the weights is one). Then click OK.

It is useful to think of $\{\hat{m}_t\}$ in (1.5.5) as a process obtained from $\{X_t\}$ by application of a linear operator or linear filter $\hat{m}_t = \sum_{j=-\infty}^{\infty} a_j X_{t-j}$ with weights $a_j = (2q + 1)^{-1}$, $-q \leq j \leq q$. This particular filter is a **low-pass** filter in the sense that it takes the

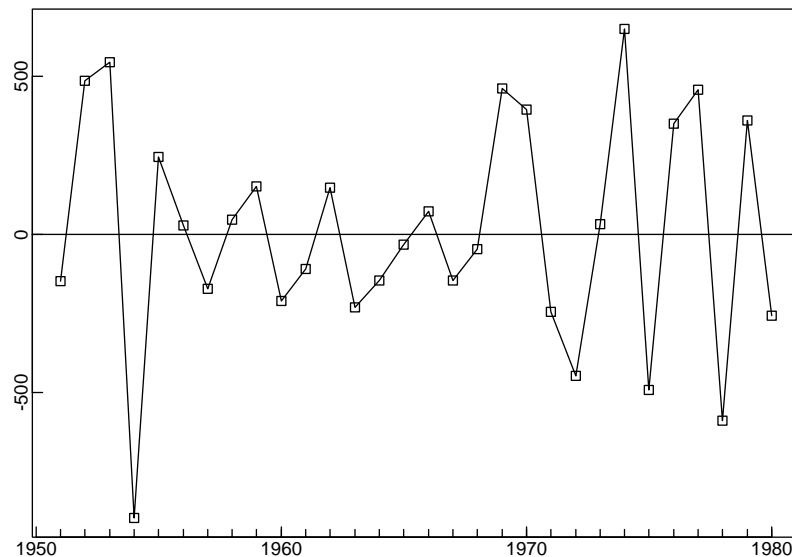


Figure 1-19

Residuals $\hat{Y}_t = X_t - \hat{m}_t$ after subtracting the 5-term moving average from the strike data

Figure 1-20
Smoothing with a
low-pass linear filter.



data $\{X_t\}$ and removes from it the rapidly fluctuating (or high frequency) component $\{\hat{Y}_t\}$ to leave the slowly varying estimated trend term $\{\hat{m}_t\}$ (see Figure 1.20).

The particular filter (1.5.5) is only one of many that could be used for smoothing. For large q , provided that $(2q + 1)^{-1} \sum_{j=-q}^q Y_{t-j} \approx 0$, it not only will attenuate noise but at the same time will allow linear trend functions $m_t = c_0 + c_1 t$ to pass without distortion (see Problem 1.11). However, we must beware of choosing q to be too large, since if m_t is not linear, the filtered process, although smooth, will not be a good estimate of m_t . By clever choice of the weights $\{a_j\}$ it is possible (see Problems 1.12–1.14 and Section 4.3) to design a filter that will not only be effective in attenuating noise in the data, but that will also allow a larger class of trend functions (for example all polynomials of degree less than or equal to 3) to pass through without distortion. The Spencer 15-point moving average is a filter that passes polynomials of degree 3 without distortion. Its weights are

$$a_j = 0, \quad |j| > 7,$$

with

$$a_j = a_{-j}, \quad |j| \leq 7,$$

and

$$[a_0, a_1, \dots, a_7] = \frac{1}{320} [74, 67, 46, 21, 3, -5, -6, -3]. \quad (1.5.6)$$

Applied to the process (1.5.2) with $m_t = c_0 + c_1 t + c_2 t^2 + c_3 t^3$, it gives

$$\sum_{j=-7}^7 a_j X_{t-j} = \sum_{j=-7}^7 a_j m_{t-j} + \sum_{j=-7}^7 a_j Y_{t-j} \approx \sum_{j=-7}^7 a_j m_{t-j} = m_t,$$

where the last step depends on the assumed form of m_t (Problem 1.12). Further details regarding this and other smoothing filters can be found in Kendall and Stuart (1976), Chapter 46.

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

$$\hat{m}_t = \alpha X_t + (1 - \alpha) \hat{m}_{t-1}, \quad t = 2, \dots, n, \quad (1.5.7)$$

and

$$\hat{m}_1 = X_1 \quad (1.5.8)$$

can be computed using ITSM by selecting Smooth>Exponential and specifying the value of α . Application of (1.5.7) and (1.5.8) is often referred to as exponential smoothing, since the recursions imply that for $t \geq 2$, $\hat{m}_t = \sum_{j=0}^{t-2} \alpha(1-\alpha)^j X_{t-j} + (1-\alpha)^{t-1} X_1$, a weighted moving average of X_t, X_{t-1}, \dots , with weights decreasing exponentially (except for the last one).

(c) *Smoothing by elimination of high-frequency components.* The option Smooth>FFT in the program ITSM allows us to smooth an arbitrary series by elimination of the high-frequency components of its Fourier series expansion (see Section 4.2). This option was used in Example 1.1.4, where we chose to retain the fraction $f = .035$ of the frequency components of the series in order to estimate the underlying signal. (The choice $f = 1$ would have left the series unchanged.)

Example 1.5.2

In Figures 1.21 and 1.22 we show the results of smoothing the strike data by exponential smoothing with parameter $\alpha = 0.4$ (see (1.5.7)) and by high-frequency elimination with $f = 0.4$, i.e., by eliminating a fraction 0.6 of the Fourier components at the top of the frequency range. These should be compared with the simple 5-term moving average smoothing shown in Figure 1.18. Experimentation with different smoothing parameters can easily be carried out using the program ITSM. The exponentially smoothed value of the last observation is frequently used to forecast the next data value. The program automatically selects an optimal value of α for this purpose if α is specified as -1 in the exponential smoothing dialog box. \square

(d) *Polynomial fitting.* In Section 1.3.2 we showed how a trend of the form $m_t = a_0 + a_1 t + a_2 t^2$ can be fitted to the data $\{x_1, \dots, x_n\}$ by choosing the parameters

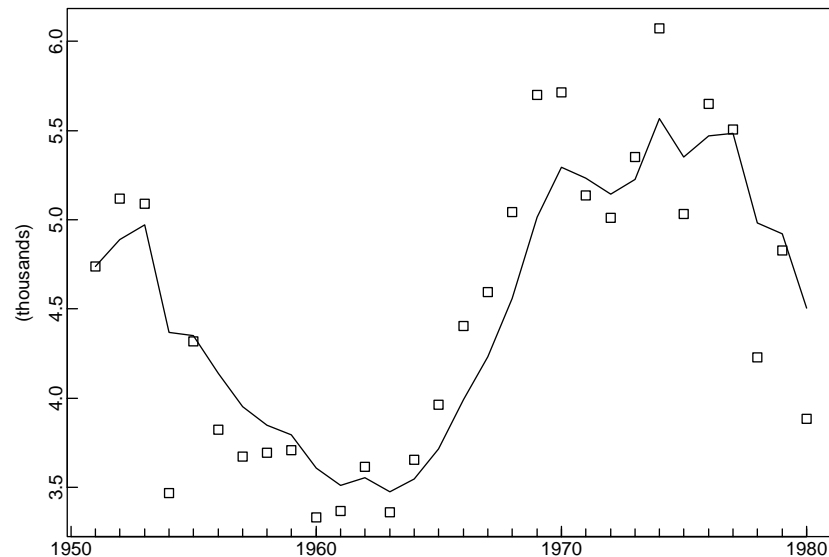


Figure 1-21
Exponentially smoothed
strike data with $\alpha = 0.4$.

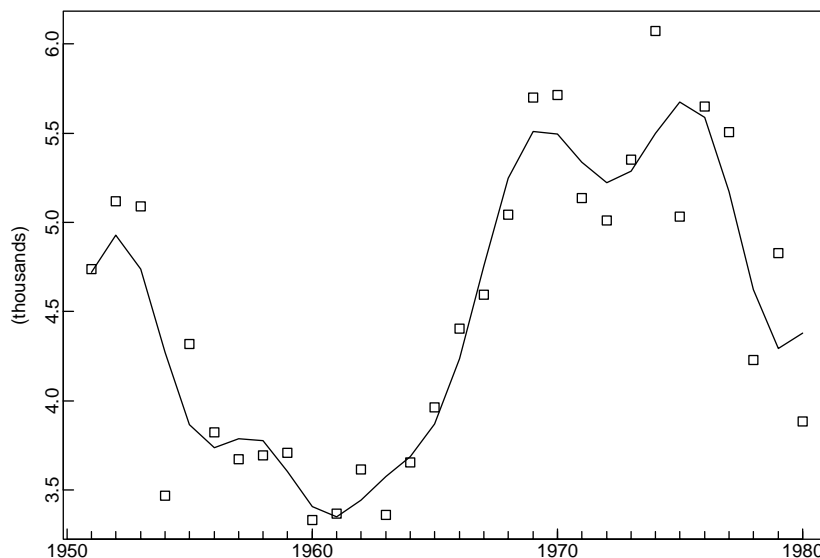


Figure 1-22
Strike data smoothed
by elimination of high
frequencies with $f = 0.4$.

a_0 , a_1 , and a_2 to minimize the sum of squares, $\sum_{t=1}^n (x_t - m_t)^2$ (see Example 1.3.4). The method of least squares estimation can also be used to estimate higher-order polynomial trends in the same way. The Regression option of ITSM allows least squares fitting of polynomial trends of order up to 10 (together with up to four harmonic terms; see Example 1.3.6). It also allows generalized least squares estimation (see Section 6.6), in which correlation between the residuals is taken into account.

Method 2: Trend Elimination by Differencing

Instead of attempting to remove the noise by smoothing as in Method 1, we now attempt to eliminate the trend term by differencing. We define the lag-1 difference operator ∇ by

$$\nabla X_t = X_t - X_{t-1} = (1 - B)X_t, \quad (1.5.9)$$

where B is the backward shift operator,

$$BX_t = X_{t-1}. \quad (1.5.10)$$

Powers of the operators B and ∇ are defined in the obvious way, i.e., $B^j(X_t) = X_{t-j}$ and $\nabla^j(X_t) = \nabla(\nabla^{j-1}(X_t))$, $j \geq 1$, with $\nabla^0(X_t) = X_t$. Polynomials in B and ∇ are manipulated in precisely the same way as polynomial functions of real variables. For example,

$$\begin{aligned} \nabla^2 X_t &= \nabla(\nabla(X_t)) = (1 - B)(1 - B)X_t = (1 - 2B + B^2)X_t \\ &= X_t - 2X_{t-1} + X_{t-2}. \end{aligned}$$

If the operator ∇ is applied to a linear trend function $m_t = c_0 + c_1t$, then we obtain the constant function $\nabla m_t = m_t - m_{t-1} = c_0 + c_1t - (c_0 + c_1(t-1)) = c_1$. In the same way any polynomial trend of degree k can be reduced to a constant by application of the operator ∇^k (Problem 1.10). For example, if $X_t = m_t + Y_t$, where $m_t = \sum_{j=0}^k c_j t^j$ and Y_t is stationary with mean zero, application of ∇^k gives

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

a stationary process with mean $k!c_k$. These considerations suggest the possibility, given any sequence $\{x_t\}$ of data, of applying the operator ∇ repeatedly until we find a sequence $\{\nabla^k x_t\}$ that can plausibly be modeled as a realization of a stationary process. It is often found in practice that the order k of differencing required is quite small, frequently one or two. (This relies on the fact that many functions can be well approximated, on an interval of finite length, by a polynomial of reasonably low degree.)

Example 1.5.3

Applying the operator ∇ to the population values $\{x_t, t = 1, \dots, 20\}$ of Figure 1.5, we find that two differencing operations are sufficient to produce a series with no apparent trend. (To carry out the differencing using ITSM, select `Transform>Difference`, enter the value 1 for the differencing lag, and click OK.) This replaces the original series $\{x_t\}$ by the once-differenced series $\{x_t - x_{t-1}\}$. Repetition of these steps gives the twice-differenced series $\nabla^2 x_t = x_t - 2x_{t-1} + x_{t-2}$, plotted in Figure 1.23. Notice that the magnitude of the fluctuations in $\nabla^2 x_t$ increases with the value of x_t . This effect can be suppressed by first taking natural logarithms, $y_t = \ln x_t$, and then applying the operator ∇^2 to the series $\{y_t\}$. (See also Figures 1.1 and 1.17.) \square

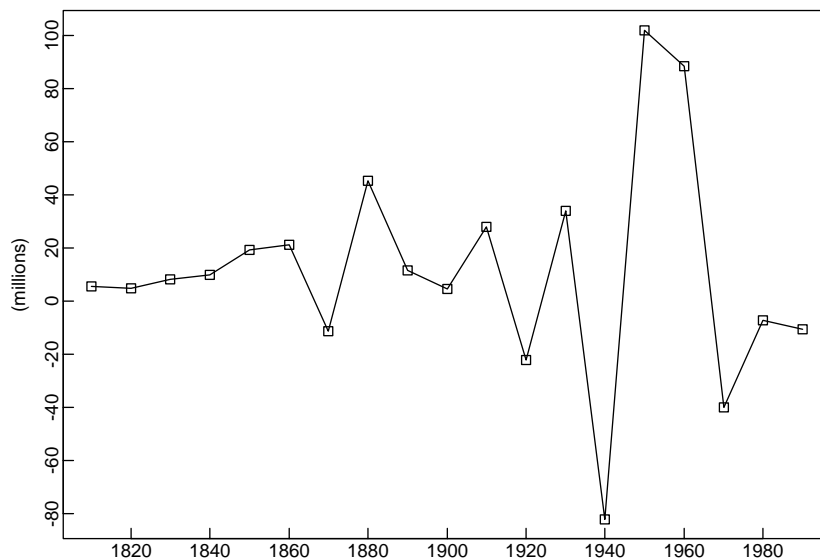


Figure 1-23

The twice-differenced series derived from the population data of Figure 1.5.

1.5.2 Estimation and Elimination of Both Trend and Seasonality

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.

Classical Decomposition Model

$$X_t = m_t + s_t + Y_t, \quad t = 1, \dots, n, \quad (1.5.11)$$

where $EY_t = 0$, $s_{t+d} = s_t$, and $\sum_{j=1}^d s_j = 0$.

We shall illustrate these methods with reference to the accidental deaths data of Example 1.1.3, for which the period d of the seasonal component is clearly 12.

Method S1: Estimation of Trend and Seasonal Components

The method we are about to describe is used in the Transform>Classical option of ITSM.

Suppose we have observations $\{x_1, \dots, x_n\}$. 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

$$\hat{m}_t = (0.5x_{t-q} + x_{t-q+1} + \dots + x_{t+q-1} + 0.5x_{t+q})/d, \quad q < t \leq n - q. \quad (1.5.12)$$

If the period is odd, say $d = 2q + 1$, then we use the simple moving average (1.5.5).

The second step is to estimate the seasonal component. For each $k = 1, \dots, d$, we compute the average w_k of the deviations $\{(x_{k+jd} - \hat{m}_{k+jd}), q < k + jd \leq n - q\}$. Since these average deviations do not necessarily sum to zero, we estimate the seasonal component s_k as

$$\hat{s}_k = w_k - d^{-1} \sum_{i=1}^d w_i, \quad k = 1, \dots, d, \quad (1.5.13)$$

and $\hat{s}_k = \hat{s}_{k-d}$, $k > d$.

The *deseasonalized* data is then defined to be the original series with the estimated seasonal component removed, i.e.,

$$d_t = x_t - \hat{s}_t, \quad t = 1, \dots, n. \quad (1.5.14)$$

Finally, we reestimate the trend from the deseasonalized data $\{d_t\}$ using one of the methods already described. The program ITSM allows you to fit a least squares polynomial trend \hat{m} to the deseasonalized series. In terms of this reestimated trend and the estimated seasonal component, the estimated noise series is then given by

$$\hat{Y}_t = x_t - \hat{m}_t - \hat{s}_t, \quad t = 1, \dots, n.$$

The reestimation of the trend is done in order to have a parametric form for the trend that can be extrapolated for the purposes of prediction and simulation.

Example 1.5.4 Figure 1.24 shows the deseasonalized accidental deaths data obtained from ITSM by reading in the series DEATHS.TSM, selecting Transform>Classical, checking *only* the box marked Seasonal Fit, entering 12 for the period, and clicking OK. The estimated seasonal component \hat{s}_t , shown in Figure 1.25, is obtained by selecting Transform>Show Classical Fit. (Except for having a mean of zero, this estimate is very similar to the harmonic regression function with frequencies $2\pi/12$ and $2\pi/6$ displayed in Figure 1.11.) The graph of the deseasonalized data suggests the presence of an additional quadratic trend function. In order to fit such a trend to the deseasonalized data, select Transform>Undo Classical to retrieve the original data and then select Transform>Classical and check the boxes marked Seasonal Fit and Polynomial Trend, entering 12 for the period and selecting Quadratic for the trend. Then click OK and you will obtain the trend function

$$\hat{m}_t = 9952 - 71.82t + 0.8260t^2, \quad 1 \leq t \leq 72.$$

At this point the data stored in ITSM consists of the estimated noise

$$\hat{Y}_t = x_t - \hat{m}_t - \hat{s}_t, \quad t = 1, \dots, 72,$$

obtained by subtracting the estimated seasonal and trend components from the original data. \square

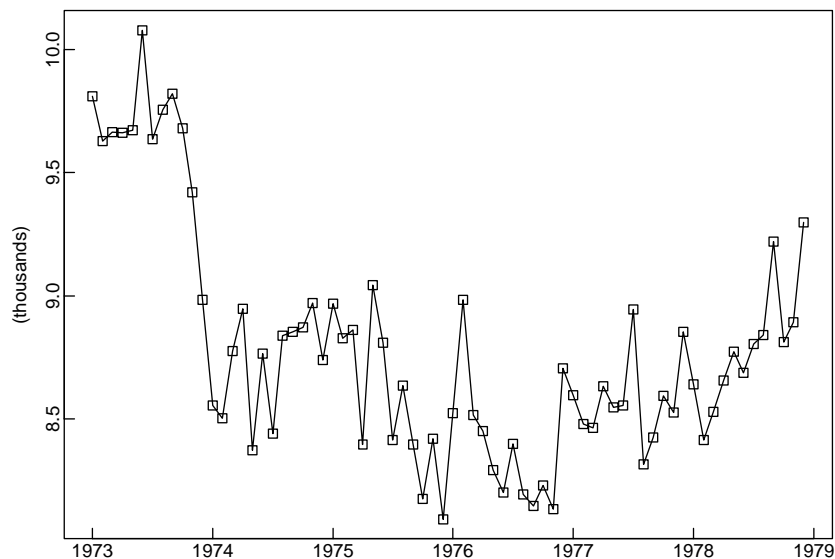


Figure 1-24
The deseasonalized
accidental deaths
data from ITSM.

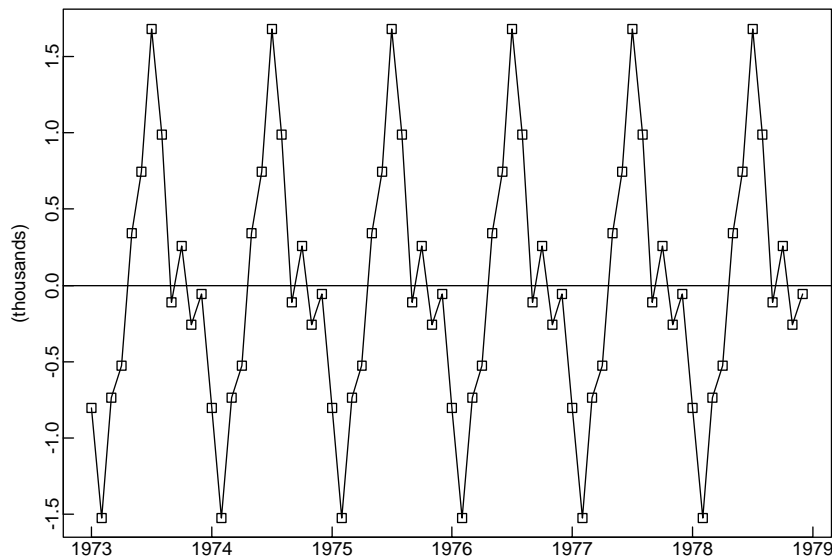


Figure 1-25
The estimated seasonal component of the accidental deaths data from ITSM.

Method S2: Elimination of Trend and Seasonal Components by Differencing

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

$$\nabla_d X_t = X_t - X_{t-d} = (1 - B^d)X_t. \quad (1.5.15)$$

(This operator should not be confused with the operator $\nabla^d = (1 - B)^d$ defined earlier.)

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 1.5.5

Figure 1.26 shows the result of applying the operator ∇_{12} to the accidental deaths data. The graph is obtained from ITSM by opening DEATHS.TSM, selecting Transform>Difference, entering lag 12, and clicking OK. The seasonal component evident in Figure 1.3 is absent from the graph of $\nabla_{12}x_t$, $13 \leq t \leq 72$. However, there still appears to be a nondecreasing trend. If we now apply the operator ∇ to $\{\nabla_{12}x_t\}$ by again selecting Transform>Difference, this time with lag one, we obtain the graph

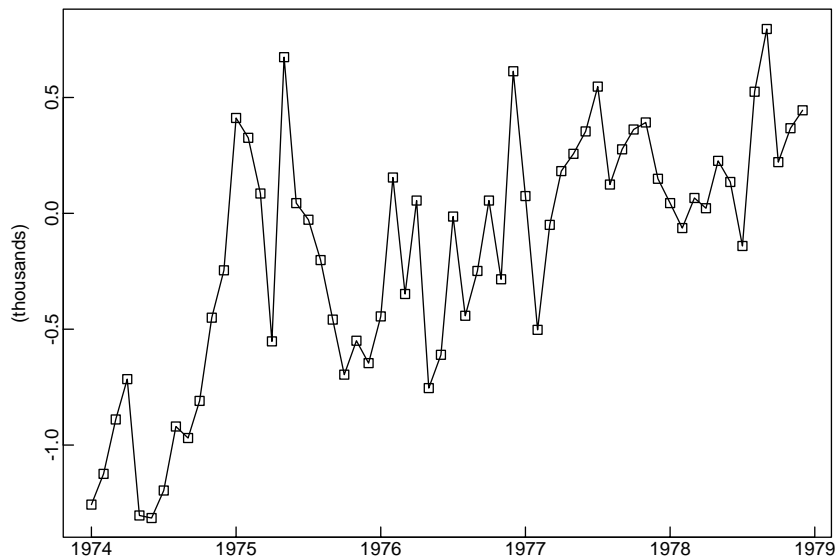


Figure 1-26
The differenced series
 $\{\nabla_{12}x_t, t = 13, \dots, 72\}$
derived from the monthly
accidental deaths
 $\{x_t, t = 1, \dots, 72\}$.

of $\nabla\nabla_{12}x_t$, $14 \leq t \leq 72$, shown in Figure 1.27, which has no apparent trend or seasonal component. In Chapter 5 we shall show that this doubly differenced series can in fact be well represented by a stationary time series model. \square

In this section we have discussed a variety of methods for estimating and/or removing trend and seasonality. The particular method chosen for any given data set will depend on a number of factors including whether or not estimates of the

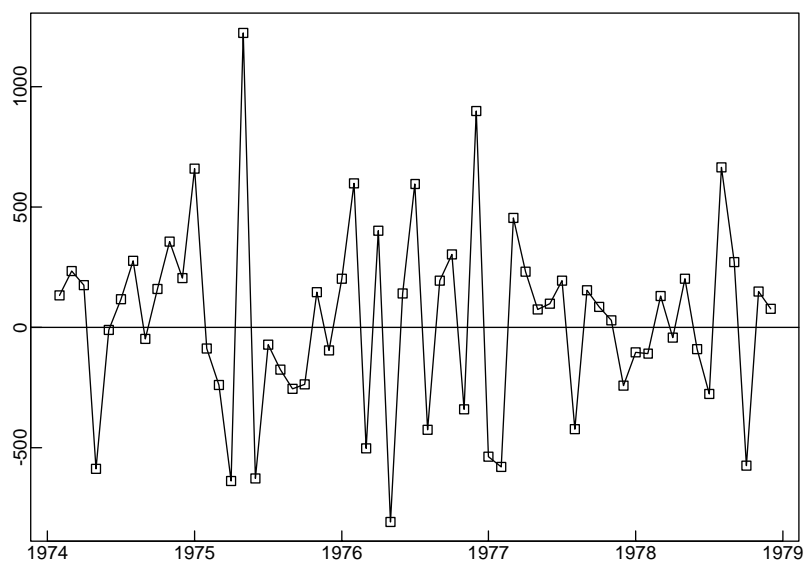


Figure 1-27
The differenced series
 $\{\nabla\nabla_{12}x_t, t = 14, \dots, 72\}$
derived from the monthly
accidental deaths
 $\{x_t, t = 1, \dots, 72\}$.

components of the series are required and whether or not it appears that the data contain a seasonal component that does not vary with time. The program ITSM allows two options under the Transform menu:

1. “classical decomposition,” in which trend and/or seasonal components are estimated and subtracted from the data to generate a noise sequence, and
2. “differencing,” in which trend and/or seasonal components are removed from the data by repeated differencing at one or more lags in order to generate a noise sequence.

A third option is to use the Regression menu, possibly after applying a Box–Cox transformation. Using this option we can (see Example 1.3.6)

3. fit a sum of harmonics and a polynomial trend to generate a noise sequence that consists of the residuals from the regression.

In the next section we shall examine some techniques for deciding whether or not the noise sequence so generated differs significantly from iid noise. If the noise sequence *does* have sample autocorrelations significantly different from zero, then we can take advantage of this serial dependence to forecast future noise values in terms of past values by modeling the noise as a stationary time series.

1.6 Testing the Estimated Noise Sequence

The objective of the data transformations described in Section 1.5 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 simple tests for checking the hypothesis that the residuals from Section 1.5 are observed values of independent and identically distributed random variables. If they are, then our work is done. If not, then we must use the theory of stationary processes to be developed in later chapters to find a more appropriate model.

(a) *The sample autocorrelation function.* For large n , the sample autocorrelations of an iid sequence Y_1, \dots, Y_n with finite variance are approximately iid with distribution $N(0, 1/n)$ (see TSTM p. 222). Hence, if y_1, \dots, 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. The bounds $\pm 1.96/\sqrt{n}$ are automatically plotted when the sample autocorrelation function is computed by the program ITSM.

(b) *The portmanteau test.* Instead of checking to see whether each sample autocorrelation $\hat{\rho}(j)$ falls inside the bounds defined in (a) above, it is also possible to consider the single statistic

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

If Y_1, \dots, Y_n is a finite-variance iid sequence, then by the same result used in (a), Q is approximately distributed as the sum of squares of the independent $N(0, 1)$ random variables, $\sqrt{n}\hat{\rho}(j)$, $j = 1, \dots, h$, i.e., as chi-squared with 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 $Q > \chi_{1-\alpha}^2(h)$, where $\chi_{1-\alpha}^2(h)$ is the $1 - \alpha$ quantile of the chi-squared distribution with h degrees of freedom. The program ITSM conducts a refinement of this test, formulated by Ljung and Box (1978), in which Q is replaced by

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

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

Another portmanteau test, formulated by McLeod and Li (1983), can be used as a further test for the iid hypothesis, since if the data are iid, then the squared data are also iid. It is based on the same statistic used for the Ljung–Box test, except that the sample autocorrelations of the data are replaced by the sample autocorrelations of the *squared* data, $\hat{\rho}_{WW}(h)$, giving

$$Q_{\text{ML}} = n(n+2) \sum_{k=1}^h \hat{\rho}_{WW}^2(k)/(n-k).$$

The hypothesis of iid data is then rejected at level α if the observed value of Q_{ML} is larger than the $1 - \alpha$ quantile of the $\chi^2(h)$ distribution.

(c) *The turning point test.* If y_1, \dots, y_n is a sequence of observations, we say that there is a turning point at time i , $1 < i < n$, 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 $\frac{2}{3}$, the expected value of T is

$$\mu_T = E(T) = 2(n - 2)/3.$$

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

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

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 of $T - \mu_T$ much smaller than zero indicates a positive correlation between neighboring observations. For an iid sequence with n large, it can be shown that

$$T \text{ is approximately } N(\mu_T, \sigma_T^2).$$

This means we can 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. (A commonly used value of α is .05, for which the corresponding value of $\Phi_{1-\alpha/2}$ is 1.96.)

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

$$\mu_S = ES = \frac{1}{2}(n - 1).$$

It can also be shown, under the same assumption, that

$$\sigma_S^2 = \text{Var}(S) = (n + 1)/12,$$

and that for large n ,

$$S \text{ is approximately } N(\mu_S, \sigma_S^2).$$

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.* The rank 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 $j > i$, $i = 1, \dots, n - 1$. There is a total of $\binom{n}{2} = \frac{1}{2}n(n - 1)$ pairs (i, j) such that $j > i$. For an iid sequence $\{Y_1, \dots, Y_n\}$, each event $\{Y_j > Y_i\}$ has probability $\frac{1}{2}$, and the mean of P is therefore

$$\mu_P = \frac{1}{4}n(n - 1).$$

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

$$\sigma_P^2 = n(n-1)(2n+5)/72$$

and that for large n ,

$$P \text{ is approximately } N(\mu_P, \sigma_P^2)$$

(see Kendall and Stuart, 1976). A large positive (negative) value of $P - \mu_P$ indicates the presence of an increasing (decreasing) trend in the data. The assumption that $\{y_j\}$ is a sample from an iid sequence is therefore rejected at level $\alpha = 0.05$ if $|P - \mu_P|/\sigma_P > \Phi_{1-\alpha/2} = 1.96$.

(f) *Fitting an autoregressive model.* A further test that can be carried out using the program ITSM is to fit an autoregressive model to the data using the Yule–Walker algorithm (discussed in Section 5.1.1) and choosing the order which minimizes the AICC statistic (see Section 5.5). A selected order equal to zero suggests that the data is white noise.

(g) *Checking for normality.* If the noise process is Gaussian, i.e., if all of its joint distributions are normal, then stronger conclusions can be drawn when a model is fitted to the data. The following test enables us to check whether it is reasonable to assume that observations from an iid sequence are also Gaussian.

Let $Y_{(1)} < Y_{(2)} < \dots < Y_{(n)}$ be the order statistics of a random sample Y_1, \dots, Y_n from the distribution $N(\mu, \sigma^2)$. If $X_{(1)} < X_{(2)} < \dots < X_{(n)}$ are the order statistics from a $N(0, 1)$ sample of size n , then

$$EY_{(j)} = \mu + \sigma m_j,$$

where $m_j = EX_{(j)}$, $j = 1, \dots, n$. The graph of the points $(m_1, Y_{(1)}), \dots, (m_n, Y_{(n)})$ is called a Gaussian **qq plot** and can be displayed in ITSM by clicking on the yellow button labeled QQ. If the normal assumption is correct, the Gaussian qq plot should be approximately linear. Consequently, the squared correlation of the points $(m_i, Y_{(i)})$, $i = 1, \dots, n$, should be near 1. The assumption of normality is therefore rejected if the squared correlation R^2 is sufficiently small. If we approximate m_i by $\Phi^{-1}((i-.5)/n)$ (see Mage, 1982 for some alternative approximations), then R^2 reduces to

$$R^2 = \frac{(\sum_{i=1}^n (Y_{(i)} - \bar{Y}) \Phi^{-1}(\frac{i-.5}{n}))^2}{\sum_{i=1}^n (Y_{(i)} - \bar{Y})^2 \sum_{i=1}^n (\Phi^{-1}(\frac{i-.5}{n}))^2},$$

where $\bar{Y} = n^{-1}(Y_1 + \dots + Y_n)$. Percentage points for the distribution of R^2 , assuming normality of the sample values, are given by Shapiro and Francia (1972) for sample sizes $n < 100$. For $n = 200$, $P(R^2 < .987) = .05$ and $P(R^2 < .989) = .10$. For larger values of n the Jarque-Bera test for normality can be used (see Section 5.3.3).

Example 1.6.1

If we did not know in advance how the signal plus noise data of Example 1.1.4 were generated, we might suspect that they came from an iid sequence. We can check this hypothesis with the aid of the tests (a)–(f) introduced above.

(a) The sample autocorrelation function (Figure 1.28) is obtained from ITSM by opening the project SIGNAL.TSM and clicking on the second yellow button at the top of the ITSM window. Observing that 25% of the autocorrelations are outside the bounds $\pm 1.96/\sqrt{200}$, we reject the hypothesis that the series is iid.

The remaining tests (b), (c), (d), (e), and (f) are performed by choosing the option Statistics>Residual Analysis>Tests of Randomness. (Since no model has been fitted to the data, the residuals are the same as the data themselves.)

(b) The sample value of the Ljung–Box statistic Q_{LB} with $h = 20$ is 51.84. Since the corresponding p -value (displayed by ITSM) is $.00012 < .05$, we reject the iid hypothesis at level .05. The p -value for the McLeod–Li statistic Q_{ML} is 0.717. The McLeod–Li statistic does therefore not provide sufficient evidence to reject the iid hypothesis at level .05.

(c) The sample value of the turning-point statistic T is 138, and the asymptotic distribution under the iid hypothesis (with sample size $n = 200$) is $N(132, 35.3)$. Thus $|T - \mu_T|/\sigma_T = 1.01$, corresponding to a computed p -value of .312. On the basis of the value of T there is therefore not sufficient evidence to reject the iid hypothesis at level .05.

(d) The sample value of the difference-sign statistic S is 101, and the asymptotic distribution under the iid hypothesis (with sample size $n = 200$) is $N(99.5, 16.7)$. Thus $|S - \mu_S|/\sigma_S = 0.38$, corresponding to a computed p -value of 0.714. On the basis of the value of S there is therefore not sufficient evidence to reject the iid hypothesis at level .05.

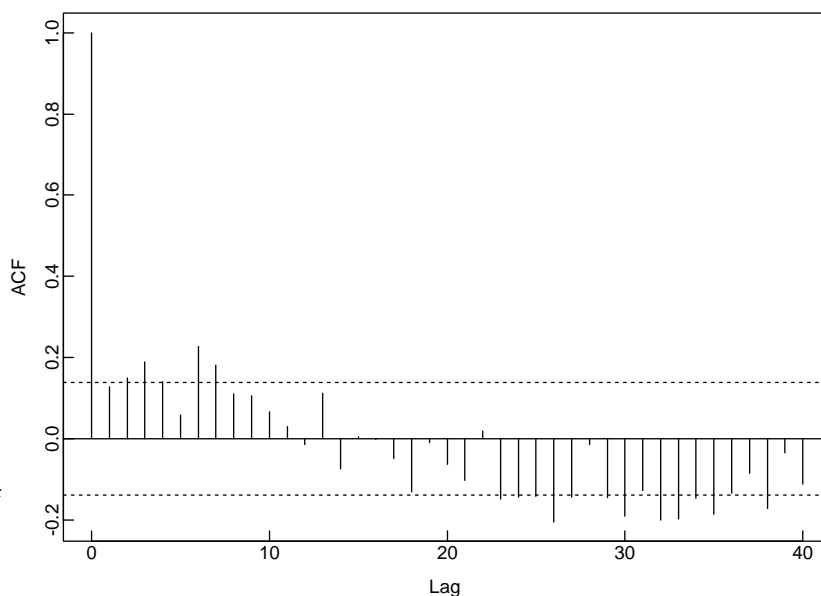


Figure 1-28

The sample autocorrelation function for the data of Example 1.1.4 showing the bounds $\pm 1.96/\sqrt{n}$.

(e) The sample value of the rank statistic P is 10310, and the asymptotic distribution under the iid hypothesis (with $n = 200$) is $N(9950, 2.239 \times 10^5)$. Thus $|P - \mu_P|/\sigma_P = 0.76$, corresponding to a computed p -value of 0.447. On the basis of the value of P there is therefore not sufficient evidence to reject the iid hypothesis at level .05.

(f) The minimum-AICC Yule–Walker autoregressive model for the data is of order seven, supporting the evidence provided by the sample ACF and Ljung–Box tests against the iid hypothesis.

Thus, although not all of the tests detect significant deviation from iid behavior, the sample autocorrelation, the Ljung–Box statistic, and the fitted autoregression provide strong evidence against it, causing us to reject it (correctly) in this example. \square

The general strategy in applying the tests described in this section is to check them all and to proceed with caution if any of them suggests a serious deviation from the iid hypothesis. (Remember that as you increase the number of tests, the probability that *at least one* rejects the null hypothesis when it is true increases. You should therefore not necessarily reject the null hypothesis on the basis of one test result only.)

Problems

1.1. Let X and Y be two random variables with $E(Y) = \mu$ and $EY^2 < \infty$.

a. Show that the constant c that minimizes $E(Y - c)^2$ is $c = \mu$.

b. Deduce that the random variable $f(X)$ that minimizes $E[(Y - f(X))^2|X]$ is

$$f(X) = E[Y|X].$$

c. Deduce that the random variable $f(X)$ that minimizes $E(Y - f(X))^2$ is also

$$f(X) = E[Y|X].$$

1.2. (Generalization of Problem 1.1.) Suppose that X_1, X_2, \dots is a sequence of random variables with $E(X_i^2) < \infty$ and $E(X_i) = \mu$.

a. Show that the random variable $f(X_1, \dots, X_n)$ that minimizes $E[(X_{n+1} - f(X_1, \dots, X_n))^2|X_1, \dots, X_n]$ is

$$f(X_1, \dots, X_n) = E[X_{n+1}|X_1, \dots, X_n].$$

b. Deduce that the random variable $f(X_1, \dots, X_n)$ that minimizes $E[(X_{n+1} - f(X_1, \dots, X_n))^2]$ is also

$$f(X_1, \dots, X_n) = E[X_{n+1}|X_1, \dots, X_n].$$

c. If X_1, X_2, \dots is iid with $E(X_i^2) < \infty$ and $EX_i = \mu$, where μ is known, what is the minimum mean squared error predictor of X_{n+1} in terms of X_1, \dots, X_n ?

- d. Under the conditions of part (c) show that the best linear unbiased estimator of μ in terms of X_1, \dots, X_n is $\bar{X} = \frac{1}{n}(X_1 + \dots + X_n)$. ($\hat{\mu}$ said to be an unbiased estimator of μ if $E\hat{\mu} = \mu$ for all μ .)
- e. Under the conditions of part (c) show that \bar{X} is the best linear predictor of X_{n+1} that is unbiased for μ .
- f. If X_1, X_2, \dots is iid with $E(X_i^2) < \infty$ and $EX_i = \mu$, and if $S_0 = 0$, $S_n = X_1 + \dots + X_n$, $n = 1, 2, \dots$, what is the minimum mean squared error predictor of S_{n+1} in terms of S_1, \dots, S_n ?

- 1.3.** Show that a strictly stationary process with $E(X_i^2) < \infty$ is weakly stationary.
- 1.4.** Let $\{Z_t\}$ be a sequence of independent normal random variables, each with mean 0 and variance σ^2 , and let a , b , and c be constants. Which, if any, of the following processes are stationary? For each stationary process specify the mean and autocovariance function.
- $X_t = a + bZ_t + cZ_{t-2}$
 - $X_t = Z_1 \cos(ct) + Z_2 \sin(ct)$
 - $X_t = Z_t \cos(ct) + Z_{t-1} \sin(ct)$
 - $X_t = a + bZ_0$
 - $X_t = Z_0 \cos(ct)$
 - $X_t = Z_t Z_{t-1}$

- 1.5.** Let $\{X_t\}$ be the moving-average process of order 2 given by

$$X_t = Z_t + \theta Z_{t-2},$$

where $\{Z_t\}$ is WN(0, 1).

- Find the autocovariance and autocorrelation functions for this process when $\theta = .8$.
 - Compute the variance of the sample mean $(X_1 + X_2 + X_3 + X_4)/4$ when $\theta = .8$.
 - Repeat (b) when $\theta = -.8$ and compare your answer with the result obtained in (b).
- 1.6.** Let $\{X_t\}$ be the AR(1) process defined in Example 1.4.5.
- Compute the variance of the sample mean $(X_1 + X_2 + X_3 + X_4)/4$ when $\phi = .9$ and $\sigma^2 = 1$.
 - Repeat (a) when $\phi = -.9$ and compare your answer with the result obtained in (a).
- 1.7.** If $\{X_t\}$ and $\{Y_t\}$ are uncorrelated stationary sequences, i.e., if X_r and Y_s are uncorrelated for every r and s , show that $\{X_t + Y_t\}$ is stationary with autocovariance function equal to the sum of the autocovariance functions of $\{X_t\}$ and $\{Y_t\}$.

1.8. Let $\{Z_t\}$ be IID $N(0, 1)$ noise and define

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

- Show that $\{X_t\}$ is $WN(0, 1)$ but not $iid(0, 1)$ noise.
- Find $E(X_{n+1}|X_1, \dots, X_n)$ for n odd and n even and compare the results.

1.9. Let $\{x_1, \dots, x_n\}$ be observed values of a time series at times $1, \dots, n$, and let $\hat{\rho}(h)$ be the sample ACF at lag h as in Definition 1.4.4.

- If $x_t = a + bt$, where a and b are constants and $b \neq 0$, show that for each fixed $h \geq 1$,

$$\hat{\rho}(h) \rightarrow 1 \text{ as } n \rightarrow \infty.$$

- If $x_t = c \cos(\omega t)$, where c and ω are constants ($c \neq 0$ and $\omega \in (-\pi, \pi]$), show that for each fixed h ,

$$\hat{\rho}(h) \rightarrow \cos(\omega h) \text{ as } n \rightarrow \infty.$$

1.10. If $m_t = \sum_{k=0}^p c_k t^k$, $t = 0, \pm 1, \dots$, show that ∇m_t is a polynomial of degree $p - 1$ in t and hence that $\nabla^{p+1} m_t = 0$.

1.11. Consider the simple moving-average filter with weights $a_j = (2q + 1)^{-1}$, $-q \leq j \leq q$.

- If $m_t = c_0 + c_1 t$, show that $\sum_{j=-q}^q a_j m_{t-j} = m_t$.
- If Z_t , $t = 0, \pm 1, \pm 2, \dots$, are independent random variables with mean 0 and variance σ^2 , show that the moving average $A_t = \sum_{j=-q}^q a_j Z_{t-j}$ is “small” for large q in the sense that $EA_t = 0$ and $\text{Var}(A_t) = \sigma^2/(2q + 1)$.

1.12. a. Show that a linear filter $\{a_j\}$ passes an arbitrary polynomial of degree k without distortion, i.e., that

$$m_t = \sum_j a_j m_{t-j}$$

for all k th-degree polynomials $m_t = c_0 + c_1 t + \dots + c_k t^k$, if and only if

$$\begin{cases} \sum_j a_j = 1 & \text{and} \\ \sum_j j^r a_j = 0, & \text{for } r = 1, \dots, k. \end{cases}$$

- Deduce that the Spencer 15-point moving-average filter $\{a_j\}$ defined by (1.5.6) passes arbitrary third-degree polynomial trends without distortion.

- 1.13.** Find a filter of the form $1 + \alpha B + \beta B^2 + \gamma B^3$ (i.e., find α , β , and γ) that passes linear trends without distortion and that eliminates arbitrary seasonal components of period 2.
- 1.14.** Show that the filter with coefficients $[a_{-2}, a_{-1}, a_0, a_1, a_2] = \frac{1}{9}[-1, 4, 3, 4, -1]$ passes third-degree polynomials and eliminates seasonal components with period 3.
- 1.15.** Let $\{Y_t\}$ be a stationary process with mean zero and let a and b be constants.
- If $X_t = a + bt + s_t + Y_t$, where s_t is a seasonal component with period 12, show that $\nabla \nabla_{12} X_t = (1 - B)(1 - B^{12})X_t$ is stationary and express its autocovariance function in terms of that of $\{Y_t\}$.
 - If $X_t = (a + bt)s_t + Y_t$, where s_t is a seasonal component with period 12, show that $\nabla_{12}^2 X_t = (1 - B^{12})^2 X_t$ is stationary and express its autocovariance function in terms of that of $\{Y_t\}$.
- 1.16.** (Using ITSM to smooth the strikes data.) Double-click on the ITSM icon, select File>Project>Open>Univariate, click OK, and open the file STRIKES.TSM. The graph of the data will then appear on your screen. To smooth the data select Smooth>Moving Ave, Smooth>Exponential, or Smooth>FFT. Try using each of these to reproduce the results shown in Figures 1.18, 1.21, and 1.22.
- 1.17.** (Using ITSM to plot the deaths data.) In ITSM select File>Project>Open>Univariate, click OK, and open the project DEATHS.TSM. The graph of the data will then appear on your screen. To see a histogram of the data, click on the sixth yellow button at the top of the ITSM window. To see the sample autocorrelation function, click on the second yellow button. The presence of a strong seasonal component with period 12 is evident in the graph of the data and in the sample autocorrelation function.
- 1.18.** (Using ITSM to analyze the deaths data.) Open the file DEATHS.TSM, select Transform>Classical, check the box marked Seasonal Fit, and enter 12 for the period. Make sure that the box labeled Polynomial Fit is not checked, and click, OK. You will then see the graph (Figure 1.24) of the deseasonalized data. This graph suggests the presence of an additional quadratic trend function. To fit such a trend to the deseasonalized data, select Transform>Undo Classical to retrieve the original data. Then select Transform>Classical and check the boxes marked Seasonal Fit and Polynomial Trend, entering 12 for the period and Quadratic for the trend. Click OK and you will obtain the trend function

$$\hat{m}_t = 9952 - 71.82t + 0.8260t^2, \quad 1 \leq t \leq 72.$$

At this point the data stored in ITSM consists of the estimated noise

$$\hat{Y}_t = x_t - \hat{m}_t - \hat{s}_t, \quad t = 1, \dots, 72,$$

obtained by subtracting the estimated seasonal and trend components from the original data. The sample autocorrelation function can be plotted by clicking on the second yellow button at the top of the ITSM window. Further tests for dependence can be carried out by selecting the options `Statistics>Residual Analysis>Tests of Randomness`. It is clear from these that there is substantial dependence in the series $\{Y_t\}$.

To forecast the data without allowing for this dependence, select the option `Forecasting>ARMA`. Specify 24 for the number of values to be forecast, and the program will compute forecasts based on the assumption that the estimated seasonal and trend components are true values and that $\{Y_t\}$ is a white noise sequence with zero mean. (This is the default model assumed by ITSM until a more complicated stationary model is estimated or specified.) The original data are plotted with the forecasts appended.

Later we shall see how to improve on these forecasts by taking into account the dependence in the series $\{Y_t\}$.

1.19. Use a text editor, e.g., WORDPAD or NOTEPAD, to construct and save a text file named TEST.TSM, which consists of a single column of 30 numbers, $\{x_1, \dots, x_{30}\}$, defined by

$$x_1, \dots, x_{10} : 486, 474, 434, 441, 435, 401, 414, 414, 386, 405;$$

$$x_{11}, \dots, x_{20} : 411, 389, 414, 426, 410, 441, 459, 449, 486, 510;$$

$$x_{21}, \dots, x_{30} : 506, 549, 579, 581, 630, 666, 674, 729, 771, 785.$$

This series is in fact the sum of a quadratic trend and a period-three seasonal component. Use the program ITSM to apply the filter in Problem 1.14 to this time series and discuss the results.

(Once the data have been typed, they can be imported directly into ITSM by copying and pasting to the clipboard, and then in ITSM selecting `File>Project>New>Univariate`, clicking on OK and selecting `File>Import Clipboard`.)