The Analysis of Time Series: An Introduction

Second Edition

C. Chatfield

The Analysis of Time Series: An Introduction

C. CHATFIELD

Senior Lecturer in Statistics,
The University of Bath

SECOND EDITION

1980
LONDON NEW YORK
CHAPMAN AND HALL
150TH ANNIVERSARY

Preface to Second Edition

This book provides an introduction to the analysis of time series, a subject which is of increasing interest at the present time. The book can be used as a text for an undergraduate or postgraduate course in time series, or it can be used for self-tuition by research workers. In writing the book I was conscious of the fact that several books had already been written on theoretical aspects of time-series analysis. So my aim was to provide a comprehensible introduction to a rather difficult subject which bridges the gap between theory and practice. Enough theory is given to introduce the concepts of time-series analysis and to make the book mathematically interesting. In addition practical problems are considered so as to help the reader tackle the analysis of real data.

Although the book is primarily an introductory text, I have nevertheless added appropriate references to further reading and to more advanced topics so that the interested reader can pursue his studies if he wishes. These references are mainly to comprehensible and readily accessible sources rather than to the original attributive references.

The book assumes a knowledge of basic probability theory and elementary statistical inference.

One difficulty in writing an introductory textbook is that many practical problems contain at least one feature which is 'non-standard' and these cannot all be envisaged in a book of reasonable length. Thus the reader who has grasped the basic concepts of time-series analysis should always be prepared to use his common-sense in tackling a problem.

Example 1 of Section 5.5 is a typical situation where common-sense has to be applied and also stresses the fact that the first step in any time-series analysis should always be to plot the data.

The second edition has much the same structure as the first edition. The text has been clarified and updated in many places, particularly in Chapters 3 and 4. Some sections such as Sections 1.5, 2.3, 4.2.2, 4.6 and 9.1 have been completely rewritten. But these changes have been made without altering the Section numbering or the page numbers up to p. 227. This should be helpful to teachers who are used to the first edition and should also keep down costs and new typesetting errors. Chapter 10 has been substantially rewritten, and the references have been completely revised. The subtitle of the book has been changed to indicate the introductory nature of the book.

I am indebted to many people for assistance in the preparation of this book including V. Barnett, D. R. Cox, K. V. Diprose, R. Fenton, P. R. Fisk, H. Neave, J. Marshall, P. Newbold, M. P. G. Pepper, M. B. Priestley, D. L. Prothero and C. M. Theobald. Of course, any errors, omissions or obscurities which remain are entirely my responsibility. The author will be glad to hear from any reader who wishes to make constructive comments.

Finally it is a pleasure to thank Mrs Jean Honebon for typing the first edition and Mrs Doreen Faulds for typing the changes for the second edition in such an efficient way.

School of Mathematics, Christopher Chatfield
Bath University,
September 1979.

Abbreviations

AR	Autoregressive
MA	Moving Average
ARMA	Mixed Autoregressive Moving Average
ARIMA	Autoregressive Integrated Moving Average
ac.f	Autocorrelation function
acv.f.	Autocovariance function
FFT	Fast Fourier Transform
$N(\mu,\sigma^2)$	A normally distributed random variable, mean μ, variance σ^2.
χ^2_ν	A chi-squared random variable with ν degrees of freedom

Notation

∇	The difference operator such that $\nabla x_t = x_t - x_{t-1}$.
B	The backward shift operator such that $Bx_t = x_{t-1}$.
E	Expected value or expectation

Contents

Preface to Second Edition *page* xii

Abbreviations xiv

1 **Introduction** 1
1.1 Examples 1
1.2 Terminology 5
1.3 Objectives of time-series analysis 6
1.4 Approaches to time-series analysis 9
1.5 Review of books on time-series 10

2 **Simple descriptive techniques** 12
2.1 Types of variation 12
2.2 Stationary time series 14
2.3 Time plot 14
2.4 Transformations 14
2.5 Analysing series which contain a trend 16
2.5.1 Curve-fitting 16
2.5.2 Filtering 17
2.5.3 Differencing 21
2.6 Seasonal fluctuations 21
2.7 Autocorrelation 23
2.7.1 The correlogram 25
2.7.2 Interpreting the correlogram 25
2.8 Other tests of randomness 31
Exercises 31

3 **Probability models for time series** 33
3.1 Stochastic processes 33
3.2 Stationary processes 35
3.2.1 Second-order stationarity 37
3.3 The autocorrelation function 37
3.4 Some useful stochastic processes 39
3.4.1 A purely random process 39
3.4.2 Random walk 40
3.4.3 Moving average processes 41
3.4.4 Autoregressive processes 44
3 4.5 Mixed models 50
3.4.6 Integrated models 51
3.4.7 The general linear process 52
3.4.8 Continuous processes 52
3.5 The Wold decomposition theorem 54
Exercises 55

4 **Estimation in the time domain** 60
4.1 Estimating the autocovariance and autocorrelation functions 60
4.1.1 Interpreting the correlogram 63
4.1.2 Ergodic theorems 65
4.2 Fitting an autoregressive process 65
4.2.1 Estimating the parameters of an autoregressive process 65
4.2.2 Determining the order of an autoregressive process 69
4.3 Fitting a moving average process 70
4.3.1 Estimating the parameters of a moving average process 70
4.3.2 Determining the order of a moving average process 73
4.4 Estimating the parameters of a mixed model 73
4.5 Estimating the parameters of an integrated model 74
4.6 The Box–Jenkins seasonal model 74

4.7 Residual analysis 76
4.8 General remarks on model building 79
Exercises 81

5 **Forecasting** 82
5.1 Introduction 82
5.2 Univariate procedures 84
5.2.1 Extrapolation of trend curves 84
5.2.2 Exponential smoothing 85
5.2.3 Holt–Winters forecasting procedure 87
5.2.4 Box–Jenkins forecasting procedure 89
5.2.5 Stepwise autoregression 93
5.2.6 Other methods 93
5.3 Multivariate procedures 94
5.3.1 Multiple regression 95
5.3.2 Econometric models 96
5.3.3 Box–Jenkins method 97
5.4 A comparison of forecasting procedures 98
5.5 Some examples 101
5.6 Prediction theory 107
Exercises 108

6 **Stationary processes in the frequency domain** 110
6.1 Introduction 110
6.2 The spectral distribution function 111
6.3 The spectral density function 116
6.4 The spectrum of a continuous process 119
6.5 Examples 120
Exercises 125

7 **Spectral analysis** 127
7.1 Fourier analysis 127
7.2 A simple sinusoidal model 128
7.2.1 The Nyquist frequency 131
7.3 Periodogram analysis 133

7.3.1 The relationship between the periodogram and the autocovariance function 136
7.3.2 Properties of the periodogram 137
7.4 Spectral analysis: some consistent estimation procedures 138
7.4.1 Transforming the truncated autocovariance function 139
7.4.2 Hanning 142
7.4.3 Hamming 143
7.4.4 Smoothing the periodogram 143
7.4.5 The fast Fourier transform 145
7.5 Confidence intervals for the spectrum 149
7.6 A comparison of different estimation procedures 150
7.7 Analysing a continuous time series 155
7.8 Discussion 158
7.9 An example 165
Exercises 168

8 Bivariate processes 169
8.1 Cross-covariance and cross-correlation functions 169
8.1.1 Examples 171
8.1.2 Estimation 172
8.1.3 Interpretation 174
8.2 The cross-spectrum 174
8.2.1 Examples 178
8.2.2 Estimation 180
8.2.3 Interpretation 184
Exercises 185

9 Linear systems 186
9.1 Introduction 186
9.2 Linear systems in the time domain 187
9.2.1 Some examples 188

9.2.2 The impulse response function 192
9.2.3 The step response function 192
9.3 Linear systems in the frequency domain 193
9.3.1 The frequency response function 193
9.3.2 Gain and phase diagrams 198
9.3.3 Some examples 200
9.3.4 General relation between input and output 203
9.3.5 Linear systems in series 210
9.3.6 Design of filters 211
9.4 Identification of linear systems 213
9.4.1 Estimating the frequency response function 215
9.4.2 The Box–Jenkins approach 219
9.4.3 Systems involving feedback 223
Exercises 226

10 Some other topics 228

Appendix I The Fourier, Laplace and Z-transforms 236

Appendix II The Dirac Delta Function 241

Appendix III Covariance 243

References 245

Answers to exercises 256

Author Index 262

Subject Index 265

CHAPTER 1

Introduction

A time series is a collection of observations made sequentially in time. Examples occur in a variety of fields, ranging from economics to engineering, and methods of analysing time series constitute an important area of statistics.

1.1 Examples

We begin with some examples of the sort of time series which arise in practice

(a) Economic time series Many time series arise in economics, including such series as share prices on successive days, export totals in successive months, average incomes in successive months, company profits in successive years; and so on. Fig. 1.1 shows part of the classic Beveridge wheat price index series which consists of the average wheat price in nearly 50 places in various countries measured in successive years from 1500 to 1869. The complete series is given by Anderson (1971). This series is of particular interest to economic historians, and, when analysed (Granger and Hughes, 1971), shows clear evidence of an important cycle with a period of roughly 13.3 years.

(b) Physical time series Many types of time series occur in the physical sciences, particularly in meteorology, marine science, and geophysics. Examples are rainfall on successive

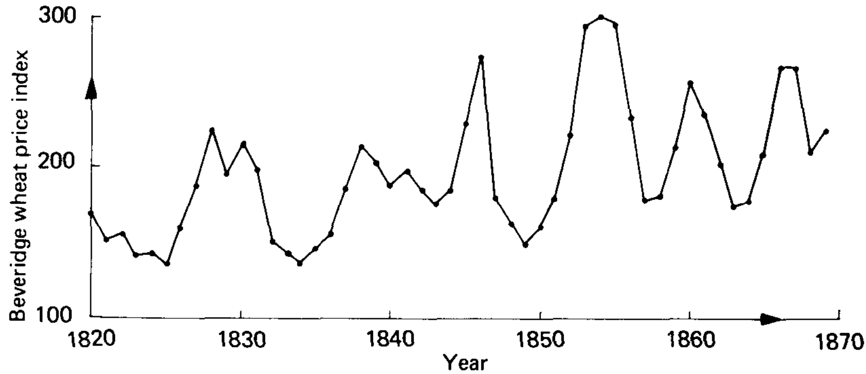

Figure 1.1 *Part of the Beveridge wheat price index series.*

days, and air temperature measured in successive hours, days or months. Fig. 1.2 shows the air temperature at Recife, in Brazil, measured over 10 years, where the individual observations are averaged over periods of one month.

Some mechanical recorders take measurements continuously and produce a continuous trace rather than observations at discrete intervals of time. For example in some laboratories it is important to keep temperature and humidity as constant as possible and so devices are installed to measure these variables continuously. Some examples of continuous traces are given in Chapter 7 in Fig. 7.4.

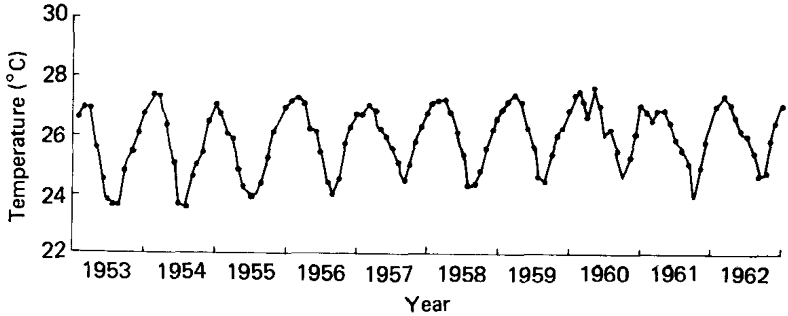

Figure 1.2 *Average air temperature at Recife, Brazil, in successive months.*

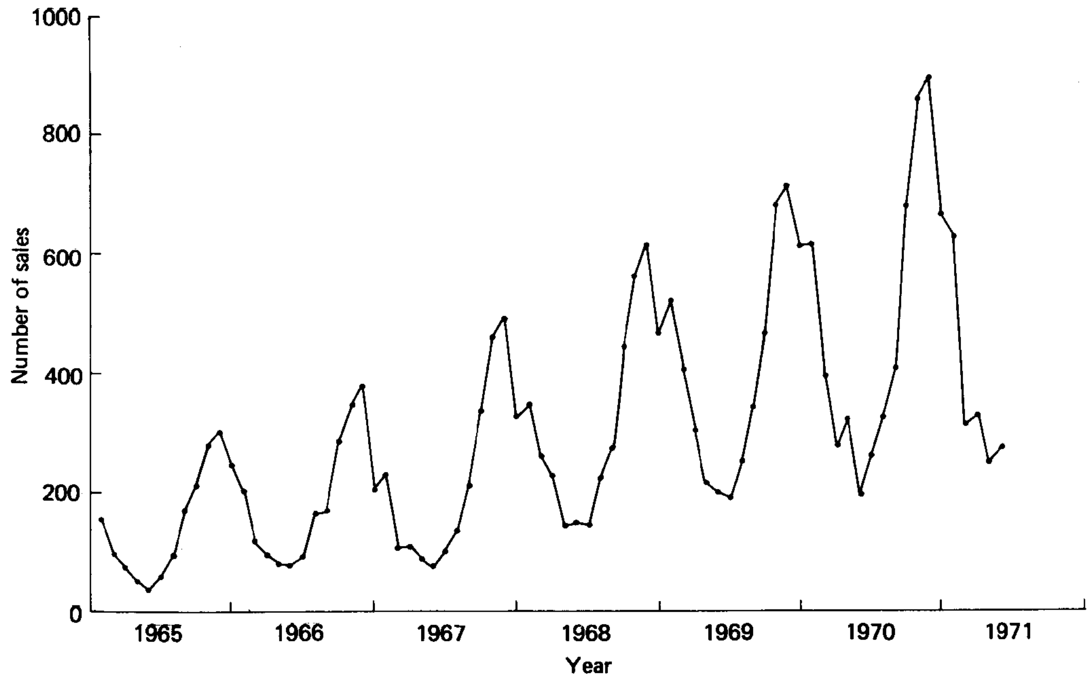

Figure 1.3 *Sales of a certain engineering company in successive months.*

Continuous recorders often smooth the data to some extent as they are insensitive to high frequency variation.

(c) Marketing time series The analysis of sales figures in successive weeks or months is an important problem in commerce. Fig. 1.3, taken from Chatfield and Prothero (1973a), shows the sales of an engineering product by a certain company in successive months over a seven-year period. Marketing data have many similarities to economic data. It is often important to forecast future sales so as to plan production. It is also of interest to examine the relationship between sales and other time series such as advertising expenditure in successive time periods.

(d) Demographic time series Time series occur in the study of population. An example is the population of England and Wales measured annually. Demographers want to predict changes in population for as long as ten or twenty years into the future (e.g. Brass, 1974).

(e) Process control In process control, the problem is to detect changes in the performance of a manufacturing process by measuring a variable which shows the quality of the process. These measurements can be plotted against time as in Fig. 1.4. When the measurements stray too far from some target value, appropriate corrective action should be

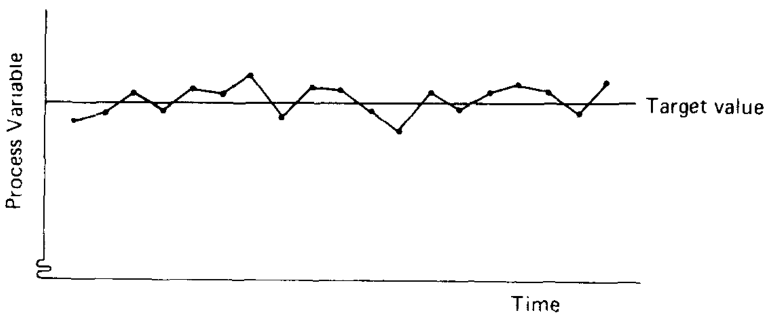

Figure 1.4 A process control chart.

taken to control the process. Special techniques have been developed for this type of time-series problem and the reader is referred to a book on statistical quality control (e.g. Wetherill, 1977).

(f) Binary processes A special type of time series arises when observations can take one of only two values, usually denoted by 0 and 1 (see Fig. 1.5). Time series of this type, called binary processes, occur particularly in communication theory. For example the position of a switch, either 'on' or 'off', could be recorded as one or zero respectively.

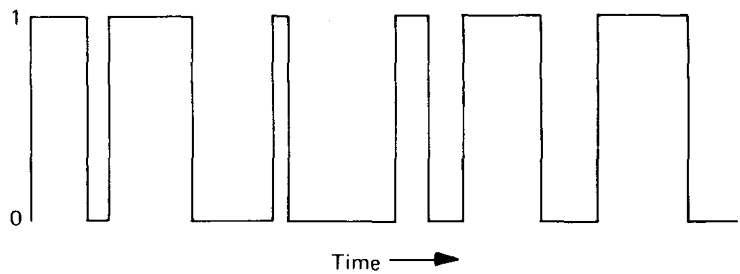

Figure 1.5 A realization of a binary process.

(g) Point processes A different type of time series occurs when we consider a series of events occurring 'randomly' in time. For example we could record the dates of major railway disasters. A series of events of this type is often called a *point process* (see Fig. 1.6). For observations of this type, we are interested in the distribution of the number of events occurring in a given time-period and also in the distribution of time intervals between events. Methods of analysing data of this type are discussed by Cox and Lewis (1966), and will not be specifically discussed in this book.

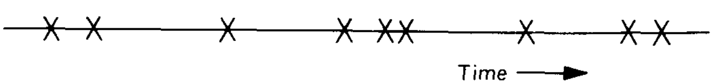

Figure 1.6 A realization of a point process. (X denotes an event.)

1.2 Terminology

A time series is said to be *continuous* when observations are made continuously in time as in Figs. 1.5 and 7.4. The term continuous is used for series of this type even when the measured variable can only take a discrete set of values, as in Fig. 1.5. A time series is said to be *discrete* when observations are taken only at specific times, usually equally spaced. The term discrete is used for series of this type even when the measured variable is a continuous variable.

In this book we are mainly concerned with discrete time series, where the observations are taken at equal intervals, We will also consider continuous time series, and in Chapter 10 will briefly consider discrete time series where the observations are taken at unequal intervals of time.

Discrete time series can arise in several ways. Given a continuous time series, we could read off the values at equal intervals of time to give a discrete series called a *sampled* series. Another type of discrete series occurs when a variable does not have an instantaneous value but we can *aggregate* (or accumulate) the values over equal intervals of time. Examples of this type are exports measured monthly and rainfall measured daily. Finally some time series are inherently discrete, an example being the dividend paid by a company to shareholders in successive years.

Much statistical theory is concerned with random samples of independent observations. The special feature of time-series analysis is the fact that successive observations are usually *not* independent and that the analysis must take into account the time *order* of the observations. When successive observations are dependent, future values may be predicted from past observations. If a time series can be predicted exactly, it is said to be *deterministic*. But most time series are *stochastic* in that the future is only partly determined by past values. For stochastic series exact predictions are impossible and must be replaced by the idea that future values have a probability distribution which is conditioned by a knowledge of past values.

1.3 Objectives of time-series analysis

There are several possible objectives in analysing a time series. These objectives may be classified as description, explanation, prediction and control, and will be considered in turn.

(a) Description When presented with a time series, the first step in the analysis is usually to plot the data and to obtain simple descriptive measures of the main properties of the series as described in Chapter 2. For example, looking at Fig. 1.3 it can be seen that there is a regular seasonal effect with sales 'high' in winter and 'low' in summer. It also looks as though annual sales are increasing (i.e. show an upward trend). For some series, the variation is dominated by such 'obvious' features, and a fairly simple model, which only attempts to describe trend and seasonal variation, may be perfectly adequate to describe the variation in the time series. For other series, more sophisticated techniques will be required to provide an adequate analysis. Then a more complex model will be constructed, such as the various types of stochastic process described in Chapter 3.

This book devotes a greater amount of space to the more advanced techniques, but this does not mean that elementary descriptive techniques are unimportant. Anyone who tries to analyse a time series, without plotting it first, is asking for trouble. Not only will a graph show up trend and seasonal variation, but it also enables one to look for 'wild' observations or *outliers* which do not appear to be consistent with the rest of the data. The treatment of outliers is a complex subject in which common-sense is more important