

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 SOME REPRESENTATIVE TIME SERIES

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. Examples include share prices on successive days, export totals in successive months, average incomes in successive months, company profits in successive years, and so on.

Figure 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. This series is of particular interest to economic historians. The complete series is tabulated by Anderson (1971).

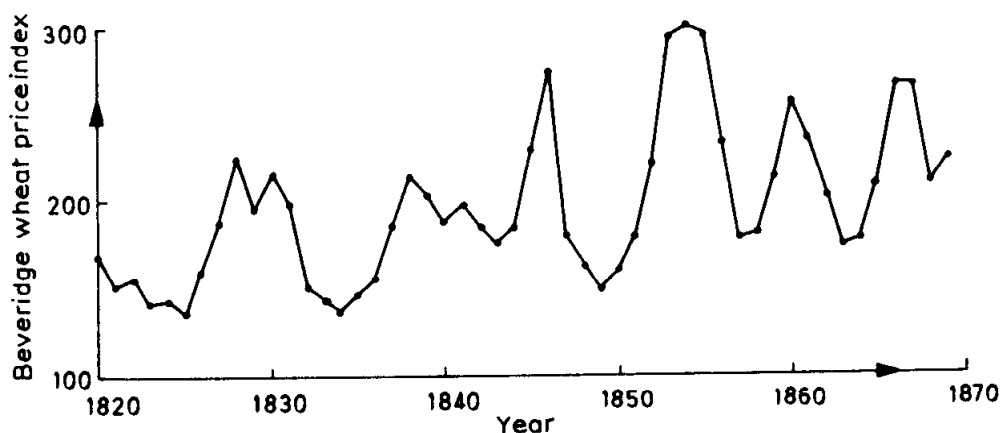


Figure 1.1 Part of the Beveridge wheat price index series.

2 Introduction

(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 days, and air temperature measured in successive hours, days or months. Figure 1.2 shows the air temperature at Recife, in Brazil, averaged over successive months. These data are tabulated and analysed in Example D.1.

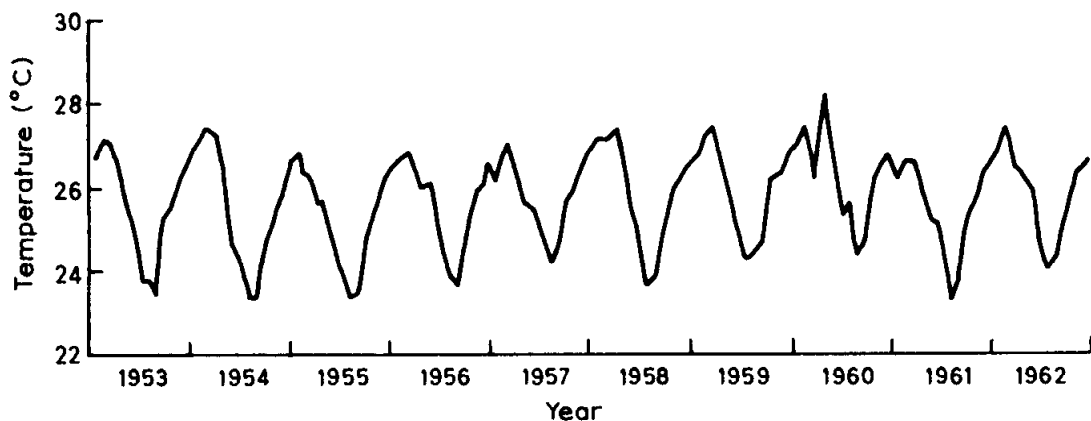


Figure 1.2 Average air temperature at Recife, Brazil, in successive months.

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 Figure 7.4. In order to analyse such series, it may be helpful to sample or digitize them at equal intervals of time.

(c) *Marketing time series*

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

(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

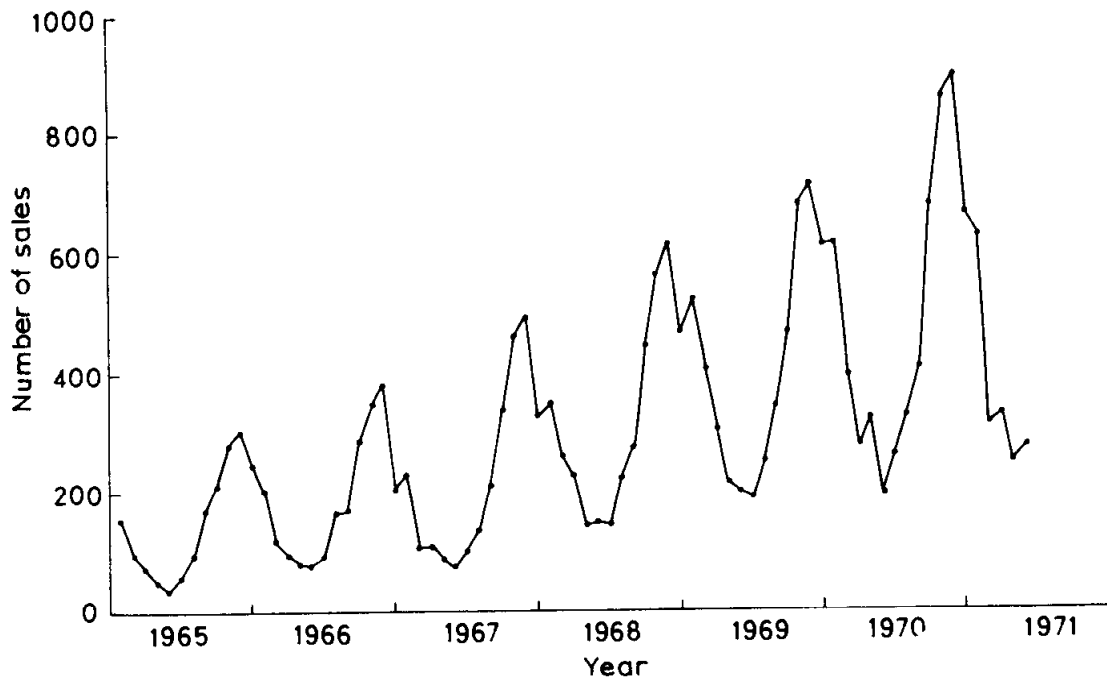


Figure 1.3 Sales of a certain engineering company in successive months.

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 Figure 1.4. When the measurements stray too far from some target value, appropriate corrective action should be taken to control the process. Special techniques

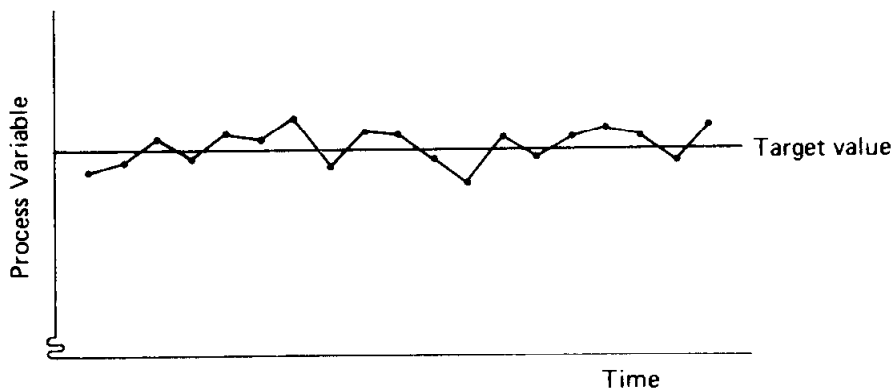


Figure 1.4 A process control chart.

4 Introduction

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 and Brown, 1991).

(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 Figure 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.

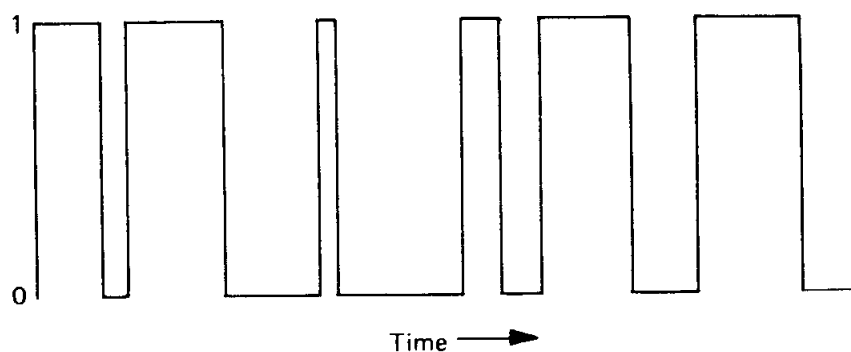


Figure 1.5 A realization of a binary process.

(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 Figure 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 will not be specifically discussed in this book (see for example Cox and Lewis, 1966; Cox and Isham, 1980).

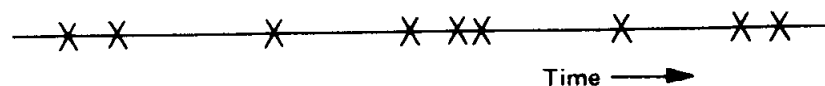


Figure 1.6 A realization of a point process (\times denotes an event).

1.2 TERMINOLOGY

A time series is said to be **continuous** when observations are made continuously in time as in Figures 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 Figure 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 also consider continuous time series more briefly, and in Section 13.5.4 we give some references regarding the analysis of discrete time series taken at unequal intervals of time.

Discrete time series can arise in several ways. Given a continuous time series, we could read off (or digitize) 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, so that 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 Figure 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

6 Introduction

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 as important as theory. The 'outlier' may be a perfectly valid but extreme observation which may for example indicate that the data are not normally distributed. Alternatively, the outlier may be a freak observation arising, for example, when a recording device goes wrong or when a strike severely affects sales. In the latter case, the outlier needs to be adjusted in some way before further analysis of the data. **Robust** methods (e.g. Martin, 1983) are designed to be insensitive to outliers.

Another feature to look for in the graph of the time series is the possible presence of turning points, where, for example, an upward trend has suddenly changed to a downward trend. If there is a turning point, different models may have to be fitted to the two parts of the series.

(b) *Explanation*

When observations are taken on two or more variables, it may be possible to use the variation in one time series to explain the variation in another series. This may lead to a deeper understanding of the mechanism which generated a given time series.

Multiple regression models may be helpful here. In Chapter 9 we also consider the analysis of what are called **linear systems**. A linear system converts an input series to an output series by a linear operation. Given observations on the input and output to a linear system (see Figure 1.7), one wants to assess the properties of the linear system. For example it is of interest to see how sea level is affected by temperature and pressure, and to see how sales are affected by price and economic conditions.

(c) *Prediction*

Given an observed time series, one may want to predict the future values of the series. This is an important task in sales forecasting, and in the analysis of economic and industrial time series. Many writers, including myself, use the

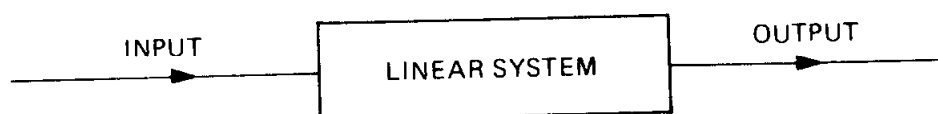


Figure 1.7 Schematic representation of a linear system.

terms 'prediction' and 'forecasting' interchangeably, but some authors do not. For example Brown (1963) uses 'prediction' to describe subjective methods and 'forecasting' to describe objective methods, whereas Brass (1974) uses 'forecast' to mean any kind of looking into the future, and 'prediction' to denote a systematic procedure for doing so.

Prediction is closely related to **control** problems in many situations. For example if one can predict that a manufacturing process is going to move off target, then appropriate corrective action can be taken.

(d) *Control*

When a time series is generated which measures the 'quality' of a manufacturing process, the aim of the analysis may be to control the process. Control procedures are of several different kinds. In statistical quality control, the observations are plotted on control charts and the controller takes action as a result of studying the charts. A more sophisticated control strategy has been described by Box *et al.* (1994). A stochastic model is fitted to the series, future values of the series are predicted, and then the input process variables are adjusted so as to keep the process on target. Many other contributions to control theory have been made by control engineers and mathematicians rather than statisticians. This topic is rather outside the scope of this book but is briefly introduced in Section 13.4.

1.4 APPROACHES TO TIME-SERIES ANALYSIS

This book will describe various approaches to time-series analysis. In Chapter 2 we will describe simple descriptive techniques, which consist of plotting the data and looking for trends, seasonal fluctuations, and so on. Chapter 3 introduces a variety of probability models for time series, while Chapter 4 discusses ways of fitting these models to time series. The major diagnostic tool which is used in Chapter 4 is a function called the **autocorrelation** function which helps to describe the evolution of a process through time. Inference based on this function is often called an analysis in the **time domain**.

Chapter 5 discusses a variety of forecasting procedures. This chapter is not a prerequisite for the rest of the book and the reader may, if he wishes, proceed from Chapter 4 to Chapter 6.

Chapter 6 introduces a function called the **spectral density** function which describes how the variation in a time series may be accounted for by cyclic components at different frequencies. Chapter 7 shows how to estimate this function, a procedure which is called **spectral analysis**. Inference based on the spectral density function is often called an analysis in the **frequency domain**.

Chapter 8 discusses the analysis of two time series, while Chapter 9 extends

8 Introduction

this work by considering linear systems in which one series is regarded as the input, while the other series is regarded as the output.

Chapter 10 introduces an important class of models, called **state-space models**. It also introduces the **Kalman filter**, which is a general method of updating the best estimate of the 'signal' in a time series in the presence of noise.

Chapters 11 and 12 give brief introductions to non-linear and multivariate time-series models while Chapter 13 surveys several other topics, with appropriate references, to give a flavour of recent research developments.

1.5 REVIEW OF BOOKS ON TIME SERIES

This section gives a brief review of some other relevant books on time series. The literature has expanded considerably in recent years and a selective approach is necessary.

Alternative general introductory texts include Harvey (1993), Kendall and Ord (1990), and Wei (1990), while Diggle (1990) is aimed primarily at biostatisticians, Gottman (1981) at social scientists, and Mills (1990) at economists.

There are many more advanced texts including Hannan (1970), Anderson (1971), Fuller (1976), Brillinger (1981), Brockwell and Davis (1991) and Priestley (1981). The latter is particularly strong on spectral analysis and multivariate time series modelling. Kendall, Stuart and Ord (1983), now in its fourth edition, is also a valuable reference source, but note that earlier editions are somewhat dated.

The classic book by Box and Jenkins (1970) describes an approach to time-series analysis, forecasting and control which is based on a particular class of models, usually called ARIMA models. This important book is not really suitable for the beginner, who is recommended to read Chapters 3–5 in this book, Vandaele (1983) or Wei (1990). The 1976 revised edition of Box and Jenkins (1970) was virtually unchanged, but the new third edition (Box *et al.*, 1994), with G. Reinsel as third author, has Chapters 12 and 13 completely rewritten and so we refer to the 1994 edition in regard to this new material on intervention analysis, outlier detection and process control. However, the first eleven chapters have a very similar structure, though there are some additions such as new material on ARMA model estimation and testing for unit roots. For historical precedence and reader convenience we continue to refer to the 1970 edition for the material on ARIMA models, identification, estimation and forecasting.

Additional books are referenced as appropriate in later chapters.