Chapter 1

Introduction

This chapter introduces the main issues involved in modelling time series. An overview of the proposed methodology is given without going into any technical details.

1.1 The nature of time series

The time series shown in Figure 1.1.1 consists of observations on the quarterly consumption of coal in the UK by 'Other final users', a group which includes public administration, commerce and agriculture. It is typical of many economic and social time series. Its salient characteristics are a trend, which represents the long-run movements in the series, and a seasonal pattern which repeats itself more or less every year. A model of the series will need to capture these characteristics. There are many ways in which such a model may be formulated, but a useful starting point is to assume that the series may be decomposed in the following way:

$$\text{Observed series} = \text{trend} + \text{seasonal} + \text{irregular} \tag{1.1.1}$$

where the 'irregular' component reflects non-systematic movements in the series. The model is an additive one. A multiplicative form,

$$\text{Observed series} = \text{trend} \times \text{seasonal} \times \text{irregular} \tag{1.1.2}$$

may often be more appropriate, as indeed it is in the case of the series shown in Figure 1.1.1. However, a multiplicative model may be handled within the additive framework by the simple expedient of taking logarithms.

There are two reasons for wishing to model a univariate time series. The first is to provide a *description* of the series in terms of its components of interest. One may, for example, wish to examine the trend in order to see the main movements which have taken place in the series. The seasonal behaviour of the series may also be of interest and for some purposes it may be desirable to extract the seasonal component to produce a seasonally adjusted series. Traditionally such operations have been carried out without recourse to a statistical model. However, it can be shown that for many of these 'model-free' procedures there is a well-defined statistical model for which the procedure in question is optimal.

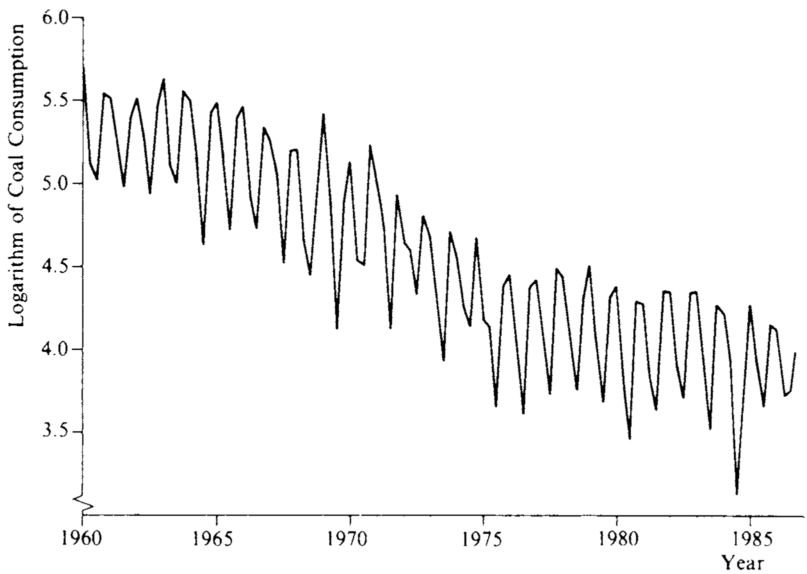


Fig. 1.1.1 UK coal consumption by 'Other final users'.

The advantage of an explicit statistical model is not only that it makes the underlying assumptions clear but that, if properly formulated, it has the flexibility to represent adequately the movements in time series which may have widely differing properties. Hence it is likely to yield a better description of the series and its components. The other motive underlying the construction of a univariate time series model is the *prediction* of future observations. As a rule, the model used for description should also be used as the basis for forecasting, and the fact that a sensible description of the series is an aim of model-building acts as a discipline for selecting models which are likely to be successful at forecasting.

Time series may contain other components. Figure 1.1.2 shows an annual series, US Real Gross National Product (GNP) over the period 1909–70. There is a clear trend, but, in addition, the earlier part of the series shows marked cyclical behaviour as the economy moves from boom to recession and back again. Indeed, we would probably have known this from the economic history of the period without even looking at the graph. Incorporating a cyclical component in a model for US Real GNP will therefore play an important role in providing a description of this series, at least in its early stages. The fact that the properties of the series appear to change shortly after the end of the Second World War illustrates another aspect of economic and social time series, namely that their properties do not necessarily remain the same over time.

A *structural time series model* is one which is set up in terms of components which have a direct interpretation. A univariate structural model is not intended

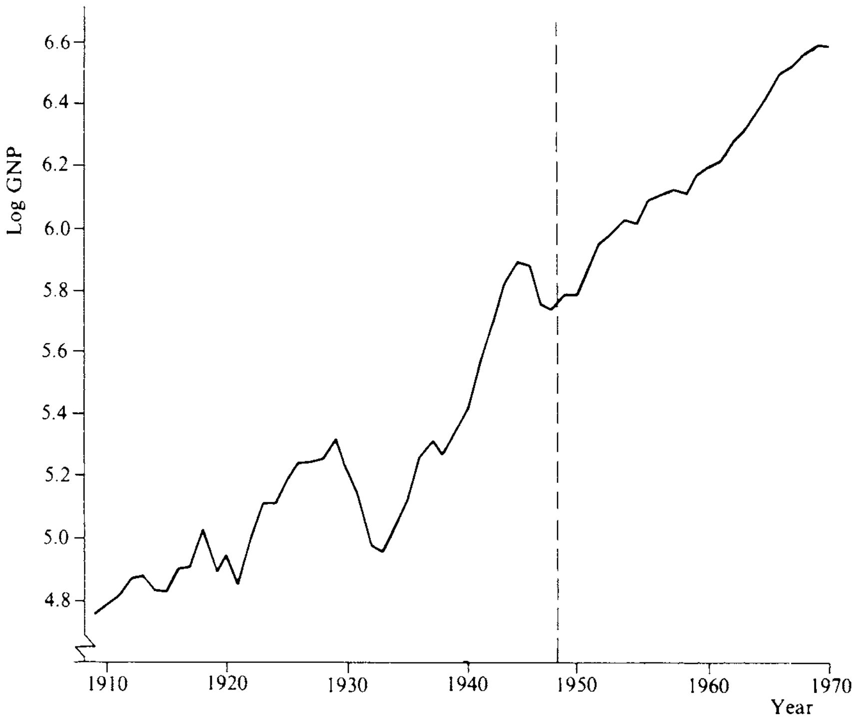


Fig. 1.1.2 US Real GNP

to represent the underlying data generation process. Rather it aims to present the 'stylised facts' of a series in terms of a decomposition into components such as trend, seasonal and cycle. These quantities are of interest in themselves. Furthermore, they highlight the features of a series which must be accounted for by a properly formulated behavioural model. Prediction from a univariate model is naive in the sense that it is just an extrapolation of past movements. Nevertheless, it is often quite effective and it provides a yardstick against which the performance of more elaborate models may be assessed.

The statistical formulation of the trend component in a structural model needs to be flexible enough to allow it to respond to general changes in the direction of the series. A trend is not seen as a deterministic function of time about which the series is constrained to move for ever more. In a similar way the seasonal component must be flexible enough to respond to changes in the seasonal pattern. A structural time series model therefore needs to be set up in such a way that its components are stochastic; in other words, they are regarded as being driven by random disturbances. The statistical framework for handling such models is outlined in section 1.4.

Fig. 1.1.3 Goals scored by England against Scotland at Hampden Park.

The series in Figure 1.1.3 records the number of goals scored by England against Scotland in international football matches at Hampden Park in Glasgow. This is an example of *count data*; by their nature the observations must be non-negative integers and this has implications for the statistical assumptions upon which a model should be based. The structure of the series, however, is clearly one where the underlying level moves up and down over time according to the relative strengths of the two teams. Given this level, the actual number of goals scored on the day depends on a mixture of inspiration and luck. Had the data been of the result of each match, that is a win, a loss or a draw, yet another statistical model would have been appropriate, but the underlying structure would still have been essentially the same.

1.2 Explanatory variables and intervention analysis

The employment–output equation provides a means of forecasting employment, given predictions of the future value of output. The economic theory discussed in Ball and St Cyr (1966), Nickell (1984) and Harvey *et al.* (1986), and the references therein, suggests that employment can be regarded as being dependent on firms' output expectations. This enables an equation to be constructed in which the level of employment is explained by current and past levels of output, by employment in the previous time period and by the capital stock and technical progress. These last two factors are not only difficult to measure, but are also difficult to separate conceptually. If they could be measured, their combined effect would yield a measure of productivity, and the employment–output equation could be formulated as:

$$\text{Employment} = \text{productivity effect} + \text{output effect} + \text{disturbance term} \tag{1.2.1}$$

The introduction of lagged values of employment and output into this equation would allow for dynamic effects. The essential point, though, is that the model could be handled by standard classical regression procedures. This is not the case once it is accepted that productivity cannot be measured directly. The productivity effect in (1.2.1) must then be proxied by a trend component, so that the equation takes the form:

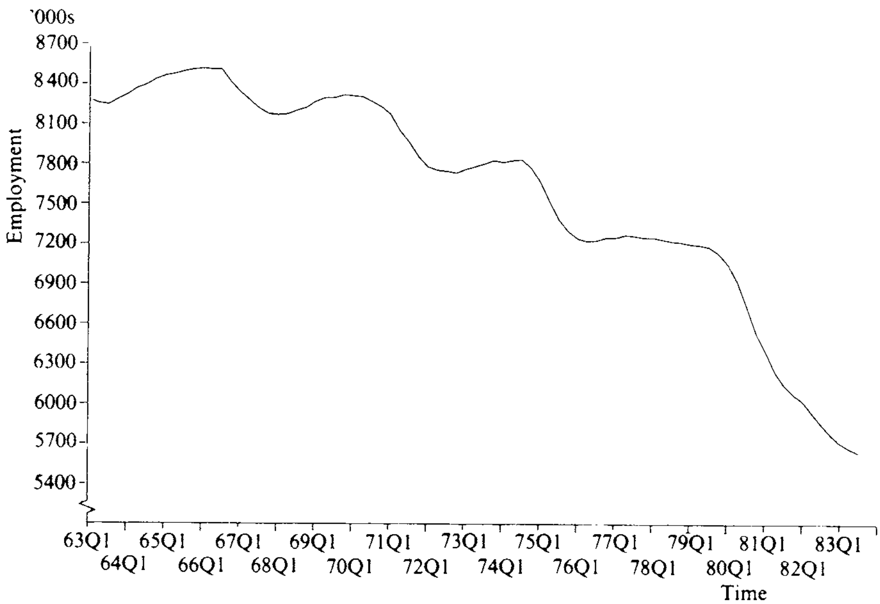
$$\text{Employment} = \text{trend} + \text{output effect} + \text{disturbance term} \tag{1.2.2}$$

Like the trend component in a univariate model, the trend component in (1.2.1) must be stochastic in order to allow for changes in the extent and influence of productivity.

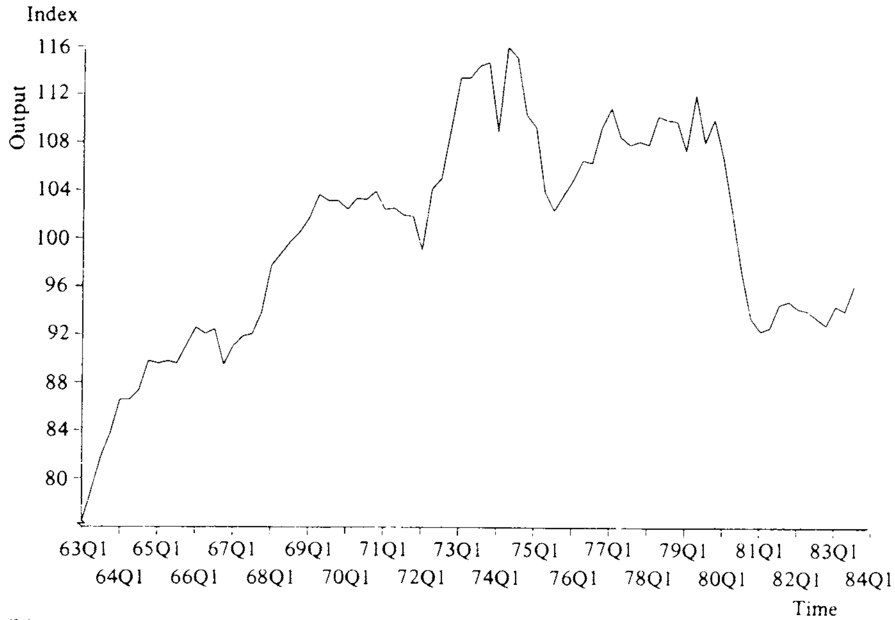
Figure 1.2.1(a and b) shows employment and output in UK manufacturing on a quarterly basis from 1963Q1 to 1983Q3; the data are seasonally adjusted. Employment is measured in thousands, while output is an index with 1980 = 100. Both series exhibit short-term movements which appear to be correlated. On the other hand, even without the prior knowledge that productivity changes influence employment, it is clear from Figure 1.2.1 that the level of output cannot possibly account for all the long-run movements in employment. For most of the period in question output shows a tendency to rise while employment goes down, hence the need for the introduction of a trend component into the model in order to account for the discrepancy.

Another example of a series in which the level can be only partly explained by explanatory variables is the one shown in Figure 1.2.2. The series is the number of car drivers in Great Britain killed and seriously injured (KSI) each month from January 1969 to December 1984. Explanatory variables such as the car traffic index, which measures the number of kilometres travelled by cars in a month, and the real price of petrol can be introduced into a model. However, other variables which affect road accidents, such as the quality of the roads, are difficult to quantify, and this means that some of the long-term movements in the series can only be proxied by a stochastic trend. In a similar way, the seasonal effects in the series are, to some extent, a reflection of weather conditions and the consumption of alcohol, the latter presumably having a very marked effect on the figures in December. While some attempt might be made to introduce explanatory variables into the model, it is unlikely that they could successfully account for all the seasonal movements. This being the case, there is still a role for a seasonal component in the model in the same way as there is a role for a trend.

An interesting feature of the car drivers KSI series is the sharp drop in its level in February 1983. As will be shown in section 7.5, this may be attributed to the introduction of the seat-belt law which became effective on 31 January 1983. Indeed it was this law which provided the original motivation for studying the



(a)



(b)

Fig. 1.2.1 Employment and output in UK manufacturing.

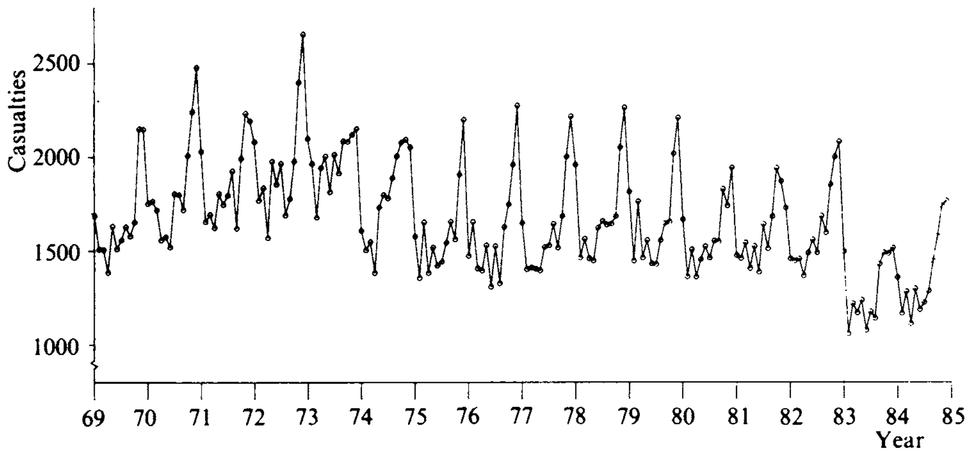


Fig. 1.2.2 Car drivers killed and seriously injured in Great Britain.

series (Durbin and Harvey, 1985; Harvey and Durbin, 1986). The effect of an event or a policy change on a series can be measured by bringing a dummy explanatory variable into the model. This is known as *intervention analysis*, a term which was introduced into the literature by Box and Tiao (1975). Intervention analysis may be carried out both with and without other explanatory variables present.

1.3 Multivariate models

The series in Figure 1.3.1 are indices showing the paid minutes of telephone calls from Australia to three other countries in the world. Each of these series could be modelled separately. However, because they are subject to similar influences, it is possible to model them jointly. When a multivariate structural model is set up to cope with this kind of situation, the disturbances in the various components are assumed to be correlated across the different series. Explanatory variables can also be introduced into the model, the rationale for doing so being exactly as given in the previous section.

The data shown in Figure 1.3.1 are a *cross-section* of time series. Data of this kind may arise in a variety of applications. We may be dealing with observations on a set countries or states, or on the sales of products of a similar type produced by a firm. The data may also be on a sample of households or individuals whose behaviour is followed over time. This is known as *panel data*.

When the data consist of cross-sections of time series, the individual elements do not interact directly with each other. This contrasts with a situation in which there are behavioural relationships between a set of variables. Such a situation

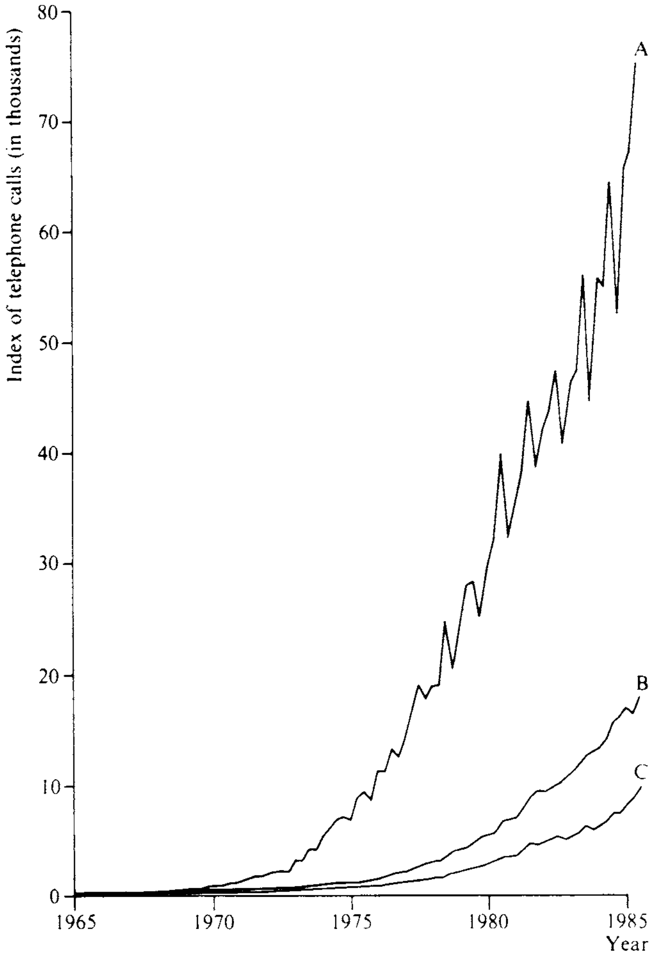


Fig. 1.3.1 Paid minutes of telephone calls from Australia to three other countries.

obtains in the two series shown in Figure 1.3.2. These series record the numbers of mink and muskrat furs traded annually by the Hudson's Bay Company in Canada from 1848 to 1909. There is known to be a prey–predator relationship between these two animals and this affects the population dynamics of both of them. Constructing a multivariate time series model for the two series reflects these dynamic interactions. Dynamic interactions are also apparent when multivariate models are constructed for economic variables, such as income, consumption and investment, which are known to affect each other. The class of models needed to handle situations of this kind is much wider than that required for cross-sections of time series. The type of model to be considered in a particular

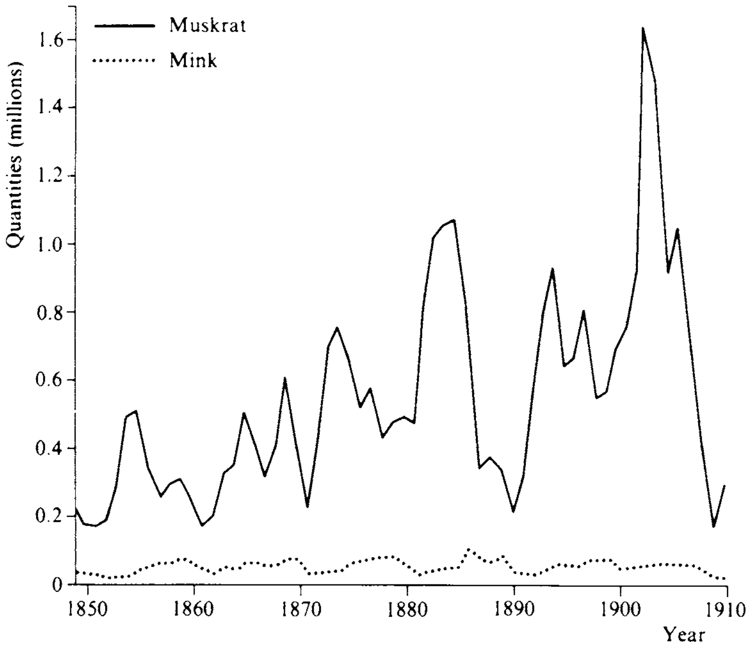


Fig. 1.3.2 Mink and muskrat furs traded by the Hudson's Bay Company.

situation will depend on the amount of information available, in terms of prior theoretical knowledge and data on relevant variables. If there is a satisfactory level of such information, it should be possible to construct an econometric model or, more specifically, a simultaneous equation model. The interest in this book, however, is on situations where a lack of information precludes the construction of a fully specified econometric model.

An interesting aspect of multivariate structural time series models concerns the possibility of components, such as the trend, being common to more than one series. This leads to a class of *dynamic factor models*. Setting up factors in terms of unobserved components produces a model which is not only tractable, but has a useful interpretation. A common trend leads to series being *co-integrated*. If two series are co-integrated, they must move together in the long term, and such a property is an important one to bear in mind, particularly when considering certain macroeconomic variables such as income and consumption.

Current practice in dealing with sets of variables which interact with each other is either to build a fully specified model, or to abandon any structure completely and set up a vector autoregression. The multivariate time series models described in this book are not only more parsimonious than vector autoregressions, but they also offer the possibility of imposing co-integrating restrictions, and taking account of explanatory variables.

1.4 Statistical treatment

A model for equation (1.1.1) could be formulated as a regression with explanatory variables consisting of a time trend and a set of seasonal dummies. This would be inadequate. The necessary flexibility may, however, be achieved by letting the regression coefficients change over time. A similar treatment may be accorded to other components, such as cycles and day-of-the-week effects. *The principal structural time series models are therefore nothing more than regression models in which the explanatory variables are functions of time and the parameters are time-varying.* Given this interpretation, the addition of observable explanatory variables is a natural extension. Furthermore the use of a regression framework opens the way to a unified model selection methodology for econometric and time series models.

The key to handling structural time series models is the *state space form*, with the state of the system representing the various unobserved components such as trends and seasonals. Once in state space form (SSF), the *Kalman filter* provides the means of updating the state as new observations become available. Predictions are made by extrapolating these components into the future. Various *smoothing* algorithms, all of which are related to the Kalman filter, can be used for obtaining the best estimate of the state at any point within the sample. This can be valuable for examining the way in which a component such as the trend has evolved in the past. Figure 1.4.1 shows the smoothed estimates of the trend for the UK coal consumption data over the period up to, and including, the last quarter of 1982. The extrapolation of the trend over the last four years is also shown. This extrapolation is roughly horizontal reflecting the halt in the decline in coal consumption, which the model estimated using the data up to 1982 has picked up. The sharp dip in the series in 1984 is due to special circumstances, namely a prolonged miners' strike.

Prediction and smoothing can only be carried out once the parameters governing the stochastic movements of the state variables have been estimated. The estimation of these parameters, which are known as *hyperparameters*, is itself based on the Kalman filter. This is because the likelihood function can be expressed in terms of one-step-ahead prediction errors, and these prediction errors emerge as a by-product of the filter. When a model is linear and time-invariant, estimation of the hyperparameters can also be carried out in the frequency domain. Again this is a maximum likelihood approach, and for many structural models it has important theoretical and computational advantages. However, the state space framework permits a much richer class of models than can be handled by frequency-domain methods. For example, it provides a vehicle for modelling non-linear effects and structural change.

Another important feature of the state space form is that it can cope with missing observations and temporal aggregation. It also permits the extension of time series models so as to make allowance for more subtle data irregularities